Time Series: Linear Processes, Cointegration and Structural Breaks

Seung Hyun Kim

August 11, 2024

Stationary Linear Processes

We start by studying the asymptotic properties of stationary linear processes. After the deriving more general results, we will study the Vector Autoregressive (VAR) processes. Throughout, we implicitly let $(\Omega, \mathcal{H}, \mathbb{P})$ be our probability space.

1.1 Asymptotic Theory for Martingales

We state here some central limit theorems and laws of large numbers involving martingale difference sequences, which are sequences of dependent variables that are uncorrelated with one another. They are named this way because their partial sum process defines a martingale.

1.1.1 Martingale Difference Arrays

An array $\{Z_{T,t}\}_{1 \le t \le k(T), T \in N_+}$ of real random variables is said to be a martingale difference array with respect to the filtration array $\{\mathcal{F}_{T,t}\}_{1 \le t \le k(T), T \in N_+}$ if

- $Z_{T,t}$ is $\mathcal{F}_{T,t}$ measurable and integrable for any $T \in N_+$ and $1 \le t \le k(T)$

- For any $T \in N_+$ and $1 \le t \le k(T)$,

$$\mathbb{E}\left[Z_{T,t} \mid \mathcal{F}_{T,t-1}\right] = 0,$$

  where $\mathcal{F}_{T,0}$ is taken to be the trivial σ-algebra.

Before presenting a CLT for martingale difference arrays, we state some preliminary results:

Lemma The following hold true:

i) For any $x \in \mathbb{R}$,

$$\left|\exp(ix) - \left(1 + ix - \frac{x^2}{2}\right)\right| \leq \min\left(|x|^3, |x|^2\right).$$

ii) For any $z \in \mathbb{C}$,

$$|\exp(z) - (1+z)| \leq |z|^2 \exp(|z|).$$

Proof) i) Choose any $x \in \mathbb{R}$, and let $n \in N_+$. Since

$$\frac{\partial}{\partial s}\left(-\frac{1}{n+1}(x-s)^{n+1}\exp(is)\right) = (x-s)^n \exp(is) - \frac{i}{n+1}(x-s)^{n+1}\exp(is)$$

on $\mathbb{R}$, we can see that

$$\int_0^x (x-s)^n \exp(is)ds - \frac{i}{n+1}\int_0^x (x-s)^{n+1}\exp(is)ds = \frac{1}{n+1}x^{n+1},$$

or equivalently,

$$\int_0^x (x-s)^n \exp(is)ds = \frac{1}{n+1}x^{n+1} + \frac{i}{n+1}\int_0^x (x-s)^{n+1}\exp(is)ds.$$

Putting $n = 1$ reveals that

$$i - i \cdot \exp(ix) = x + i \cdot \int_0^x (x-s)\exp(is)ds,$$

or that

$$\exp(ix) = 1 + ix - \int_0^x (x-s)\exp(is)ds.$$

The result for $n = 2$ now shows us that

$$\int_0^x (x-s)\exp(is)ds = \frac{x^2}{2} + \frac{i}{2}\int_0^x (x-s)^2 \exp(is)ds,$$

or that

$$\exp(ix) = 1 + ix - \frac{x^2}{2} - \frac{i}{2}\int_0^x (x-s)^2 \exp(is)ds$$

We therefore have the upper bound

$$\left|\exp(ix) - \left(1 + ix - \frac{x^2}{2}\right)\right| \leq \frac{1}{2}\left|\int_0^x (x-s)^2 \exp(is)ds\right|.$$

To obtain the first bound, note that, if $x \geq 0$, then

$$\left|\int_0^x (x-s)^2 \exp(is)ds\right| \leq \int_0^x (x-s)^2 ds = \frac{x^3}{3}.$$

On the other hand, if $x < 0$, then

$$\left|\int_0^x (x-s)^2 \exp(is)ds\right| \leq \int_x^0 (x-s)^2 ds = -\frac{x^3}{3},$$

so that

$$\left|\exp(ix) - \left(1 + ix - \frac{x^2}{2}\right)\right| \leq \frac{|x|^3}{6} \leq |x|^3.$$

It is slightly trickier to obtain the second bound. From the relationship

$$\int_0^x (x-s)\exp(is)ds = \frac{x^2}{2} + \frac{i}{2}\int_0^x (x-s)^2 \exp(is)ds,$$

we can see that

$$\frac{1}{2}\left|\int_0^x (x-s)^2 \exp(is)ds\right| = \left|\frac{i}{2}\int_0^x (x-s)^2 \exp(is)ds\right| \leq \left|\int_0^x (x-s)\exp(is)ds\right| + \frac{x^2}{2}.$$

If $x \geq 0$, then

$$\left|\int_0^x (x-s)\exp(is)ds\right| \leq \int_0^x |x-s|ds = \int_0^x (x-s)ds = \frac{x^2}{2},$$

while if $x < 0$, then

$$\left|\int_0^x (x-s)\exp(is)ds\right| \leq \int_x^0 |x-s|ds = \int_x^0 (s-x)ds = \frac{x^2}{2}.$$

Therefore,

$$\frac{1}{2}\left|\int_0^x (x-s)^2 \exp(is)ds\right| \leq \frac{x^2}{2} \leq x^2,$$

and we have

$$\left|\exp(ix) - \left(1 + ix - \frac{x^2}{2}\right)\right| \leq \min(|x|^3, |x|^2).$$

ii) Choose any $z \in \mathbb{C}$. Then,

$$\begin{aligned}\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} &= 1 + z + \sum_{n=2}^{\infty} \frac{z^n}{n!} \\ &= 1 + z + \frac{z^2}{\cdot} \left(\sum_{n=2}^{\infty} \frac{z^{n-2}}{n!} \right),\end{aligned}$$

so that

$$|\exp(z) - (1 + z)| \leq |z|^2 \cdot \left(\sum_{n=2}^{\infty} \frac{|z|^{n-2}}{n!} \right) \leq |z|^2 \cdot \left(\sum_{n=0}^{\infty} \frac{|z|^n}{n!} \right) = |z|^2 \exp(|z|).$$

Q.E.D.

We are now ready to present a CLT for martingale difference arrays:

Theorem (CLT for Martingale Difference Arrays)
Let $\{Z_{T,t}\}_{T\in N_+,1\leq t\leq k(T)}$ be a square integrable (or L^2) martingale difference array with respect to the filtration array $\mathcal{F}=\{\mathcal{F}_{T,t}\}_{1\leq t\leq k(T),T\in N_+}$. Define $\sigma^2_{T,t}=\mathbb{E}\left[Z^2_{T,t}\mid\mathcal{F}_{T,t-1}\right]$ for any $T\in N_+$ and $1\leq t\leq k(T)$, and let

$$V_T=\sum_{t=1}^{k(T)}\sigma^2_{T,t}$$

for any $T\in N_+$. Assume that:

i) $V_T\xrightarrow{p}1$ as $T\to\infty$.

ii) **(The Lindeberg Condition)** For any $\epsilon>0$,

$$\lim_{T\to\infty}\sum_{t=1}^{k(T)}\mathbb{E}\left[\left|Z^2_{T,t}\right|\cdot I_{\{|Z_{T,t}|>\epsilon\}}\right]=0.$$

Then, we have

$$\sum_{t=1}^{k(T)}Z_{T,t}\xrightarrow{d}N(0,1).$$

Proof) We proceed in small steps.

Part 1: Bounding V_T

We first modify the array $\{Z_{T,t}\}_{T\in N_+,1\leq t\leq k(T)}$ so that the sum of conditional variances V_T is bounded. Define $\{V_{T,t}\}_{T\in N_+,1\leq t\leq k(T)}$ as

$$V_{T,t}=\sum_{s=1}^{t}\sigma^2_{T,s}$$

for any $T\in N_+$, $1\leq t\leq k(T)$, so that $V_{T,T}=V_T$, and let $\{Y_{T,t}\}_{T\in N_+,1\leq t\leq k(T)}$ be defined as

$$Y_{T,t}=Z_{T,t}\cdot I_{\{V_{T,t}\leq 2\}}$$

for any $T\in N_+$ and $1\leq t\leq k(T)$. We now establish some properties of $\{Y_{T,t}\}_{T\in N_+,1\leq t\leq k(T)}$:

- **Martingale Difference Array**
  It is clear that $\{Y_{T,t}\}_{T\in N_+,1\leq t\leq k(T)}$ is a martingale difference array with respect to $\mathcal{F}$; for any $T\in N_+$ and $1\leq t\leq k(T)$, since $V_{T,t}$ is $\mathcal{F}_{T,t-1}$-measurable, $Y_{T,t}$ is

clearly $\mathcal{F}_{T,t}$ measurable, integrable due to the integrability of $Z_{T,t}$, and

$$\mathbb{E}\left[Y_{T,t} \mid \mathcal{F}_{T,t-1}\right] = \mathbb{E}\left[Z_{T,t} \mid \mathcal{F}_{T,t-1}\right] \cdot I_{\{V_{T,t} \leq 2\}} = 0.$$

– **Square Integrable**
By the square integrability of $Z_{T,t}$, each $Y_{T,t}$ is also square integrable. Furthermore, defining

$$\Gamma_T = \sum_{t=1}^{T} \underbrace{\mathbb{E}\left[Y_{T,t}^2 \mid \mathcal{F}_{T,t-1}\right]}_{\rho_{T,t}^2},$$

since

$$\mathbb{E}\left[Y_{T,t}^2 \mid \mathcal{F}_{T,t-1}\right] = \mathbb{E}\left[Z_{T,t}^2 \mid \mathcal{F}_{T,t-1}\right] \cdot I_{\{V_{T,t} \leq 2\}} = \sigma_{T,t}^2 \cdot I_{\{V_{T,t} \leq 2\}},$$

we can see that

$$\Gamma_T = \sum_{t=1}^{T} \sigma_{T,t}^2 \cdot I_{\{V_{T,t} \leq 2\}}.$$

This implies that $\Gamma_T \leq 2$ on Ω.
Analgously to $V_{T,t}$, we define

$$\Gamma_{T,t} = \sum_{s=1}^{t} \rho_{T,s}^2$$

for any $T \in N_+$ and $1 \leq t \leq k(T)$.

– **Convergence of Γ_T**
Moreover,

$$\begin{aligned} |V_T - \Gamma_T| = V_T - \Gamma_T &= \sum_{t=1}^{T} \sigma_{T,t}^2 \cdot I_{\{V_{T,t} > 2\}} \\ &\leq \left(\sum_{t=1}^{T} \sigma_{T,t}^2\right) \cdot I_{\{V_T > 2\}} = V_T \cdot I_{\{V_T > 2\}}. \end{aligned}$$

By assumption, $V_T \xrightarrow{p} 1$, and for any $\delta > 0$,

$$\mathbb{P}\left(V_T \cdot I_{\{V_T > 2\}} > \delta\right) \leq \mathbb{P}\left(V_T > 2\right) \leq \mathbb{P}\left(|V_T - 1| > 1\right),$$

so that

$$\lim_{T \to \infty} \mathbb{P}\left(V_T \cdot I_{\{V_T > 2\}} > \delta\right) = 0.$$

This holds for any $\delta > 0$, so

$$V_T \cdot I_{\{V_T > 2\}} \xrightarrow{p} 0$$

and we have

$$\Gamma_T - V_T \xrightarrow{p} 0$$

as $T \to \infty$ as well. Therefore, we can conclude that

$$\Gamma_T \xrightarrow{p} 1.$$

– **The Lindeberg Condition**
For any $T \in N_+$ and $\epsilon > 0$,

$$Y_{T,t} \leq Z_{T,t}$$

and thus

$$\mathbb{E}\left[Y_{T,t}^2 \cdot I_{\{|Y_{T,t}| > \epsilon\}}\right] \leq \mathbb{E}\left[Z_{T,t}^2 \cdot I_{\{|Z_{T,t}| > \epsilon\}}\right]$$

for $1 \leq t \leq k(T)$, so that

$$\sum_{t=1}^{k(T)} \mathbb{E}\left[Y_{T,t}^2 \cdot I_{\{|Y_{T,t}| > \epsilon\}}\right] \leq \sum_{t=1}^{k(T)} \mathbb{E}\left[Z_{T,t}^2 \cdot I_{\{|Z_{T,t}| > \epsilon\}}\right].$$

The right hand side goes to 0 as $T \to \infty$, so $\{Y_{T,t}\}_{T \in N_+, 1 \leq t \leq k(T)}$ satisfies the Lindeberg condition

$$\lim_{T \to \infty} \sum_{t=1}^{k(T)} \mathbb{E}\left[Y_{T,t}^2 \cdot I_{\{|Y_{T,t}| > \epsilon\}}\right] = 0.$$

We have thus shown that $\{Y_{T,t}\}_{T \in N_+, 1 \leq t \leq k(T)}$ possesses all the same properties as $\{Z_{T,t}\}_{T \in N_+, 1 \leq t \leq k(T)}$, with the added property that

$$\sum_{t=1}^{T} \mathbb{E}\left[Y_{T,t}^2 \mid \mathcal{F}_{T,t-1}\right] \leq 2$$

for any $T \in N_+$. Furthermore, since

$$\left|\sum_{t=1}^{k(T)} Z_{T,t} - \sum_{t=1}^{k(T)} Y_{T,t}\right| = \left|\sum_{t=1}^{k(T)} Z_{T,t} \cdot I_{\{V_{T,t} > 2\}}\right|$$

$$\leq \left(\sum_{t=1}^{k(T)} |Z_{T,t}| \right) \cdot I_{\{V_T > 2\}},$$

and

$$\mathbb{P}\left(\left(\sum_{t=1}^{k(T)} |Z_{T,t}| \right) \cdot I_{\{V_T > 2\}} > \delta \right) \leq \mathbb{P}\left(V_T > 2\right)$$

for any $\delta > 0$, we can see that

$$\sum_{t=1}^{k(T)} Z_{T,t} - \sum_{t=1}^{k(T)} Y_{T,t} \xrightarrow{p} 0.$$

Therefore, if we can show that

$$\sum_{t=1}^{k(T)} Y_{T,t} \xrightarrow{d} N(0,1),$$

then by Slutsky's theorem, we can prove the claim of the theorem.

Part 2: The Characteristic Function of $\sum_{t=1}^{k(T)} Y_{T,t}$

To show that the partial sums $\sum_{t=1}^{k(T)} Y_{T,t}$ converge in distribution to the standard normal distribution, we make use of the continuity theorem and show that their characteristic functions converge to that of the desired distribution. To that end, denote by φ_T the characteristic function of

$$S_T = \sum_{t=1}^{k(T)} Y_{T,t}$$

For any $r \in \mathbb{R}$,

$$\begin{aligned}
\left| \varphi_T(r) - \exp\left(-\frac{r^2}{2}\right) \right| &= \left| \mathbb{E}\left[\exp(ir \cdot S_T)\right] - \exp\left(-\frac{r^2}{2}\right) \right| \\
&\leq \left| \mathbb{E}\left[\exp(ir \cdot S_T)\right] - \mathbb{E}\left[\exp(ir \cdot S_T) \exp\left(\frac{r^2 \Gamma_T}{2}\right) \exp\left(-\frac{r^2}{2}\right) \right] \right| \\
&\quad + \left| \mathbb{E}\left[\exp(ir \cdot S_T) \exp\left(\frac{r^2 \Gamma_T}{2}\right) \exp\left(-\frac{r^2}{2}\right) \right] - \exp\left(-\frac{r^2}{2}\right) \right| \\
&\leq \mathbb{E}\left| 1 - \exp\left(\frac{r^2 \Gamma_T}{2}\right) \exp\left(-\frac{r^2}{2}\right) \right| + \left| \mathbb{E}\left[\exp(ir \cdot S_T) \exp\left(\frac{r^2 \Gamma_T}{2}\right) - 1 \right] \right|.
\end{aligned}$$

Because $\Gamma_T \xrightarrow{p} 1$, by the continuous mapping theorem

$$1 - \exp\left(\frac{r^2\Gamma_T}{2}\right)\exp\left(-\frac{r^2}{2}\right) \xrightarrow{p} 0.$$

Furthermore, for any $T \in N_+$, $0 \leq \Gamma_T \leq 2$ on Ω, so that the sequence

$$\left\{1 - \exp\left(\frac{r^2\Gamma_T}{2}\right)\exp\left(-\frac{r^2}{2}\right)\right\}_{T\in N_+}$$

is L^p-bounded for any $p \in [1, +\infty)$. By implication, the sequence is uniformly integrable, which, together with the convergence in probability result above, implies that

$$1 - \exp\left(\frac{r^2\Gamma_T}{2}\right)\exp\left(-\frac{r^2}{2}\right) \xrightarrow{L^1} 0,$$

or equivalently,

$$\mathbb{E}\left|1 - \exp\left(\frac{r^2\Gamma_T}{2}\right)\exp\left(-\frac{r^2}{2}\right)\right| \to 0$$

as $T \to \infty$.

Therefore, it remains to show that

$$\left|\mathbb{E}\left[\exp(ir \cdot S_T)\exp\left(\frac{r^2\Gamma_T}{2}\right) - 1\right]\right| \to 0$$

as $T \to \infty$ for the characteristic function of S_T to converge to that of the standard normal distribution as $T \to \infty$.

Part 3: Decomposing the Second Term

We first express the term

$$\left|\mathbb{E}\left[\exp(ir \cdot S_T)\exp\left(\frac{r^2\Gamma_T}{2}\right) - 1\right]\right|$$

as a telescoping sum, that is,

$$\begin{aligned}&\mathbb{E}\left[\exp(ir \cdot S_T)\exp\left(\frac{r^2\Gamma_T}{2}\right)\right] - 1\\ &\qquad = \sum_{t=1}^{k(T)} \mathbb{E}\left[\exp(ir \cdot S_{T,t})\exp\left(\frac{r^2\Gamma_{T,t}}{2}\right) - \exp(ir \cdot S_{T,t-1})\exp\left(\frac{r^2\Gamma_{T,t-1}}{2}\right)\right],\end{aligned}$$

where we define

$$S_{T,t} = \sum_{s=1}^{t} Y_{T,s}$$

for $1 \leq t \leq k(T)$. For any $1 \leq t \leq k(T)$, using the law of iterated expectations, we can see that

$$\begin{aligned}
&\mathbb{E}\left[\exp(ir \cdot S_{T,t})\exp\left(\frac{r^2\Gamma_{T,t}}{2}\right) - \exp(ir \cdot S_{T,t-1})\exp\left(\frac{r^2\Gamma_{T,t-1}}{2}\right)\right] \\
&\qquad = \mathbb{E}\left[\exp(ir \cdot S_{T,t-1})\exp\left(\frac{r^2\Gamma_{T,t}}{2}\right)\mathbb{E}\left[\exp(ir \cdot Y_{T,t}) - \exp\left(-\frac{r^2\rho_{T,t}^2}{2}\right)\Big|\mathcal{F}_{T,t-1}\right]\right].
\end{aligned}$$

Therefore, using the fact that $\Gamma_{T,t}$ is bounded above by 2 on Ω, we have

$$\begin{aligned}
&\left|\mathbb{E}\left[\exp(ir \cdot S_T)\exp\left(\frac{r^2\Gamma_T}{2}\right)\right] - 1\right| \\
&\qquad \leq \exp\left(r^2\right) \cdot \sum_{t=1}^{k(T)} \mathbb{E}\left|\mathbb{E}\left[\exp(ir \cdot Y_{T,t}) \mid \mathcal{F}_{T,t-1}\right] - \exp\left(-\frac{r^2\rho_{T,t}^2}{2}\right)\right|.
\end{aligned}$$

Part 4: Finding an Upper Bound for the Second Term

For any $1 \leq t \leq k(T)$, the previous lemma tells us that

$$\begin{aligned}
\left|\mathbb{E}\left[\exp(ir \cdot Y_{T,t}) \mid \mathcal{F}_{T,t-1}\right] - \left(1 - \frac{r^2\rho_{T,t}^2}{2}\right)\right| &\leq \mathbb{E}\left[\left|\exp(ir \cdot Y_{T,t}) - \left(1 + ir \cdot Y_{T,t} - \frac{r^2Y_{T,t}^2}{2}\right)\right|\Big|\mathcal{F}_{T,t-1}\right] \\
&\leq \mathbb{E}\left[\min(|rY_{T,t}|^3, |rY_{T,t}|^2) \mid \mathcal{F}_{T,t-1}\right].
\end{aligned}$$

It can now be seen that, for any $\epsilon > 0$,

$$\begin{aligned}
&\left|\mathbb{E}\left[\exp(ir \cdot Y_{T,t}) \mid \mathcal{F}_{T,t-1}\right] - \left(1 - \frac{r^2\sigma_{T,t}^2}{2}\right)\right| \\
&\qquad \leq \mathbb{E}\left[|rY_{T,t}|^3 \cdot I_{\{|Y_{T,t}| \leq \epsilon\}} \mid \mathcal{F}_{T,t-1}\right] + \mathbb{E}\left[|rY_{T,t}|^2 \cdot I_{\{|Y_{T,t}| > \epsilon\}} \mid \mathcal{F}_{T,t-1}\right] \\
&\qquad \leq \epsilon|r|^3 \cdot \mathbb{E}\left[Y_{T,t}^2 \cdot I_{\{|Y_{T,t}| \leq \epsilon\}} \mid \mathcal{F}_{T,t-1}\right] + r^2 \cdot \mathbb{E}\left[Y_{T,t}^2 \cdot I_{\{|Y_{T,t}| > \epsilon\}} \mid \mathcal{F}_{T,t-1}\right] \\
&\qquad \leq \epsilon|r|^3 \cdot \rho_{T,t}^2 + r^2 \cdot \mathbb{E}\left[Y_{T,t}^2 \cdot I_{\{|Y_{T,t}| > \epsilon\}} \mid \mathcal{F}_{T,t-1}\right].
\end{aligned}$$

Similarly, the second result in the previous lemma implies

$$\begin{aligned}\left|\exp\left(-\frac{r^2\rho_{T,t}^2}{2}\right)-\left(1-\frac{r^2\rho_{T,t}^2}{2}\right)\right| &\le \left|\frac{r^2\rho_{T,t}^2}{2}\right|^2\cdot\exp\left(\frac{r^2\rho_{T,t}^2}{2}\right)\\ &\le \frac{r^4}{4}\exp\left(r^2\right)\rho_{T,t}^2\cdot\left(\max_{1\le s\le k(T)}\rho_{T,s}^2\right),\end{aligned}$$

where we used the fact that

$$\rho_{T,t}^2 \le \Gamma_T \le 2.$$

By implication,

$$\begin{aligned}&\left|\mathbb{E}\left[\exp(ir\cdot Y_{T,t})\mid \mathcal{F}_{T,t-1}\right]-\exp\left(-\frac{r^2\rho_{T,t}^2}{2}\right)\right|\\ &\quad\le \epsilon|r|^3\cdot\rho_{T,t}^2+r^2\cdot\mathbb{E}\left[Y_{T,t}^2\cdot I_{\{|Y_{T,t}|>\epsilon\}}\mid \mathcal{F}_{T,t-1}\right]+\frac{r^4}{4}\exp\left(r^2\right)\cdot\rho_{T,t}^2\cdot\left(\max_{1\le s\le k(T)}\rho_{T,s}^2\right),\end{aligned}$$

and as such

$$\begin{aligned}&\left|\mathbb{E}\left[\exp(ir\cdot S_T)\exp\left(\frac{r^2\Gamma_T}{2}\right)\right]-1\right|\\ &\le \exp\left(r^2\right)\epsilon|r|^3\cdot\mathbb{E}\left[\Gamma_T\right]+\exp\left(r^2\right)r^2\cdot\sum_{t=1}^{k(T)}\mathbb{E}\left[Y_{T,t}^2\cdot I_{\{|Y_{T,t}|>\epsilon\}}\right]+\exp\left(2r^2\right)\frac{r^4}{4}\mathbb{E}\left[\Gamma_T\cdot\max_{1\le s\le k(T)}\rho_{T,s}^2\right]\\ &\le \underbrace{2\exp\left(r^2\right)|r|^3\cdot\epsilon}_{I}+\underbrace{\exp\left(r^2\right)r^2\cdot\sum_{t=1}^{k(T)}\mathbb{E}\left[Y_{T,t}^2\cdot I_{\{|Y_{T,t}|>\epsilon\}}\right]}_{II}+\underbrace{\exp\left(2r^2\right)\frac{r^4}{2}\mathbb{E}\left[\max_{1\le s\le k(T)}\rho_{T,s}^2\right]}_{III}.\end{aligned}$$

Part 5: The Convergence of the Second Term

The Lindeberg condition ensures that II converges to 0.

As for III, note that

$$\begin{aligned}\rho_{T,s}^2 &= \mathbb{E}\left[Y_{T,s}^2\cdot I_{\{|Y_{T,s}|\le\epsilon\}}\mid \mathcal{F}_{T,s-1}\right]+\mathbb{E}\left[Y_{Tst}^2\cdot I_{\{|Y_{T,s}|>\epsilon\}}\mid \mathcal{F}_{T,s-1}\right]\\ &\le \epsilon^2+\mathbb{E}\left[Y_{T,s}^2\cdot I_{\{|Y_{T,s}|>\epsilon\}}\mid \mathcal{F}_{T,s-1}\right]\\ &\le \epsilon^2+\sum_{t=1}^{k(T)}\mathbb{E}\left[Y_{T,t}^2\cdot I_{\{|Y_{T,t}|>\epsilon\}}\mid \mathcal{F}_{T,t-1}\right]\end{aligned}$$

for any $1 \leq s \leq k(T)$, so

$$\max_{1\leq s\leq k(T)} \rho_{T,s}^2 \leq \epsilon^2 + \sum_{t=1}^{k(T)} \mathbb{E}\left[Y_{T,t}^2 \cdot I_{\{|Y_{T,t}|>\epsilon\}} \mid \mathcal{F}_{T,t-1}\right].$$

It follows that

$$\mathbb{E}\left[\max_{1\leq s\leq k(T)} \rho_{T,s}^2\right] \leq \epsilon^2 + \sum_{t=1}^{k(T)} \mathbb{E}\left[Y_{T,t}^2 \cdot I_{\{|Y_{T,t}|>\epsilon\}}\right],$$

and by the Lindeberg condition,

$$\limsup_{T\to\infty} \mathbb{E}\left[\max_{1\leq s\leq k(T)} \rho_{T,s}^2\right] \leq \epsilon^2.$$

Therefore, the limit supremum of the term III is bounded above by $\exp(2r^2)\frac{r^4}{2}\cdot\epsilon^2$, so that

$$\limsup_{T\to\infty} \left|\mathbb{E}\left[\exp(ir\cdot S_T)\exp\left(\frac{r^2\Gamma_T}{2}\right)\right] - 1\right| \leq \left[2\exp\left(r^2\right)|r|^3\cdot\epsilon + \exp\left(2r^2\right)\frac{r^4}{2}\cdot\epsilon^2\right].$$

Since this holds for any $\epsilon > 0$, it follows that

$$\lim_{T\to\infty} \left|\mathbb{E}\left[\exp(ir\cdot S_T)\exp\left(\frac{r^2\Gamma_T}{2}\right)\right] - 1\right| = 0.$$

We have shown that

$$\lim_{T\to\infty} \left|\varphi_T(r) - \exp\left(-\frac{r^2}{2}\right)\right| = 0,$$

and because $r \in \mathbb{R}$ was chosen arbitrarily, by the continuity theorem we may conclude that

$$S_T \xrightarrow{d} N(0,1).$$

Q.E.D.

1.1.2 Martingale Difference Sequences

We now turn our attention to martingale difference sequences instead of arrays. A sequence $\{Y_t\}_{t\in\mathbb{Z}}$ of n-dimensional random vectors is said to be an n-dimensional martingale difference sequence (MDS) with respect to the filtration $\mathcal{F}=\{\mathcal{F}_t \mid t\in\mathbb{Z}\}$ if:

- Y_t is $\mathcal{F}_t$-measurable and integrable for any $t\in\mathbb{Z}$
- $\mathbb{E}[Y_t]=\mathbf{0}$ for any $t\in\mathbb{Z}$
- For any $t\in\mathbb{Z}$,

$$\mathbb{E}[Y_t \mid \mathcal{F}_{t-1}]=\mathbf{0}.$$

Given an n-dimensional MDS $\{Y_t\}_{t\in\mathbb{Z}}$, it can easily be seen that $\{\alpha' Y_t\}_{t\in\mathbb{Z}}$ is a univariate MDS for any $\alpha\in\mathbb{R}^n$. Furthermore, given an univariate MDS $\{y_t\}_{t\in\mathbb{Z}}$ with respect to the filtration $\mathcal{F}=\{\mathcal{F}_t \mid t\in\mathbb{Z}\}$, we can always define a martingale difference array by defining

$$Z_{T,t}=y_t \quad \text{and} \quad \mathcal{F}_{T,t}=\mathcal{F}_t$$

for any $T\in N_+$ and $1\le t\le T=k(T)$.

To obtain a workable version of the martingale difference array CLT for martingale difference sequences, we require the following law of large numbers, adapted from Andrews (1988).

Theorem (A Martingale WLLN)
Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional martingale difference sequence with respect to the filtration $\mathcal{F}=\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$ such that $\{|Y_t|^p \mid t\in\mathbb{Z}\}$ is uniformly integrable for some $1\le p\le 2$. Then,

$$\frac{1}{T}\sum_{t=1}^{T} Y_t \xrightarrow{L^p} \mathbf{0}.$$

Proof) Choose any $\epsilon>0$. By uniform integrability,

$$\lim_{b\to\infty}\sup_{t\in\mathbb{Z}}\mathbb{E}\left[|Y_t|^p\cdot I_{\{|Y_t|^p>b\}}\right]=0,$$

so there exists a $B>0$ such that

$$\sup_{t\in\mathbb{Z}}\mathbb{E}\left[|Y_t|^p\cdot I_{\{|Y_t|^p>B\}}\right]<\left(\frac{\epsilon}{4}\right)^p.$$

For any $t\in\mathbb{Z}$, define

$$e_t=Y_t\cdot I_{\{|Y_t|^p\le B\}} \quad \text{and}$$
$$u_t=Y_t\cdot I_{\{|Y_t|^p>B\}}.$$

Then, $Y_t = e_t + u_t$, and we have

$$\mathbb{E}[Y_t \mid \mathcal{F}_{t-1}] = \mathbf{0} = \mathbb{E}[e_t \mid \mathcal{F}_{t-1}] + \mathbb{E}[u_t \mid \mathcal{F}_{t-1}].$$

Furthermore, the sequence $\{e_t - \mathbb{E}[e_t \mid \mathcal{F}_{t-1}]\}_{t\in\mathbb{Z}}$ defines an n-dimensional MDS with respect to $\mathcal{F}$, since both e_t and $\mathbb{E}[e_t \mid \mathcal{F}_{t-1}]$ are integrable random vectors and

$$\mathbb{E}[e_t - \mathbb{E}[e_t \mid \mathcal{F}_{t-1}] \mid \mathcal{F}_{t-1}] = \mathbf{0}.$$

For any $T \in N_+$, we now have

$$\begin{aligned}\left\|\frac{1}{T}\sum_{t=1}^{T} Y_t\right\|_p &\leq \left\|\frac{1}{T}\sum_{t=1}^{T}(e_t - \mathbb{E}[e_t \mid \mathcal{F}_{t-1}])\right\|_p + \frac{1}{T}\sum_{t=1}^{T}\|u_t - \mathbb{E}[u_t \mid \mathcal{F}_{t-1}]\|_p \\ &\leq \left\|\frac{1}{T}\sum_{t=1}^{T}(e_t - \mathbb{E}[e_t \mid \mathcal{F}_{t-1}])\right\|_p + \frac{1}{T}\sum_{t=1}^{T}\left(\|u_t\|_p + \|\mathbb{E}[u_t \mid \mathcal{F}_{t-1}]\|_p\right)\end{aligned}$$

by Minkowski's inequality. Note that, for any random vector $X \in L^p(\mathcal{H},\mathbb{P})$, Jensen's inequality implies that

$$(\mathbb{E}|X|^p)^{\frac{2}{p}} \leq \mathbb{E}|X|^2,$$

so that $\|X\|_p \leq \|X\|_2$. Likewise, the conditional version of Jensen's inequality tells us that, for any $t \in \mathbb{Z}$,

$$\begin{aligned}\|\mathbb{E}[u_t \mid \mathcal{F}_{t-1}]\|_p &= (\mathbb{E}|\mathbb{E}[u_t \mid \mathcal{F}_{t-1}]|^p)^{\frac{1}{p}} \\ &\leq (\mathbb{E}|u_t|^p)^{\frac{1}{p}} = \|u_t\|_p.\end{aligned}$$

It follows that

$$\left\|\frac{1}{T}\sum_{t=1}^{T} Y_t\right\|_p \leq \left\|\frac{1}{T}\sum_{t=1}^{T}(e_t - \mathbb{E}[e_t \mid \mathcal{F}_{t-1}])\right\|_2 + \frac{2}{T}\sum_{t=1}^{T}\|u_t\|_p.$$

Since

$$u_t = Y_t \cdot I_{\{|Y_t|^p > B\}},$$

by assumption we have

$$\mathbb{E}|u_t|^p = \mathbb{E}\left[|Y_t|^p \cdot I_{\{|Y_t|^p > B\}}\right],$$

and as such

$$\sup_{t\in\mathbb{Z}} \|u_t\|_p \le \left(\sup_{t\in\mathbb{Z}} \mathbb{E}\left[|Y_t|^p \cdot I_{\{|Y_t|^p > B\}}\right]\right)^{\frac{1}{p}} < \frac{\epsilon}{4},$$

which implies

$$\left\|\frac{1}{T}\sum_{t=1}^{T} Y_t\right\|_p \le \left\|\frac{1}{T}\sum_{t=1}^{T}(e_t - \mathbb{E}[e_t \mid \mathcal{F}_{t-1}])\right\|_2 + \frac{\epsilon}{2}.$$

On the other hand, since martingale difference sequences are pairwise uncorrelated,

$$\begin{aligned}
\mathbb{E}\left|\frac{1}{T}\sum_{t=1}^{T}(e_t - \mathbb{E}[e_t \mid \mathcal{F}_{t-1}])\right|^2 &= \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}|e_t - \mathbb{E}[e_t \mid \mathcal{F}_{t-1}]|^2 \\
&\le \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}\left[(|e_t| + |\mathbb{E}[e_t \mid \mathcal{F}_{t-1}]|)^2\right] \\
&\le \frac{1}{T^2}\sum_{t=1}^{T}(\|e_t\|_2 + \|\mathbb{E}[e_t \mid \mathcal{F}_{t-1}]\|_2)^2 \\
&\qquad \text{(Minkowski's inequality)} \\
&\le \frac{4}{T^2}\sum_{t=1}^{T}\mathbb{E}|e_t|^2. \\
&\qquad \text{(Conditional version of Jensen's inequality)}
\end{aligned}$$

By definition,

$$\mathbb{E}|e_t|^2 = \mathbb{E}\left[|Y_t|^2 \cdot I_{\{|Y_t|^p \le B\}}\right] \le B^{\frac{2}{p}}\mathbb{P}(|Y_t|^p \le B) \le B^{\frac{2}{p}},$$

so we have

$$\left\|\frac{1}{T}\sum_{t=1}^{T} Y_t\right\|_p \le \frac{2B^{\frac{1}{p}}}{\sqrt{T}} + \frac{\epsilon}{2}.$$

Choose $N \in N_+$ so that $\frac{2B^{\frac{1}{p}}}{\sqrt{T}} < \frac{\epsilon}{2}$ for any $T \ge N$; this N depends on B and ϵ, and because our choice of B depends only on ϵ, so does N. We can now see that, for any $T \ge N$,

$$\left\|\frac{1}{T}\sum_{t=1}^{T} Y_t\right\|_p < \epsilon.$$

This holds for any $\epsilon > 0$, so by definition

$$\lim_{T\to\infty}\left\|\frac{1}{T}\sum_{t=1}^{T} Y_t\right\|_p = 0.$$

Q.E.D.

We now state and prove a CLT for (possibly multivariate) martingale difference sequences:

Theorem (CLT for Martingale Difference Sequences)
Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional martingale difference sequence with respect to the filtration $\mathcal{F} = \{\mathcal{F}_t\}_{t\in\mathbb{Z}}$. Suppose $\{Y_t\}_{t\in\mathbb{Z}}$ satisfies the following properties:

i) $\{|Y_t| \mid t \in \mathbb{Z}\}$ is L^p-bounded for some $p > 2$.

ii) There exists a positive definite matrix $Q \in \mathbb{R}^{n\times n}$ such that

$$\frac{1}{T}\sum_{t=1}^{T} Y_t Y_t' \xrightarrow{p} Q.$$

Then, as $T \to \infty$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Y_t \xrightarrow{p} N\left[\mathbf{0}, Q\right].$$

Proof) We make use of the Cramer-Wold device to show this result. Choose any non-zero $\alpha \in \mathbb{R}^n$, and define $Z_t = \alpha' Y_t$ for any $t \in \mathbb{Z}$. As stated earlier, $\{Z_t\}_{t\in\mathbb{Z}}$ is a univariate MDS with respect to $\mathcal{F}$ satisfying

$$\sigma_T^2 = \frac{1}{T}\sum_{t=1}^{T} Z_t^2 = \alpha'\left(\frac{1}{T}\sum_{t=1}^{T} Y_t Y_t'\right)\alpha \xrightarrow{p} \alpha' Q \alpha = \sigma^2.$$

Here, $\sigma^2 > 0$ becausae Q is positive definite and α is non-zero. Furthermore, $\{Z_t\}_{t\in\mathbb{Z}}$ is L^p-bounded, since

$$\mathbb{E}|Z_t|^p \leq |\alpha|^p \cdot \mathbb{E}|Y_t|^p$$

for any $t \in \mathbb{Z}$ by the Cauchy-Schwarz inequality.

Defining

$$Z_{T,t} = \frac{Z_t}{\sigma\sqrt{T}} \quad \text{and} \quad \mathcal{F}_{T,t} = \mathcal{F}_t$$

for any $T \in N_+$ and $1 \leq t \leq T = k(T)$, we obtain the martingale difference array $\{Z_{T,t}\}_{T\in N_+, 1\leq t\leq k(T)}$ with respect to the filtration array $\{\mathcal{F}_{T,t} \mid T \in N_+, 1 \leq t \leq k(T)\}$. This martingale difference array is clearly square integrable, due to the L^p-boundedness of $\{Z_t\}_{t\in\mathbb{Z}}$ and the fact that $p > 2$. We now verify the conditions of the CLT for martingale difference arrays:

– **Convergence of Sum of Variances**

For any $T \in N_+$, define

$$V_T = \sum_{t=1}^{T} \mathbb{E}\left[Z_{T,t}^2 \mid \mathcal{F}_{T,t-1}\right] = \frac{1}{\sigma^2 \cdot T} \sum_{t=1}^{T} \mathbb{E}\left[Z_t^2 \mid \mathcal{F}_{t-1}\right].$$

We saw earlier that

$$\sigma_T^2 = \frac{1}{T} \sum_{t=1}^{T} Z_t^2 \xrightarrow{p} \sigma^2.$$

If we can show that $\frac{\sigma_T^2}{\sigma^2} - V_T \xrightarrow{p} 0$, then we will obtain the desired result $V_T \xrightarrow{p} 1$.

To this end, define

$$x_t = Z_t^2 - \mathbb{E}\left[Z_t^2 \mid \mathcal{F}_{t-1}\right]$$

for any $t \in \mathbb{Z}$. $\{x_t\}_{t\in\mathbb{Z}}$ defines a martingale difference sequence with respect to the filtration $\mathcal{F}$, since each x_t is clearly $\mathcal{F}_t$-measurable, integrable with mean 0, and

$$\mathbb{E}\left[x_t \mid \mathcal{F}_{t-1}\right] = \mathbb{E}\left[Z_t^2 \mid \mathcal{F}_{t-1}\right] - \mathbb{E}\left[Z_t^2 \mid \mathcal{F}_{t-1}\right] = 0.$$

We noted above that $\{Z_t\}_{t\in\mathbb{Z}}$ was L^p-bounded; because $p > 2$, we can see that

$$\mathbb{E}|Z_t|^p = \mathbb{E}\left|Z_t^2\right|^{\frac{p}{2}}$$

for any $t \in \mathbb{Z}$, which tells us that $\{Z_t^2\}_{t\in\mathbb{Z}}$ is $L^{\frac{p}{2}}$-bounded, where $\frac{p}{2} > 1$. By implication, it is uniformly integrable, which implies that $\{x_t\}_{t\in\mathbb{Z}}$ is also a uniformly integrable martingale difference sequence. By the martingale WLLN proved earlier,

$$\frac{1}{T} \sum_{t=1}^{T} x_t \xrightarrow{L^1} 0,$$

from which it can be inferred that

$$\sigma_T^2 - \sigma^2 V_T = \frac{1}{T} \sum_{t=1}^{T} \left(Z_t - \mathbb{E}\left[Z_t^2 \mid \mathcal{F}_{t-1}\right]\right) = \frac{1}{T} \sum_{t=1}^{T} x_t \xrightarrow{p} 0.$$

Therefore,

$$V_T \xrightarrow{p} 1.$$

– **The Lindeberg Condition**

We can also show that $\{Z_{T,t}\}_{T \in N_+, 1 \le t \le k(T)}$ satisfies the Lindeberg condition. By the L^p-boundedness of $\{Z_t\}_{t \in \mathbb{Z}}$, there exists an $M < +\infty$ such that

$$\mathbb{E}|Z_t|^p < M$$

for any $t \in \mathbb{Z}$, which implies that

$$\sum_{t=1}^{k(T)} \mathbb{E}|Z_{T,t}|^p = \frac{1}{\sigma^p T^{\frac{p}{2}}} \sum_{t=1}^{k(T)} \mathbb{E}|Z_t|^p \le \sigma^{-p} T^{1-\frac{p}{2}}.$$

Since $\frac{p}{2} > 1$, taking $T \to \infty$ on both sides yields

$$\lim_{T \to \infty} \sum_{t=1}^{k(T)} \mathbb{E}|Z_{T,t}|^p = 0,$$

which is actually equivalent to Lyapunov's condition.

It now remains to show that Lyapunov's condition implies Lindeberg's. For any $\epsilon > 0$, if $|Z_{T,t}| > \epsilon$, then

$$|Z_{T,t}|^p = |Z_{T,t}|^2 \cdot |Z_{T,t}|^{p-2} > |Z_{T,t}|^2 \cdot \epsilon^{p-2},$$

since $p - 2 > 0$; this means that

$$\epsilon^{p-2} \cdot |Z_{T,t}|^2 \cdot I_{\{|Z_{T,t}| > \epsilon\}} \le |Z_{T,t}|^p \cdot I_{\{|Z_{T,t}| > \epsilon\}} \le |Z_{T,t}|^p.$$

Therefore,

$$\sum_{t=1}^{k(T)} \mathbb{E}\left[|Z_{T,t}|^2 \cdot I_{\{|Z_{T,t}| > \epsilon\}}\right] \le \epsilon^{2-p} \cdot \sum_{t=1}^{k(T)} \mathbb{E}|Z_{T,t}|^p,$$

so taking $T \to \infty$ on both sides yields

$$\lim_{T \to \infty} \sum_{t=1}^{k(T)} \mathbb{E}\left[|Z_{T,t}|^2 \cdot I_{\{|Z_{T,t}| > \epsilon\}}\right] = 0.$$

We have thus seen that the two conditions in the CLT for martingale difference arrays are satisfied. As per that theorem, then, we can conclude that

$$\frac{1}{\sigma\sqrt{T}} \sum_{t=1}^{T} Z_t = \sum_{t=1}^{k(T)} Z_{T,t} \xrightarrow{d} N(0,1).$$

By Slutsky's theorem, this then implies that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t \xrightarrow{d} N(0,\sigma^2),$$

or equivalently, for some n-dimensional normally distributed random vector Z with variance Q,

$$\alpha'\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Y_t\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t \xrightarrow{d} \alpha' Z.$$

Thish holds for any non-zero $\alpha \in \mathbb{R}^n$, so by the Cramer-Wold device,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Y_t \xrightarrow{d} Z \sim N\left[\mathbf{0}, Q\right].$$

Q.E.D.

1.2 Asymptotic Theory for Stationary Processes

The previous section derived asymptotic results for martingale difference sequences, which are fundamentally dependent but uncorrelated sequences. In time series analysis, we must also often deal with processes that are correlated, so we present here some asymptotic results for stationary processes that are possibly serially correlated.

1.2.1 Stationary Processes

An n-dimensional process $\{Y_t\}_{t\in\mathbb{Z}}$ is said to be a strictly stationary process if, for any $t \in \mathbb{Z}$ and $0 \le \tau_1 < \cdots < \tau_k$, the distribution of

$$(Y_{t+\tau_1}, \cdots, Y_{t+\tau_k})$$

does not depend on t. This implies that any strictly stationary process $\{Y_t\}_{t\in\mathbb{Z}}$ is identically distributed, since we can take $\tau_1 = 0$ and see that the distribution of $Y_t = Y_{t+\tau_1}$ is the same across all t. Compared to an i.i.d. process, we are strengthening the identical distribution condition to compensate for relaxing the independence condition.

There is a weaker form of stationarity that is frequently observed. Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional process. We say that it is weakly stationary if:

- **Mean Stationarity**
  $\{Y_t\}_{t\in\mathbb{Z}}$ is an L^1 process, and there exists a $\mu \in \mathbb{R}^n$ such that $\mathbb{E}[Y_t] = \mu$ for any $t \in \mathbb{Z}$

- **Covariance Stationarity**
  $\{Y_t\}_{t\in\mathbb{Z}}$ is an L^2 process, and there exists a function $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$ such that, for any $t, \tau \in \mathbb{Z}$, we have

$$\mathrm{Cov}[Y_t, Y_{t-\tau}] = \mathbb{E}\left[(Y_t - \mathbb{E}[Y_t])(Y_{t-\tau} - \mathbb{E}[Y_{t-\tau}])'\right] = \Gamma(\tau).$$

The function Γ is called the autocovariance function of $\{Y_t\}_{t\in\mathbb{Z}}$, and each $\Gamma(\tau)$ the autocovariances of the process. Note that, for any $\tau \in \mathbb{Z}$,

$$\Gamma(\tau) = \mathbb{E}[(Y_t - \mu)(Y_{t-\tau} - \mu)'] = (\mathbb{E}[(Y_{t-\tau} - \mu)(Y_t - \mu)'])' = \Gamma(-\tau)',$$

that is, $\Gamma(-\tau) = \Gamma(\tau)'$. We usually assume that the variance, $\Gamma(0)$, of $\{Y_t\}_{t\in\mathbb{Z}}$ is positive definite.

A weakly stationary process that we often encounter are white noise processes, which are essentially pairwise uncorrelated processes with mean zero. An n-dimensional process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is said to be an n-dimensional white noise process if it has mean 0 and

$$\mathbb{E}\left[\varepsilon_t \varepsilon_{t-\tau}'\right] = \begin{cases} \Sigma & \text{if } \tau = 0 \\ O & \text{otherwise} \end{cases}$$

for any $t, \tau \in \mathbb{Z}$ and some $\Sigma \in \mathbb{R}^{n \times n}$. It is trivially weakly stationary.

1.2.2 m-dependent Processes

Given a dependent process $\{Y_t\}_{t\in\mathbb{Z}}$, we want to find conditions under which it satisfies a form of the CLT. One such condition is stationarity; even though it is not independent, $\{Y_t\}_{t\in\mathbb{Z}}$ should retain some form of the identical distribution property. A second is limited dependence, that is, the degree to which any two entries in $\{Y_t\}_{t\in\mathbb{Z}}$ are dependent must fall to 0 as the number of observatitons between the entries increases. We will investigate one form of limited dependence, m-dependence, in this section.

An n-dimensional process $\{Y_t\}_{t\in\mathbb{Z}}$ is said to be m-dependent for some $m \in N_+$ if, for any $t \in \mathbb{Z}$, the collections $\{Y_s \mid s \leq t\}$ and $\{Y_s \mid s > t+m\}$ are independent. In other words, if $\{Y_t\}_{t\in\mathbb{Z}}$ is an m-dependent process, then any set of variables in $\{Y_t\}_{t\in\mathbb{Z}}$ that are more than m periods apart are independent.

In most applications, we wish to relax the assumption that observations more than m periods apart are independent, and just assume uncorrelatedness. One way to do this is through the concept of weak m-dependence, which is a generalization of martingale difference sequences. An n-dimensional process $\{Y_t\}_{t\in\mathbb{Z}}$ is said to be weakly m-dependent if

- $\{Y_t\}_{t\in\mathbb{Z}}$ is integrable with mean zero
- Letting $\mathcal{F}$ be the filtration generated by $\{Y_t\}_{t\in\mathbb{Z}}$,

  $$\mathbb{E}\left[Y_t \mid \mathcal{F}_{t-m}\right] = \mathbf{0}$$

  for any $t \in \mathbb{Z}$.

This is a weaker form of m-dependence because any two observations of a weakly m-dependent process that are at least m periods apart are uncorrelated; to see this, note that, for any $k \geq m$,

$$\mathbb{E}\left[Y_t Y_{t-k}'\right] = \mathbb{E}\left[\mathbb{E}\left[Y_t \mid \mathcal{F}_{t-m}\right] \cdot Y_{t-k}'\right] = \mathbf{0},$$

where the first equality follows because Y_{t-k} is $\mathcal{F}_{t-m}$-measurable. Note that any m-dependent process $\{Y_t\}_{t\in\mathbb{Z}}$ is also weakly m-dependent.

We say that the n-dimensional process $\{Y_t\}_{t\in\mathbb{Z}}$ is second-order weakly m-dependent if

- $\{Y_t\}_{t\in\mathbb{Z}}$ is square integrable
- For any $\tau \in N_+$, the elements of $\{Y_t Y_{t-\tau} - \mathbb{E}\left[Y_t Y'_{t-\tau}\right]\}_{t\in\mathbb{Z}}$ that are more than $m+\tau$-periods apart are uncorrelated.

Again, m-dependence implies second order weak m-dependence. To see this, let $\{Y_t\}_{t\in\mathbb{Z}}$ be a square integrable and m-dependent process. Then the elements of the sequence $\{Y_t Y'_{t-\tau} - \mathbb{E}\left[Y_t Y'_{t-\tau}\right]\}_{t\in\mathbb{Z}}$ are $m+\tau$-dependent for any $\tau \in \mathbb{N}$; this is because, for any $t \in \mathbb{Z}$, the elements of $Y_t, Y_{t-\tau}$ are independent of those of Y_{t+k}, Y_{t+k} by m-dependence for any $k \geq m+\tau$.

We refer to weak m-dependence as first order weak m-dependence, to differentiate it from second-order m-dependence.

1.2.3 CLT for Stationary and m-dependent Processes

The main goal in this section is to show that any weakly stationary process $\{Y_t\}_{t\in\mathbb{Z}}$ that is weakly m-dependent in both the first and second orders satisfies a version of the CLT. Note that this includes, as a special case, the CLT for weakly stationary and m-dependent processes.

We first require a lemma before continuing. The setting is given as follows: we have a double sequence $\{Y_{Tk}\}_{T,k\in N_+}$ such that $\{Y_{Tk}\}_{T\in N_+}$ converges weakly to some Y_k for any $k \in N_+$. We also know that $\{Y_k\}_{k\in N_+}$ converges weakly to some Y. Graphically, we have

$$\begin{array}{ccccc} Y_{11} & Y_{12} & Y_{13} & \cdots & \\ Y_{21} & Y_{22} & Y_{23} & \cdots & \\ Y_{31} & Y_{32} & Y_{33} & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \\ \downarrow & \downarrow & \downarrow & \cdots & \\ Y_1 & Y_2 & Y_3 & \rightarrow & Y \end{array}$$

Intuitively, a diagonal argument of sorts would seem to suggest that $\{Y_{kk}\}_{k\in N_+}$ converges weakly to Y. The lemma confirms this intuition by stating that, if there exists a process $\{X_T\}_{T\in N_+}$ that is diagonal to $\{Y_{Tk}\}_{T\in N_+}$ at the limit, then $\{X_T\}_{T\in N_+}$ should converge weakly to Y. The specific definition of "diagonal at the limit" is given as

$$\lim_{k\to\infty}\left(\limsup_{T\to\infty}\mathbb{P}\left(|X_T - Y_{Tk}| > \epsilon\right)\right) = 0$$

for any $\epsilon > 0$. Heuristically, this tells us that $X_T - Y_{Tk} \xrightarrow{p} 0$ when the limits on T and k are taken sequentially, so that, for large T, X_T represents an element of the sequence $\{Y_{Tk}\}_{k\in N_+}$.

The formal statement and proof are given below:

Lemma (Diagonal Argument for Weak Convergence)
Let $\{Y_{Tk}\}_{T,k\in N_+}$ be a sequence of n-dimensional random vectors such that:

i) For any $k \in N_+$, there exists a random vector Y_k such that $Y_{Tk} \xrightarrow{d} Y_k$ as $T \to \infty$.

ii) There exists a random vector Y such that $Y_k \xrightarrow{d} Y$ as $k \to \infty$.

iii) There exists a process $\{X_T\}_{T\in N_+}$ of n-dimensional random vectors such that, for any $\epsilon > 0$,

$$\lim_{k\to\infty}\left(\limsup_{T\to\infty}\mathbb{P}(|X_T - Y_{Tk}| > \epsilon)\right) = 0.$$

Then, $X_T \xrightarrow{d} Y$ as $T \to \infty$.

Proof) By the continuity theorem, we can reduce weak convergence to the pointwise convergence of the corresponding characteristic functions. For any n-dimensional random vector X, we denote by $\varphi_X : \mathbb{R}^n \to \mathbb{C}$ the characteristic function of X. The assumptions above then tell us that, for any $r \in \mathbb{R}^n$,

$$\lim_{T\to\infty}\varphi_{Y_{Tk}}(r) = \varphi_{Y_k}(r) \quad \text{for any } k \in N_+$$
$$\lim_{k\to\infty}\varepsilon_{Y_k}(r) = \varepsilon_Y(r).$$

Note that, for any $r \in \mathbb{R}^n$,

$$|\varphi_{X_T}(r) - \varphi_Y(r)| \leq |\varphi_{X_T}(r) - \varphi_{Y_{Tk}}(r)| + |\varphi_{Y_{Tk}}(r) - \varphi_{Y_k}(r)| + |\varphi_{Y_k}(r) - \varphi_Y(r)|.$$

Thus, if we can show that

$$\lim_{k\to\infty}\left(\limsup_{T\to\infty}|\varphi_{X_T}(r) - \varphi_{Y_{Tk}}(r)|\right) = 0,$$

then taking $T \to \infty$ and $k \to \infty$ successively to the above inequality implies that

$$\limsup_{T\to\infty}|\varphi_{X_T}(r) - \varphi_Y(r)| = 0,$$

which would imply that $\varphi_{X_T}(r) \to \varphi_Y(r)$ and thus complete the proof.

We now use the third assumption above to show that $\lim_{k\to\infty}(\limsup_{T\to\infty}|\varphi_{X_T}(r) - \varphi_{Y_{Tk}}(r)|) = 0$. By definition, for any $r \in \mathbb{R}^n$ and $T, k \in N_+$,

$$\begin{aligned}|\varphi_{X_T}(r) - \varphi_{Y_{Tk}}(r)| &= |\mathbb{E}[\exp(i\cdot r'X_T)] - \mathbb{E}[\exp(i\cdot r'Y_{Tk})]| \\ &\leq \mathbb{E}|\exp(i\cdot r'X_T) - \exp(i\cdot r'Y_{Tk})|.\end{aligned}$$

Choose any $\epsilon > 0$. By the uniform continuity of the mapping $x \mapsto \exp(i \cdot r'x)$ on $\mathbb{R}^n$, there exists a $\delta > 0$ such that

$$\left|\exp(i \cdot r'x) - \exp(i \cdot r'y)\right| < \epsilon$$

for any $x, y \in \mathbb{R}^n$ such that $|x - y| \leq \delta$. Therefore,

$$\begin{aligned}\left|\exp(i \cdot r'X_T) - \exp(i \cdot r'Y_{Tk})\right| &= \left|\exp(i \cdot r'X_T) - \exp(i \cdot r'Y_{Tk})\right| \cdot I_{\{|X_T - Y_{Tk}| > \delta\}} \\ &\quad + \left|\exp(i \cdot r'X_T) - \exp(i \cdot r'Y_{Tk})\right| \cdot I_{\{|X_T - Y_{Tk}| \leq \delta\}} \\ &\leq \left|\exp(i \cdot r'X_T) - \exp(i \cdot r'Y_{Tk})\right| \cdot I_{\{|X_T - Y_{Tk}| > \delta\}} + \epsilon \cdot I_{\{|X_T - Y_{Tk}| \leq \delta\}} \\ &\leq 2 \cdot I_{\{|X_T - Y_{Tk}| > \delta\}} + \epsilon \cdot I_{\{|X_T - Y_{Tk}| \leq \delta\}},\end{aligned}$$

which implies that

$$\begin{aligned}\mathbb{E}\left|\exp(i \cdot r'X_T) - \exp(i \cdot r'Y_{Tk})\right| &\leq 2 \cdot \mathbb{P}\left(|X_T - Y_{Tk}| > \delta\right) + \epsilon \cdot \mathbb{P}\left(|X_T - Y_{Tk}| \leq \delta\right) \\ &\leq 2 \cdot \mathbb{P}\left(|X_T - Y_{Tk}| > \delta\right) + \epsilon.\end{aligned}$$

Due to the assumption that

$$\lim_{k \to \infty} \left(\limsup_{T \to \infty} \mathbb{P}\left(|X_T - Y_{Tk}| > \delta\right) \right) = 0,$$

taking $T \to \infty$ and $k \to \infty$ successively yields

$$\limsup_{k \to \infty} \left(\limsup_{T \to \infty} |\varphi_{X_T}(r) - \varphi_{Y_{Tk}}(r)| \right) \leq \epsilon.$$

This holds for any $\epsilon > 0$, so we have

$$\lim_{k \to \infty} \left(\limsup_{T \to \infty} |\varphi_{X_T}(r) - \varphi_{Y_{Tk}}(r)| \right) = 0,$$

which completes the proof.

Q.E.D.

We can now present the CLT for stationary m-dependent processes. The basic idea of the proof is simple. Given a stationary m-dependent process $\{Y_t\}_{t \in \mathbb{Z}}$, we can divide the partial sum $\sum_{t=1}^{T} Y_t$ into two blocks. The first block is the sum of the collections of $k > m$ consecutive elements of $\{Y_t\}_{t \in \mathbb{Z}}$ separated by m observations, while the second collects the sum of the observations that separate the entries in the first block. Since the entries of the first blocks are independent, we can apply the CLT for iid processes to show that it converges to a normal distribution, while the entries in the second block comprise m observations each and are also independent ($k > m$), so that we can show that it converges to 0 using Chebyshev's inequality.

The formal statement and proof are given below:

Theorem (CLT for Stationary m-dependent Processes)
Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a mean zero L^4-bounded n-dimensional process that is weakly stationary and weakly m-dependent in both the first and second orders. Let $\{Y_t\}_{t\in\mathbb{Z}}$ have the autocovariance function $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$, and assume that the sum of the autocovariances is positive definite. Then,

$$\sqrt{T}\cdot \bar{Y}_T := \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Y_t \xrightarrow{d} N\left[\mathbf{0},\ \sum_{j=-m}^{m}\Gamma(j)\right].$$

Proof) Define $V = \sum_{j=-m}^{m}\Gamma(j)$. For any $k \in N_+$ such that $k > m$, define the process $\{A_{k,i}\}_{i\in\mathbb{Z}}$ and $\{B_{k,i}\}_{i\in\mathbb{Z}}$ as

$$\begin{aligned} A_{k,i} &= Y_{(i-1)(k+m)+1} + \cdots + Y_{ik+(i-1)m} \\ B_{k,i} &= Y_{ik+(i-1)m+1} + \cdots + Y_{i(k+m)}. \end{aligned}$$

for any $i \in \mathbb{Z}$. Note that the entries comprising $\{A_{k,i}\}_{i\in N_+}$ are separated by m observations, which are collected in $\{B_{k,i}\}_{i\in N_+}$.

We will decompose the partial sum of the Y_t into two blocks, the first one being the partial sum of the $A_{k,i}$ and the second of $B_{k,i}$. Afterward, we show that the first block converges in distribution while the second one converges in probability to 0. We now proceed in steps:

Step 1: Asymptotic Results for the First Block

Let $\mathcal{G}$ be the filtration generated by $\{Y_t\}_{t\in\mathbb{Z}}$ and $\mathcal{F}$ that generated by $\{A_{k,i}\}_{i\in\mathbb{Z}}$. By first order weak m-dependence, we have

$$\mathbb{E}\left[Y_t \mid \mathcal{F}_{t-m}\right] = \mathbf{0}$$

for any $t \in \mathbb{Z}$. We now show that $\{A_{k,i}\}_{i\in\mathbb{Z}}$ is an MDS with respect to $\mathcal{G}$. The integrability of $\{A_{k,i}\}_{i\in\mathbb{Z}}$ follows from that of $\{Y_t\}_{t\in\mathbb{Z}}$, and it is $\mathcal{G}$-adapted by definition. Furthermore, it has mean 0 for any $i \in \mathbb{Z}$ since each Y_t has mean zero. Finally, for any $i \in \mathbb{Z}$, since $\mathcal{G}_i \subset \mathcal{F}_{i(k+m)_j-m}$ for any $j \in \mathbb{N}$, we can see that

$$\begin{aligned} \mathbb{E}\left[A_{k,i+1} \mid \mathcal{G}_i\right] &= \sum_{j=1}^{m}\mathbb{E}\left[Y_{i(k+m)+j} \mid \mathcal{G}_i\right] \\ &= \sum_{j=1}^{m}\mathbb{E}\left[\mathbb{E}\left[Y_{i(k+m)+j} \mid \mathcal{F}_{i(k+m)+j-m}\right] \mid \mathcal{G}_i\right] = \mathbf{0}. \end{aligned}$$

Thus, by definition, $\{A_{k,i}\}_{i\in\mathbb{Z}}$ is an MDS with respect to $\mathcal{G}$.

The assumption that $\{Y_t\}_{t\in\mathbb{Z}}$ has bounded fourth moments implies that $\{A_{k,i}\}_{i\in\mathbb{Z}}$ does as well. In addition, since $\{Y_t\}_{t\in\mathbb{Z}}$ is second-order weak m-dependent, for any $\tau \in \mathbb{N}$ elements of the sequence $\{Y_t Y'_{t-\tau} - \Gamma(\tau)\}_{t\in\mathbb{Z}}$ that are more than $m+\tau$ periods apart are uncorrelated. Thus, for any $1 \le j \le k$ and $\tau \in \mathbb{N}$ chosen so that $j-\tau \ge 0$, the sequence $\{Y_{(i-1)(k+m)+j} Y_{(i-1)(k+m)+j-\tau} - \Gamma(\tau)\}_{i\in\mathbb{Z}}$ is pairwise uncorrelated; each adjacent pair of entries in the sequence are $k+m$ periods apart, where $k+m \ge m+\tau$.

Since $\{Y_{(i-1)(k+m)+j} Y_{(i-1)(k+m)+j-\tau} - \Gamma(\tau)\}_{i\in\mathbb{Z}}$ also has bounded second moments due to the boundedness of the fourth moments of $\{Y_t\}_{t\in\mathbb{Z}}$, by the WLLN for uncorrelated sequences with finite second moments, we have

$$\frac{1}{r}\sum_{i=1}^{r} Y_{(i-1)(k+m)+j} Y'_{(i-1)(k+m)+j-\tau} \xrightarrow{p} \Gamma(\tau)$$

as $r \to \infty$.

From the additivity of convergence in probability, it now follows that

$$\begin{aligned}
&\frac{1}{r}\sum_{i=1}^{r} A_{k,i}A'_{k,i} \\
&\qquad = \sum_{j=0}^{k}\frac{1}{r}\sum_{i=1}^{r} Y_{(i-1)(k+m)+j} Y'_{(i-1)(k+m)+j} \\
&\quad + \sum_{j=1}^{k}\left[\frac{1}{r}\sum_{i=1}^{r} Y_{(i-1)(k+m)+j} Y'_{(i-1)(k+m)+j-1} + \frac{1}{r}\sum_{i=1}^{r} Y_{(i-1)(k+m)+j-1} Y'_{(i-1)(k+m)+j}\right] \\
&\quad + \cdots + \left[\frac{1}{r}\sum_{i=1}^{r} Y_{(i-1)(k+m)+k} Y'_{(i-1)(k+m)+1} + \frac{1}{r}\sum_{i=1}^{r} Y_{(i-1)(k+m)+1} Y'_{(i-1)(k+m)+k}\right] \\
&\qquad \xrightarrow{p} k\Gamma(0) + \sum_{j=-m}^{m}(k-|j|)\Gamma(j),
\end{aligned}$$

where we used the fact that $\Gamma(\tau) = O$ for any $\tau > m$. Here, the right hand side is a positive semidefinite matrix because it equals the positive semidefinite matrix $\mathbb{E}\left[A_{k,i}A'_{k,i}\right]$. Defining $V_k \in \mathbb{R}^{n\times n}$ as

$$V_k = \frac{k}{k+m}\Gamma(0) + \sum_{j=1}^{m}\frac{k-|j|}{k+m}\left(\Gamma(j)+\Gamma(j)'\right),$$

Since $V_k \to V$ as $k \to \infty$, V_k is a positive definite function, and the determinant is a continuous function on the space of $n \times n$ matrices, for large enough k the determinant of V_k is positive. This indicates, by the positive semidefiniteness of V_k, that V_k is positive definite, and by extension that the probability limit of $\frac{1}{r}\sum_{i=1}^{r} A_{k,i}A'_{k,i}$ is positive definite.

So far, for large enough $k > m$, we have shown that $\{A_{k,i}\}_{i\in\mathbb{Z}}$ is an MDS with respect to $\mathcal{G}$ that:

- Has bounded fouth moments
- Satsifies

$$\frac{1}{r}\sum_{i=1}^{r} A_{k,i}A'_{k,i} \xrightarrow{p} k\Gamma(0) + \sum_{j=-m}^{m}(k-|j|)\Gamma(j)$$

  where $k\Gamma(0)+\sum_{j=-m}^{m}(k-|j|)\Gamma(j)$ is a positive definite $n \times n$ matrix.

Therefore, by the MDS CLT, we have

$$\frac{1}{\sqrt{r}}\sum_{i=1}^{r} A_{k,i} \xrightarrow{d} X_k,$$

as $r \to \infty$, where

$$X_k \sim N\left[\mathbf{0}, \quad k\Gamma(0) + \sum_{j=-m}^{m}(k-|j|)\Gamma(j)\right].$$

Step 2: Partitioning the Partial Sum of Y_t

Now choose any $T \in N_+$ such that $T \geq k+m$. Denoting $r = \lfloor \frac{T}{k+m} \rfloor \in N_+$, define X_{Tk} and Z_{Tk} as

$$\begin{aligned}
X_{Tk} &= \frac{1}{\sqrt{T}}\sum_{i=1}^{r} A_{k,i} \\
Z_{Tk} &= \frac{1}{\sqrt{T}}\sum_{i=1}^{r} B_{k,i} \\
C_{Tk} &= \frac{1}{\sqrt{T}}\left(Y_{r(k+m)+1} + \cdots + Y_T\right).
\end{aligned}$$

Since $r \leq \frac{T}{k+m} < r+1$, we have $r(k+m) \leq T$ and $T < r(k+m)+(k+m)$, so that $T - r(k+m) < k+m$. This indicates that C_{Tk} comprises at most $k+m-1$ entries.

By construction,

$$\sqrt{T}\cdot\bar{Y}_T = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Y_t = X_{Tk} + Z_{Tk} + C_{Tk}.$$

$r \to \infty$ as $T \to \infty$, so as $T \to \infty$,

$$\frac{1}{\sqrt{T}}\sum_{i=1}^{r} A_{k,i} = \frac{\sqrt{r}}{\sqrt{T}}\left(\frac{1}{\sqrt{r}}\sum_{i=1}^{r} A_{k,i}\right)$$

$$\xrightarrow{d} \frac{1}{\sqrt{k+m}} X_k \sim N\left[\mathbf{0}, V_k\right].$$

We already showed above that $V_k \to V$ as $k \to \infty$. Defining X as a random vector such that $X \sim N[\mathbf{0}, V]$, the characteristic function of $\frac{1}{\sqrt{k+m}} X_k$ converges to that of X as $k \to \infty$, which implies, by the continuity theorem, that $\frac{1}{\sqrt{k+m}} X_k \xrightarrow{d} X$ as $k \to \infty$.

Step 3: Convergence of the Second Block

We have so far shown that

$$X_{Tk} \xrightarrow{d} \frac{1}{\sqrt{k+m}} X_k \quad \text{as } T \to \infty$$
$$\frac{1}{\sqrt{k+m}} X_k \xrightarrow{d} X \quad \text{as } k \to \infty.$$

It remains to show that

$$\lim_{k \to \infty} \left(\limsup_{T \to \infty} \mathbb{P}\left(\left| \sqrt{T} \cdot \bar{Y}_T - X_{Tk} \right| > \epsilon \right) \right) = 0$$

for any $\epsilon > 0$ to be able to apply the previous lemma and conclude that $\sqrt{T} \cdot \bar{Y}_T \xrightarrow{d} X$. To this end, note that

$$\sqrt{T} \cdot \bar{Y}_T - X_{Tk} = Z_{Tk} + C_{Tk}.$$

Z_{Tk} is the sum of mean zero random vectors, so it also has mean zero. Furthermore, the entries in the process $\{B_{k,i}\}_{i \in \mathbb{Z}}$ are separated by $k > m$ observations, so by the weak m-dependence of $\{Y_t\}_{t \in \mathbb{Z}}$, $\{B_{k,i}\}_{i \in \mathbb{Z}}$ is pairwise uncorrelated. This implies that

$$\begin{aligned} \mathbb{E}|Z_{Tk}|^2 &= \mathbb{E}\left| \frac{1}{\sqrt{T}} \sum_{i=1}^{r} B_{k,i} \right|^2 \\ &= \frac{1}{T} \sum_{i=1}^{r} \operatorname{tr}\left(\operatorname{Var}\left[B_{k,i}\right] \right). \end{aligned}$$

The variance of each $B_{k,i}$ is identical due to weak stationarity, and equal to

$$\operatorname{Var}\left[B_{k,i}\right] = m\Gamma(0) + \sum_{j=1}^{m-1} (m - |j|) \left[\Gamma(j) + \Gamma(j)'\right].$$

It follows that, for any $\epsilon > 0$,

$$\mathbb{P}\left(|Z_{Tk}| > \frac{\epsilon}{2} \right) \leq \frac{4}{\epsilon^2} \mathbb{E}|Z_{Tk}|^2$$

$$= \frac{r}{T} \operatorname{tr}\left(m\Gamma(0) + \sum_{j=1}^{m-1} (m - |j|) \left[\Gamma(j) + \Gamma(j)' \right] \right),$$

so taking $T \to \infty$ on both sides yields

$$\limsup_{T\to\infty} \mathbb{P}\left(|Z_{Tk}| > \frac{\epsilon}{2} \right) \le \frac{4}{\epsilon^2} \cdot \frac{1}{k+m} \operatorname{tr}\left(m\Gamma(0) + \sum_{j=1}^{m-1} (m - |j|) \left[\Gamma(j) + \Gamma(j)' \right] \right).$$

By a similar line of reasoning, since there are never more than $k+m$ entries comprising C_{Tk}, taking $T \to \infty$ on both sides of

$$\mathbb{E}|C_{Tk}|^2 = \frac{1}{T} \operatorname{tr}\left(\operatorname{Var}\left[Y_{r(k+m)+1} + \cdots + Y_T \right] \right),$$

yields $C_{Tk} \xrightarrow{p} \mathbf{0}$. Putting the results together, we have

$$\begin{aligned} \limsup_{T\to\infty} \mathbb{P}\left(\left| \sqrt{T} \cdot \bar{Y}_T - X_{Tk} \right| > \epsilon \right) &\le \limsup_{T\to\infty} \mathbb{P}\left(|Z_{Tk}| > \frac{\epsilon}{2} \right) + \limsup_{T\to\infty} \mathbb{P}\left(|C_{Tk}| > \frac{\epsilon}{2} \right) \\ &\le \frac{4}{\epsilon^2} \cdot \frac{1}{k+m} \operatorname{tr}\left(m\Gamma(0) + \sum_{j=1}^{m-1} (m - |j|) \left[\Gamma(j) + \Gamma(j)' \right] \right). \end{aligned}$$

Therefore, taking $k \to \infty$ on both sides gives us the result

$$\lim_{k\to\infty} \left(\limsup_{T\to\infty} \mathbb{P}\left(\left| \sqrt{T} \cdot \bar{Y}_T - X_{Tk} \right| > \epsilon \right) \right) = 0.$$

This holds for any $\epsilon > 0$, so by the previous lemma,

$$\sqrt{T} \cdot \bar{Y}_T \xrightarrow{d} X \sim N\left[\mathbf{0}, V \right].$$

Q.E.D.

1.2.4 Application: GMM Estimation under Serially Correlated Errors

The CLT for m-dependent processes derived above is quite useful when studying GMM estimation under limited serial correlation. The exposition in this section is based on Cumby et al (1983).

The General Model

We consider the following general model. Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional process and Θ a convex and open subset of $\mathbb{R}^k$. For $L \geq k$, let $g : \mathbb{R}^n \times \Theta \to \mathbb{R}^L$ be a function such that, for any $x \in \mathbb{R}^n$, $g(x, \cdot)$ is differentiable. Letting θ_0 denote the true parameter value, suppose that the moment condition

$$\mathbb{E}\left[g(Y_t, \theta_0)\right] = \mathbf{0}$$

holds for any $t \in \mathbb{Z}$. Then, the GMM estimator of θ is found as the minimizer of the objective function

$$Q_T(\theta) = \left(\sum_{t=1}^{T} g(Y_t, \theta)\right)' W_T \left(\sum_{t=1}^{T} g(Y_t, \theta)\right),$$

where $\{W_T\}_{T\in N_+}$ is a sequence of $L \times L$ random matrices that converge in probability to some positive definite $W \in \mathbb{R}^{L\times L}$.

Suppose $\hat{\theta}_T$ is the GMM estimator of θ. Then, the first order condition for minimization tells us that

$$\frac{1}{2}\frac{\partial Q_T(\hat{\theta}_T)}{\partial \theta} = \left(\sum_{t=1}^{T} \frac{\partial g(Y_t, \hat{\theta}_T)}{\partial \theta'}\right)' W_T \left(\sum_{t=1}^{T} g(Y_t, \hat{\theta}_T)\right) = \mathbf{0}.$$

The asymptotic distribution of the GMM estimator (assuming that it exists in Θ) is derived on the basis of the following assumptions:

1) **Population Moment Condition**
   There exists a continuous function $g : \mathbb{R}^n \times \Theta \to \mathbb{R}^L$, continuously differentiable with respect to its second argument, such that, for any $t \in \mathbb{Z}$,

$$\mathbb{E}\left[g(Y_t, \theta_0)\right] = \mathbf{0}.$$

2) **Weighting Matrix**
   There exists a sequence $\{W_T\}_{T\in N_+}$ of positive definite $L \times L$ matrices that converges in

probability to the positive definite $L \times L$ matrix W.

3) **Consistency of GMM Estimator**
   Assume that the GMM estimator is consistent for θ_0, that is, $\hat{\theta}_T \xrightarrow{p} \theta_0$.

4) **Consistency of First Derivative**
   There exists a full rank matrix $G \in \mathbb{R}^{L \times k}$ such that, for any consistent estimator $\tilde{\theta}_T$ of θ_0,

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g(Y_t, \tilde{\theta}_T)}{\partial \theta'} \xrightarrow{p} G.$$

5) **CLT for Sample Moment Condition**
   The process $\{g(Y_t, \theta_0)\}_{t \in \mathbb{Z}}$ satisfies the CLT

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g(Y_t, \theta_0) \xrightarrow{d} N\left[\mathbf{0}, V\right]$$

   for some positive definite $L \times L$ matrix V.

Under these assumptions, we know that the GMM estimator has the following asymptotic distribution:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N\left[\mathbf{0}, (G'WG)^{-1}G'WVWG(G'WG)^{-1}\right],$$

where $V = \sum_{j=-m}^{m}\Gamma(j)$. This result utilizes the stochastic mean value theorem; for details, consult the document on that topic.

We also showed that the optimal weighting matrix is $W^* = V^{-1}$, in which case the asymptotic variance becomes

$$plim\left[\sqrt{T}(\hat{\theta}_T - \theta_0)\right] = (G'V^{-1}G)^{-1}.$$

2SLS Estimation

The preceding result can be applied to an instrumental variables regression framework as follows. Let $\{y_t\}_{t\in\mathbb{Z}}$ and $\{X_t\}_{t\in\mathbb{Z}}$ be a univariate and k-dimensional process, respectively, such that

$$y_t = X_t'\beta_0 + u_t$$

for any $t \in \mathbb{Z}$ and some $\beta_0 \in \mathbb{R}^k$, where the error process $\{u_t\}_{t\in\mathbb{Z}}$ is a mean zero weakly stationary m-dependent process with bounded fourth moments and autocovariance function $\gamma : \mathbb{Z} \to \mathbb{R}$. We let $\Theta = \mathbb{R}^k$ be the open and convex parameter space. Suppose that, while X_t and u_t are correlated, we instead have the moment conditions

$$\mathbb{E}[Z_t u_t] = \mathbf{0}$$

for some L-dimensional process $\{Z_t\}_{t\in\mathbb{Z}}$ where $L \geq k$.

To derive the 2SLS estimator of $\hat{\beta}_T$, we cast this regression model into a GMM framework. Define the $n = 1+k+L$ -dimensional process $\{Y_t\}_{t\in\mathbb{Z}}$ as

$$Y_t = (y_t, X_t', Z_t')'$$

for any $t \in \mathbb{Z}$, and the function $g : \mathbb{R}^n \times \Theta \to \mathbb{R}^L$ as

$$g((y,x,z),\beta) = z(y - x'\beta)$$

for any $(y,x,z) \in \mathbb{R}^n$ and $\beta \in \Theta$. g is a continuous function on $\mathbb{R}^n \times \Theta$ with derivative

$$\frac{\partial g((y,x,z),\beta)}{\partial \beta'} = -zx'$$

with respect to β. Note that each entry in $\frac{\partial g}{\partial \beta'}$ is continuous on $\mathbb{R}^n \times \Theta$, so that g is continuously differentiable with respect to β. The moment conditions can now be written as

$$\mathbb{E}[g(Y_t, \beta_0)] = \mathbf{0}.$$

Given a sequence $\{W_T\}_{T\in N_+}$ of $L \times L$ matrices that converges to some postive definite weight matrix W, the GMM objective function is defined as

$$\begin{aligned} Q_T(\beta) &= \left(\sum_{t=1}^{T} g(Y_t,\beta)\right)' W_T \left(\sum_{t=1}^{T} g(Y_t,\beta)\right) \\ &= \left(\sum_{t=1}^{T} Z_t(y_t - X_t'\beta)\right)' W_T \left(\sum_{t=1}^{T} Z_t(y_t - X_t'\beta)\right). \end{aligned}$$

Letting $\hat{\beta}_T$ be the GMM estimator of β, the first order condition for minimization tells us that

$$\frac{1}{2}\frac{\partial Q_T(\hat{\beta}_T)}{\partial \beta} = \left(\sum_{t=1}^{T} X_t Z_t'\right) W_T \left(\sum_{t=1}^{T} Z_t(y_t - X_t'\hat{\beta}_T)\right) = \mathbf{0},$$

or that

$$\begin{aligned}\hat{\beta}_T &= \left[\left(\sum_{t=1}^{T} X_t Z_t'\right) W_T \left(\sum_{t=1}^{T} Z_t X_t'\right)\right]^{-1} \left(\sum_{t=1}^{T} X_t Z_t'\right) W_T \left(\sum_{t=1}^{T} Z_t y_t\right) \\ &= \beta_0 + \left[\left(\sum_{t=1}^{T} X_t Z_t'\right) W_T \left(\sum_{t=1}^{T} Z_t X_t'\right)\right]^{-1} \left(\sum_{t=1}^{T} X_t Z_t'\right) W_T \left(\sum_{t=1}^{T} Z_t u_t\right).\end{aligned}$$

To derive the asymptotic properties of $\hat{\beta}_T$, we make the following assumptions:

1) **Backward-Looking Exogeneity**
   We strengthen the identification condition by assuming that u_t is independent of the current and past values of the instrument, that is, of $Z_t, Z_{t-1}, \cdots$.

2) **Relevance Condition**
   There exists a full rank matrix $Q_{zx} \in \mathbb{R}^{L\times k}$ such that

$$\frac{1}{T}\sum_{t=1}^{T} Z_t X_t' \xrightarrow{p} Q_{zx}.$$

   Similarly, there exists a positive definite matrix $Q_{zz} \in \mathbb{R}^{L\times L}$ such that

$$\frac{1}{T}\sum_{t=1}^{T} Z_t Z_t' \xrightarrow{p} Q_{zz}.$$

3) **Stationarity and Limited Dependence of Errors**
   The process $\{Z_t u_t\}_{t\in\mathbb{Z}}$ is a weakly stationary L^4-bounded and m-dependent process with autocovariance function $\Gamma : \mathbb{Z} \to \mathbb{R}^{L\times L}$.

The first two conditions are standard for GMM models. The third assumption is unique to a model with serially correlated errors. We take a brief moment to justify its inclusion.

Initially, the assumption that $\{Z_t u_t\}_{t\in\mathbb{Z}}$ is stationary is standard in the literature. Andrews (1991) points out that, if it is non-stationary, then the standard asymptotic results do not hold and we must make use of the asymptotic theory of unit root processes, which is detailed in a later section. Even heuristically, both the error process $\{u_t\}_{t\in\mathbb{Z}}$ and the instrument process $\{Z_t\}_{t\in\mathbb{Z}}$ are likely to be stationary if y_t is stationary and we use lagged values of stationary time series as our instruments. It follows naturally that their product, $\{Z_t u_t\}_{t\in\mathbb{Z}}$, should reasonably be stationary as well.

The more troublesome assumption is that of m-dependence. Realistically, a weaker form of limited dependence could be used, such as the strong mixing or mixingale assumption to be introduced below. However, m-dependence can be justified in a number of situations that arise naturally in economics; here we present a few.

- **Non-Serially Correlated Errors**
  Suppose initially that the errors are not serially correlated. In this case, assumption 3) can be relaxed, since we can show that $\{Z_t u_t\}_{t\in\mathbb{Z}}$ is a MDS when the error process is iid and the instruments have bounded fourth moments (this is similar to the line of reasoning we use to study VAR models; consult the section on Vector autoregressions for details). Since martingale difference sequences are pairwise uncorrelated, we can slightly strengthen this result and assume that $\{Z_t u_t\}_{t\in\mathbb{Z}}$ is independent, then we are essentially assuming that $\{Z_t u_t\}_{t\in\mathbb{Z}}$ is 0-dependent.

- **m-dependent Errors**
  Suppose now that the errors are serially correlated but exhibit limited dependence, namely m-dependence. This situation arises frequently in the literature, for instance, when modeling rational expectations models (as in Cumby et al. (1983)), or more generally when the error process follows a finite order MA process.

  To see how the m-dependence of the errors can be extended to the m-dependence of $\{Z_t u_t\}_{t\in\mathbb{Z}}$, we show that $\{Z_t u_t\}_{t\in\mathbb{Z}}$ is in fact weakly m-dependent when the errors are m-dependent.

  Note initially that $\{Z_t u_t\}_{t\in\mathbb{Z}}$ is a mean zero process by the identification assumption. Define the filtration $\mathcal{F} = \{\mathcal{F}_t \mid t \in \mathbb{Z}\}$ as

$$\mathcal{F}_t = \sigma\{Z_{t+m}, \cdots, u_t, \cdots\}$$

  for any $t \in \mathbb{Z}$. Then, $\{Z_t u_t\}_{t\in\mathbb{Z}}$ is an $\mathcal{F}$-adapted process with mean zero such that

$$\mathbb{E}\left[Z_t u_t \mid \mathcal{F}_{t-m}\right] = Z_t \cdot \mathbb{E}\left[u_t \mid \mathcal{F}_{t-m}\right] = \mathbf{0}$$

  where the first equality follows because Z_t is $\mathcal{F}_t$-measurable, and the second equality because u_t is independent of $\mathcal{F}_{t-m}$ due to the backward looking exogeneity and m-dependence assumptions. The filtration generated by $\{Z_t u_t\}_{t\in\mathbb{Z}}$ is contained in $\mathcal{F}$ due to the $\mathcal{F}$-adaptedness of $\{Z_t u_t\}_{t\in\mathbb{Z}}$, so by definition it is weakly m-dependent in the first order.

  Thus, it is not too much of a stretch, in this case, to assume the stronger condition that $\{Z_t u_t\}_{t\in\mathbb{Z}}$ is m-dependent.

In any case, the assumption that $\{Z_t u_t\}_{t\in\mathbb{Z}}$ is a stationary m-dependent process is not too unreasonable when modeling serially correlated errors. In light of the content in the previous section, assumption 3 implies that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t u_t \xrightarrow{d} N\left[\mathbf{0}, \sum_{j=-m}^{m}\Gamma(j)\right].$$

Now we show that, under the above assumptions, the GMM estimator of β is consistent and asymptotically normal.

We already saw that the first assumption in the general model is satisfied. In addition, since

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g(Y_t,\beta)}{\partial \beta'} = \frac{1}{T}\sum_{t=1}^{T} Z_t X_t'$$

for any $\beta \in \Theta$, for any sequence $\{\tilde{\beta}_T\}_{T\in N_+}$ such that$\tilde{\beta}_T \xrightarrow{p} \beta_0$, we have

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g(Y_t,\tilde{\beta}_T)}{\partial \beta'} = \frac{1}{T}\sum_{t=1}^{T} Z_t X_t' \xrightarrow{p} Q_{zx}$$

as $T \to \infty$. This shows us that the fourth assumption in the general model is satisfied.

We also showed above that

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g(Y_t,\beta_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t(y_t - X_t'\beta_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t u_t \\ \xrightarrow{d} N\left[\mathbf{0}, \sum_{j=-m}^{m}\Gamma(j)\right].\end{aligned}$$

Thus, the final assumption in the general model is also satisfied. This actually implies that

$$\frac{1}{T}\sum_{t=1}^{T} Z_t u_t \xrightarrow{p} \mathbf{0},$$

which in turn tells us that

$$\begin{aligned}\hat{\beta}_T - \beta_0 = \left[\left(\frac{1}{T}\sum_{t=1}^{T} X_t Z_t'\right) W_T \left(\frac{1}{T}\sum_{t=1}^{T} Z_t X_t'\right)\right]^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} X_t Z_t'\right) W_T \left(\frac{1}{T}\sum_{t=1}^{T} Z_t u_t\right) \\ \xrightarrow{p} (Q_{zx}' W Q_{zx})^{-1} Q_{zx}' W \mathbf{0} = \mathbf{0}.\end{aligned}$$

The GMM estimator of β is consistent, so that the third assumption in the general model is also satisfied.

Since all 5 assumptions in the general model are satisfied, we can see that, as per the conclusion of the general model,

$$\sqrt{T}\left(\hat{\beta}_T-\beta_0\right)\xrightarrow{d}N\left[\mathbf{0},\left(Q_{zx}'WQ_{zx}\right)^{-1}Q_{zx}'W\left(\sum_{j=-m}^{m}\Gamma(j)\right)WQ_{zx}\left(Q_{zx}'WQ_{zx}\right)^{-1}\right].$$

The optimal weighting matrix is now seen to be

$$W=\left(\sum_{j=-m}^{m}\Gamma(j)\right)^{-1},$$

under which the asymptotic distribution of the 2SLS estimator becomes

$$\sqrt{T}\left(\hat{\beta}_T-\beta_0\right)\xrightarrow{d}N\left[\mathbf{0},\left[Q_{zx}'\left(\sum_{j=-m}^{m}\Gamma(j)\right)^{-1}Q_{zx}\right]^{-1}\right].$$

To obtain the asymptotic variance and, indeed, to compute $\hat{\beta}_T$, we require a consistent estimator W_T of $\left(\sum_{j=-m}^{m}\Gamma(j)\right)^{-1}$. This can be achieved by defining W_T^{-1} as a consistent estimator of the long run variance $\sum_{j=-m}^{m}\Gamma(j)$ using the HAC method laid out in Andrews (1991).

However, we run into another problem at this point. The construction of W_T^{-1} requires the 2SLS residuals, which themselves depend on W_T. To resolve this issue, $\hat{\beta}_T$ and W_T can be computed iteratively according to the following algorithm:

Step 0: Loading initial value
Put

$$W_T=I_L \text{ or } \frac{1}{T}\sum_{t=1}^{T}Z_tZ_t'.$$

This becomes our weighting matrix for the 0th iteration.

Step 1: Computing 2SLS Estimator
For any $i\in N_+$, given the weighting matrix $W_T^{(i-1)}$ at the $i-1$th iteration, we compute the ith 2SLS estimator

$$\hat{\beta}_T^{(i)}=\left[\left(\sum_{t=1}^{T}X_tZ_t'\right)W_T^{(i-1)}\left(\sum_{t=1}^{T}Z_tX_t'\right)\right]^{-1}\left(\sum_{t=1}^{T}X_tZ_t'\right)W_T^{(i-1)}\left(\sum_{t=1}^{T}Z_ty_t\right)$$

and procure the ith residual process

$$\hat{u}_t^{(i)}=y_t-X_t'\hat{\beta}_T^{(i)}$$

Step 2: Computing Weighting Matrix

Compute the ith weighting matrix as the consistent estimator of $\sum_{j=-m}^{m}\Gamma(j)$ given the residuals obtained above; one possible candidate is

$$(W_T^{(i)})^{-1} = \hat{\Gamma}(0) + \sum_{j=1}^{S_T} k\left(\frac{j}{S_T}\right)\left(\hat{\Gamma}(j) + \hat{\Gamma}(j)'\right)$$

where $k(\cdot)$ is an appropriate chosen kernel, S_T is the truncation lag window and

$$\hat{\Gamma}(j) = \frac{1}{T}\sum_{t=j+1}^{T} Z_t \hat{u}_t^{(i)} \hat{u}_{t-j}^{(i)} Z'_{t-j}$$

for $0 \leq j \leq T-1$.

Step 3: Convergence Criterion

If some convergence criterion is met (usually conerning the distance between the iterates of the 2SLS estimator or the weighting matrix) then terminate the process. Otherwise, return to step 1.

1.2.5 Mixing Processes

So far, we have shown that, if a weakly stationary process satisfies certain moment conditions and its entries are independent with observations that are sufficiently far away, then it satisfies a form of the CLT. However, the assumption that the dependence between the entries peters out so rapidly is unrealistic; indeed, it is not even by satisfied the most basic AR processes. In this section, we introduce a weaker form of limited dependence; simply put, it allows entries in the process to be independent "at the limit".

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional process. For any sub σ-algebra $\mathcal{F}$ and $\mathcal{G}$ of $\mathcal{H}$, define $\alpha(\mathcal{F},\mathcal{G})$ as

$$\alpha(\mathcal{F},\mathcal{G}) = \sup_{A\in\mathcal{F},B\in\mathcal{G}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| \in [0,+\infty].$$

Defining $\mathcal{F}_{-\infty}^{k} = \sigma\{Y_k, Y_{k-1}, \cdots\}$ and $\mathcal{F}_{k+\tau}^{\infty} = \sigma\{Y_{k+\tau}, Y_{k+\tau+1}, \cdots\}$ for any $k\in\mathbb{Z}$ and $\tau\in N_+$, define

$$\alpha(\tau) = \sup_{k\in\mathbb{Z}} \alpha\left(\mathcal{F}_{-\infty}^{k}, \mathcal{F}_{k+\tau}^{\infty}\right) \in [0,+\infty].$$

We say that $\{Y_t\}_{t\in\mathbb{Z}}$ is α-mixing, or strong mixing, if

$$\lim_{\tau\to\infty} \alpha(\tau) = 0.$$

That is, the dependence of any two groups of entries in $\{Y_t\}_{t\in\mathbb{Z}}$ that are τ periods apart uniformly decreases to 0 as $\tau\to\infty$. If $\{Y_t\}_{t\in\mathbb{Z}}$ is strictly stationary, then we can equivalently define

$$\alpha(\tau) = \alpha\left(\mathcal{F}_{-\infty}^{0}, \mathcal{F}_{\tau}^{\infty}\right).$$

1.3 Linear Processes and the Wold Representation

In this section we study (possibly multivariate) weakly stationary time series, with a focus on causal linear processes and the Wold representation theorem. Throughout, we let the matrix norm $\|\cdot\|$ be the trace norm (for more on trace norms, refer to the notes on factor models).

1.3.1 Multidimensional L^p Spaces

For any $p \in [1, +\infty)$ and a measure space $(E, \mathcal{E}, \mu)$, we define the n-dimensional L^p-norm of a measurable function $f : E \to \mathbb{R}^n$ as the L^p-norm of the non-negative measurable function $|f| : E \to \mathbb{R}_+$, that is,

$$\|f\|_{n,p} := \||f|\|_p = \left(\int_E |f|^p d\mu \right)^{\frac{1}{p}}.$$

$\|\cdot\|_{n,p}$ satifies Hölder's and Minkowski's inequalities because $\|\cdot\|_p$ does: for any $p, q \in (1, +\infty)$ such that $1 = \frac{1}{p} + \frac{1}{q}$, and measurable $f, g : E \to \mathbb{R}^n$,

$$\|f'g\|_1 = \int_E |f'g| d\mu \le \int_E |f||g| d\mu \le \||f|\|_p \cdot \||g|\|_q = \|f\|_{n,p} \cdot \|g\|_{n,q}$$

by the Cauchy-Schwarz inequality and the univariate version of Hölder's inequality.

Similarly, for any $p \in [1, +\infty)$ and measurable $f, g : E \to \mathbb{R}^n$,

$$\|f + g\|_{n,p} \le \||f| + |g|\|_p \le \||f|\|_p + \||g|\|_p = \|f\|_{n,p} + \|g\|_{n,p}.$$

Let $L^p_n(\mathcal{E}, \mu)$ be the collection of all $\mathcal{E}$-measurable function $f : E \to \mathbb{R}^n$ such that $\|f\|_{n,p} < +\infty$. $L^p_n(\mathcal{E}, \mu)$ is clearly a vector space over the real field.

Multidimensional L^p Norms are Norms

We can also see that $\|\cdot\|_{n,p}$ is a norm on $L^p_n(\mathcal{E}, \mu)$, given that we identify μ-almost everywhere equal functions:

- For any $f \in L^p_n(\mathcal{E}, \mu)$, suppose $\|f\|_{n,p} = 0$. Then, $|f| = 0$ μ-almost everywhere on E, which implies that $f = 0$ μ-almost everywhere on E. Conversely, if $f = 0$ μ-almost everywhere on E, then $|f| = 0$ μ-a.e. and thus $\|f\|_{n,p} = 0$.

- For any $z \in \mathbb{R}$ and $f \in L_n^p(\mathcal{E},\mu)$,

$$\|z \cdot f\|_{n,p} = \||z| \cdot |f|\|_p = |z| \cdot \|f\|_{n,p}.$$

- For any $f, g \in L_n^p(\mathcal{E},\mu)$,

$$\|f+g\|_{n,p} \leq \||f|+|g|\|_p \leq \|f\|_{n,p} + \|g\|_{n,p}.$$

Multidimensional L^p Spaces are Banach Spaces

We can see that the space $(L_n^p(\mathcal{E},\mu), \|\cdot\|_{n,p})$ is a Banach space, just as in the univariate case: this follows easily from the Riesz-Fischer theorem for the univariate case. Choose any sequence $\{f_k\}_{k\in N_+}$ that is Cauchy in $(L_n^p(\mathcal{E},\mu), \|\cdot\|_{n,p})$. For any $1 \leq i \leq n$, the sequence of coordinates $\{f_{ik}\}_{k\in N_+}$ is a Cauchy sequence in the Banach space $(L^p(\mathcal{E},\mu), \|\cdot\|_p)$, so there exists a $g_i \in L^p(\mathcal{E},\mu)$ such that

$$\lim_{k\to\infty} \|f_{ik} - g_i\|_p = 0.$$

Define $g = (g_1, \cdots, g_n)$. Then, because g is measurable and

$$\|g\|_{n,p} = \||g|\|_p \leq \left\|\sum_{i=1}^n |g_i|\right\|_p \leq \sum_{i=1}^n \|g_i\|_p < +\infty,$$

we can see that $g \in L_n^p(\mathcal{E},\mu)$. Furthermore,

$$\|f_k - g\|_{n,p} \leq \sum_{i=1}^n \|f_{ik} - g_i\|_p$$

for any $k \in N_+$, so we have

$$\lim_{k\to\infty} \|f_k - g\|_{n,p} = 0$$

and thus $f_k \xrightarrow{L^p} g$. This shows us that $(L_n^p(\mathcal{E},\mu), \|\cdot\|_{n,p})$ is a Banach space.

Multidimensional L^2 Space is a Hilbert Space

Finally, we can define an inner product on the space $(L_2^p(\mathcal{E},\mu), \|\cdot\|_{n,p})$. Define the function $\langle\cdot,\cdot\rangle_{n,2}$: $L_2^p(\mathcal{E},\mu)\times L_2^p(\mathcal{E},\mu)\to\mathbb{R}$ as

$$\langle f,g\rangle_{n,2}=\int_E f'gd\mu$$

for any $f,g\in L_2^p(\mathcal{E},\mu)$; this is well-defined, since

$$\int_E |f'g|\mu=\|f'g\|_1\leq\|f\|_{n,2}\cdot\|g\|_{n,2}<+\infty.$$

$\langle\cdot,\cdot\rangle_{n,2}$ is an inner product on $L_2^p(\mathcal{E},\mu)$, given that we identify functions equal μ-a.e.:

- For any $z\in\mathbb{R}$ and $f,g,h\in L_2^p(\mathcal{E},\mu)$,

$$\langle z\cdot f+g,h\rangle_{n,2}=\int_E(z\cdot f+g)'hd\mu=z\cdot\int_E f'hd\mu+\int_E g'hd\mu=z\cdot\langle f,h\rangle_{n,2}+\langle g,h\rangle_{n,2}.$$

- For any $f,g\in L_2^p(\mathcal{E},\mu)$,

$$\langle f,g\rangle_{n,2}=\int_E f'gd\mu=\int_E g'fd\mu=\langle g,f\rangle_{n,2}.$$

- For any $f\in L_2^p(\mathcal{E},\mu)$,

$$\langle f,f\rangle_{n,2}=\int_E|f|^2d\mu\geq 0.$$

  If $\langle f,f\rangle_{n,2}=0$, then $|f|^2=0$ μ-a.e. and thus $f=0$ μ-a.e.

$\|\cdot\|_{n,2}$ is the norm induced by $\langle\cdot,\cdot\rangle_{n,2}$, so $(L_n^2(\mathcal{E},\mu),\langle\cdot,\cdot\rangle_{n,2})$ is a Hilbert space over the real field. Below, for notational brevity, we denote

$$L_n^p(\mathcal{E},\mu)=L^p(\mathcal{E},\mu),\quad \|\cdot\|_{n,p}=\|\cdot\|_p\quad\text{and}\quad\langle\cdot,\cdot\rangle_{n,2}=\langle\cdot,\cdot\rangle_2$$

when the dimension n is determined.

1.3.2 Linear Processes

We first show that we can construct autocorrelated weakly stationary processes given a white noise process and a sequence of coefficients that satisfy weak regularity conditions:

Theorem (Convergence of Linear Processes under Square Summability)
Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be an n-dimensional white noise process with variance $\Sigma \in \mathbb{R}^{n\times n}$, and $\{\Psi_j\}_{j\in\mathbb{Z}}$ a sequence of $n \times n$ real matrices such that

$$\sum_{j=-\infty}^{\infty} \operatorname{tr}\left(\Psi_j \Sigma \Psi_j'\right) < +\infty.$$

Then, for any $\mu \in \mathbb{R}^n$ and $t \in \mathbb{Z}$, the partial sum process

$$\left\{\mu + \sum_{j=-m}^{m} \Psi_j \cdot \varepsilon_{t-j}\right\}_{m\in N_+}$$

is L^2 and converges in L^2 to some random vector Y_t.

The L^2 process $\{Y_t\}_{t\in\mathbb{Z}}$ defined as above is weakly stationary with mean μ and autocovariance function $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$ defined as

$$\Gamma(\tau) = \sum_{j=-\infty}^{\infty} \Psi_j \Sigma \Psi_{j-\tau}'$$

for any $\tau \in \mathbb{Z}$.

Proof) For any $t \in \mathbb{Z}$ and $m \in N_+$, define

$$Y_{mt} = \mu + \sum_{j=-m}^{m} \Psi_j \cdot \varepsilon_{t-j}.$$

Note that $\mathbb{E}[Y_{mt}] = \mu$. For any $\tau \in \mathbb{Z}$ and $m \in N_+$, we have

$$\mathbb{E}\left[(Y_{mt} - \mu)(Y_{m,t-\tau} - \mu)'\right] = \sum_{j=-m}^{m}\sum_{i=-m}^{m} \Psi_j \cdot \mathbb{E}\left[\varepsilon_{t-j}\varepsilon_{t-\tau-i}'\right] \cdot \Psi_i'.$$

For any $m, k \in N_+$ such that $m > k$, note that

$$\begin{aligned}\|Y_{mt} - Y_{kt}\|_2^2 = \mathbb{E}|Y_{mt} - Y_{kt}|^2 = \mathbb{E}\left|\sum_{j=-m}^{-k-1} \Psi_j \cdot \varepsilon_{t-j} + \sum_{j=k+1}^{m} \Psi_j \cdot \varepsilon_{t-j}\right|^2 \\ = \sum_{j=-m}^{-k-1} \operatorname{tr}\left(\Psi_j \Sigma \Psi_j'\right) + \sum_{j=k+1}^{m} \operatorname{tr}\left(\Psi_j \Sigma \Psi_j'\right).\end{aligned}$$

The right hand side goes to 0 as $m, k \to \infty$, so it follows that $\{Y_{mt}\}_{m\in N_+}$ is Cauchy in

$L^2(\mathcal{H}, \mathbb{P})$. Because $(L^2(\mathcal{H}, \mathbb{P}), \langle \cdot, \cdot \rangle_2)$ is a Hilbert space, there exists a $Y_t \in L^2(\mathcal{H}, \mathbb{P})$ such that

$$Y_{mt} \xrightarrow{L^2} Y_t$$

as $m \to \infty$.

It remains to show that $\{Y_t\}_{t \in \mathbb{Z}}$ is weakly stationary.
For any $t \in \mathbb{Z}$ and $m \in N_+$,

$$|\mathbb{E}[Y_t] - \mu| = |\mathbb{E}[Y_t] - \mathbb{E}[Y_{mt}]| \leq \mathbb{E}|Y_t - Y_{mt}| \leq \|Y_t - Y_{mt}\|_2$$

by Jensen's inequality, so taking $m \to \infty$ on both sides shows us that

$$|\mathbb{E}[Y_t] - \mu| = 0,$$

that is, $\mathbb{E}[Y_t] = \mu$. Therefore, $\{Y_t\}_{t \in \mathbb{Z}}$ is mean stationary.

Define $Z_t = Y_t - \mu$ and $Z_{mt} = Y_{mt} - \mu$ for any $t \in \mathbb{Z}$ and $m \in N_+$. For any $t \in \mathbb{Z}$, $\tau \in \mathbb{Z}$ and $m \in N_+$,

$$\begin{aligned}
\left\| \mathbb{E}\left[Z_{mt} Z'_{m,t-\tau}\right] - \mathbb{E}\left[Z_t Z'_{t-\tau}\right] \right\| &\leq \mathbb{E}\|(Z_{mt} - Z_t)(Z_{m,t-\tau} - Z_{t-\tau})'\| + \mathbb{E}\|Z_t (Z_{m,t-\tau} - Z_{t-\tau})'\| \\
&\quad + \mathbb{E}\|(Z_{mt} - Z_t) Z'_{t-\tau}\| \\
&\leq \mathbb{E}\left[|Z_{mt} - Z_t| \cdot |Z_{m,t-\tau} - Z_{t-\tau}|\right] + \mathbb{E}\left[|Z_t| \cdot |Z_{m,t-\tau} - Z_{t-\tau}|\right] \\
&\quad + \mathbb{E}\left[|Z_{mt} - Z_t| \cdot |Z_{t-\tau}|\right] \\
&\leq \|Z_{mt} - Z_t\|_2 \cdot \|Z_{m,t-\tau} - Z_{t-\tau}\|_2 + \|Z_t\|_2 \cdot \|Z_{m,t-\tau} - Z_{t-\tau}\|_2 \\
&\quad + \|Z_{mt} - Z_t\|_2 \cdot \|Z_{t-\tau}\|_2,
\end{aligned}$$

where the second inequality follows from the trace norm inequality and the third from Hölder's inequality. Therefore, taking $m \to \infty$ on both sides yields

$$\lim_{m \to \infty} \mathbb{E}\left[Z_{mt} Z'_{m,t-\tau}\right] = \mathbb{E}\left[Z_t Z'_{t-\tau}\right].$$

The existence of the limit is proved alongside the fact that the limit is $\mathbb{E}\left[Z_t Z'_{t-\tau}\right]$. Given the value of $\mathbb{E}\left[Z_{mt} Z'_{m,t-\tau}\right]$ for large m, we can see that

$$\begin{aligned}
\operatorname{Cov}[Y_t, Y_{t-\tau}] &= \mathbb{E}\left[Z_t Z'_{t-\tau}\right] \\
&= \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} \Psi_j \cdot \mathbb{E}\left[\varepsilon_{t-j} \varepsilon'_{t-\tau-i}\right] \cdot \Psi'_i = \sum_{j=-\infty}^{\infty} \Psi_j \Sigma \Psi'_{j-\tau}.
\end{aligned}$$

Therefore, $\{Y_t\}_{t \in \mathbb{Z}}$ is also covariance stationary and has the specified autocovariance function.

Q.E.D.

In light of the above theorem, we can define a new class of weakly stationary processes. An n-dimensional process $\{Y_t\}_{t\in\mathbb{Z}}$ is said to be a linear process if it satisfies the following conditions:

- There exists a square-summable sequence $\{\Psi_j\}_{j\in\mathbb{Z}}$ of $n \times n$ matrices, that is,

$$\sum_{j=-\infty}^{\infty} \operatorname{tr}\left(\Psi_j' \Sigma \Psi_j\right) < +\infty.$$

- There exists a $\mu \in \mathbb{R}^n$ and an n-dimensional white noise process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ with covariance matrix Σ such that

$$Y_t = \mu + \Psi(L)\varepsilon_t = \mu + \sum_{j=-\infty}^{\infty} \Psi_j \cdot \varepsilon_{t-j}$$

  for any $t \in \mathbb{Z}$, where $\sum_{j=-\infty}^{\infty} \Psi_j \cdot \varepsilon_{t-j}$ is the L^2-limit of the corresponding partial sum process.

If $\mu = \mathbf{0}$ above, then we call $\{Y_t\}_{t\in\mathbb{Z}}$ a zero-mean linear process.

The previous theorem shows us that $\{Y_t\}_{t\in\mathbb{Z}}$ is a weakly stationary process with mean μ and autocovariance

$$\Gamma(\tau) = \sum_{j=-\infty}^{\infty} \Psi_j \Sigma \Psi_{j-\tau}$$

for any $\tau \in \mathbb{Z}$.

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional linear process with mean $\mu \in \mathbb{R}^n$, innovation process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ and square summable linear filter $\{\Psi_j\}_{j\in\mathbb{Z}}$. We say that $\{Y_t\}_{t\in\mathbb{Z}}$ is a causal linear process if

$$\Psi_j = O \quad \text{for any } j < 0.$$

This means that Y_t is a function only of the current and past innovations; in this case, we write

$$Y_t = \mu + \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j},$$

where the equality denotes the L^2-convergence of the partial sum process.

A stronger condition than square summability is absolute summability, which states that

$$\sum_{j=-\infty}^{\infty} \|\Psi_j\| < +\infty.$$

Under the absolute summability of $\{\Psi_j\}_{j\in\mathbb{Z}}$, the convergence of

$$\left\{\mu + \sum_{j=-m}^{m} \Psi_j \cdot \varepsilon_{t-j}\right\}_{m\in N_+}$$

to Y_t holds both in L^2 and almost surely. We establish these claims below:

Lemma (Absolute Summability implies Square Summability)
Let $\{\Psi_j\}_{j\in\mathbb{Z}}$ be an absolutely summable sequence of $n \times n$ real matrices. Then, it is also square summable under any positive definite covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$.

Proof) Suppose that

$$\sum_{j=-\infty}^{\infty} \|\Psi_j\| < +\infty.$$

Then, note that

$$\begin{aligned} 0 \le \sum_{j=-\infty}^{\infty} \operatorname{tr}\left(\Psi_j \Sigma \Psi_j'\right) &= \sum_{j=-\infty}^{\infty} \operatorname{tr}\left(\Sigma \cdot (\Psi_j' \Psi_j)\right) \\ &\le \sum_{j=-\infty}^{\infty} \|\Sigma\| \cdot \left\|\Psi_j' \Psi_j\right\| \le \|\Sigma\| \cdot \left(\sum_{j=-\infty}^{\infty} \|\Psi_j\|^2\right). \end{aligned}$$

For any $m \in N_+$,

$$\begin{aligned} \sum_{j=-m}^{m} \|\Psi_j\|^2 &\le \left(\max_{-m\le j\le m} \|\Psi_j\|\right) \cdot \left(\sum_{j=-m}^{m} \|\Psi_j\|\right) \\ &\le \left(\sum_{j=-\infty}^{\infty} \|\Psi_j\|\right)^2 < +\infty. \end{aligned}$$

Therefore,

$$\sum_{j=-\infty}^{\infty} \|\Psi_j\|^2 < +\infty$$

and we have

$$\sum_{j=-\infty}^{\infty} \operatorname{tr}\left(\Psi_j \Sigma \Psi_j'\right) < +\infty.$$

Q.E.D.

Theorem (Convergence of Linear Processes under Absolute Summability)
Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be an n-dimensional mean-zero square integrable and weakly stationary time series with absolutely summable autocovariance function $G:\mathbb{Z}\to\mathbb{R}^{n\times n}$. Let $\{\Psi_j\}_{j\in\mathbb{Z}}$ a sequence of $n\times n$ real matrices such that

$$\sum_{j=-\infty}^{\infty}\|\Psi_j\|<+\infty.$$

Then, for any $\mu\in\mathbb{R}^n$ and $t\in\mathbb{Z}$, the partial sum process

$$\Big\{\mu+\sum_{j=-m}^{m}\Psi_j\cdot\varepsilon_{t-j}\Big\}_{m\in N_+}$$

is in L^2 and converges almost surely and in L^2 to some random vector Y_t.

The L^2 process $\{Y_t\}_{t\in\mathbb{Z}}$ defined as above is weakly stationary with mean μ and autocovariance function $\Gamma:\mathbb{Z}\to\mathbb{R}^{n\times n}$ defined as

$$\Gamma(\tau)=\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\Psi_i\cdot G(\tau+j-i)\cdot\Psi_j'$$

for any $\tau\in\mathbb{Z}$.

Proof) **Step 1: Almost Sure Convergence**

For any $t\in\mathbb{Z}$ and $m\in N_+$, define

$$Y_{mt}=\mu+\sum_{j=-m}^{m}\Psi_j\cdot\varepsilon_{t-j}.$$

Define c as the counting measure on $\mathbb{Z}$, and the function $f:\mathbb{Z}\times\Omega\to[0,+\infty)$ as

$$f(j,\omega)=|\Psi_j\cdot\varepsilon_{t-j}(\omega)|$$

for any $(j,\omega)\in\mathbb{Z}\times\Omega$. Since f is a non-negative measurable function relative to the product σ-algebra on $\mathbb{Z}\times\Omega$, by Fubini's theorem for non-negative functions,

$$\begin{aligned}\mathbb{E}\left[\sum_{j=-\infty}^{\infty}|\Psi_j\cdot\varepsilon_{t-j}|\right]&=\int_\Omega\int_\mathbb{Z}f(\tau,\omega)dc(\tau)d\mathbb{P}(\omega)\\&=\int_\mathbb{Z}\int_\Omega f(\tau,\omega)d\mathbb{P}(\omega)dc(\tau)\\&=\sum_{j=-\infty}^{\infty}\mathbb{E}|\Psi_j\cdot\varepsilon_{t-j}|\\&\leq\sum_{j=-\infty}^{\infty}\|\Psi_j\|\cdot\|\varepsilon_{t-j}\|_2=\mathrm{tr}\,(G(0))^{\frac{1}{2}}\cdot\left(\sum_{j=-\infty}^{\infty}\|\Psi_j\|\right)<+\infty.\end{aligned}$$

By the finiteness property of non-negative functions, it now follows that

$$\sum_{j=-\infty}^{\infty} |\Psi_j \cdot \varepsilon_{t-j}| < +\infty$$

almost surely, which implies that the series

$$\sum_{j=-\infty}^{\infty} \Psi_j \cdot \varepsilon_{t-j}$$

converges almost surely. Define Y_t as

$$Y_t = \sum_{j=-\infty}^{\infty} \Psi_j \cdot \varepsilon_{t-j}$$

on the subset of Ω on which the series converges, and $\mathbf{0}$ otherwise. Then, Y_t is a random vector such that

$$Y_{mt} \overset{a.s.}{\to} Y_t$$

as $m \to \infty$.

Step 2: Mean Square Convergence

To see that the convergence is in L^2 as well, we follow the steps of the previous theorem almost step for step. For any $m, k \in N_+$ such that $m \geq k$,

$$\begin{aligned}
\|Y_{mt} - Y_{kt}\|_2 &= \left\| \sum_{j=-m}^{-k-1} \Psi_j \cdot \varepsilon_{t-j} + \sum_{j=k+1}^{m} \Psi_j \cdot \varepsilon_{t-j} \right\|_2 \\
&\leq \sum_{j=-m}^{-k-1} \|\Psi_j\| \cdot \|\varepsilon_{t-j}\|_2 + \sum_{j=k+1}^{m} \|\Psi_j\| \cdot \|\varepsilon_{t-j}\|_2 \\
&\leq \operatorname{tr}(G(0))^{\frac{1}{2}} \sum_{k<|j|\leq m} \|\Psi_j\|.
\end{aligned}$$

Since $\{\Psi_j\}_{j\in\mathbb{Z}}$ is absolutely summable, sending $m, k \to \infty$ on both sides yields

$$\lim_{m,k\to\infty} \|Y_{mt} - Y_{kt}\|_2 = 0.$$

This tells us that $\{Y_{mt}\}_{m\in N_+}$ is Cauchy in the L^2 metric, and by the completeness of the n-dimensional L^2 space, there exists a $Z_t \in L^2(\mathcal{H}, \mathbb{P})$ such that $Y_{mt} \overset{L^2}{\to} Z_t$ as $m \to \infty$.

Step 3: Equivalence between L^2 and Almost Sure Limit

It remains to prove that $Z_t = Y_t$. Note that

$$\begin{aligned}\mathbb{E}|Y_t - Z_t| &\leq \mathbb{E}|Z_t - Y_{mt}| + \mathbb{E}|Y_{mt} - Y_t| \\ &\leq \|Z_t - Y_{mt}\|_2 + \mathbb{E}|Y_{mt} - Y_t|\end{aligned}$$

for any $m \in N_+$. Since

$$\begin{aligned}|Y_{mt} - Y_t| &\leq |Y_{mt}| + |Y_t| \\ &\leq \sum_{j=-m}^{m} |\Psi_j \cdot \varepsilon_{t-j}| + |Y_t| \leq 2 \cdot \sum_{j=-\infty}^{\infty} |\Psi_j \cdot \varepsilon_{t-j}|\end{aligned}$$

for any $m \in N_+$, where

$$\mathbb{E}\left[\sum_{j=-\infty}^{\infty} |\Psi_j \cdot \varepsilon_{t-j}|\right] < +\infty,$$

and $Y_{mt} \overset{a.s.}{\to} Y_t$ as $m \to \infty$, by the DCT we have

$$\lim_{m\to\infty} \mathbb{E}|Y_{mt} - Y_t| = 0.$$

Thus,

$$\mathbb{E}|Y_t - Z_t| = 0,$$

and $Y_t = Z_t$ almost surely.

Step 4: Mean Stationarity

For any $t \in \mathbb{Z}$ and $m \in N_+$,

$$\mathbb{E}[Y_{mt}] = \mu,$$

and since

$$|\mu - \mathbb{E}[Y_t]| \leq \mathbb{E}|Y_{mt} - Y_t| \leq \|Y_{mt} - Y_t\|_2,$$

sending $m \to \infty$ shows us that $\mathbb{E}[Y_t] = \mu$; $\{Y_t\}_{t\in\mathbb{Z}}$ is a mean ergodic L^2 process.

Step 5: Covariance Stationarity

Finally, for any $t, \tau \in \mathbb{Z}$ and $m \in N_+$,

$$\mathbb{E}\left[(Y_{mt} - \mu)(Y_{m,t-\tau} - \mu)'\right] = \sum_{i=-m}^{m} \sum_{j=-m}^{m} \Psi_i \cdot G(\tau + j - i) \cdot \Psi_j';$$

since

$$\sum_{i=-m}^{m} \sum_{j=-m}^{m} \left\| \Psi_i \cdot G(\tau + j - i) \cdot \Psi_j' \right\| \leq \left(\sum_{j=-\infty}^{\infty} \|\Psi_j\| \right)^2 \left(\sum_{j=-\infty}^{\infty} \|G(j)\| \right) < +\infty,$$

$\mathbb{E}\left[(Y_{mt} - \mu)(Y_{m,t-\tau} - \mu)'\right]$ converges to

$$\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \Psi_i \cdot G(\tau + j - i) \cdot \Psi_j'$$

as $m \to \infty$.

We saw in the proof of the previous theorem that

$$\lim_{m \to \infty} \mathbb{E}\left[(Y_{mt} - \mu)(Y_{m,t-\tau} - \mu)'\right] = \mathbb{E}\left[(Y_t - \mu)(Y_{t-\tau} - \mu)'\right] := \Gamma(\tau)$$

since $Y_{mt} \xrightarrow{L^2} Y_t$ for any $t \in \mathbb{Z}$. Therefore,

$$\Gamma(\tau) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \Psi_i \cdot G(\tau + j - i) \cdot \Psi_j'$$

and thus $\{Y_t\}_{t \in \mathbb{Z}}$ is weakly stationary.

Q.E.D.

The assumption of absolutely summable coefficients thus allows us to extend the result of the previous theorem to arbitrary weakly stationary innovation processes. In particular, the absolute summability of the autocovariances of the underlying innovation process implies that the linear process constructed using the process is also weakly stationary. A useful shorthand for this result is

$$Y_t = \mu + \Psi(L)\varepsilon_t := \mu + \sum_{j=-\infty}^{\infty} \Psi_j \cdot \varepsilon_{t-j},$$

where $\Psi(L)$ is the lag polynomial written as

$$\sum_{j=-\infty}^{\infty} \Psi_j \cdot L^j.$$

If $\Psi(L)$ and $\Theta(L)$ are two lag polynomials corresponding to the absolutely summable coefficients $\{\Psi_j\}_{j\in\mathbb{Z}}$ and $\{\Theta_j\}_{j\in\mathbb{Z}}$ and satisfy

$$\Psi(L)\Theta(L)\varepsilon_t = \varepsilon_t,$$

then we denote

$$\Theta(L) = \Psi(L)^{-1}.$$

1.3.3 Ergodicity of Linear Processes

Here we present two law of large numbers for linear processes dealing with first and second moments, respectively.
We first state a mean ergodicity result for arbitrary weakly stationary processes satisfying certain regularity conditions:

Theorem (Mean Ergodicity of Weakly Stationary Processes)
Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional weakly stationary process with mean $\mu \in \mathbb{R}^n$ and autocovariance function $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$. Suppose that $\{Y_t\}_{t\in\mathbb{Z}}$ has trace summable autocovariances, that is,

$$\sum_{\tau=-\infty}^{\infty} \mathrm{tr}\,(\Gamma(\tau)) < +\infty.$$

Then, the sample mean of $\{Y_t\}_{t\in\mathbb{Z}}$ converges in L^2 to μ, that is,

$$\frac{1}{T}\sum_{t=1}^{T} Y_t \xrightarrow{L^2} \mu.$$

Proof) For any $T \in N_+$, define

$$\bar{Y}_T = \frac{1}{T}\sum_{t=1}^{T} Y_t,$$

and note that

$$\begin{aligned}
\mathbb{E}\left|\bar{Y}_T - \mu\right|^2 &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left[(Y_t-\mu)'(Y_s-\mu)\right] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{tr}\,(\Gamma(t-s)) \\
&\le \frac{1}{T^2}\sum_{t=1}^{T}\left[\sum_{\tau=-\infty}^{\infty}\mathrm{tr}\,(\Gamma(\tau))\right] = \frac{1}{T}\sum_{\tau=-\infty}^{\infty}\mathrm{tr}\,(\Gamma(\tau))\,.
\end{aligned}$$

Since the sum on the right is finite by assumption, taking $T \to \infty$ on both sides yields

$$\lim_{T\to\infty} \mathbb{E}\left|\bar{Y}_T - \mu\right|^2 = 0,$$

and by definition $\bar{Y}_T \xrightarrow{L^2} \mu$.
Q.E.D.

The following utilizes the preceding result to derive an ergodicity result for linear processes under absolute summability.

Theorem (Mean Ergodicity of Linear Processes)
Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be an n-dimensional white noise process with positive definite covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$, $\{\Psi_j\}_{j\in\mathbb{Z}}$ an absolutely summable sequence of $n\times n$ matrices, and $\{Y_t\}_{t\in\mathbb{Z}}$ the causal linear process defined, for any $t\in\mathbb{Z}$ and some $\mu\in\mathbb{R}^n$, as

$$Y_t = \mu + \sum_{j=-\infty}^{\infty} \Psi_j \cdot \varepsilon_{t-j}.$$

Let $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$ be the autocovariance function of $\{Y_t\}_{t\in\mathbb{Z}}$. Then,

$$\sum_{\tau=-\infty}^{\infty} \|\Gamma(\tau)\| < +\infty$$

and

$$\frac{1}{T}\sum_{t=1}^{T} Y_t \xrightarrow{L^2} \mu.$$

Proof) By the results in the previous section, we know that $\{Y_t\}_{t\in\mathbb{Z}}$ is a weakly stationary process with mean μ and

$$\Gamma(\tau) = \sum_{j=-\infty}^{\infty} \Psi_j \Sigma \Psi'_{j-\tau}$$

for any $\tau\in\mathbb{Z}$. We can easily establish the trace summability of the autocovariances:

$$\begin{aligned}\sum_{\tau=-\infty}^{\infty} \operatorname{tr}(\Gamma(\tau)) &\leq \sum_{\tau=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} \operatorname{tr}\left(\Psi_j \Sigma \Psi'_{j-\tau}\right)\\ &\leq \|\Sigma\| \cdot \sum_{\tau=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} \left\|\Psi'_{j-\tau}\Psi_j\right\|\\ &\leq \|\Sigma\| \cdot \left(\sum_{j=-\infty}^{\infty} \|\Psi_j\|\right)^2 < +\infty.\end{aligned}$$

This in turn implies the absolute summability of the autocovariances, since for any symmetric matrix $A\in\mathbb{R}^{n\times n}$,

$$\|A\| = \operatorname{tr}\left(A'A\right)^{\frac{1}{2}} \leq \operatorname{tr}(A).$$

Finally, by the preceding theorem, we can see that

$$\frac{1}{T}\sum_{t=1}^{T} Y_t \xrightarrow{L^2} \mu.$$

Q.E.D.

The next result concerns the ergodicity of linear processes for second moments, under a few different regularity conditions imposed on the innovation process.

Theorem (Covariance Ergodicity for Linear Processes)
Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be an n-dimensional white noise process with positive definite covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$, $\{\Psi_j\}_{j\in\mathbb{Z}}$ an absolutely summable sequence of $n\times n$ matrices, and $\{u_t\}_{t\in\mathbb{Z}}$ the mean zero linear process defined as

$$u_t = \sum_{j=-\infty}^{\infty} \Psi_j \cdot \varepsilon_{t-j}$$

for any $t\in\mathbb{Z}$. Let $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$ be the autocovariance function of $\{u_t\}_{t\in\mathbb{Z}}$, which is absolutely summable due to the absolute summability of $\{\Psi_j\}_{j\in\mathbb{Z}}$. Suppose $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is i.i.d. Then, for any $h \geq 0$,

$$\frac{1}{T}\sum_{t=1}^{T} u_t u'_{t-h} \xrightarrow{p} \Gamma(h).$$

Proof) Choose some $h \geq 0$. Then, for any $T \in N_+$ we can write

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} u_t u'_{t-h} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} \Psi_k \varepsilon_{t-k}\varepsilon'_{t-h-l}\Psi_l \\ &= \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} \Psi_k \left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t-k}\varepsilon'_{t-h-l}\right)\Psi'_l,\end{aligned}$$

where the equality holds almost surely due ot the absolute summability of $\{\Psi_j\}_{j\in\mathbb{Z}}$. We know that

$$\Gamma(h) = \mathbb{E}\left[u_t u'_{t-h}\right] = \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} \Psi_k \underbrace{\mathbb{E}\left[\varepsilon_{t-k}\varepsilon'_{t-h-l}\right]}_{\Sigma^{kl}} \Psi'_l,$$

where

$$\Sigma^{kl} = \begin{cases} \Sigma & \text{if } k = h+l \\ O & \text{if } k \neq h+l \end{cases}$$

by the uncorrelatedness of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, so that

$$\begin{aligned}\left|\frac{1}{T}\sum_{t=1}^{T} u_t u'_{t-h} - \Gamma(h)\right| &= \left|\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} \Psi_k \left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t-k}\varepsilon'_{t-h-l} - \Sigma^{kl}\right)\Psi'_l\right| \\ &\leq \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} \|\Psi_k\|\cdot\|\Psi_l\|\cdot\left\|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t-k}\varepsilon'_{t-h-l} - \Sigma^{kl}\right\|\end{aligned}$$

almost surely. Defining

$$A_{kl,T} = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t-k}\varepsilon'_{t-h-l} - \Sigma^{kl}$$

for any $k,l \in \mathbb{Z}$ and $T \in N_+$, we can write

$$\left|\frac{1}{T}\sum_{t=1}^{T}u_t u'_{t-h} - \Gamma(h)\right| \le \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\|\Psi_k\|\cdot\|\Psi_l\|\cdot\|A_{kl,T}\|$$

almost surely, and by Fubini's theorem for non-negative integrands,

$$\mathbb{E}\left|\frac{1}{T}\sum_{t=1}^{T}u_t u'_{t-h} - \Gamma(h)\right| \le \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\|\Psi_k\|\cdot\|\Psi_l\|\cdot\mathbb{E}\|A_{kl,T}\|.$$

Let c be the counting measure on $\mathbb{Z}^2$. Defining the function $f_T : \mathbb{Z}^2 \to \mathbb{R}$ as

$$f_T(k,l) = \|\Psi_k\|\cdot\|\Psi_l\|\cdot\mathbb{E}\|A_{kl,T}\|$$

for any $k,l \in \mathbb{Z}$, we can write

$$\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\|\Psi_k\|\cdot\|\Psi_l\|\cdot\mathbb{E}\|A_{kl,T}\| = \int_{\mathbb{Z}^2} f_T dc.$$

Note that, for any $k,l \in \mathbb{Z}$,

$$\begin{aligned}\mathbb{E}\|A_{kl,T}\| &\le \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\|\varepsilon_{t-k}\epsilon'_{t-h-l}\| + \left\|\Sigma^{kl}\right\| \\ &\le 2\left\|\Sigma^{kl}\right\| \le 2\|\Sigma\|.\end{aligned}$$

This tells us that, defining $g : \mathbb{Z}^2 \to \mathbb{R}_+$ as

$$g(k,l) = 2\|\Psi_k\|\cdot\|\Psi_l\|\cdot\|\Sigma\|$$

for any $k,l \in \mathbb{Z}$, we have

$$|f_T| \le g,$$

for any $T \in N_+$, and g satisfies

$$\begin{aligned}\int_{\mathbb{Z}^2} g dc &= 2\|\Sigma\|\cdot\left(\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\|\Psi_k\|\cdot\|\Psi_l\|\right) \\ &= 2\|\Sigma\|\cdot\left(\sum_{k=-\infty}^{\infty}\|\Psi_k\|\right)^2 < +\infty\end{aligned}$$

by the absolute summability of $\{\Psi_j\}_{j\in\mathbb{Z}}$.

Suppose each $\mathbb{E}\|A_{kl,T}\|$ converges to 0 as $T\to\infty$. Then,

$$\lim_{T\to\infty} f_T(k,l) = 0$$

for any $k,l\in\mathbb{Z}$, so that, by the DCT,

$$\lim_{T\to\infty}\int_{\mathbb{Z}^2} f_T dc = 0,$$

and because $\mathbb{E}\left|\frac{1}{T}\sum_{t=1}^{T} u_t u_{t-h}' - \Gamma(h)\right|$ is dominated by $\int_{\mathbb{Z}^2} f_T dc$ for any $T\in N_+$, we have

$$\lim_{T\to\infty}\mathbb{E}\left|\frac{1}{T}\sum_{t=1}^{T} u_t u_{t-h}' - \Gamma(h)\right| = 0,$$

that is,

$$\frac{1}{T}\sum_{t=1}^{T} u_t u_{t-h}' \xrightarrow{L^1} \Gamma(h).$$

As such, it remains to see that each $\mathbb{E}\|A_{kl,T}\|$ does indeed converge to 0. Given $k,l\in\mathbb{Z}$, we can consider two cases.

i) $k=h+1$

   Then,

$$\begin{aligned}\|A_{kl,T}\| &= \left\|\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{t-k}\varepsilon_{t-k}' - \Sigma)\right\| \\ &\le \sum_{i=1}^{n}\sum_{j=1}^{n}\left|\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{i,t-k}\varepsilon_{j,t-k} - \Sigma_{ij})\right|\end{aligned}$$

   for any $T\in N_+$. Choose any $1\le i,j\le n$. Because $\{\varepsilon_t\varepsilon_t'\}_{t\in\mathbb{Z}}$ is i.i.d. and $\mathbb{E}[\varepsilon_t\varepsilon_t'] = \Sigma$ for any $t\in\mathbb{Z}$, $\{\varepsilon_{it}\varepsilon_{jt} - \Sigma_{ij}\}_{t\in\mathbb{Z}}$ is a uniformly integrable process, which in turn implies that it is L^1-bounded. Furthermore, since $\{\varepsilon_{it}\varepsilon_{jt} - \Sigma_{ij}\}_{t\in\mathbb{Z}}$ is i.i.d, with mean 0, it is also a martingale difference sequence with respect to the filtration it generates, and by the WLLN for martingale difference sequences, we can conclude that

$$\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{i,t-k}\varepsilon_{j,t-k} - \Sigma_{ij}) \xrightarrow{L^1} 0.$$

   This holds for any $1\le i,j\le n$, and

$$\mathbb{E}\|A_{kl,T}\| \le \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left|\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{i,t-k}\varepsilon_{j,t-k} - \Sigma_{ij})\right|,$$

so we have

$$\mathbb{E}\|A_{kl,T}\| \to 0$$

as $T \to \infty$.

ii) $k \neq h+1$

Then,

$$\|A_{kl,T}\| \leq \sum_{i=1}^{n}\sum_{j=1}^{n}\left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{i,t-k}\varepsilon_{j,t-h-l}\right|$$

for any $T \in N_+$. For any $1 \leq i,j \leq n$, the sequence $\{\varepsilon_{i,t-k}\varepsilon_{j,t-h-l}\}_{t\in\mathbb{Z}}$ has mean 0. Assume without loss of generality that $k < h+l$, and let $\mathcal{F}$ be the filtration on $\mathbb{Z}$ generated by $\{\varepsilon_{t-k}\}_{t\in\mathbb{Z}}$. Then, for any $t \in \mathbb{Z}$, since $\varepsilon_{j,t-h-l}$ is $\mathcal{F}_{t-1}$-measurable $(t-h-l < t-k)$, we have

$$\begin{aligned}\mathbb{E}\left[\varepsilon_{i,t-k}\varepsilon_{j,t-h-l} \mid \mathcal{F}_{t-1}\right] &= \mathbb{E}\left[\varepsilon_{i,t-k} \mid \mathcal{F}_{t-1}\right]\cdot\varepsilon_{j,t-h-l}\\ &= \mathbb{E}\left[\varepsilon_{i,t-k}\right]\cdot\varepsilon_{j,t-h-l} = 0,\end{aligned}$$

where the second inequality follows because ε_{t-k} is independent of $\mathcal{F}_{t-1}$. By definition, $\{\varepsilon_{i,t-k}\varepsilon_{j,t-h-l}\}_{t\in\mathbb{Z}}$ is a martingale difference sequence with respect to $\mathcal{F}$. Furthermore, because $\varepsilon_{i,t-k}$ and $\varepsilon_{j,t-h-l}$ are independent for any $t \in \mathbb{Z}$,

$$\mathbb{E}|\varepsilon_{i,t-k}\varepsilon_{j,t-h-l}|^2 = \left(\mathbb{E}|\varepsilon_{i,t-k}|^2\right)\left(\mathbb{E}|\varepsilon_{j,t-h-l}|^2\right) = \Sigma_{ii}\Sigma_{jj} < +\infty,$$

so that $\{\varepsilon_{i,t-k}\varepsilon_{j,t-h-l}\}_{t\in\mathbb{Z}}$ is L^2-bounded and thus uniformly integrable.

We can now apply the WLLN for martingale difference sequences to conclude that

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{i,t-k}\varepsilon_{j,t-h-l} \xrightarrow{L^1} 0,$$

and since

$$\mathbb{E}\|A_{kl,T}\| \leq \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{i,t-k}\varepsilon_{j,t-h-l}\right|,$$

we have

$$\mathbb{E}\|A_{kl,T}\| \to 0$$

as $T \to \infty$ as well.

Q.E.D.

1.3.4 The Wold Representation

In this section we provide a rationale for focusing on linear processes. It is shown that any weakly stationary time series that satisfies some light regularity conditions can be written as the sum of a linear process and a predictable process, the latter whose definition will be stated shortly. This result, called the Wold representation theorem, provides the theoretical justification for studying linear processes, and is also why we like to focus on vector autoregressions; they are simply linear processes with appropriately truncated lags.

We first introduce some definitions. Let $\{u_1,\cdots,u_T\} \subset L^2(\mathcal{H},\mathbb{P})$ be a collection of n-dimensional square integrable random vectors, and $X \in L^2(\mathcal{H},\mathbb{P})$ another n-dimensional random vector. Usually, when we refer to the projection of X on the span of $\{u_1,\cdots,u_T\}$, we refer to the linear combination of $\{u_1,\cdots,u_T\}$ that best approximates X in the L^2-norm. In what follows, we call the linear projection of X on the span of $\{u_1,\cdots,u_T\}$ the random vector

$$\hat{P}(X \mid u_1,\cdots,u_T) = \begin{pmatrix} \hat{X}_1 \\ \vdots \\ \hat{X}_n \end{pmatrix},$$

where, for each $1 \le i \le n$, $\hat{X}_i$ is the orthogonal projection of X_i on V, and the vector space V is defined as the span of the collection of (univariate) random variables $\{u_{it} \mid 1 \le i \le n, 1 \le t \le T\}$. Since $\hat{X}_i$ is contained in the univariate L^2-space for each $1 \le i \le n$, the linear projection $\hat{P}(X \mid u_1,\cdots,u_T)$ is itself a square integrable n-dimensional random vector.

From the properties of orthogonal projections onto linear subspaces spanned by a finite set, we can see that there exist matrices $A_1,\cdots,A_T \in \mathbb{R}^{n\times n}$ such that

$$\hat{P}(X \mid u_1,\cdots,u_T) = \sum_{t=1}^{T} A_t \cdot u_t.$$

Clearly, $\hat{P}(X \mid u_1,\cdots,u_T)$ is a more precise approximation to X than the orthogonal projection of X on the span of $\{u_1,\cdots,u_T\}$. Note that linear projections and orthogonal projections are identical when $n=1$.

The same idea can be extended to the span of infinite sets as well. For any collection $\{u_t\}_{t\in N_+} \subset L^2(\mathcal{H},\mathbb{P})$ of n-dimensional square integrable random vectors, we define

$$\hat{P}(X \mid u_1,u_2,\cdots) = \begin{pmatrix} proj_{\overline{W}} X_1 \\ \vdots \\ proj_{\overline{W}} X_n \end{pmatrix},$$

where

$$W = \operatorname{span}(\{u_{it} \mid 1 \le i \le n, t \in N_+\});$$

the only difference with the finite case is that each coordinate of X is projected onto the closure of W instead of W itself. The reason for this, made clearer below, is that only when we project onto the closure of W is a unique orthogonal projection guaranteed to exist.

We first show some preliminary results concerning linear projections, using a limit result that holds for arbitrary Hilbert spaces.

Lemma (Limiting Behavior of Projections in Hilbert Spaces)
Let $(V, \langle\cdot,\cdot\rangle)$ be an arbitrary Hilbert space over the real field, $\|\cdot\|$ the norm induced by $\langle\cdot,\cdot\rangle$, and d the metric induced by $\|\cdot\|$. Let $\{W_k\}_{k\in N_+}$ be an increasing sequence of linear subspaces of V, and denote the union of these subpsaces by $W = \bigcup_k W_k$. Fix some $x \in V$, and for each $k \in N_+$, let there exists a unique orthogonal projection y_k of x onto W_k.

Then, the sequence $\{y_k\}_{k\in N_+}$ converges in d to some $y \in \overline{W}$, where y is the unique orthogonal projection of x on $\overline{W}$.

Proof) We first establish that there exists a unique orthogonal projection of x onto $\overline{W}$. It is immediately clear that $\overline{W}$ is closed. To establish that it is a linear subspace of V over the real field, choose any $w, z \in \overline{W}$ and $a \in \mathbb{R}$. Then, there exist sequences $\{w_k\}_{k\in N_+}$ and $\{z_k\}_{k\in N_+}$ that are contained in W and converge in d to w and z. For any $k \in N_+$, there exist $m_1, m_2 \in N_+$ such that $w_k \in W_{m_1}$ and $z_k \in W_{m_2}$; since $\{W_m\}_{m\in N_+}$ is an increasing sequence of linear subspaces of V,

$$a \cdot w_k + z_k \in W_{\max(m_1,m_2)} \subset W.$$

Thus, $\{a \cdot w_k + z_k\}_{k\in N_+}$ is a sequence in W that converges in d to $a \cdot w + z$; since $\overline{W}$ is closed, $a \cdot w + z \in \overline{W}$, which establishes that $\overline{W}$ is closed under linear combinations. Furthermore, the zero vector is contained in each W_k (since they are linear subspaces) and therefore in $\overline{W}$ as well, proving that $\overline{W}$ is a linear subspace of V.

$\overline{W}$ is a closed and convex (this follows from linearity) subset of the Hilbert space V. By the Hilbert projection theorem, there exists a unique $y \in \overline{W}$ such that y is the orthogonal projection of x on W, that is,

$$\|x - y\| \le \|x - z\| \quad \forall z \in \overline{W}.$$

It remains to prove that this y is the limit of the sequence $\{y_k\}_{k\in N_+}$.

Note that

$$\overline{W} = \bigcup_k \overline{W_k};$$

the inclusion $\overline{W} \subset \bigcup_k \overline{W_k}$ holds because $\bigcup_k \overline{W_k}$ is a closed set containing W, while the fact that $\overline{W_k W}$ for any $k \in N_+$ implies the reverse inclusion. Since $y \in \overline{W}$, and $\{\overline{W_k}\}_{k \in N_+}$ is an increasing sequence of subsets of V, this means that there exists an $N \in N_+$ such that $y \in \overline{W_k}$ for any $k \geq N$. For any $k \geq N$, choose $z_k \in W_k$ such that

$$\|y - z_k\| < \frac{1}{k};$$

it is possible to make this choice because $y \in \overline{W_k}$. For $k < N$, choose any $z_k \in W_k$; then, the sequence $\{z_k\}_{k \in N_+}$ satisfies

$$z_k \in W_k \quad \forall k \in N_+ \quad \text{and} \quad z_k \to y \quad \text{in } d$$

For any $k \in N_+$, $z_k \in W_k$ and y_k is defined as the orthogonal projection of x on W_k, so we have the inequality

$$\|x - y_k\| \leq \|x - z_k\|.$$

Furthermore, because each $y_k \in W_k \subset \overline{W}$ and y is the unique orthogonal projection of x on $\overline{W}$,

$$\|x - y\| \leq \|x - y_k\|.$$

In other words, $\|x - y\| \leq \|x - y_k\| \leq \|x - z_k\|$ for any $k \in N_+$, so sending $k \to \infty$ on both sides yields

$$\lim_{k \to \infty} \|x - y_k\| = \|x - y\|.$$

Likewise, defining $v_k = \frac{y + y_k}{2} \in \overline{W}$ for any $k \in N_+$,

$$\|x - y\| \leq \|x - v_k\| \leq \frac{1}{2}\|x - y\| + \frac{1}{2}\|x - y_k\|$$

for any $k \in N_+$, so sending $k \to \infty$ on both sides yields

$$\lim_{k \to \infty} \|x - v_k\| = \|x - y\|.$$

Now we use the parallelogram law on inner product spaces to show that $y_k \to y$ in the metric d. By the parallelogram law, for any $k \in N_+$,

$$2 \cdot \|x - y_k\|^2 + 2 \cdot \|x - y\|^2 = \|y_k - y\|^2 + 4 \cdot \left\|x - \frac{y_k + y}{2}\right\|^2,$$

so that rearranging terms yields

$$\|y_k - y\|^2 = 2 \cdot \|x - y_k\|^2 + 2 \cdot \|x - y\|^2 - 4 \cdot \|x - v_k\|^2.$$

Sending $k \to \infty$ on both sides then yields

$$\lim_{k\to\infty} \|y_k - y\|^2 = 0,$$

so that $y_k \to y$ in the metric d.

Q.E.D.

Corollary Let $\{u_t\}_{t\in\mathbb{Z}} \subset L_n^2(\mathcal{H},\mathbb{P})$ and $X \in L_n^2(\mathcal{H},\mathbb{P})$. Then, for any $t \in \mathbb{Z}$,

$$\hat{P}(X \mid u_t, \cdots, u_{t-k}) \xrightarrow{L^2} \hat{P}(X \mid u_t, u_{t-1}, \cdots)$$

as $k \to \infty$.

Proof) Fix any $t \in \mathbb{Z}$. For any $k \in N_+$, define the linear subspace

$$W_k = \text{span}\left(\{u_{i,t-j} \mid 1 \le i \le n, 0 \le j \le k\}\right)$$

of the univariate L^2 space $L^2(\mathcal{H},\mathbb{P})$. Then, $\{W_k\}_{k\in N_+}$ is an increasing sequence of linear subspaces such that

$$W = \bigcup_k W_k = \text{span}\left(\{u_{i,t-j} \mid 1 \le i \le n, j \in \mathbb{N}\}\right).$$

For any $1 \le i \le n$, let $\hat{X}_i$ be the orthogonal projection of X_i onto $\overline{W}$ in the L^2-norm, and let $\hat{X}_i^{(k)}$ denote the orthogonal projection of X_i onto W_k in the L^2-norm for any $k \in N_+$. By definition,

$$\hat{P}(X \mid u_t, \cdots, u_{t-k}) = \begin{pmatrix} \hat{X}_1^{(k)} \\ \vdots \\ \hat{X}_n^{(k)} \end{pmatrix} \quad \forall k \in N_+ \quad \text{and} \quad \hat{P}(X \mid u_t, u_{t-1}, \cdots) = \begin{pmatrix} \hat{X}_1 \\ \vdots \\ \hat{X}_n \end{pmatrix}.$$

By the lemma above, $\hat{X}_i^{(k)} \xrightarrow{L^2} \hat{X}_i$ for each $1 \le i \le n$. Since

$$\begin{aligned} \left\|\hat{P}(X \mid u_t, u_{t-1}, \cdots) - \hat{P}(X \mid u_t, \cdots, u_{t-k})\right\|_{n,2} &= \left\|\left\|\hat{P}(X \mid u_t, u_{t-1}, \cdots) - \hat{P}(X \mid u_t, \cdots, u_{t-k})\right\|\right\|_2 \\ &\le \sum_{i=1}^n \left\|\hat{X}_i^{(k)} - \hat{X}_i\right\|_2 \end{aligned}$$

by Minkowski's inequality and the definition of the n-dimensional L^2-norm, taking $k \to \infty$ on both sides yields

$$\hat{P}(X \mid u_t, \cdots, u_{t-k}) \xrightarrow{L^2} \hat{P}(X \mid u_t, u_{t-1}, \cdots).$$

Q.E.D.

The next result shows us how to find the coefficients of linear projections on finite sets.

Lemma Let $\{u_1, \cdots, u_h\} \subset L^2(\mathcal{H}, \mathbb{P})$ be a sequence of n-dimensional square integrable random vectors such that

$$\mathbb{E}\left[u_i u_j'\right] = \begin{cases} \Sigma & \text{if } i = j \\ O & \text{if } i \neq j \end{cases}$$

for some positive definite $n \times n$ matrix Σ, and let $X \in L^2(\mathcal{H}, \mathbb{P})$ be an n-dimensional square integrable random vector. Then,

$$\hat{P}(X \mid u_1, \cdots, u_h) = \sum_{j=1}^{h} \mathbb{E}\left[X \cdot u_j'\right] \Sigma^{-1} \cdot u_j.$$

Proof) There exist $A_1, \cdots, A_h \in \mathbb{R}^{n \times n}$ such that

$$\hat{P}(X \mid u_1, \cdots, u_h) = \sum_{j=1}^{h} A_j \cdot u_j.$$

Since each coordinate of $\hat{P}(X \mid u_1, \cdots, u_h)$ is the orthogonal projection of X_i onto the linear subspace

$$W = \text{span}\left(\{u_{it} \mid 1 \leq i \leq n, 1 \leq t \leq h\}\right),$$

by the characterization of orthogonal projections we have

$$\mathbb{E}\left[(X - \hat{P}(X \mid u_1, \cdots, u_h))u_j'\right] = O$$

for any $1 \leq j \leq h$. This implies that

$$\mathbb{E}\left[X \cdot u_j'\right] = \mathbb{E}\left[\hat{P}(X \mid u_1, \cdots, u_h) \cdot u_j'\right] = \sum_{i=1}^{h} A_i \cdot \mathbb{E}\left[u_i u_j'\right] = A_j \cdot \Sigma$$

for any $1 \leq j \leq h$, which implies that

$$\hat{P}(X \mid u_1, \cdots, u_h) = \sum_{j=1}^{h} \mathbb{E}\left[X \cdot u_j'\right] \Sigma^{-1} \cdot u_j.$$

Q.E.D.

We now state and prove the main result of this section:

Theorem (Wold Representation Theorem)
Let $\{X_t\}_{t\in\mathbb{Z}}$ be an n-dimensional square integrable and weakly stationary process with mean zero and autocovariance function $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$. Assume that

$$\begin{pmatrix} \Gamma(0) & \cdots & \Gamma(h) \\ \vdots & \ddots & \vdots \\ \Gamma(h)' & \cdots & \Gamma(0) \end{pmatrix} \in \mathbb{R}^{nh\times nh}$$

is nonsingular for any $h \in N_+$, and that $X_t - \hat{P}(X_t \mid X_{t-1}, X_{t-2}, \cdots)$ has positive definite variance for some $t \in \mathbb{Z}$.

Then, there exists an n-dimensional white noise process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ with positive definite covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$, and a sequence $\{\Psi_j\}_{j\in\mathbb{N}}$ of $n \times n$ matrices such that

$$\sum_{j=0}^{\infty} \operatorname{tr}\left(\Psi_j \cdot \Sigma \cdot \Psi_j'\right) < +\infty$$

and

$$X_t = \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j} + \eta_t,$$

for any $t \in \mathbb{Z}$, where $\{\eta_t\}_{t\in\mathbb{Z}}$ satisfies

$$\hat{P}\left(\eta_t \mid X_{t-1}, X_{t-2}, \cdots\right) = \eta_t.$$

Proof) We first start by constructing the white noise process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. This process is defined as the difference of X_t and the linear projection of X_t on past values $X_{t-1}, X_{t-2}, \cdots$; it thus collects the one-period ahead linear forecast errors. Then, the linear process part of X_t is constructed as the linear projection of X_t on these projection errors. In effect, X_t is represented as a collection of updates following projection errors (the linear process part) and a part corresponding to the projection (η_t).

Step 1: Constructing the White Noise Process

As stated above, ε_t is defined as the one-period ahead linear forecast error of X_t, that is,

$$\varepsilon_t = X_t - \hat{P}(X_t \mid X_{t-1}, X_{t-2}, \cdots).$$

The difficult part is showing that $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a white noise process with positive definite covariance matrix.

For any $t \in \mathbb{Z}$, we established above that

$$\hat{P}(X_t \mid X_{t-1}, X_{t-2}, \cdots)$$

is the L^2-limit of the sequence $\{\hat{P}(X_t \mid X_{t-1}, \cdots, X_{t-h})\}_{h \in N_+}$. This implies that

$$\varepsilon_{h,t} = X_t - \hat{P}(X_t \mid X_{t-1}, \cdots, X_{t-h}) \xrightarrow{L^2} X_t - \hat{P}(X_t \mid X_{t-1}, X_{t-2}, \cdots) = \varepsilon_t$$

as $h \to \infty$. By the definition of the linear projection as a collection of orthogonal projections of the coordinates of X_t, we have

$$\mathbb{E}\left[\varepsilon_{h,t} \cdot X'_{t-i}\right] = O$$

for any $1 \leq i \leq h$ and $h \in N_+$, and likewise,

$$\mathbb{E}\left[\varepsilon_t \cdot X'_{t-i}\right] = O$$

for any $i \in N_+$.

Fix $h \in N_+$, and note that there exist $A_{1,t}^{(h)}, \cdots, A_{h,t}^{(h)} \in \mathbb{R}^{n \times n}$ such that

$$\hat{P}(X_t \mid X_{t-1}, \cdots, X_{t-h}) = \sum_{i=1}^{h} A_{i,t}^{(h)} \cdot X_{t-i}.$$

Then, for any $1 \leq j \leq h$,

$$\begin{aligned} O = \mathbb{E}\left[\varepsilon_{h,t} X'_{t-j}\right] &= \mathbb{E}\left[X_t \cdot X'_{t-j}\right] - \sum_{i=1}^{h} A_{i,t}^{(h)} \cdot \mathbb{E}\left[X_{t-i} \cdot X'_{t-j}\right] \\ &= \Gamma(j) - \sum_{i=1}^{h} A_{i,t}^{(h)} \cdot \Gamma(j-i). \end{aligned}$$

Collecting these equations, we end up with the equation

$$\begin{pmatrix} A_{1,t}^{(h)} & \cdots & A_{h,t}^{(h)} \end{pmatrix} \begin{pmatrix} \Gamma(0) & \cdots & \Gamma(h-1) \\ \vdots & \ddots & \vdots \\ \Gamma(h-1)' & \cdots & \Gamma(0) \end{pmatrix} = \begin{pmatrix} \Gamma(1) & \cdots & \Gamma(h) \end{pmatrix}.$$

By assumption, the matrix $nh \times nh$ matrix $\begin{pmatrix} \Gamma(0) & \cdots & \Gamma(h-1) \\ \vdots & \ddots & \vdots \\ \Gamma(h-1)' & \cdots & \Gamma(0) \end{pmatrix}$ is non-singular, so the coefficients are given as

$$\begin{pmatrix} A_{1,t}^{(h)} & \cdots & A_{h,t}^{(h)} \end{pmatrix} = \begin{pmatrix} \Gamma(1) & \cdots & \Gamma(h) \end{pmatrix} \begin{pmatrix} \Gamma(0) & \cdots & \Gamma(h-1) \\ \vdots & \ddots & \vdots \\ \Gamma(h-1)' & \cdots & \Gamma(0) \end{pmatrix}^{-1}$$

This implies that $A_{i,t}^{(h)}$ does not depend on the time t; we denote $A_{i,t}^{(h)} = A_i^{(h)}$ for any $1 \le i \le h$. We can now show that $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a white noise process as follows:

- **Mean of ε_t**

  For any $t \in \mathbb{Z}$,

$$\mathbb{E}[\varepsilon_{h,t}] = \sum_{i=1}^{h} A_i^{(h)} \cdot \mathbb{E}[X_{t-i}] = \mathbf{0},$$

  and because $\varepsilon_{h,t} \overset{L^2}{\to} \varepsilon_t$ as $h \to \infty$,

$$\mathbb{E}[\varepsilon_t] = \lim_{h\to\infty} \mathbb{E}[\varepsilon_{h,t}] = \mathbf{0}.$$

- **Covariance of $\varepsilon_{t-\tau}$ and X_t**

  For any $t \in \mathbb{Z}$, $\tau \in \mathbb{Z}$ and $h \in N_+$,

$$\begin{aligned}
\mathbb{E}\left[X_t \cdot \varepsilon'_{h,t-\tau}\right] &= \mathbb{E}\left[X_t \cdot X'_{t-\tau}\right] - \mathbb{E}\left[X_t \cdot \hat{P}(X_{t-\tau} \mid X_{t-\tau-1}, \cdots, X_{t-\tau-h})'\right] \\
&= \Gamma(\tau) - \sum_{i=1}^{h} \Gamma(\tau+i) \cdot A_i^{(h)\prime} \\
&= \Gamma(\tau) - \underbrace{\begin{pmatrix} \Gamma(\tau+1) & \cdots & \Gamma(\tau+h) \end{pmatrix} \begin{pmatrix} \Gamma(0) & \cdots & \Gamma(h-1) \\ \vdots & \ddots & \vdots \\ \Gamma(h-1)' & \cdots & \Gamma(0) \end{pmatrix}^{-1} \begin{pmatrix} \Gamma(1)' \\ \vdots \\ \Gamma(h)' \end{pmatrix}}_{G_\tau^{(h)}}.
\end{aligned}$$

  Note that $G_\tau^{(h)}$ depends only on τ and h. The left hand side converges to $\mathbb{E}[X_t \cdot \varepsilon_{t-\tau}]$ as $h \to \infty$ because $\varepsilon_{h,t-\tau} \overset{L^2}{\to} \varepsilon_{t-\tau}$, so it follows that

$$\lim_{h\to\infty} G_\tau^{(h)} = G_\tau := \Gamma(\tau) - \mathbb{E}\left[X_t \cdot \varepsilon'_{t-\tau}\right],$$

  and we can see that $\mathbb{E}\left[X_t \cdot \varepsilon'_{t-\tau}\right]$ does not depend on t.

  As a special case,

$$\mathbb{E}\left[X_t \cdot \varepsilon'_t\right] = \Gamma(0) - G_0.$$

- **The Covariance of ε_t and ε_s, $t \neq s$**

  First note that, for any $h \in N_+$ and $t, s \in \mathbb{Z}$ such that $t \le s$,

$$\mathbb{E}\left[\hat{P}(X_t \mid X_{t-1}, \cdots, X_{t-h}) \cdot \varepsilon'_s\right] = \sum_{i=1}^{h} A_i^{(h)} \cdot \mathbb{E}\left[X_{t-i}\varepsilon'_s\right] = O.$$

Since $\hat{P}(X_t \mid X_{t-1}, \cdots, X_{t-h}) \xrightarrow{L^2} \hat{P}(X_t \mid X_{t-1}, X_{t-2}, \cdots)$ as $h \to \infty$, the equality above implies that

$$\mathbb{E}\left[\hat{P}(X_t \mid X_{t-1}, X_{t-2}, \cdots) \cdot \varepsilon_s'\right] = \lim_{h\to\infty} \mathbb{E}\left[\hat{P}(X_t \mid X_{t-1}, \cdots, X_{t-h}) \cdot \varepsilon_s'\right] = O.$$

Now choose any $t, s \in \mathbb{Z}$, and assume without loss of generality that $t < s$. Since

$$\mathbb{E}\left[\varepsilon_t \varepsilon_s'\right] = \mathbb{E}\left[X_t \varepsilon_s'\right] - \mathbb{E}\left[\hat{P}(X_t \mid X_{t-1}, X_{t-2}, \cdots) \cdot \varepsilon_s'\right],$$

and $\mathbb{E}\left[X_t \varepsilon_s'\right] = O$ by the property of linear projections, we have

$$\mathbb{E}\left[\varepsilon_t \varepsilon_s'\right] = O.$$

$\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a pairwise uncorrelated sequence.

- **The Variance of ε_t**
  For any $t \in \mathbb{Z}$ and $h \in N_+$, note that

$$\begin{aligned}\mathbb{E}\left[\varepsilon_t \varepsilon_t'\right] &= \mathbb{E}\left[X_t \cdot \varepsilon_t'\right] - \mathbb{E}\left[\hat{P}(X_t \mid X_{t-1}, X_{t-2}, \cdots) \cdot \varepsilon_t'\right] \\ &= \mathbb{E}\left[X_t \cdot \varepsilon_t'\right] = \Gamma(0) - G_0 := \Sigma,\end{aligned}$$

  where the second equality uses the result that $\mathbb{E}\left[\hat{P}(X_t \mid X_{t-1}, X_{t-2}, \cdots) \cdot \varepsilon_s'\right] = O$ for any $t \le s$.

We have shown that $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a white noise process with covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$. Since Σ is the covariance matrix of every $X_t - \hat{P}(X_t \mid X_{t-1}, X_{t-2}, \cdots)$, which we assumed to be positive definite, Σ is positive definite.

Step 2: Constructing the Linear Process

Now that we have the desired white noise process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, we now construct the linear process part of X_t.

For any $t \in \mathbb{Z}$ and $h \in N_+$, define

$$Y_t = \hat{P}(X_t \mid \varepsilon_t, \varepsilon_{t-1}, \cdots)$$

and

$$Y_{h,t} = \hat{P}(X_t \mid \varepsilon_t, \cdots, \varepsilon_{t-h}).$$

By the above lemma,

$$
\begin{aligned}
Y_{h,t} &= \sum_{j=0}^{h} \mathbb{E}\left[X_t \cdot \varepsilon'_{t-j}\right] \Sigma^{-1} \cdot \varepsilon_{t-j} \\
&= \sum_{j=0}^{h} \underbrace{(\Gamma(j) - G_j)\,\Sigma^{-1}}_{\Psi_j} \cdot \varepsilon_{t-j}.
\end{aligned}
$$

We now show that the sequence $\{Y_{h,t}\}_{h \in N_+}$ converges in L^2 to Y_t:

- **Square Summability of $\{\Psi_j\}_{j \in \mathbb{N}}$**

  For any $h \in N_+$,

$$
\begin{aligned}
\mathbb{E}|X_t - Y_{h,t}|^2 &= \operatorname{tr}\left(\mathbb{E}\left[(X_t - Y_{h,t})(X_t - Y_{h,t})'\right]\right) \\
&= \operatorname{tr}\left(\mathbb{E}\left[\left(X_t - \sum_{j=0}^{h} \Psi_j \cdot \varepsilon_{t-j}\right)\left(X_t - \sum_{j=0}^{h} \Psi_j \cdot \varepsilon_{t-j}\right)'\right]\right) \\
&= \operatorname{tr}(\Gamma(0)) - \operatorname{tr}\left(\sum_{j=0}^{h} \Psi_j \cdot \mathbb{E}\left[\varepsilon_{t-j} X'_t\right] - \sum_{j=0}^{h} \mathbb{E}\left[X_t \varepsilon'_{t-j}\right] \Psi'_j + \sum_{j=0}^{h}\sum_{i=0}^{h} \Psi_j \mathbb{E}\left[\varepsilon_{t-j}\varepsilon'_{t-i}\right] \Psi'_i\right).
\end{aligned}
$$

  Since

$$
\mathbb{E}\left[X_t \cdot \varepsilon'_{t-j}\right] = \Gamma(j) - G_j = \Psi_j \cdot \Sigma,
$$

  we have

$$
\mathbb{E}|X_t - Y_{h,t}|^2 = \operatorname{tr}(\Gamma(0)) - \sum_{j=0}^{h} \operatorname{tr}\left(\Psi_j \Sigma \Psi'_j\right).
$$

  The left hand side is always non-negative, so

$$
\sum_{j=0}^{h} \operatorname{tr}\left(\Psi_j \Sigma \Psi'_j\right) \le \operatorname{tr}(\Gamma(0)) < +\infty.
$$

  This holds for any $h \in N_+$, and since each $\operatorname{tr}\left(\Psi_j \Sigma \Psi'_j\right)$, being the trace of a positive semidefinite matrix, is non-negative, taking $h \to \infty$ on both sides reveals that

$$
\sum_{j=0}^{\infty} \operatorname{tr}\left(\Psi_j \Sigma \Psi'_j\right) \le \operatorname{tr}(\Gamma(0)) < +\infty.
$$

- **The L^2-limit of $\{Y_{h,t}\}_{h \in N_+}$**

  $\{\Psi_j\}_{j \in \mathbb{N}}$ is a square summable sequence of $n \times n$ matrices, so by the results on linear processes established above,

$$Y_{h,t} = \sum_{j=0}^{h} \Psi_j \cdot \varepsilon_{t-j} \xrightarrow{L^2} Z_t$$

for some n-dimensional $Z_t \in L^2(\mathcal{H}, \mathbb{P})$. Since $Y_{h,t} \xrightarrow{L^2} Y_t$ by the result on Hilbert spaces established just above, the uniqueness of L^2-limits now tells us that $Y_t = Z_t$ almost surely, or that $\{Y_t\}_{t\in\mathbb{Z}}$ is the mean zero linear process defined as

$$Y_t = \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j}$$

for any $t \in \mathbb{Z}$.

We have just shown that X_t can be written as the sum of a linear process and an error component as follows:

$$X_t = Y_t + (X_t - Y_t) = \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j} + \eta_t.$$

The proof is completed by showing that the linear projection of η_t on $X_{t-1}, X_{t-2}, \cdots$ is itself.

Step 3: Establishing Predictability of η_t

By definition,

$$\eta_t = X_t - Y_t = X_t - \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j}.$$

For any $h \in N_+$,

$$Y_{h,t} - \varepsilon_t = \sum_{j=1}^{h} \Psi_j \cdot \varepsilon_{t-j};$$

this tells us that each coordinate of $Y_t - \varepsilon_t$ is contained in the linear subspace

$$W_h = \text{span}\left(\{\varepsilon_{i,t-j} \mid 1 \leq i \leq n, 1 \leq j \leq h\}\right),$$

which is itself contained in the closure of the linear subspace

$$W = \bigcup_h W_h.$$

Since $Y_t - \varepsilon_t$ is the L^2-limit of the sequence $\{Y_{h,t} - \varepsilon_t\}_{h \in N_+}$, it follows from the closedness of $\overline{W}$ that each coordinate of $Y_t - \varepsilon_t$ is contained in $\overline{W}$.

For any $j \in N_+$,

$$\varepsilon_{t-j} = X_{t-j} - \hat{P}(X_{t-j} \mid X_{t-j-1}, X_{t-j-2}, \cdots);$$

the same line of reasoning as above leads us to conclude that each coordinate of ε_{t-j} is contained in the closure of the linear subspace

$$V = \text{span}\left(\{X_{i,t-j} \mid 1 \leq i \leq n, j \in N_+\}\right).$$

Thus, $\overline{W} \subset \overline{V}$, and we can conclude that each coordinate of $Y_t - \varepsilon_t$ is contained in $\overline{V}$.

Finally, since

$$\begin{aligned}\eta_t = X_t - Y_t &= (X_t - \varepsilon_t) - (Y_t - \varepsilon_t) \\ &= \hat{P}(X_t \mid X_{t-1}, X_{t-2}, \cdots) - (Y_t - \varepsilon_t),\end{aligned}$$

each coordinate of η_t is contained in $\overline{V}$. This leads us to conclude that the orthogonal projection of each η_{it} on $\overline{V}$ is η_{it} itself, and since the linear projection of η_t on $X_{t-1}, X_{t-2}, \cdots$ is the collection of the orthogonal projections of each η_{it} on $\overline{V}$, we have the result

$$\eta_t = \hat{P}(\eta_t \mid X_{t-1}, X_{t-2}, \cdots).$$

Q.E.D.

Above, we have shown that any zero-mean weakly stationary time series $\{X_t\}_{t \in \mathbb{Z}}$ subject to very mild assumptions can be decomposed into two parts: a square summable linear process part $\sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j}$ and a predictable part η_t. In the construction of the linear process, we saw that it was nothing more than the weighted average of all the one-period ahead linear forecast errors, while the predictable part η_t is "predictable" in the sense that it can be perfectly linearly forecast using past values of X_t. Heuristically, this means that any X_t can be expressed as the sum of an expected/linearly predicted component and an unpredictable component, the latter of which is a function of the linear forecasts up to time t.

1.4 Asymptotic Theory for Linear Processes

Here we present an important asymptotic result pertaining to linear processes with i.i.d. innovations that have finite fourth moments. The analysis necessitates the use of the vectorization operator and the Kronecker product, two devices that will come in handy very often in the analysis to come.

1.4.1 Vectorization and Kronecker Product

Let $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{p\times q}$. The Kronecker product $A \bigotimes B$ of A and B is the $mp \times pq$ matrix defined as

$$A \bigotimes B = \begin{pmatrix} A_{11}B & \cdots & A_{1n}B \\ \vdots & \ddots & \vdots \\ A_{m1}B & \cdots & A_{mn}B \end{pmatrix}.$$

Clearly, the Kronecker product is not commutative. The following are additional properties of the Kronecker product:

Lemma (Properties of the Kronecker Product)
Let $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{p\times q}$, $C \in \mathbb{C}^{k\times l}$, and $D \in \mathbb{C}^{s\times r}$. Then, the following hold true:

i) $(A \bigotimes B) \bigotimes C = A \bigotimes (B \bigotimes C)$

ii) If $m = p$, $n = q$, $k = s$ and $l = r$,

$$(A+B) \bigotimes (C+D) = A \bigotimes C + B \bigotimes C + A \bigotimes D + B \bigotimes D$$

iii) If $n = p$ and $l = s$, then

$$(A \bigotimes C)(B \bigotimes D) = AB \bigotimes CD$$

iv) If A and B are nonsingular square matrices, then

$$(A \bigotimes B)^{-1} = (A^{-1} \bigotimes B^{-1})$$

v) If A and B are square matrices, then

$$\operatorname{tr}\left(A \bigotimes B\right) = \operatorname{tr}(A)\operatorname{tr}(B)$$

vi) Let A be an $m \times m$ square matrix with eigenvalues $\lambda_1, \cdots, \lambda_m \in \mathbb{C}$, and B a $p \times p$ square matrix with eigenvalues $\mu_1, \cdots, \mu_p \in \mathbb{C}$. Then, the eigenvalues of $A \bigotimes B$ are collected in the set

$$\{\lambda_i\mu_j \mid 1 \le i \le m, 1 \le j \le p\}.$$

vii) Let A be an $m \times m$ square matrix and B a $p \times p$ square matrix. Then,

$$\left|A \bigotimes B\right| = |A|^p |B|^m.$$

Proof) i) This follows immediately by noting that

$$\begin{aligned}(A \bigotimes B) \bigotimes C &= \begin{pmatrix} A_{11}B & \cdots & A_{1n}B \\ \vdots & \ddots & \vdots \\ A_{m1}B & \cdots & A_{mn}B \end{pmatrix} \bigotimes C \\ &= \begin{pmatrix} A_{11}(B \otimes C) & \cdots & A_{1n}(B \otimes C) \\ \vdots & \ddots & \vdots \\ A_{m1}(B \otimes C) & \cdots & A_{mn}(B \otimes C) \end{pmatrix} = A \bigotimes (B \bigotimes C).\end{aligned}$$

ii) This follows immediately by noting that

$$(A+B) \bigotimes (C+D) = \begin{pmatrix} (A_{11}+B_{11})(C+D) & \cdots & (A_{1n}+B_{1n})(C+D) \\ \vdots & \ddots & \vdots \\ (A_{m1}+B_{m1})(C+D) & \cdots & (A_{mn}+B_{mn})(C+D) \end{pmatrix}$$

$$= \begin{pmatrix} A_{11}C + B_{11}C + A_{11}D + B_{11}D & \cdots & A_{1n}C + B_{1n}C + A_{1n}D + B_{1n}D \\ \vdots & \ddots & \vdots \\ A_{m1}C + B_{m1}C + A_{m1}D + B_{m1}D & \cdots & A_{mn}C + B_{mn}C + A_{mn}D + B_{mn}D \end{pmatrix}$$

$$= A \bigotimes C + B \bigotimes C + A \bigotimes D + B \bigotimes D.$$

iii) This also follows immediately by noting that

$$\begin{aligned}(A \bigotimes C)(B \bigotimes D) &= \begin{pmatrix} A_{11}C & \cdots & A_{1n}C \\ \vdots & \ddots & \vdots \\ A_{m1}C & \cdots & A_{mn}C \end{pmatrix} \begin{pmatrix} B_{11}D & \cdots & B_{1q}D \\ \vdots & \ddots & \vdots \\ B_{n1}D & \cdots & B_{nq}D \end{pmatrix} \\ &= \begin{pmatrix} (\sum_{i=1}^n A_{1i}B_{i1})CD & \cdots & (\sum_{i=1}^n A_{1i}B_{iq})CD \\ \vdots & \ddots & \vdots \\ (\sum_{i=1}^n A_{mi}B_{i1})CD & \cdots & (\sum_{i=1}^n A_{mi}B_{iq})CD \end{pmatrix} \\ &= AB \bigotimes CD.\end{aligned}$$

iv) Suppose A and B are nonsingular square matrices. Then, by the preceding result,

$$(A \bigotimes B)(A^{-1} \bigotimes B^{-1}) = (I_m \bigotimes I_p) = I_{mp},$$

so that the inverse of $A \bigotimes B$ exists and is given as

$$(A \bigotimes B)^{-1} = (A^{-1} \bigotimes B^{-1}).$$

v) Suppose A and B are again nonsingular square matrices. Then,

$$\operatorname{tr}(A \bigotimes B) = \sum_{i=1}^{m} A_{ii} \operatorname{tr}(B) = \operatorname{tr}(A) \operatorname{tr}(B).$$

vi) Let the Schur decompositions of A and B be given as

$$A = PDP^{-1} \quad \text{and} \quad B = Q\Lambda Q^{-1},$$

where P, Q are unitary matrices, so that their inverses are their conjugate transposes, and D, Λ are upper triangular matrices. Since the characteristic polynomial of A and B are given by

$$\begin{aligned} ch_A(z) &= |A - z \cdot I_m| = |D - z \cdot I_m| = \prod_{i=1}^{m} (D_{ii} - z) \\ ch_B(z) &= |B - z \cdot I_p| = |\Lambda - z \cdot I_p| = \prod_{i=1}^{p} (\Lambda_{ii} - z). \end{aligned}$$

Therefore, the diagonal entries of D and Λ are the eigenvalues of A and B.

We can now see that

$$\begin{aligned} A \bigotimes B &= (PDP^{-1}) \bigotimes (Q\Lambda Q^{-1}) \\ &= (P \bigotimes Q)((DP^{-1}) \bigotimes (\Lambda Q^{-1})) \\ &= (P \bigotimes Q)(D \bigotimes \Lambda)(P^{-1} \bigotimes Q^{-1}) \\ &= (P \bigotimes Q)(D \bigotimes \Lambda)(P \bigotimes Q)^{-1}. \end{aligned}$$

Therefore, the characteristic polynomial of $A \bigotimes B$ is given as

$$\begin{aligned} ch_{A \bigotimes B}(z) &= \left|(P \bigotimes Q)(D \bigotimes \Lambda)(P \bigotimes Q)^{-1} - z \cdot I_{mp}\right| \\ &= \left|(D \bigotimes \Lambda) - z \cdot I_{mp}\right| \end{aligned}$$

$$= \prod_{i=1}^{m} \prod_{j=1}^{p} (D_{ii}\Lambda_{jj} - z)$$
$$= \prod_{i=1}^{m} \prod_{j=1}^{p} (\lambda_i \mu_j - z)$$

The eigenvalues of $A \bigotimes B$ are thus collected in the set

$$\{\lambda_i \mu_j \mid 1 \leq i \leq m, 1 \leq j \leq p\}.$$

vii) In light of the above result, we can see that

$$\left| A \bigotimes B \right| = ch_{A \bigotimes B}(0) = \prod_{i=1}^{m} \lambda_i \left(\prod_{j=1}^{p} \mu_j \right)$$
$$= \left(\prod_{i=1}^{m} \lambda_i \right)^p \left(\prod_{j=1}^{p} \mu_j \right)^m = |A|^p |B|^m.$$

Q.E.D.

The vectorization operator stacks the columns of a matrix to transform it into a vector. Formally, for any $A \in \mathbb{C}^{m\times n}$, the vectorization of A is the mn-dimensional vector defined as

$$\mathrm{vec}(A) = \begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix},$$

where $A_1, \cdots, A_n \in \mathbb{C}^m$ are the columns of A. The vec operator has very useful properties, especially in conjunction with the Kronecker product.

Lemma (Properties of the Vectorization Operator)
Let $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{p\times q}$, $C \in \mathbb{C}^{k\times l}$, and $D \in \mathbb{C}^{s\times r}$. Then, the following hold true:

i) If $n = q = 1$, we have

$$\mathrm{vec}(AB') = B \bigotimes A.$$

ii) If $m = p$ and $n = q$, we have

$$\mathrm{vec}(A)'\,\mathrm{vec}(B) = \mathrm{tr}(A'B).$$

iii) If $n = p$ and $q = k$, we have

$$\mathrm{vec}(ABC) = (C' \bigotimes A)\mathrm{vec}(B).$$

iv) Suppose the matrix product $ABCD$ is well-defined. Then,

$$\mathrm{tr}(ABCD) = \mathrm{vec}(D')'(C' \bigotimes A)\mathrm{vec}(B)$$

Proof) i) If A, B are column vectors, then

$$\mathrm{vec}(AB') = \mathrm{vec}\begin{pmatrix} A_{11}B_{11} & \cdots & A_{11}B_{p1} \\ \vdots & \ddots & \vdots \\ A_{m1}B_{11} & \cdots & A_{m1}B_{p1} \end{pmatrix}$$

$$= \begin{pmatrix} A_{11}B_{11} \\ \vdots \\ A_{m1}B_{11} \\ \vdots \\ A_{11}B_{p1} \\ \vdots \\ A_{m1}B_{p1} \end{pmatrix} = \begin{pmatrix} B_{11}A \\ \vdots \\ B_{p1}A \end{pmatrix} = B \bigotimes A.$$

ii) If A and B are matrices sharing the same dimensions, then

$$\operatorname{vec}(A)' \operatorname{vec}(B) = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}B_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{m} A'_{ji}B_{ij} = \operatorname{tr}(A'B).$$

iii) Suppose $n = p$ and $q = k$, so that the matrix product ABC is well-defined. Letting $B_1, \cdots, B_q \in \mathbb{C}^p$ be the columns of B, and $\{e_1, \cdots, e_p\} \subset \mathbb{R}^p$ the standard basis of $\mathbb{R}^p$,

$$B = \sum_{i=1}^{q} B_i e_i'.$$

It follows that

$$\begin{aligned}
\operatorname{vec}(ABC) &= \sum_{i=1}^{q} \operatorname{vec}(AB_i \cdot e_i' C) \\
&= \sum_{i=1}^{q} (C'e_i) \bigotimes (AB_i) \\
&= (C' \bigotimes A) \left(\sum_{i=1}^{q} (e_i \bigotimes B_i) \right) \\
&= (C' \bigotimes A) \left(\sum_{i=1}^{q} \operatorname{vec}(B_i e_i') \right) \\
&= (C' \bigotimes A) \operatorname{vec}(B),
\end{aligned}$$

where we used the preceding result on vectorization and one of the properties of the Kronecker product.

iv) Suppose that the matrix product $ABCD$ is well-defined. In this case, the previous results tell us that

$$\operatorname{vec}(D')' (C' \bigotimes A) \operatorname{vec}(B) = \operatorname{vec}(D')' \operatorname{vec}(ABC) = \operatorname{tr}(DABC) = \operatorname{tr}(ABCD).$$

Q.E.D.

1.4.2 Linear Processes and the Martingale Difference CLT

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional absolutely summable causal linear process with n-dimensional innovation process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. In many cases we will be interested in the asymptotic distribution of the quantity

$$\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} Y_{t-h}\varepsilon_t'$$

for any $1 \le h \le p$. It turns out that the vectorization of the above quantity follows an asymptotically normal distribution when appropriate regularity conditions are imposed on $\{\varepsilon_t\}_{t\in\mathbb{Z}}$.

Suppose $\{Y_t\}_{t\in\mathbb{Z}}$ has absolutely summable coefficients $\{\Psi_j\}_{j\in\mathbb{N}}$ and mean $\mu \in \mathbb{R}^n$, so that

$$Y_t = \mu + \Psi(L)\varepsilon_t$$

for any $t \in \mathbb{Z}$. Let $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$ be the autocovariance function of $\{Y_t\}_{t\in\mathbb{Z}}$.

Often, primarily when autoregressions are involved, we want to investigate the properties of sums involving the $np+1$-dimensional process $\{X_t\}_{t\in\mathbb{Z}}$, where $p \in N_+$ and

$$X_t = \begin{pmatrix} 1 \\ Y_{t-1} \\ \vdots \\ Y_{t-p} \end{pmatrix}$$

for any $t \in \mathbb{Z}$. The mean of X_t is given as

$$\bar{\mu} := \mathbb{E}[X_t] = \begin{pmatrix} 1 & \mu' & \cdots & \mu' \end{pmatrix}' \in \mathbb{R}^{np+1}.$$

We now make the following assumptions:

A1. Nonsingular Autocovariances

We assume that the matrix

$$Q = \begin{pmatrix} 1 & \mu' & \cdots & \mu' \\ \mu & \Gamma(0)+\mu\mu' & \cdots & \Gamma(p-1)+\mu\mu' \\ \vdots & \vdots & \ddots & \vdots \\ \mu & \Gamma(p-1)'+\mu\mu' & \cdots & \Gamma(0)+\mu\mu' \end{pmatrix}$$

is nonsingular.

A2. I.I.D. Innovations

This is our core assumption. We assume that $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is an i.i.d. process with positive definite covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ and finite fourth moments.

This implies, by absolute summability, that $\{Y_t\}_{t\in\mathbb{Z}}$ also has finite fourth moments and thus bounded second moments.

A3. Stronger Fourth Moment Assumption

In this assumption, we assume that $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ also has finite $4+2\eta$ moments for some $\eta > 0$.

We define parameters related to the skewness and kurtosis of the innovation process as follows:

$$\kappa_3 = \mathbb{E}\left[\varepsilon_t \left(\varepsilon_t' \bigotimes \varepsilon_t'\right)\right] \in \mathbb{R}^{n\times n^2}$$

and

$$\kappa_4 = \mathbb{E}\left[\varepsilon_t\varepsilon_t' \bigotimes \varepsilon_t\varepsilon_t'\right] \in \mathbb{R}^{n^2\times n^2}.$$

These quantities are designed so that, for any $a,b,c,d \in \{1,\cdots,n\}$ and $t \in \mathbb{Z}$,

$$\mathbb{E}\left[\varepsilon_{at}\varepsilon_{bt}\varepsilon_{ct}\right] = e_a' \cdot \kappa_3 \cdot (e_b \bigotimes e_c)$$
$$\mathbb{E}\left[\varepsilon_{at}\varepsilon_{bt}\varepsilon_{ct}\varepsilon_{dt}\right] = (e_a' \bigotimes e_b') \cdot \kappa_4 \cdot (e_c \bigotimes e_d),$$

where $\{e_1,\cdots,e_n\} \subset \mathbb{R}^n$ is the standard basis of $\mathbb{R}^n$.

The following is our main result:

Theorem (Asymptotic Results for Linear Processes with IID Errors)
Under assumptions A1 and A2, Y_{t-h} is independent of ε_t for any $t \in \mathbb{Z}$ and $h > 0$, and the following hold true:

$$\frac{1}{T}\sum_{t=p+1}^{T} X_t \xrightarrow{p} \bar{\mu}$$

$$\frac{1}{T}\sum_{t=p+1}^{T} X_t X_t' \xrightarrow{p} Q$$

$$\frac{1}{T}\sum_{t=p+1}^{T} \varepsilon_t\varepsilon_t' \xrightarrow{p} \Sigma$$

$$\frac{1}{T}\sum_{t=p+1}^{T} \operatorname{vec}\left(X_t\varepsilon_t'\right)\operatorname{vec}\left(X_t\varepsilon_t'\right)' \xrightarrow{p} \Sigma \bigotimes Q$$

$$\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} \operatorname{vec}\left(X_t\varepsilon_t'\right) \xrightarrow{d} N\left[\mathbf{0}, \Sigma \bigotimes Q\right].$$

If, in addition to assumptions A1 and A2, assumption A3 holds, then the following also hold true:

$$\frac{1}{T}\sum_{t=p+1}^{T}\begin{pmatrix}\operatorname{vec}(X_t\varepsilon_t')\\ \operatorname{vec}(\varepsilon_t\varepsilon_t'-\Sigma)\end{pmatrix}\begin{pmatrix}\operatorname{vec}(X_t\varepsilon_t')\\ \operatorname{vec}(\varepsilon_t\varepsilon_t'-\Sigma)\end{pmatrix}'\xrightarrow{p}\begin{pmatrix}\Sigma\bigotimes Q & (I_n\bigotimes\bar{\mu})\kappa_3\\ \kappa_3'(I_n\bigotimes\bar{\mu}') & \kappa_4-\Sigma\bigotimes\Sigma\end{pmatrix}$$

$$\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\begin{pmatrix}\operatorname{vec}(X_t\varepsilon_t')\\ \operatorname{vec}(\varepsilon_t\varepsilon_t'-\Sigma)\end{pmatrix}\xrightarrow{d}N\left[\mathbf{0},\begin{pmatrix}\Sigma\bigotimes Q & (I_n\bigotimes\bar{\mu})\kappa_3\\ \kappa_3'(I_n\bigotimes\bar{\mu}') & \kappa_4-\Sigma\bigotimes\Sigma\end{pmatrix}\right].$$

Proof) We first establish the convergence results pertaining to X_t. For any $0 < h \le p$, $\{Y_{t-h} - \mu\}_{t\in\mathbb{Z}}$ is a mean zero causal linear process with absolutely summable coefficients and i.i.d. innovations, so by the mean and covariance ergodicity of linear processes,

$$\frac{1}{T}\sum_{t=p+1}^{T}Y_{t-h}\xrightarrow{p}\mu$$

and

$$\frac{1}{T}\sum_{t=p+1}^{T}(Y_{t-h}-\mu)(Y_{t-h-j}-\mu)'\xrightarrow{p}\Gamma(j)$$

for any $0 \le j \le p-h$, which establishes that

$$\frac{1}{T}\sum_{t=p+1}^{T}Y_{t-h}Y_{t-h-j}'\xrightarrow{p}\Gamma(j)+\mu\mu'$$

for any $0 \le j \le p-h$. Putting together these results reveals that

$$\frac{1}{T}\sum_{t=p+1}^{T}X_t\xrightarrow{p}\bar{\mu}\quad\text{and}\quad\frac{1}{T}\sum_{t=p+1}^{T}X_tX_t'\xrightarrow{p}Q.$$

Let $\mathcal{F} = \{\mathcal{F}_t \mid t \in \mathbb{Z}\}$ be the filtration generated by $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. First note that, for any $t \in \mathbb{Z}$, X_t is $\mathcal{F}_{t-1}$-measurable. To see this, choose any $t \in \mathbb{Z}$ and $h > 0$, and note that

$$\mu+\sum_{j=0}^{m}\Psi_j\cdot\varepsilon_{t-h-j}\xrightarrow{a.s.}Y_{t-h}$$

as $m \to \infty$ due to the absolute summability of $\{\Psi_j\}_{j\in\mathbb{N}}$. For any $m \in N_+$, $\mu + \sum_{j=0}^{m}\Psi_j \cdot \varepsilon_{t-h-j}$, being an affine function of the innovations $\varepsilon_{t-h}, \cdots, \varepsilon_{t-h-m}$, is clearly $\mathcal{F}_{t-1}$-measurable, since $\mathcal{F}_{t-1}$ is the σ-algebra generated by the collection $\{\varepsilon_s \mid s \le t-1\}$. The (almost sure) pointwise limit of a sequence of $\mathcal{F}_{t-1}$-measurable functions is itself $\mathcal{F}_{t-1}$-measurable, so it follows that Y_{t-h} is $\mathcal{F}_{t-1}$-measurable. This holds for any $h > 0$, so X_t is also a $\mathcal{F}_{t-1}$-measurable random vector. The independence of ε_t and $\mathcal{F}_{t-1}$ implies that ε_t is also independent of X_t for any $t \in \mathbb{Z}$.

Note that we can express

$$\operatorname{vec}(X_t\varepsilon_t') = (\varepsilon_t \bigotimes X_t) = \left(I_n \bigotimes X_t\right)\varepsilon_t \quad \text{and} \quad \operatorname{vec}(\varepsilon_t\varepsilon_t' - \Sigma) = (\varepsilon_t \bigotimes \varepsilon_t) - \operatorname{vec}(\Sigma)$$

for any $t \in \mathbb{Z}$. Note also that, for any matrix $A \in \mathbb{R}^{m\times k}$,

$$\left\|A \bigotimes I_q\right\|^2 = q\sum_{i=1}^{m}\sum_{j=1}^{k}|A_{ij}|^2 = q\cdot\|A\|^2$$

and, similarly, $\|I_q \bigotimes A\|^2 = q\|A\|^2$.

We now show that $\{\operatorname{vec}(X_t\varepsilon_t')\}_{t\in\mathbb{Z}}$ is a martingale difference sequences with respect to $\mathcal{F}$ with bounded fourth moments under A1 and A2. Likewise, under assumption A3, $\{\operatorname{vec}(\varepsilon_t\varepsilon_t' - \Sigma)\}_{t\in\mathbb{Z}}$ is a martingale difference sequences with respect to $\mathcal{F}$ with bounded $2+\eta$ moments.

1) $\{\operatorname{vec}(X_t\varepsilon_t')\}_{t\in\mathbb{Z}}$
   Suppose assumptions A1 and A2 hold. For any $t \in \mathbb{Z}$, since

$$\begin{aligned}\left|\varepsilon_t \bigotimes X_t\right| &= \left|(I_n \bigotimes X_t)\varepsilon_t\right| \\ &\le \left\|I_n \bigotimes X_t\right\||\varepsilon_t| \\ &= \sqrt{n}|\varepsilon_t||X_t|,\end{aligned}$$

   we can see that

$$\begin{aligned}\mathbb{E}\left|\varepsilon_t \bigotimes X_t\right|^4 &\le n^2\mathbb{E}\left[|\varepsilon_t|^4|X_t|^4\right] \\ &= n^2\left(\mathbb{E}|X_t|^4\right)\left(\mathbb{E}|\varepsilon_t|^4\right) \qquad \text{(Independence of } X_{it} \text{ and } \varepsilon_{jt}\text{)} \\ &\le n^2\left(\sup_{s\in\mathbb{Z}}\mathbb{E}|X_s|^4\right)\left(\sup_{s\in\mathbb{Z}}\mathbb{E}|\varepsilon_s|^4\right) < +\infty,\end{aligned}$$

   where the second to last inequality follows from Hölder's inequality, and the last from the fact that $\{Y_t\}_{t\in\mathbb{Z}}$, and thus $\{X_t\}_{t\in\mathbb{Z}}$, has bounded fourth moments. Thus, $\{\operatorname{vec}(X_t\varepsilon_t')\}_{t\in\mathbb{Z}}$ has bounded fourth moments.
   Consequently, $\{\operatorname{vec}(X_t\varepsilon_t')\}_{t\in\mathbb{Z}}$ is integrable, and since X_t and ε_t are both $\mathcal{F}_t$-measurable for any $t \in \mathbb{Z}$, $\{\operatorname{vec}(X_t\varepsilon_t')\}_{t\in\mathbb{Z}}$ is $\mathcal{F}$-adapted. Furthermore, for any $t \in \mathbb{Z}$,

$$\begin{aligned}\mathbb{E}\left[\operatorname{vec}(X_t\varepsilon_t') \mid \mathcal{F}_{t-1}\right] &= \mathbb{E}\left[\varepsilon_t \bigotimes X_t \mid \mathcal{F}_{t-1}\right] \\ &= \mathbb{E}\left[(\varepsilon_t \bigotimes I_{np+1}) \mid \mathcal{F}_{t-1}\right]\cdot\left(I_n \bigotimes X_t\right) \\ &= \mathbb{E}\left[\varepsilon_t \bigotimes I_{np+1}\right]\cdot\left(I_n \bigotimes X_t\right) = \mathbf{0},\end{aligned}$$

   where the second equality uses the fact that X_t is $\mathcal{F}_{t-1}$-measurable and the third one follows from the independence of ε_t and $\mathcal{F}_{t-1}$. Taking expectations on both sides now reveals, via the law of iterated expectations, that $\{\text{vec}(X_t\varepsilon_t')\}_{t\in\mathbb{Z}}$ is a zero mean process, and by definition it is an MDS with respect to $\mathcal{F}$ that also has bounded fourth moments.

2) $\{\text{vec}(\varepsilon_t\varepsilon_t' - \Sigma)\}_{t\in\mathbb{Z}}$
Now assume A3 in addition to A1 and A2. For any $t \in \mathbb{Z}$, since

$$\begin{aligned}\left|\varepsilon_t \bigotimes \varepsilon_t\right| &\le \left\|\varepsilon_t \bigotimes I_n\right\| \left\|I_n \bigotimes \varepsilon_t\right\| \\ &= n|\varepsilon_t|^2,\end{aligned}$$

we have

$$\begin{aligned}\mathbb{E}\left|\varepsilon_t \bigotimes \varepsilon_t\right|^{2+\eta} &\le n^{2+\eta}\mathbb{E}|\varepsilon_t|^{4+2\eta} \\ &\le n^{2+\eta}\left(\sup_{s\in\mathbb{Z}} \mathbb{E}|\varepsilon_s|^{4+2\eta}\right) < +\infty.\end{aligned}$$

Thus, $\{\text{vec}(\varepsilon_t\varepsilon_t' - \Sigma)\}_{t\in\mathbb{Z}}$ has bounded $2+\eta$ moments.
This also shows that $\{\text{vec}(\varepsilon_t\varepsilon_t' - \Sigma)\}_{t\in\mathbb{Z}}$ is integrable, and since ε_t is $\mathcal{F}_t$-measurable for any $t \in \mathbb{Z}$, $\{\text{vec}(\varepsilon_t\varepsilon_t' - \Sigma)\}_{t\in\mathbb{Z}}$ is $\mathcal{F}$-adapted. Furthermore, for any $t \in \mathbb{Z}$,

$$\begin{aligned}\mathbb{E}\left[\text{vec}(\varepsilon_t\varepsilon_t' - \Sigma) \mid \mathcal{F}_{t-1}\right] &= \mathbb{E}\left[\varepsilon_t \bigotimes \varepsilon_t \mid \mathcal{F}_{t-1}\right] - \text{vec}(\Sigma) \\ &= \mathbb{E}\left[\varepsilon_t \bigotimes \varepsilon_t\right] - \text{vec}(\Sigma) = \mathbf{0},\end{aligned}$$

where the second equality uses the independence of ε_t and $\mathcal{F}_{t-1}$. Taking expectations on both sides now reveals, via the law of iterated expectations, that $\{\text{vec}(\varepsilon_t\varepsilon_t' - \Sigma)\}_{t\in\mathbb{Z}}$ is a zero mean process, and by definition it is an MDS with respect to $\mathcal{F}$ that is $L^{2+\eta}$-bounded.

From the above results, we are able to conclude that

$$\{\text{vec}(X_t\varepsilon_t') \mid t \in \mathbb{Z}\}$$

is an MDS with respect to $\mathcal{F}$ with bounded fourth moments under A1 and A2, and, if we assume A3 as well, that

$$\left\{\begin{pmatrix}\text{vec}(X_t\varepsilon_t') \\ \text{vec}(\varepsilon_t\varepsilon_t' - \Sigma)\end{pmatrix} \middle| t \in \mathbb{Z}\right\}$$

is an MDS with respect to $\mathcal{F}$ with bounded $2+\eta$ moments.

It remains to investigate the asymptotic variance of the above processes to implement the MDS CLT. Assume A1 to A3. Note first that we can write

$$\begin{aligned}
\operatorname{vec}(X_t\varepsilon_t')\operatorname{vec}(X_t\varepsilon_t')' &= \left(\varepsilon_t \bigotimes X_t\right)\left(\varepsilon_t' \bigotimes X_t'\right) = \left(\varepsilon_t\varepsilon_t' \bigotimes X_tX_t'\right), \\
\operatorname{vec}(X_t\varepsilon_t')\operatorname{vec}(\varepsilon_t\varepsilon_t' - \Sigma)' &= \left(I_n \bigotimes X_t\right)\varepsilon_t\left[\left(\varepsilon_t' \bigotimes \varepsilon_t'\right) - \operatorname{vec}(\Sigma)'\right] \\
\operatorname{vec}(\varepsilon_t\varepsilon_t' - \Sigma)\operatorname{vec}(\varepsilon_t\varepsilon_t' - \Sigma)' &= \left(\varepsilon_t\varepsilon_t' \bigotimes \varepsilon_t\varepsilon_t'\right) - \operatorname{vec}(\Sigma)\left(\varepsilon_t' \bigotimes \varepsilon_t'\right) \\
&\quad - \left(\varepsilon_t \bigotimes \varepsilon_t\right)\operatorname{vec}(\Sigma)' + \operatorname{vec}(\Sigma)\operatorname{vec}(\Sigma)'
\end{aligned}$$

for any $t \in \mathbb{Z}$.

Note that

$$\begin{aligned}
\mathbb{E}\left\|(\varepsilon_t\varepsilon_t' - \Sigma) \bigotimes X_tX_t'\right\| &\leq \left(\mathbb{E}\left\|(\varepsilon_t\varepsilon_t' - \Sigma) \bigotimes I_{np+1}\right\|\right)\left(\mathbb{E}\left\|I_n \bigotimes X_t\right\|\right) \\
&\qquad\qquad \text{(Independence of } \varepsilon_t \text{ and } X_t\text{)} \\
&= \sqrt{n(np+1)}\,(\mathbb{E}\|\varepsilon_t\varepsilon_t' - \Sigma\|)\,(\mathbb{E}|X_t|) \\
&\leq \sqrt{n(np+1)}\left(\mathbb{E}|\varepsilon_t|^2 + \|\Sigma\|\right)(\mathbb{E}|X_t|) \\
&= \sqrt{n(np+1)}\,(\operatorname{tr}(\Sigma) + \|\Sigma\|)\,(\mathbb{E}|X_t|)
\end{aligned}$$

for any $t \in \mathbb{Z}$. $\{X_t\}_{t\in\mathbb{Z}}$ has bounded fourth moments and thus first moments, so it follows that $\{(\varepsilon_t\varepsilon_t' - \Sigma) \bigotimes X_tX_t'\}_{t\in\mathbb{Z}}$ is L^1-bounded.

In addition,

$$\begin{aligned}
\mathbb{E}\left[(\varepsilon_t\varepsilon_t' - \Sigma) \bigotimes X_tX_t' \mid \mathcal{F}_{t-1}\right] &= \mathbb{E}\left[(\varepsilon_t\varepsilon_t' - \Sigma) \bigotimes I_{np+1} \mid \mathcal{F}_{t-1}\right] \cdot (I_n \bigotimes X_tX_t') \\
&= \mathbb{E}\left[(\varepsilon_t\varepsilon_t' - \Sigma) \bigotimes I_{np+1}\right] \cdot (I_n \bigotimes X_tX_t') = O
\end{aligned}$$

for any $t \in \mathbb{Z}$, so each element in the sequence of matrices

$$\left\{(\varepsilon_t\varepsilon_t' - \Sigma) \bigotimes X_tX_t'\right\}_{t\in\mathbb{Z}}$$

is an MDS with respect to $\mathcal{F}$ that is L^1-bounded. By the martingale WLLN,

$$\frac{1}{T}\sum_{t=p+1}^{T}\operatorname{vec}(X_t\varepsilon_t')\operatorname{vec}(X_t\varepsilon_t')' - \left[\Sigma \bigotimes \frac{1}{T}\sum_{t=1}^{T}X_tX_t'\right] \xrightarrow{L^1} O.$$

Using the fact that $\frac{1}{T}\sum_{t=1}^{T}X_tX_t' \xrightarrow{p} Q$, we now have the result

$$\frac{1}{T}\sum_{t=p+1}^{T}\operatorname{vec}(X_t\varepsilon_t')\operatorname{vec}(X_t\varepsilon_t')' \xrightarrow{p} \Sigma \bigotimes Q.$$

This result was derived without the need for assumption A3.

For any $t \in \mathbb{Z}$,

$$\mathbb{E}\left[\varepsilon_t\left[\left(\varepsilon_t'\bigotimes\varepsilon_t'\right)-\mathrm{vec}\left(\Sigma\right)'\right]\right]=\mathbb{E}\left[\varepsilon_t\left(\varepsilon_t'\bigotimes\varepsilon_t'\right)\right]-\mathbb{E}\left[\varepsilon_t\right]\cdot\mathrm{vec}\left(\Sigma\right)'=\mathbb{E}\left[\varepsilon_t\left(\varepsilon_t'\bigotimes\varepsilon_t'\right)\right]=\kappa_3.$$

Consider the sequence

$$\left\{\left(I_n\bigotimes X_t\right)\left(\varepsilon_t\left[\left(\varepsilon_t'\bigotimes\varepsilon_t'\right)-\mathrm{vec}\left(\Sigma\right)'\right]-\kappa_3\right)\right\}_{t\in\mathbb{Z}}.$$

For any $t \in \mathbb{Z}$,

$$\begin{aligned}
&\mathbb{E}\left\|\left(I_n\bigotimes X_t\right)\left(\varepsilon_t\left[\left(\varepsilon_t'\bigotimes\varepsilon_t'\right)-\mathrm{vec}\left(\Sigma\right)'\right]-\kappa_3\right)\right\| \\
&\qquad\le\mathbb{E}\left[\left\|I_n\bigotimes X_t\right\|\cdot\left|\varepsilon_t\right|\cdot\left|\varepsilon_t\bigotimes\varepsilon_t-\mathrm{vec}\left(\Sigma\right)\right|\right]+\left(\mathbb{E}\left\|I_n\bigotimes X_t\right\|\right)\left\|\kappa_3\right\| \\
&\qquad=\left(\mathbb{E}\left\|I_n\bigotimes X_t\right\|\right)\left[\mathbb{E}\left[\left|\varepsilon_t\right|\cdot\left|\varepsilon_t\bigotimes\varepsilon_t-\mathrm{vec}\left(\Sigma\right)\right|\right]-\left\|\kappa_3\right\|\right] \\
&\qquad\le\sqrt{n}\left(\mathbb{E}\left|X_t\right|\right)\left(\mathrm{tr}\left(\Sigma\right)^{\frac{1}{2}}\left(\mathbb{E}\left|\varepsilon_t\bigotimes\varepsilon_t-\mathrm{vec}\left(\Sigma\right)\right|^2\right)^{\frac{1}{2}}+\left\|\kappa_3\right\|\right).
\end{aligned}$$

Since $\{X_t\}_{t\in\mathbb{Z}}$ and $\{\varepsilon_t\bigotimes\varepsilon_t-\mathrm{vec}\left(\Sigma\right)\}_{t\in\mathbb{Z}}$ have bounded $2+\eta$ moments, it follows that the expression on the right hand side is bounded above.

Furthermore, for any $t \in \mathbb{Z}$,

$$\begin{aligned}
&\mathbb{E}\left[\left(I_n\bigotimes X_t\right)\left(\varepsilon_t\left[\left(\varepsilon_t'\bigotimes\varepsilon_t'\right)-\mathrm{vec}\left(\Sigma\right)'\right]-\kappa_3\right)\mid\mathcal{F}_{t-1}\right] \\
&\qquad=\left(I_n\bigotimes X_t\right)\cdot\mathbb{E}\left[\varepsilon_t\left[\left(\varepsilon_t'\bigotimes\varepsilon_t'\right)-\mathrm{vec}\left(\Sigma\right)'\right]-\kappa_3\mid\mathcal{F}_{t-1}\right] \\
&\qquad=\left(I_n\bigotimes X_t\right)\cdot\mathbb{E}\left[\varepsilon_t\left[\left(\varepsilon_t'\bigotimes\varepsilon_t'\right)-\mathrm{vec}\left(\Sigma\right)'\right]-\kappa_3\right]=O.
\end{aligned}$$

It follows that each element in the sequence $\left\{(I_n\bigotimes X_t)\left(\varepsilon_t\left[(\varepsilon_t'\bigotimes\varepsilon_t')-\mathrm{vec}\left(\Sigma\right)'\right]-\kappa_3\right)\right\}_{t\in\mathbb{Z}}$ is an MDS with respect to $\mathcal{F}$ that is L^1-bounded.

By the martingale WLLN,

$$\frac{1}{T}\sum_{t=p+1}^{T}\mathrm{vec}\left(X_t\varepsilon_t'\right)\mathrm{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma\right)'-\left(I_n\bigotimes\frac{1}{T}\sum_{t=p+1}^{T}X_t\right)\kappa_3\xrightarrow{p}O.$$

Using the fact that $\frac{1}{T}\sum_{t=p+1}^{T}X_t\xrightarrow{p}\bar{\mu}$, we have

$$\frac{1}{T}\sum_{t=p+1}^{T}\mathrm{vec}\left(X_t\varepsilon_t'\right)\mathrm{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma\right)'\xrightarrow{p}\left(I_n\bigotimes\bar{\mu}\right)\kappa_3.$$

Finally, since

$$\mathbb{E}\left[\operatorname{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma\right)\operatorname{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma\right)'\right]=\kappa_4-\operatorname{vec}\left(\Sigma\right)\operatorname{vec}\left(\Sigma\right)'$$

for any $t\in\mathbb{Z}$, each element in the sequence

$$\{\operatorname{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma\right)\operatorname{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma\right)'-\left(\kappa_4-\operatorname{vec}\left(\Sigma\right)\operatorname{vec}\left(\Sigma\right)'\right)\}_{t\in\mathbb{Z}}$$

is an i.i.d. sequence with mean zero and thus an MDS with respect to $\mathcal{F}$. They are also L^1-bounded because $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is L^4-bounded. Therefore, by the martingale WLLN,

$$\frac{1}{T}\sum_{t=p+1}^{T}\operatorname{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma\right)\operatorname{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma\right)'\xrightarrow{p}\kappa_4-\operatorname{vec}\left(\Sigma\right)\operatorname{vec}\left(\Sigma\right)'.$$

Putting these results together,

$$\frac{1}{T}\sum_{t=p+1}^{T}\begin{pmatrix}\operatorname{vec}\left(X_t\varepsilon_t'\right)\\ \operatorname{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma\right)\end{pmatrix}\begin{pmatrix}\operatorname{vec}\left(X_t\varepsilon_t'\right)\\ \operatorname{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma\right)\end{pmatrix}'\xrightarrow{p}\begin{pmatrix}\Sigma\bigotimes Q & (I_n\bigotimes\bar{\mu})\kappa_3\\ \kappa_3'(I_n\bigotimes\bar{\mu}') & \kappa_4-\Sigma\bigotimes\Sigma\end{pmatrix}.$$

By the MDS CLT, we now have

$$\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\begin{pmatrix}\operatorname{vec}\left(X_t\varepsilon_t'\right)\\ \operatorname{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma\right)\end{pmatrix}\xrightarrow{d}N\left[\mathbf{0},\begin{pmatrix}\Sigma\bigotimes Q & (I_n\bigotimes\bar{\mu})\kappa_3\\ \kappa_3'(I_n\bigotimes\bar{\mu}') & \kappa_4-\Sigma\bigotimes\Sigma\end{pmatrix}\right].$$

Since the asymptotic distribution of $\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\operatorname{vec}\left(X_t\varepsilon_t'\right)$ does not require assumption A3, it also holds when we only assume assumptions A1 and A2.

Q.E.D.

Vector Autoregressions

Here, we study the statistical properties of VAR models as a special case of weakly stationary linear processes.

An n-dimensional process $\{Y_t\}_{t\in\mathbb{Z}}$ is said to follow a (reduced-form) VAR(p) process if there exist a white noise process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, an intercept $\delta\in\mathbb{R}^n$ and coefficients $\Phi_1,\cdots,\Phi_p\in\mathbb{R}^{n\times n}$ such that

$$Y_t = \delta + \Phi_1 \cdot Y_{t-1} + \cdots + \Phi_p \cdot Y_{t-p} + \varepsilon_t$$

for any $t\in\mathbb{Z}$. Defining the lag polynomial

$$\Phi(L) = I_n - \Phi_1 \cdot L - \cdots - \Phi_p \cdot L^p,$$

we can also write

$$\Phi(L)Y_t = \delta + \varepsilon_t.$$

for any $t\in\mathbb{Z}$. The companion matrix for this VAR(p) process is defined as

$$F = \begin{pmatrix} \Phi_1 & \cdots & \Phi_{p-1} & \Phi_p \\ I_n & \cdots & O & O \\ \vdots & \ddots & \vdots & \vdots \\ O & \cdots & I_n & O \end{pmatrix} \in \mathbb{R}^{np\times np}.$$

Defining $Z_t = (Y_t', \cdots, Y_{t-p+1}')'$ and $c = (\delta', \mathbf{0}', \cdots, \mathbf{0}')$ and $u_t = (\varepsilon_t', \mathbf{0}', \cdots, \mathbf{0}')'$ for any $t\in\mathbb{Z}$, we can see that $\{Z_t\}_{t\in\mathbb{Z}}$ follows a VAR(1) process with mean reversion parameter F:

$$Z_t = c + F \cdot Z_{t-1} + u_t$$

for any $t\in\mathbb{Z}$. This shows us that a VAR process of any order can be expressed as a VAR(1) process. In particular, this tells us that the conditions under which $\{Y_t\}_{t\in\mathbb{Z}}$ is weakly stationary can simply be imposed on F. This is formally developed in the first subsection.

2.1 Conditions for Stationarity

Before we state and prove this result, we first extend the trace norm for real matrices to complex matrices. Let $A \in \mathbb{C}^{m\times n}$ be a complex $m \times n$ matrix. Then, we define

$$\|C\| = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|C_{ij}|^2\right)^{\frac{1}{2}}.$$

To see how this is a natural extension of the trace norm to complex matrices, let $A, B \in \mathbb{R}^{m\times n}$ be the real and imaginary parts of C. Then,

$$\begin{aligned}\|C\|^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}|C_{ij}|^2 &= \sum_{i=1}^{m}\sum_{j=1}^{n}|A_{ij}|^2 + \sum_{i=1}^{m}\sum_{j=1}^{n}|B_{ij}|^2 \\ &= \|A\|^2 + \|B\|^2,\end{aligned}$$

where we used the definition of the absolute value of complex numbers, which tells us that

$$\|C\| = \left(\|A\|^2 + \|B\|^2\right)^{\frac{1}{2}}.$$

Therefore, the trace norm for complex matrices is defined using the trace norm for real matrices in a similar manner to how the absolute value of complex numbers is defined.

This extension of the trace norm satisfies many properties of the trace norm for real matrices. We state a few of these below; let $z \in \mathbb{C}$, $x \in \mathbb{C}^n$, $A \in \mathbb{R}^{m\times n}$, and $B \in \mathbb{R}^{n\times p}$.

- $\|AB\| \leq \|A\| \cdot \|B\|$
  By definition,

  $$\begin{aligned}\|AB\|^2 &= \sum_{i=1}^{m}\sum_{j=1}^{p}\sum_{k=1}^{n}|A_{ik}B_{kj}|^2 \\ &\leq \sum_{i=1}^{m}\sum_{j=1}^{p}\sum_{k=1}^{n}\sum_{l=1}^{n}|A_{ik}B_{lj}|^2 = \|A\|^2\|B\|^2,\end{aligned}$$

  so that $\|AB\| \leq \|A\| \cdot \|B\|$.

- $|Ax| \leq \|A\| \cdot |x|$
  This follows immediately by noting that $\|y\| = |y|$ for any $y \in \mathbb{R}^n$ by definition.

- $\|A+B\| \leq \|A\| + \|B\|$
  This holds when A and B have the same dimensions. In this case,

  $$\|A+B\| = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|A_{ij}+B_{ij}|^2\right)^{\frac{1}{2}} \leq \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|A_{ij}|^2\right)^{\frac{1}{2}} + \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|B_{ij}|^2\right)^{\frac{1}{2}} = \|A\| + \|B\|,$$

where the inequality follows by viewing A_{ij}, B_{ij} as functions defined on the space $\{1, \cdots, m\} \times \{1, \cdots, n\}$ and applying Minkowski's inequality for L^2 spaces.

We now prove the result of interest.

Theorem (Eigenvalue Condition for Stationarity of VAR Process)
Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a VAR(p) process given as

$$Y_t = \delta + \Phi_1 \cdot Y_{t-1} + \cdots + \Phi_p \cdot Y_{t-p} + \varepsilon_t$$

for any $t \in \mathbb{Z}$, where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a white noise process with positive definite variance Σ. Let F be the companion matrix of $\{Y_t\}_{t\in\mathbb{Z}}$.

If $\{Y_t\}_{t\in\mathbb{Z}}$ is a square integrable process such that $\sup_{t\in\mathbb{Z}} \|Y_t\|_2 < +\infty$ and the eigenvalues of F lie within the unit circle, then $\{Y_t\}_{t\in\mathbb{Z}}$ is weakly stationary with a causal linear process representation

$$Y_t = \mu + \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j}$$

for some one-summable $\{\Psi_j\}_{j\in\mathbb{N}}$ and $\mu = \Psi(1)\delta$.

Proof) We first investigate the convergence properties of the companion matrix F. Let $\lambda_1, \cdots, \lambda_m \in \mathbb{C}$ be the distinct eigenvalues of F, and

$$J = \begin{pmatrix} J_1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & J_m \end{pmatrix}$$

the Jordan normal form of F, where the ith block J_i is a Jordan matrix corresponding to the ith distinct eigenvalue λ_i of F:

$$J_i = \begin{pmatrix} \lambda_i & 1 & \cdots & 0 & 0 \\ 0 & \lambda_i & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & \cdots & 0 & \lambda_i \end{pmatrix}.$$

Then, by the Jordan decomposition theorem, there exists a nonsingular matrix $\Lambda \in \mathbb{C}^{np\times np}$ such that

$$F = \Lambda J \Lambda^{-1}.$$

For any $j > 0$,

$$F^j = \Lambda J^j \Lambda^{-1} = \Lambda \begin{pmatrix} J_1^j & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & J_m^j \end{pmatrix} \Lambda^{-1}.$$

Letting J_i be a $k_i \times k_i$ matrix, where $k_i \leq np$,

$$J_i^j = \begin{pmatrix} \lambda_i^j & \begin{pmatrix} j \\ j-1 \end{pmatrix} \lambda_i^{j-1} & \cdots & \begin{pmatrix} j \\ j-k_i+1 \end{pmatrix} \lambda_i^{j-k_i+1} \\ 0 & \lambda_i^j & \cdots & \begin{pmatrix} j \\ j-k_i+2 \end{pmatrix} \lambda_i^{j-k_i+2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_i^j \end{pmatrix},$$

for any $j \geq k_i$. Since

$$\begin{pmatrix} j \\ j-k \end{pmatrix} = \frac{j!}{k!(j-k)!} = \frac{j \times (j-1) \times \cdots \times (j-k+1)}{k!} \leq j^k$$

for any $j \in N_+$ and $0 \leq k \leq j$, it follows that, for any $j \geq k_i$,

$$\begin{aligned} \left\| J_i^j \right\| &\leq \sum_{l=1}^{k_i} \sum_{q=1}^{k_i-l} \begin{pmatrix} j \\ j-q+1 \end{pmatrix} |\lambda_i|^{j-q+1} \\ &\leq \sum_{l=1}^{k_i} \sum_{q=1}^{k_i-l} j^{q-1} |\lambda_i|^{j-q+1} \\ &\leq |\lambda_i|^{j-k_i+1} \left(\sum_{l=1}^{k_i} \sum_{q=1}^{k_i-l} j^{q-1} \right) \\ &\leq |\lambda_i|^{j-k_i+1} \left(k_i \cdot \sum_{q=1}^{k_i} j^{q-1} \right) \\ &\leq k_i^2 \cdot j^{k_i-1} |\lambda_i|^{j-k_i+1}, \end{aligned}$$

where the third inequality is justified by the fact that $|\lambda_i| < 1$ by assumption. By implication, letting $k = \max_{1 \leq i \leq m} k_i$ and $\lambda = \max_{1 \leq i \leq m} |\lambda_i| < 1$,

$$\left\| J^j \right\| \leq \sum_{i=1}^{m} \left\| J_i^j \right\| \leq m k^2 j^{k-1} \lambda^{j-k+1}$$

for any $j \geq k$. Defining $\Theta_j = F^j$ for any $j \in \mathbb{N}$, since

$$\|\Theta_j\| \leq \|\Lambda\| \left\| \Lambda^{-1} \right\| \cdot \left\| J^j \right\|,$$

we have

$$(j \cdot \|\Theta_j\|)^{\frac{1}{j}} \leq \left(\|\Lambda\|\left\|\Lambda^{-1}\right\|k^2\right)^{\frac{1}{j}} \cdot \left(j^{\frac{1}{j}}\right)^k \lambda^{\frac{j-k+1}{j}}$$

for any $j \geq k_i$, so that

$$\lim_{j\to\infty} (j \cdot \|\Theta_j\|)^{\frac{1}{j}} = \lambda < 1.$$

By the root test for convergence,

$$\sum_{j=0}^{\infty} j \cdot \|\Theta_j\| < +\infty,$$

so that $\{\Theta_j\}_{j\in\mathbb{N}}$ is a one-summable and thus absolutely summable sequence of $np \times np$ matrices. The convergence of the series above also implies that $j\|\Theta_j\|$ and, by implication, $\|\Theta_j\|$ converges to 0 as $j \to \infty$. In other words,

$$\lim_{j\to\infty} \left\|F^j\right\| = 0.$$

Define $Z_t = (Y_t', \cdots, Y_{t-p+1}')'$, $c = (\delta', \mathbf{0}', \cdots, \mathbf{0}')$ and $u_t = (\varepsilon_t', \mathbf{0}', \cdots, \mathbf{0}')'$ for any $t \in \mathbb{Z}$. Then, $\{Z_t\}_{t\in\mathbb{Z}}$ follows a VAR(1) process with mean reversion parameter F

$$Z_t = c + FZ_{t-1} + u_t$$

for any $t \in \mathbb{Z}$. Additionally, we can see that

$$|Z_t| \leq \sum_{j=0}^{p-1} |Y_{t-j}|$$

for any $t \in \mathbb{Z}$, so that, by Minkowski's inequality,

$$\sup_{t\in\mathbb{Z}} \|Z_t\|_{np,2} \leq p \cdot \sup_{t\in\mathbb{Z}} \|Y_t\|_{n,2} < +\infty.$$

For any $T \in N_+$, the VAR(1) property of Z_t allows us to write

$$Z_t = \left(\sum_{j=0}^{T} \Theta_j\right) c + \sum_{j=0}^{T} \Theta_j \cdot u_{t-j} + F^{T+1} Z_{t-T-1}.$$

Since $\{\Theta_j\}_{j\in\mathbb{N}}$ is absolutely summable, the sequence

$$\left\{\sum_{j=0}^{T} \Theta_j \cdot u_{t-j}\right\}_{T\in\mathbb{N}}$$

converges almost surely and in L^2 to the square integrable random vector

$$\sum_{j=0}^{\infty} \Theta_j \cdot u_{t-j},$$

while the sequence $\{\sum_{j=0}^{T} \Theta_j\}_{T\in\mathbb{N}}$ converges to

$$\Theta(1) = \sum_{j=0}^{\infty} \Theta_j.$$

Finally, since

$$\left\|F^{T+1} Z_{t-T-1}\right\|_{np,2} \leq \left\|F^{T+1}\right\| \cdot \|Z_{t-T-1}\|_{np,2} \leq \left\|F^{T+1}\right\| \cdot \left(\sup_{s\in\mathbb{Z}} \|Z_s\|_{np,2}\right),$$

where the right hand side converges to 0 as $T \to \infty$, we can see that

$$\left\{ \left(\sum_{j=0}^{T} \Theta_j\right) c + \sum_{j=0}^{T} \Theta_j \cdot u_{t-j} + F^{T+1} Z_{t-T-1} \right\}_{T\in\mathbb{N}}$$

converges in L^2 to

$$\Theta(1)c + \sum_{j=0}^{\infty} \Theta_j \cdot u_{t-j}.$$

Since Z_t is also an L^2 limit of the above sequence, the uniqueness of mean square limits up to almost sure equivalence tells us that

$$Z_t = \Theta(1)c + \sum_{j=0}^{\infty} \Theta_j \cdot u_{t-j}.$$

almost surely. This holds for any $t \in \mathbb{Z}$, so $\{Z_t\}_{t\in\mathbb{Z}}$ is a weakly stationary causal linear process with mean $\Psi(1)c$ and one-summable coefficients $\{\Theta_j\}_{j\in\mathbb{N}}$.

Letting $\{\Psi_j\}_{j\in\mathbb{N}}$ collect the elements of $\{\Theta_j\}_{j\in\mathbb{N}}$ in the first $n\times n$ blocks, $\{\Psi_j\}_{j\in\mathbb{N}}$ is a sequence of one-summable $n\times n$ matrices such that

$$Y_t = \Psi(1)\delta + \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j}$$

for any $t \in \mathbb{Z}$.

Q.E.D.

The next lemma furnishes us with an equivalent formulation of the eigenvalue condition of the previous theorem in terms of determinants.

Lemma (Eigenvalue Condition in terms of Determinants)
Let $\Phi_1, \cdots, \Phi_p \in \mathbb{R}^{n\times n}$ and let F be the companion matrix defined as

$$F = \begin{pmatrix} \Phi_1 & \cdots & \Phi_{p-1} & \Phi_p \\ I_n & \cdots & O & O \\ \vdots & \ddots & \vdots & \vdots \\ O & \cdots & I_n & O \end{pmatrix}.$$

Then, the np eigenvalues of F lie within the unit circle if and only if all of the np roots of the polynomial

$$\det(I_n - \Phi_1 \cdot z - \cdots - \Phi_p \cdot z^p)$$

lie outside of the unit circle.

Proof) The characteristic function of F is given as

$$\det(F - \lambda \cdot I_n) = \left| \begin{pmatrix} \Phi_1 - \lambda \cdot I_n & \cdots & \Phi_{p-1} & \Phi_p \\ I_n & \cdots & O & O \\ \vdots & \ddots & \vdots & \vdots \\ O & \cdots & I_n & -\lambda \cdot I_n \end{pmatrix} \right|.$$

We proceed by induction for $\lambda \neq 0$. When $p = 1$, it is clear that

$$|\det(I_n - \lambda \cdot F)| = |\det(\Phi_1 - \lambda \cdot I_n)|.$$

Suppose that, for some $k \geq 1$,

$$|\det(I_n - \lambda \cdot F)| = \left|\det\Big(\Phi_p + \lambda \cdot \Phi_{p-1} + \cdots + \lambda^{p-1} \cdot \Phi_1 - \lambda^p \cdot I_n\Big)\right|$$

for $p = k$ and any specification of the coefficient matrices in F. Now suppose that $p = k+1$. In this case, defining

$$\tilde{F}_k = \begin{pmatrix} \Phi_1 - \lambda \cdot I_n & \cdots & \Phi_{k-1} & \Phi_k \\ I_n & \cdots & O & O \\ \vdots & \ddots & \vdots & \vdots \\ O & \cdots & I_n & -\lambda \cdot I_n \end{pmatrix},$$

$$B_k = \begin{pmatrix} \Phi_{k+1} \\ O \\ \vdots \\ O \end{pmatrix} \in \mathbb{R}^{nk\times n}$$

$$C_k = \begin{pmatrix} O & \cdots O & I_n \end{pmatrix} \in \mathbb{R}^{n \times nk},$$

we can see that

$$F - \lambda \cdot I_n = \begin{pmatrix} \tilde{F}_k & B_k \\ C_k & -\lambda \cdot I_n \end{pmatrix}.$$

By the formula for the determinant of block matrices, if $\lambda \neq 0$, then

$$\begin{aligned}
\det(F - \lambda \cdot I_n) &= \det(-\lambda \cdot I_n) \cdot \det\left(\tilde{F}_k + \frac{1}{\lambda} B_k C_k\right) \\
&= \left| \lambda^n \cdot \det\left[\begin{pmatrix} \Phi_1 - \lambda \cdot I_n & \cdots & \Phi_{k-1} & \Phi_k + \frac{1}{\lambda}\Phi_{k+1} \\ I_n & \cdots & O & O \\ \vdots & \ddots & \vdots & \vdots \\ O & \cdots & -\lambda \cdot I_n & O \\ O & \cdots & I_n & -\lambda \cdot I_n \end{pmatrix}\right] \right| \\
&= \left| \lambda^n \cdot \det\left(\frac{1}{\lambda}\Phi_{k+1} + \Phi_k + \cdots + \lambda^{k-1} \cdot \Phi_1 - \lambda^k \cdot I_n\right) \right| \\
&= \left| \det\left(\Phi_{k+1} + \lambda \cdot \Phi_k + \cdots + \lambda^k \cdot \Phi_1 - \lambda^{k+1} \cdot I_n\right) \right|,
\end{aligned}$$

where the second equality follows from the inductive hypothesis.

Therefore, when $\lambda \neq 0$, we can see that

$$\begin{aligned}
|\det(F - \lambda \cdot I_n)| &= \left| \det\left(\Phi_p + \lambda \cdot \Phi_{p-1} + \cdots + \lambda^{p-1}\Phi_1 - \lambda^p \cdot I_n\right) \right| \\
&= \frac{1}{|\lambda^p|} \cdot \left| \det\left(I_n - \Phi_1 \cdot \frac{1}{\lambda} - \cdots - \frac{1}{\lambda^p}\Phi_p\right) \right|.
\end{aligned}$$

In other words, λ is a non-zero eigenvalue of F if and only if its reciprocal is a root of $\det(I_n - \Phi_1 \cdot z - \cdots - \Phi_p \cdot z^p)$. On the other hand, if λ is a non-zero root of $\det(I_n - \Phi_1 \cdot z - \cdots - \Phi_p \cdot z^p)$, then $\frac{1}{\lambda}$ is a non-zero eigenvalue of F.

Suppose that the np roots of $\det(I_n - \Phi_1 \cdot z - \cdots - \Phi_p \cdot z^p)$ are all outside the unit circle. Then, since they are non-zero and their reciprocals are eigenvalues of F, the eigenvalues of F are all within the unit circle.

Conversely, suppose the eigenvalues of F are all within the unit circle. Assume, for the sake of contradiction, that $\det(I_n - \Phi_1 \cdot z - \cdots - \Phi_p \cdot z^p)$ has roots within or on the unit circle; these roots are non-zero since $\det(I_n - \Phi_1 \cdot z - \cdots - \Phi_p \cdot z^p) = 1$ when $z = 0$. Since the reciprocals of these roots are eigenvalues of F, this means that F has eigenvalues on or outside the unit circle, a contradiction. Thus, the roots of $\det(I_n - \Phi_1 \cdot z - \cdots - \Phi_p \cdot z^p)$ must all be outside the unit circle.

Q.E.D.

Retaining the notation of the above theorem, we have

$$\Phi(L)Y_t = \delta + \varepsilon_t$$

for any $t \in \mathbb{Z}$. The theorem tells us that, if the eigenvalues of F are all within the unit circle and the VAR process $\{Y_t\}_{t\in\mathbb{Z}}$ is L^2-bounded, then there exists a one-summable sequence $\{\Psi_j\}_{j\in\mathbb{N}}$ of $n \times n$ matrices such that

$$Y_t = \Psi(1)\delta + \Psi(L)\varepsilon_t$$

for any $t \in \mathbb{Z}$. It follows that

$$\Phi(1)\Psi(1)\delta + \Phi(L)\Psi(L)\varepsilon_t = \delta + \varepsilon_t$$

for any $t \in \mathbb{Z}$; therefore, using the notation defined earlier for linear processes,

$$\Psi(L) = \Phi(L)^{-1},$$

and we can equivalently write

$$Y_t = \Phi(1)^{-1}\delta + \Phi(L)^{-1}\varepsilon_t$$

for any $t \in \mathbb{Z}$.

Letting $\mu = \Psi(1)\delta$ be the mean of Y_t, since

$$Y_t = \delta + \Phi_1 \cdot Y_{t-1} + \cdots + \Phi_p \cdot Y_{t-p} + \varepsilon_t$$

for any $t \in \mathbb{Z}$ and $\{Y_t\}_{t\in\mathbb{Z}}$ is mean stationary, taking expectations on both sides yields

$$(I_n - \Phi_1 - \cdots - \Phi_p) \cdot \mu = \delta.$$

If $(I_n - \Phi_1 - \cdots - \Phi_p)$ is singular, then $\det(I_n - \Phi_1 - \cdots - \Phi_p) = 0$, which means that $\det(I_n - \Phi_1 \cdot z - \cdots - \Phi_p \cdot z^p)$ has a unit root. In light of the above lemma, this contradicts the fact that the eigenvalues of F are within the unit circle, so that $(I_n - \Phi_1 - \cdots - \Phi_p)$ must be nonsingular. This reveals that

$$\mu = (I_n - \Phi_1 - \cdots - \Phi_p)^{-1}\delta = \Psi(1)\delta,$$

and as such, it is consistent notation to write

$$\Psi(1) = (I_n - \Phi_1 - \cdots - \Phi_p)^{-1} = \Phi(1)^{-1}.$$

2.2 Topological Properties of Positive Definite Matrices

Now we investigate the asymptotic properties of the QMLE (equivalently, OLS) estimates of the parameters of stationary VARs. We first establish some topological facts concerning the space of all real symmetric matrices.

Define $S^{n\times n}$ as the space of all symmetric $n\times n$ matrices; we know that $(S^{n\times n}, \langle\cdot,\cdot\rangle_{tr})$ is an inner product space over the real field, where $\langle\cdot,\cdot\rangle_{tr}$ is the trace inner product defined as

$$\langle A,B\rangle_{tr} = \mathrm{tr}\,(A'B)$$

for any $A,B\in S^{n\times n}$, and that this inner product induces the trace norm $\|\cdot\|$ defined as

$$\|A\| = \mathrm{tr}\,(A'A)^{\frac{1}{2}},$$

which can be extended to encompass any real matrix $A\in\mathbb{R}^{m\times n}$ of arbitrary dimension.

Now let $PS^{n\times n}$ denote the space of all positive definite $n\times n$ matrices. This is clearly a convex cone contained in $S^{n\times n}$; for any $A,B\in PS^{n\times n}$ and $a\geq 0$, since $aA+B$ is also positive definite, it is contained in $S^{n\times n}$. We can also show that $PS^{n\times n}$ is an open subset of $S^{n\times n}$ with respect to the metric topology induced by the metric induced by the trace norm on $S^{n\times n}$:

Lemma (Set of Positive Definite Matrices is Open)
Let d_s be the metric on $S^{n\times n}$ induced by the trace norm $\|\cdot\|$. Then, $PS^{n\times n}$ is an open subset of $S^{n\times n}$ with respect to the metric d_s.

Proof) Let $\mathbb{T}^n$ be the unit circle in $\mathbb{R}^n$. We first show that the function $f: S^{n\times n}\to\mathbb{R}$ defined as

$$f(A) = \inf_{v\in\mathbb{T}^n} v'Av$$

for any $A\in S^{n\times n}$ is a continuous function. Initially, we can see that f is well-defined and takes values in $\mathbb{R}$ because $\mathbb{T}^n$ is a compact subset of $\mathbb{R}^n$ and the mapping $v\mapsto v'Av$ is continuous for any fixed $A\in S^{n\times n}$. By the extreme value theorem, this implies that, for any $A\in S^{n\times n}$, there exists a $v_A\in\mathbb{T}^n$ such that

$$f(A) = v_A'Av_A.$$

Now choose $A,B\in S^{n\times n}$. We consider two cases: initially, suppose that $f(A)\leq f(B)$. Then, since

$$v_B'Bv_B = \inf_{v\in\mathbb{T}^n} v'Bv \leq v_A'Bv_A,$$

we have

$$f(B) - f(A) = v_B' B v_B - v_A' A v_A \leq v_A' B v_A - v_A' A v_A = v_A'(B - A) v_A$$
$$= \operatorname{tr}\left((B-A) v_A v_A'\right) \leq \|B - A\| \cdot \|v_A v_A'\| \leq \|B - A\|,$$

where we used the Cauchy-Schwarz inequality and the fact that $\|v_A v_A'\| \leq |v_A|^2 = 1$. By symmetry, if $f(B) \leq f(A)$, then

$$f(A) - f(B) \leq \|A - B\|$$

as well. Therefore,

$$|f(A) - f(B)| \leq \|A - B\|,$$

and we can see that f is Lipschitz continuous with respect to the metric d_s.

Let $A \in PS^{n\times n}$. Since $v_A \in \mathbb{T}^n$ and is thus non-zero, by definition we have

$$f(A) = v_A' A v_A > 0.$$

Defining $\epsilon = f(A) > 0$, since f is continuous on $S^{n\times n}$, this means that there exists a $\delta > 0$ such that, for any $B \in S^{n\times n}$ such that $\|A - B\| < \delta$, we must also have

$$|f(A) - f(B)| < \epsilon,$$

and in particular, $f(B) > 0$. This indicates that, for any non-zero $v \in \mathbb{R}^n$,

$$v' B v = |v|^2 \cdot \left(\frac{v}{|v|}\right)' B \left(\frac{v}{|v|}\right) \geq |v|^2 \cdot v_B' B v_B = |v|^2 \cdot f(B) > 0,$$

where the last inequality follows because $|v|^2 > 0$ and $f(B) > 0$. Therefore, $B \in PS^{n\times n}$ by definition. In other words, as long as $\|A - B\| < \delta$, B is positive definite. Since A was chosen as an arbitrary positive definite matrix, this shows us that $PS^{n\times n}$ is an open subset of $S^{n\times n}$.

Q.E.D.

The operator $\text{vech}(\cdot)$ on $S^{n\times n}$ is defined as

$$\text{vech}(A) = \begin{pmatrix} A_{11} \\ \vdots \\ A_{n1} \\ A_{22} \\ \vdots \\ A_{n2} \\ \vdots \\ A_{n-1,n-1} \\ A_{n,n-1} \\ A_{nn} \end{pmatrix} \in \mathbb{R}^{n(n+1)/2}.$$

In other words, we stack the lower triangular elements of A. It is clear that $\text{vech}(\cdot)$ is a bijection from $S^{n\times n}$ onto $\mathbb{R}^{n(n+1)/2}$ and thus admits an inverse function $\text{vech}^{-1}(\cdot)$ from $\mathbb{R}^{n(n+1)/2}$ onto $S^{n\times n}$. We can also easily show that $\text{vech}^{-1}(\cdot)$ is continuous on $\mathbb{R}^{n(n+1)/2}$; for any $a, b \in \mathbb{R}^{n(n+1)/2}$, letting $A = \text{vech}^{-1}(a), B = \text{vech}^{-1}(b) \in S^{n\times n}$,

$$\left\|\text{vech}^{-1}(a) - \text{vech}^{-1}(b)\right\|^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} |A_{ij} - B_{ij}|^2 \leq 2 \cdot \sum_{j=1}^{n}\sum_{i=j}^{n} |A_{ij} - B_{ij}|^2 = 2 \cdot |a-b|^2,$$

where the inequality follows due to the symmetry of A and B.

Defining the subset $\mathcal{A}$ of $\mathbb{R}^{n(n+1)/2}$ as

$$\mathcal{A} = \text{vech}\left(PS^{n\times n}\right),$$

we can now see that

$$\mathcal{A} = \left(\text{vech}^{-1}\right)^{-1}\left(PS^{n\times n}\right),$$

that is, $\mathcal{A}$ is the inverse image of the open set $PS^{n\times n}$ under the continuous function $\text{vech}^{-1}(\cdot)$. By the definition of continuity, $\mathcal{A}$ is an open subset of $\mathbb{R}^{n(n+1)/2}$.

$\mathcal{A}$ is also convex. Choosing any $a = \text{vech}(A), b = \text{vech}(B) \in \mathcal{A}$ for some $A, B \in PS^{n\times n}$ and $t \in [0,1]$, note that

$$tA + (1-t)B \in PS^{n\times n}$$

since $PS^{n\times n}$ is a convex cone. Therefore,

$$ta + (1-t)b = \text{vech}(tA + (1-t)B) \in \mathcal{A}.$$

We have thus shown that $\mathcal{A}$ is an open convex subset of $\mathbb{R}^{n(n+1)/2}$, and as such differentiation of functions on $\mathcal{A}$ is well-defined.

2.2.1 Duplication, Elimination and Commutation Matrices

To facilitate the transition between the vech and vec operators, we define the duplication matrix $D_n \in \mathbb{R}^{n^2 \times n(n+1)/2}$ as the unique matrix such that

$$\text{vec}(A) = D_n \cdot \text{vech}(A)$$

for any $A \in S^{n\times n}$. Clearly, D_n has full rank $\frac{n(n+1)}{2}$, so that its pseudoinverse

$$D_n^+ = (D_n' D_n)^{-1} D_n' \in \mathbb{R}^{n(n+1)/2 \times n^2}$$

is well-defined and satisfies $D_n^+ D_n = I_{n(n+1)/2}$. Note that

$$D_n^+ \text{vec}(A) = \text{vech}(A)$$

for any $A \in \mathbb{R}^{n\times n}$; we call D_n^+ the elimination matrix.

Another matrix that often comes in handy is the commutation matrix. For any $m, n \in N_+$, and $A \in \mathbb{R}^{m\times n}$, the mn-dimensional vectors $\text{vec}(A)$ and $\text{vec}(A')$ contain the same elements, just arranged in a different manner. Thus, there exists a matrix $K_{mn} \in \mathbb{R}^{mn\times mn}$ such that

$$K_{mn}\text{vec}(A) = \text{vec}(A')$$

for any $A \in \mathbb{R}^{m\times n}$. We write K_n for K_{nn}.

We can construct K_{mn} by noting that A_{ij} is the $(m(j-1)+i)$th element of $\text{vec}(A)$, while it is the $(n(i-1)+j)$th element of $\text{vec}(A')$. This indicates that K_{mn} is the elementary matrix formed by interchanging the $(m(j-1)+i)$ and $(n(i-1)+j)$th columns (or equivalently, rows) of the identity matrix I_{mn}. To do this, we need only put the $(n(i-1)+j, m(j-1)+i)$ element of K_{mn} equal to 1 for any $1 \le i \le m$ and $1 \le j \le n$, and put every other element equal to 0.

Note that, by definition,

$$K_{nm}\text{vec}(A') = \text{vec}(A),$$

so that

$$K_{mn}K_{nm} \cdot \text{vec}(A') = \text{vec}(A')$$

for any $A \in \mathbb{R}^{m\times n}$. This indicates that $K_{nm} = K_{mn}^{-1}$, and since K_{mn}, being an elementary matrix, is also orthogonal, it follows that $K_{nm} = K_{mn}' = K_{mn}^{-1}$.

The following are some properties of commutation matrices:

Lemma (Properties of Commutation Matrices)
Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$. Then, the following hold true:

i) Pre-and Post-multiplying by commutation matrices exchanges the order of Kronecker products:

$$K_{pm}\left(A\bigotimes B\right) = \left(B\bigotimes A\right)K_{qn},$$
$$K_{pm}\left(A\bigotimes B\right)K_{nq} = B\bigotimes A.$$

ii) The vectorization of a Kronecker product can be written as the Kronecker product of vectorizations:

$$\text{vec}\left(A\bigotimes B\right) = (I_n\bigotimes K_{qm}\bigotimes I_p)\cdot(\text{vec}\,(A)\bigotimes \text{vec}\,(B)).$$

iii) $K_nD_n = D_n$ and $D_n^+K_n = D_n^+$.

iv) Defining $N_n := \frac{1}{2}(I_{n^2} + K_n)$, we have

$$N_n = D_nD_n^+.$$

Proof) i) For any $C \in \mathbb{R}^{q\times n}$,

$$\begin{aligned} K_{pm}\left(A\bigotimes B\right)\text{vec}\,(C) &= K_{pm}\cdot\text{vec}\,(BCA') = \text{vec}\,(AC'B') \\ &= \left(B\bigotimes A\right)\text{vec}\,(C') = \left(B\bigotimes A\right)K_{qn}\cdot\text{vec}\,(C). \end{aligned}$$

This holds for any $C \in \mathbb{R}^{q\times n}$, so we must have

$$K_{pm}\left(A\bigotimes B\right) = \left(B\bigotimes A\right)K_{qn}.$$

Since K_{nq} is the inverse of K_{qn}, we now have

$$K_{pm}\left(A\bigotimes B\right)K_{nq} = \left(B\bigotimes A\right).$$

ii) Let $\{e_1,\cdots,e_n\} \subset \mathbb{R}^n$ and $\{u_1,\cdots,u_q\} \subset \mathbb{R}^q$ be the standard bases of their respective euclidean spaces. Letting $a_1,\cdots,a_n \in \mathbb{R}^m$ be the columns of A and $b_1,\cdots,b_q \in \mathbb{R}^p$ those of B, note that we can write

$$A = \sum_{i=1}^{n} a_ie_i' \quad \text{and} \quad B = \sum_{i=1}^{q} b_iu_i'.$$

It now follows that

$$\begin{aligned}
\mathrm{vec}\left(A \bigotimes B\right) &= \sum_{i=1}^{n}\sum_{j=1}^{q} \mathrm{vec}\left(a_i e_i' \bigotimes b_j u_j'\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{q} \mathrm{vec}\left((a_i \bigotimes b_j)(e_i \bigotimes u_j)'\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{q} \left[(e_i \bigotimes u_j) \bigotimes (a_i \bigotimes b_j)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{q} \left[e_i \bigotimes (u_j \bigotimes a_i) \bigotimes b_j\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{q} \left[e_i \bigotimes \mathrm{vec}\left(a_i u_j'\right) \bigotimes b_j\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{q} (I_n \bigotimes K_{qm} \bigotimes I_p) \cdot \left[e_i \bigotimes \mathrm{vec}\left(u_j a_i'\right) \bigotimes b_j\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{q} (I_n \bigotimes K_{qm} \bigotimes I_p) \cdot \left[(e_i \bigotimes a_i) \bigotimes (u_j \bigotimes b_j)\right] \\
&= (I_n \bigotimes K_{qm} \bigotimes I_p) \left[\sum_{i=1}^{n}\sum_{j=1}^{q} \mathrm{vec}\left(a_i e_i' \bigotimes b_j u_j'\right)\right] \\
&= (I_n \bigotimes K_{qm} \bigotimes I_p) \cdot \mathrm{vec}\left(A \bigotimes B\right)
\end{aligned}$$

iii) For any $A \in S^{n\times n}$,

$$K_n D_n \mathrm{vech}\,(A) = K_n \mathrm{vec}\,(A) = \mathrm{vec}\,(A) = D_n \mathrm{vech}\,(A)\,,$$

so that $K_n D_n = D_n$. Similarly,

$$D_n^+ K_n = \left(D_n' D_n\right)^{-1} D_n' K_n = \left(D_n' D_n\right)^{-1} D_n' = D_n^+.$$

This shows us that $K_n D_n = D_n$ and $D_n^+ K_n = D_n^+$.

iv) Note first that

$$K_n D_n = D_n \quad \text{and} \quad D_n^+ K_n = D_n^+$$

imply

$$N_n D_n = D_n \quad \text{and} \quad D_n^+ N_n = D_n^+.$$

We can now show that $N_n - D_n D_n^+$ is a symmetric and idempotent matrix: it is immediately seen to be symmetric, and

$$\begin{aligned}\left(N_n - D_n D_n^+\right)^2 &= N_n^2 - N_n D_n D_n^+ - D_n D_n^+ N_n + D_n D_n^+ \\ &= N_n - D_n D_n^+,\end{aligned}$$

where we used the fact that $N_n^2 = N_n$ (which is clear from insepection). Therefore,

$$\begin{aligned}rank(N_n - D_n D_n^+) = \operatorname{tr}\left(N_n - D_n D_n^+\right) &= \operatorname{tr}(N_n) - \operatorname{tr}\left(D_n (D_n' D_n)^{-1} D_n'\right) \\ &= \frac{1}{2}\left(n^2 + n\right) - \frac{n(n+1)}{2} = 0,\end{aligned}$$

so that $N_n = D_n D_n^+$.

Q.E.D.

The last property implies that, for any $A, B \in PS^{n\times n}$,

$$\begin{aligned}D_n^+(A \bigotimes B) D_n D_n^+ (A^{-1} \bigotimes B^{-1}) D_n &= \frac{1}{2} D_n^+ (A \bigotimes B)\left(I_{n^2} + K_n\right)(A^{-1} \bigotimes B^{-1}) D_n \\ &= \frac{1}{2} D_n^+ \left(I_{n^2} + K_n\right)(A \bigotimes B)(A^{-1} \bigotimes B^{-1}) D_n \\ &= D_n^+ I_{n^2} D_n = D_n^+ D_n = I_{n(n+1)/2},\end{aligned}$$

so that

$$\left(D_n^+(A \bigotimes B) D_n\right)^{-1} = D_n^+(A^{-1} \bigotimes B^{-1}) D_n.$$

We can also infer that, for any $A, B \in \mathbb{R}^{n\times n}$,

$$\begin{aligned}(I_n \bigotimes A) D_n D_n^+ (I_n \bigotimes B) &= \frac{1}{2}(I_n \bigotimes A)(I_{n^2} + K_n)(I_n \bigotimes B) \\ &= \frac{1}{2}\left[(I_n \bigotimes AB) + (I_n \bigotimes A) K_n \cdot K_n (I_n \bigotimes B)\right] \\ &= \frac{1}{2}\left[(I_n \bigotimes AB) + K_n (A \bigotimes I_n)(B \bigotimes I_n) K_n\right] \\ &= \frac{1}{2}\left[(I_n \bigotimes AB) + (I_n \bigotimes AB)\right] = I_n \bigotimes AB.\end{aligned}$$

2.3 Maximum Likelihood Estimation

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional VAR(p) process with innovation process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ such that

$$Y_t = \delta + \Phi_1 \cdot Y_{t-1} + \cdots + \Phi_p \cdot Y_{t-p} + \varepsilon_t$$

for any $t \in \mathbb{Z}$ and some $\delta \in \mathbb{R}^n$, $\Phi_1, \cdots, \Phi_p \in \mathbb{R}^{n\times n}$. Assume that the variance $\Sigma \in \mathbb{R}^{n\times n}$ of the innovation process is positive definite.

We can collect the coefficients into the $(np+1)\times n$ matrix

$$\Pi = \begin{pmatrix} \delta' \\ \Phi_1' \\ \vdots \\ \Phi_p' \end{pmatrix}$$

and define the $np+1$-dimensional random vector X_t as

$$X_t = \begin{pmatrix} 1 \\ Y_{t-1} \\ \vdots \\ Y_{t-p} \end{pmatrix}$$

for any $t \in \mathbb{Z}$. Then, the model can be rewritten as

$$Y_t = \begin{pmatrix} \delta & \Phi_1 & \cdots & \Phi_p \end{pmatrix} \begin{pmatrix} 1 \\ Y_{t-1} \\ \vdots \\ Y_{t-p} \end{pmatrix} + \varepsilon_t = \Pi' X_t + \varepsilon_t.$$

We make the following assumptions:

A1. Stationarity

We assume that $\{Y_t\}_{t\in\mathbb{Z}}$ is a square integrable process, and that its companion matrix $F \in \mathbb{R}^{np\times np}$ has eigenvalues within the unit circle. By the stationarity results above, this implies that $\{Y_t\}_{t\in\mathbb{Z}}$ is a weakly stationary square integrable process with causal linear process representation

$$Y_t = \mu + \Psi(L)\varepsilon_t,$$

where $\{\Psi_j\}_{j\in\mathbb{N}}$ is a one-summable sequence of $n \times n$ matrices and $\mu = (I_n - \Phi_1 - \cdots - \Phi_p)^{-1}\delta$.

A2. Nonsingular Autocovariances

Letting $\Gamma : \mathbb{Z} \to \mathbb{R}^{n \times n}$ be the autocovariance function of $\{Y_t\}_{t \in \mathbb{Z}}$, we assume that the matrix

$$Q = \begin{pmatrix} 1 & \mu' & \cdots & \mu' \\ \mu & \Gamma(0)+\mu\mu' & \cdots & \Gamma(p-1)+\mu\mu' \\ \vdots & \vdots & \ddots & \vdots \\ \mu & \Gamma(p-1)'+\mu\mu' & \cdots & \Gamma(0)+\mu\mu' \end{pmatrix} \in \mathbb{R}^{(np+1)\times(np+1)}$$

is non-singular.

Since we can also write

$$Q = \mathbb{E}\left[X_t X_t'\right]$$

for any $t \in \mathbb{Z}$, we also assume, for the sample analogue of this moment condition, that $\sum_{t=1}^{T} X_t X_t'$ is almost surely nonsingular for large enough T.

A3. I.I.D. Innovations

We assume that the innovation process $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is i.i.d. with finite $4+2\eta$ moments, where $\eta > 0$:

$$\mathbb{E}|\varepsilon_t|^{4+2\eta} < +\infty.$$

As above, we define parameters related to the skewness and kurtosis of the innovation process as follows:

$$\kappa_3 = \mathbb{E}\left[\varepsilon_t \left(\varepsilon_t' \bigotimes \varepsilon_t'\right)\right] \in \mathbb{R}^{n \times n^2}$$

and

$$\kappa_4 = \mathbb{E}\left[\varepsilon_t \varepsilon_t' \bigotimes \varepsilon_t \varepsilon_t'\right] \in \mathbb{R}^{n^2 \times n^2}.$$

Furthermore, we also define

$$\bar{\mu} := \mathbb{E}\left[X_t\right] = \begin{pmatrix} 1 & \mu' & \cdots & \mu' \end{pmatrix}' \in \mathbb{R}^{np+1}.$$

Given these assumptions, preliminary asymptotic results follow immediately from the result stated in the previous section:

Theorem (Asymptotic Results for Stationary VARs with IID Errors)
Under the assumptions above, the following hold true:

$$\frac{1}{T}\sum_{t=p+1}^{T} X_t \xrightarrow{p} \bar{\mu}$$

$$\frac{1}{T}\sum_{t=p+1}^{T} X_t X_t' \xrightarrow{p} Q$$

$$\frac{1}{T}\sum_{t=p+1}^{T} \varepsilon_t \varepsilon_t' \xrightarrow{p} \Sigma$$

$$\frac{1}{T}\sum_{t=p+1}^{T} \begin{pmatrix} \mathrm{vec}\,(X_t\varepsilon_t') \\ \mathrm{vec}\,(\varepsilon_t\varepsilon_t' - \Sigma) \end{pmatrix} \begin{pmatrix} \mathrm{vec}\,(X_t\varepsilon_t') \\ \mathrm{vec}\,(\varepsilon_t\varepsilon_t' - \Sigma) \end{pmatrix}' \xrightarrow{p} \begin{pmatrix} \Sigma \bigotimes Q & (I_n \bigotimes \bar{\mu})\kappa_3 \\ \kappa_3'(I_n \bigotimes \bar{\mu}') & \kappa_4 - \Sigma \bigotimes \Sigma \end{pmatrix}$$

$$\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} \begin{pmatrix} \mathrm{vec}\,(X_t\varepsilon_t') \\ \mathrm{vec}\,(\varepsilon_t\varepsilon_t' - \Sigma) \end{pmatrix} \xrightarrow{d} N\left[\mathbf{0}, \begin{pmatrix} \Sigma \bigotimes Q & (I_n \bigotimes \bar{\mu})\kappa_3 \\ \kappa_3'(I_n \bigotimes \bar{\mu}') & \kappa_4 - \Sigma \bigotimes \Sigma \end{pmatrix}\right].$$

Proof) $\{Y_t\}_{t\in\mathbb{Z}}$ is a causal linear process with absolutely summable coefficients. In addition, its autocovariance function Γ and the innovation process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ satisfy the conditions of assumptions A1 and A2 of the previous section. Therefore, these results follow immediately.

Q.E.D.

In the special case that the errors are normally distributed, letting L be the Cholesky factor of Σ,

$$u_t := L^{-1}\varepsilon_t \sim N[\mathbf{0}, I_n],$$

making $\{u_t\}_{t\in\mathbb{Z}}$ an i.i.d. sequence of standard normally distributed random vectors. Since

$$\kappa_3 = \mathbb{E}\left[\varepsilon_t\left(\varepsilon_t' \bigotimes \varepsilon_t'\right)\right] = L \cdot \mathbb{E}\left[u_t\left(u_t' \bigotimes u_t'\right)\right](L' \bigotimes L')$$

and $\mathbb{E}[u_{it}u_{jt}u_{kt}] = 0$ for any $1 \le i,j,k \le n$, we have $\kappa_3 = O$ and thus $(I_n \bigotimes \bar{\mu})\kappa_3 = O$.

Similarly,

$$\kappa_4 = \mathbb{E}\left[\left(\varepsilon_t \bigotimes \varepsilon_t\right)\left(\varepsilon_t' \bigotimes \varepsilon_t'\right)\right] = (L \bigotimes L) \cdot \mathbb{E}\left[\left(u_t \bigotimes u_t\right)\left(u_t' \bigotimes u_t'\right)\right](L' \bigotimes L'),$$

and

$$\mathbb{E}\left[u_{it}u_{jt}u_{kt}u_{lt}\right] - = \begin{cases} 3 & \text{if } i = j = k = l \\ 1 & \text{if } i = j, k = l \text{ or } i = k, j = l \text{ or } i = l, j = k \\ 0 & \text{otherwise} \end{cases}$$

for any $1 \leq i, j, k \leq n$. We now have

$$\mathbb{E}\left[\left(u_t u_t' \bigotimes u_t u_t'\right)\right] - \left(I_n \bigotimes I_n\right) = I_{n^2} + K_n.$$

It follows that

$$\begin{aligned} \kappa_4 - \Sigma \bigotimes \Sigma &= \left(L \bigotimes L\right)\left[\mathbb{E}\left[\left(u_t u_t' \bigotimes u_t u_t'\right)\right] - \left(I_n \bigotimes I_n\right)\right]\left(L' \bigotimes L'\right) \\ &= \Sigma \bigotimes \Sigma + \left(L \bigotimes L\right) K_n \left(L' \bigotimes L'\right) \\ &= \left(I_{n^2} + K_n\right)\left(\Sigma \bigotimes \Sigma\right) = 2 D_n D_n^+ \left(\Sigma \bigotimes \Sigma\right). \end{aligned}$$

Therefore, when the errors are i.i.d. Gaussian,

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \begin{pmatrix} \operatorname{vec}\left(X_t \varepsilon_t'\right) \\ \operatorname{vec}\left(\varepsilon_t \varepsilon_t' - \Sigma\right) \end{pmatrix} \xrightarrow{d} N\left[\mathbf{0}, \begin{pmatrix} \Sigma \otimes Q & O \\ O & 2 D_n D_n^+ \left(\Sigma \otimes \Sigma\right) \end{pmatrix}\right].$$

2.3.1 Deriving the Score

Define the vector of parameters θ as

$$\theta = \begin{pmatrix} \beta \\ \gamma \end{pmatrix} \in \Theta := \mathbb{R}^{n(np+1)} \times \mathcal{A},$$

where we define $\beta = \mathrm{vec}\,(\Pi) \in \mathbb{R}^{n(np+1)}$ and $\gamma = \mathrm{vech}\,(\Sigma) \in \mathcal{A} \subset \mathbb{R}^{n(n+1)/2}$. Since the parameter space Θ is an open subset of $\mathbb{R}^{n(np+1)+n(n+1)/2}$, differentiation is defined at any point on Θ.

Suppose our sample contains observations from period 1 to T. Then, for each $p+1 \le t \le T$, the density of Y_t given its past values and initial values $Y_p, \cdots, Y_1$ is

$$f(Y_t \mid Y_{t-1}, \cdots, Y_0; \theta) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}(Y_t - \Pi' X_t)'\Sigma^{-1}(Y_t - \Pi' X_t)\Big),$$

so that the conditional Gaussian log-likelihood of the model given the initial values $Y_p, \cdots, Y_1$ is

$$\begin{aligned} l(\theta) &= \sum_{t=p+1}^{T} \log(f(Y_t \mid Y_{t-1}, \cdots, Y_0; \theta)) \\ &= -\frac{n(T-p)}{2}\log(2\pi) - \frac{T-p}{2}\log(|\Sigma|) - \frac{1}{2}\sum_{t=p+1}^{T}(Y_t - \Pi' X_t)'\Sigma^{-1}(Y_t - \Pi' X_t) \\ &= -\frac{n(T-p)}{2}\log(2\pi) - \frac{T-p}{2}\log(|\Sigma|) \\ &\quad -\frac{1}{2}\,\mathrm{tr}\left(\left[\sum_{t=p+1}^{T}(Y_t - (I_n \bigotimes X_t')\beta)(Y_t - (I_n \bigotimes X_t')\beta)'\right]\Sigma^{-1}\right) \end{aligned}$$

for any $\theta \in \Theta$. Denote

$$S(\beta) = \sum_{t=p+1}^{T}(Y_t - (I_n \bigotimes X_t')\beta)(Y_t - (I_n \bigotimes X_t')\beta)'$$

for any $\beta \in \mathbb{R}^{n(np+1)}$.

Recall the matrix derivative results

$$\begin{aligned} \frac{\partial |\Sigma|}{\partial \Sigma} &= |\Sigma| \cdot \Sigma^{-1} \\ \frac{\partial \Sigma^{-1}}{\partial x} &= -\Sigma^{-1}\frac{\partial \Sigma}{\partial x}\Sigma^{-1}. \end{aligned}$$

These can be used to conclude that

$$\begin{aligned} \frac{\partial\, \mathrm{tr}\,(S(\beta)\Sigma^{-1})}{\partial x} &= \mathrm{tr}\left(S(\beta) \cdot \frac{\partial \Sigma^{-1}}{\partial x}\right) \\ &= -\,\mathrm{tr}\left(S(\beta)\Sigma^{-1}\frac{\partial \Sigma}{\partial x}\Sigma^{-1}\right) = -\,\mathrm{tr}\left(\Sigma^{-1}S(\beta)\Sigma^{-1}\frac{\partial \Sigma}{\partial x}\right), \end{aligned}$$

so that

$$\frac{\partial \operatorname{tr}\left(S(\beta)\Sigma^{-1}\right)}{\partial \Sigma} = -\left(\Sigma^{-1}S(\beta)\Sigma^{-1}\right)' = -\Sigma^{-1}S(\beta)\Sigma^{-1}.$$

The conditional score function is now given as

$$s(\theta) = \frac{\partial l(\theta)}{\partial \theta} = \begin{pmatrix} \frac{\partial l(\theta)}{\partial \beta} \\ \frac{\partial l(\theta)}{\partial \gamma} \end{pmatrix} = \begin{pmatrix} \sum_{t=p+1}^{T}(I_n \bigotimes X_t)\Sigma^{-1}\left(Y_t - (I_n \bigotimes X_t')\beta\right) \\ -\frac{T-p}{2}\operatorname{vech}\left(\Sigma^{-1}\right) + \frac{1}{2}\operatorname{vech}\left(\Sigma^{-1}S(\beta)\Sigma^{-1}\right) \end{pmatrix}.$$

Note that, for any $v \in \mathbb{R}^n$,

$$(I_n \bigotimes X_t)\Sigma^{-1}v = (I_n \bigotimes X_t)\operatorname{vec}\left(v'\Sigma^{-1}\right) = \operatorname{vec}\left(X_t v'\Sigma^{-1}\right) = (\Sigma^{-1} \bigotimes X_t)v,$$

so that $(I_n \bigotimes X_t)\Sigma^{-1} = (\Sigma^{-1} \bigotimes X_t)$. Using this property, we can see that

$$(I_n \bigotimes X_t)\Sigma^{-1}(I_n \bigotimes X_t') = \Sigma^{-1} \bigotimes X_t X_t'$$

and

$$(I_n \bigotimes X_t)\Sigma^{-1}Y_t = (\Sigma^{-1} \bigotimes X_t)Y_t = \operatorname{vec}\left(X_t Y_t'\Sigma^{-1}\right) = (\Sigma^{-1} \bigotimes I_{np+1}) \cdot \operatorname{vec}\left(X_t Y_t'\right)$$

for any $p+1 \leq t \leq T$. Furthermore, we have

$$\operatorname{vech}\left(\Sigma^{-1}S(\beta)\Sigma^{-1}\right) = D_n^+ \cdot \operatorname{vec}\left(\Sigma^{-1}S(\beta)\Sigma^{-1}\right) = D_n^+(\Sigma^{-1} \bigotimes \Sigma^{-1})\operatorname{vec}\left(S(\beta)\right).$$

This allows us to write the score function as

$$s(\theta) = \begin{pmatrix} (\Sigma^{-1} \bigotimes I_{np+1})\operatorname{vec}\left(\sum_{t=p+1}^{T} X_t Y_t'\right) - \left(\Sigma^{-1} \bigotimes \sum_{t=p+1}^{T} X_t X_t'\right)\beta \\ \frac{1}{2} \cdot D_n^+(\Sigma^{-1} \bigotimes \Sigma^{-1})\operatorname{vec}\left(S(\beta) - (T-p)\Sigma\right) \end{pmatrix}.$$

2.3.2 The QMLE and its Asymptotic Properties

The (Quasi) MLE $\hat{\theta}_T$ of θ is defined as the unique vector such that $s(\hat{\theta}_T) = \mathbf{0}$; specifically, we have

$$\hat{\beta}_T = \left[I_n \bigotimes \left(\sum_{t=p+1}^{T} X_t X_t' \right)^{-1} \right] \text{vec} \left(\sum_{t=p+1}^{T} X_t Y_t' \right)$$

$$= \text{vec} \left(\left(\sum_{t=p+1}^{T} X_t X_t' \right)^{-1} \left(\sum_{t=p+1}^{T} X_t Y_t' \right) \right)$$

$$\hat{\Sigma}_T = \frac{1}{T-p} S(\hat{\beta}_T) = \frac{1}{T-p} \sum_{t=p+1}^{T} (Y_t - \hat{\Pi}_T' X_t)(Y_t - \hat{\Pi}_T' X_t)',$$

where the MLE of Π is given as

$$\hat{\Pi}_T = \left(\sum_{t=p+1}^{T} X_t X_t' \right)^{-1} \left(\sum_{t=p+1}^{T} X_t Y_t' \right).$$

These estimators can be interpreted as GMM estimators with empirical moment/identification conditions given as $s(\theta) = \mathbf{0}$.

Due to their formal similarity to the least squares estimators of the parameters, we can naturally hypothesize that the MLE $\hat{\theta}_T$ is consistent for the true parameters θ_0, where the 0 subscript denotes true values. We confirm below that this is indeed the case under our assumptions:

Theorem (Consistency of MLEs)
Under our assumptions, the MLEs of Π and Σ are consistent:

$$\hat{\theta}_T \xrightarrow{p} \theta_0.$$

Proof) Using the fact that $Y_t = \Pi_0' X_t + \varepsilon_t$ for any $t \in \mathbb{Z}$,

$$\hat{\Pi}_T = \left(\sum_{t=p+1}^{T} X_t X_t' \right)^{-1} \left(\sum_{t=p+1}^{T} X_t Y_t' \right)$$

$$= \Pi_0 + \left(\frac{1}{T} \sum_{t=p+1}^{T} X_t X_t' \right)^{-1} \left(\frac{1}{T} \sum_{t=p+1}^{T} X_t \varepsilon_t' \right).$$

We proved above that

$$\frac{1}{T} \sum_{t=p+1}^{T} X_t X_t' \xrightarrow{p} Q \quad \text{and} \quad \frac{1}{T} \sum_{t=p+1}^{T} X_t \varepsilon_t' \xrightarrow{p} O.$$

Therefore,

$$\hat{\Pi}_T \xrightarrow{p} \Pi_0,$$

which implies that $\hat{\beta}_T \xrightarrow{p} \beta_0$ as well.

As for $\hat{\Sigma}_T$, note that

$$\begin{aligned}
\frac{T}{T-p}\hat{\Sigma}_T &= \frac{1}{T}\sum_{t=p+1}^{T}(Y_t - \hat{\Pi}_T' X_t)(Y_t - \hat{\Pi}_T' X_t)' \\
&= \frac{1}{T}\sum_{t=p+1}^{T}\left[\left(\Pi_0 - \hat{\Pi}_T\right)' X_t + \varepsilon_t\right]\left[\left(\Pi_0 - \hat{\Pi}_T\right)' X_t + \varepsilon_t\right]' \\
&= \left(\Pi_0 - \hat{\Pi}_T\right)'\left(\frac{1}{T}\sum_{t=p+1}^{T} X_t X_t'\right)\left(\Pi_0 - \hat{\Pi}_T\right) + \left(\Pi_0 - \hat{\Pi}_T\right)'\left(\frac{1}{T}\sum_{t=p+1}^{T} X_t \varepsilon_t'\right) \\
&\quad + \left(\frac{1}{T}\sum_{t=p+1}^{T} X_t \varepsilon_t'\right)'\left(\Pi_0 - \hat{\Pi}_T\right) + \frac{1}{T}\sum_{t=p+1}^{T}\varepsilon_t\varepsilon_t' \\
&\xrightarrow{p} \Sigma_0
\end{aligned}$$

by the consistency of $\hat{\Pi}_T$ and the result, derived above, that $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is variance stationary.

Q.E.D.

We can actually show a stronger result, namely that $\hat{\theta}_T - \theta_0$ is $O_p(T^{-1/2})$ with an asymptotically normal distribution:

Theorem (Asymptotic Normality of MLEs)
Under our assumptions, the MLEs of Π and Σ have the following asymptotic distribution:

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{d} N\left[\mathbf{0}, \begin{pmatrix} \Sigma_0 \bigotimes Q^{-1} & \left(I_n \bigotimes Q^{-1}\bar{\mu}\right)\kappa_3 D_n^{+\prime} \\ D_n^{+}\kappa_3'\left(I_n \bigotimes \bar{\mu}' Q^{-1}\right) & D_n^{+}\left(\kappa_4 - \Sigma_0 \bigotimes \Sigma_0\right) D_n^{+\prime} \end{pmatrix}\right]$$

Proof) Note that

$$\hat{\beta}_T = \text{vec}\left(\hat{\Pi}_T\right) = \beta_0 + \left[I_n \bigotimes \left(\frac{1}{T}\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\right] \text{vec}\left(\frac{1}{T}\sum_{t=p+1}^{T} X_t \varepsilon_t'\right).$$

Since $\text{vec}\left(\frac{1}{T}\sum_{t=p+1}^{T} X_t \varepsilon_t'\right) = O_p(1)$, we can see that $\sqrt{T}\left(\hat{\beta}_T - \beta_0\right) = O_p(1)$. Likewise,

$$\begin{aligned}\hat{\Sigma}_T - \Sigma = &\left(\Pi_0 - \hat{\Pi}_T\right)'\left(\frac{1}{T}\sum_{t=p+1}^{T} X_t X_t'\right)\left(\Pi_0 - \hat{\Pi}_T\right) + \left(\Pi_0 - \hat{\Pi}_T\right)'\left(\frac{1}{T}\sum_{t=p+1}^{T} X_t \varepsilon_t'\right) \\ &+ \left(\frac{1}{T}\sum_{t=p+1}^{T} X_t \varepsilon_t'\right)'\left(\Pi_0 - \hat{\Pi}_T\right) + \frac{1}{T}\sum_{t=p+1}^{T}\left(\varepsilon_t \varepsilon_t' - \Sigma_0\right),\end{aligned}$$

so that

$$\begin{aligned}\text{vech}\left(\hat{\Sigma}_T\right) - \text{vech}\left(\Sigma_0\right) &= D_n^{+}\text{vec}\left(\hat{\Sigma}_T - \Sigma\right) \\ &= D_n^{+}\left[\left(\Pi_0 - \hat{\Pi}_T\right)' \bigotimes \left(\Pi_0 - \hat{\Pi}_T\right)'\right]\text{vec}\left(\frac{1}{T}\sum_{t=p+1}^{T} X_t X_t'\right) \\ &\quad + D_n^{+}\left[I_n \bigotimes \left(\Pi_0 - \hat{\Pi}_T\right)'\right]\text{vec}\left(\frac{1}{T}\sum_{t=p+1}^{T} X_t \varepsilon_t'\right) \\ &\quad + D_n^{+}\left[\left(\Pi_0 - \hat{\Pi}_T\right)' \bigotimes I_n\right]\text{vec}\left(\frac{1}{T}\sum_{t=p+1}^{T} X_t \varepsilon_t'\right) \\ &\quad + D_n^{+}\frac{1}{T}\sum_{t=p+1}^{T}\text{vec}\left(\varepsilon_t \varepsilon_t' - \Sigma_0\right).\end{aligned}$$

We can see that

$$\begin{aligned}&\sqrt{T}\left[\left(\Pi_0 - \hat{\Pi}_T\right)' \bigotimes \left(\Pi_0 - \hat{\Pi}_T\right)'\right]\text{vec}\left(\frac{1}{T}\sum_{t=p+1}^{T} X_t X_t'\right) \\ &\qquad = \left[\sqrt{T}\left(\Pi_0 - \hat{\Pi}_T\right)' \bigotimes \left(\Pi_0 - \hat{\Pi}_T\right)'\right]\text{vec}\left(\frac{1}{T}\sum_{t=p+1}^{T} X_t X_t'\right) \xrightarrow{p} \mathbf{0}\end{aligned}$$

because $\sqrt{T}\left(\Pi_0 - \hat{\Pi}_T\right)$ and $\frac{1}{T}\sum_{t=p+1}^{T} X_t X_t'$ are $O_p(1)$ while $\left(\Pi_0 - \hat{\Pi}_T\right)$ is $o_p(1)$. Since $\frac{1}{T}\sum_{t=p+1}^{T} X_t \varepsilon_t'$ is $o_p(1)$ and $\sqrt{T}\left(\Pi_0 - \hat{\Pi}_T\right)$ is $O_p(1)$, we have

$$\sqrt{T}\left[I_n \bigotimes \left(\Pi_0 - \hat{\Pi}_T\right)'\right] \text{vec}\left(\frac{1}{T}\sum_{t=p+1}^{T} X_t \varepsilon_t'\right) \xrightarrow{p} \mathbf{0}$$

as well. Putting these together,

$$\begin{aligned}
\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) &= \begin{pmatrix} \sqrt{T}\left(\hat{\beta}_T - \beta_0\right) \\ \sqrt{T}\left(\text{vech}\left(\hat{\Sigma}_T\right) - \text{vech}\left(\Sigma_0\right)\right) \end{pmatrix} \\
&= \begin{pmatrix} I_n \otimes \left(\frac{1}{T}\sum_{t=p+1}^{T} X_t X_t'\right)^{-1} & O \\ O & D_n^+ \end{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \begin{pmatrix} \text{vec}\left(X_t \varepsilon_t'\right) \\ \text{vec}\left(\varepsilon_t \varepsilon_t' - \Sigma_0\right) \end{pmatrix} + o_p(1) \\
&\xrightarrow{d} N\left[\mathbf{0}, \begin{pmatrix} \Sigma_0 \otimes Q^{-1} & \left(I_n \otimes Q^{-1}\bar{\mu}\right)\kappa_3 D_n^{+\prime} \\ D_n^+ \kappa_3' \left(I_n \otimes \bar{\mu}' Q^{-1}\right) & D_n^+ \left(\kappa_4 - \Sigma_0 \otimes \Sigma_0\right) D_n^{+\prime} \end{pmatrix}\right]
\end{aligned}$$

by the CMT.

Q.E.D.

The asymptotic results presented in the previous section also allow us to derive the asymptotic distribution of the score function at the true parameter values:

$$\begin{aligned}
\frac{1}{\sqrt{T}} s(\theta_0) &= \begin{pmatrix} \left(\Sigma_0^{-1} \otimes I_{np+1}\right) \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \text{vec}\left(X_t \varepsilon_t'\right) \\ \frac{1}{2} D_n^+ \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1}\right) \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \text{vec}\left(\varepsilon_t \varepsilon_t' - \Sigma_0\right) \end{pmatrix} \\
&= \begin{pmatrix} \Sigma_0^{-1} \otimes I_{np+1} & O \\ O & \frac{1}{2} D_n^+ \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1}\right) \end{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \begin{pmatrix} \text{vec}\left(X_t \varepsilon_t'\right) \\ \text{vec}\left(\varepsilon_t \varepsilon_t' - \Sigma_0\right) \end{pmatrix} \\
&\xrightarrow{d} N\left[\mathbf{0}, I_0\right],
\end{aligned}$$

where

$$I_0 = \begin{pmatrix} \Sigma_0^{-1} \otimes I_{np+1} & O \\ O & \frac{1}{2} D_n^+ \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1}\right) \end{pmatrix} \begin{pmatrix} \Sigma_0 \otimes Q & (I_n \otimes \bar{\mu})\kappa_3 \\ \kappa_3'(I_n \otimes \bar{\mu}') & \kappa_4 - \Sigma_0 \otimes \Sigma_0 \end{pmatrix} \begin{pmatrix} \Sigma_0^{-1} \otimes I_{np+1} & O \\ O & \frac{1}{2} D_n^+ \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1}\right) \end{pmatrix}'.$$

If the errors are i.i.d. normal,

$$I_0 = \begin{pmatrix} \Sigma_0^{-1} \otimes Q & O \\ O & \frac{1}{2} D_n^+ \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1}\right) D_n^{+\prime} \end{pmatrix}.$$

2.3.3 Deriving the Hessian

To derive the hessian, note that, for any entry x of Σ,

$$\begin{aligned}
\frac{\partial}{\partial x}\Sigma^{-1}(S(\beta)-(T-p)\Sigma)\Sigma^{-1} &= \frac{\partial \Sigma^{-1}}{\partial x}(S(\beta)-(T-p)\Sigma)\Sigma^{-1}+\Sigma^{-1}\frac{\partial(S(\beta)-(T-p)\Sigma)}{x}\Sigma^{-1} \\
&\quad +\Sigma^{-1}(S(\beta)-(T-p)\Sigma)\frac{\partial \Sigma^{-1}}{\partial x} \\
&= -\Sigma^{-1}\frac{\partial \Sigma}{\partial x}\Sigma^{-1}(S(\beta)-(T-p)\Sigma)\Sigma^{-1} \\
&\quad -\Sigma^{-1}(S(\beta)-(T-p)\Sigma)\Sigma^{-1}\frac{\partial \Sigma}{\partial x}\Sigma^{-1} \\
&\quad -(T-p)\Sigma^{-1}\frac{\partial \Sigma}{\partial x}\Sigma^{-1} \\
&= -\Sigma^{-1}\left[\frac{\partial \Sigma}{\partial x}\Sigma^{-1}(S(\beta)-(T-p)\Sigma)+(S(\beta)-(T-p)\Sigma)\Sigma^{-1}\frac{\partial \Sigma}{\partial x}+(T-p)\frac{\partial \Sigma}{\partial x}\right]\Sigma^{-1}.
\end{aligned}$$

This implies that

$$\begin{aligned}
&\frac{\partial \text{vec}\,(\Sigma^{-1}(S(\beta)-(T-p)\Sigma)\Sigma^{-1})}{\partial x} \\
&= -\left(\Sigma^{-1}\bigotimes\Sigma^{-1}\right)\cdot \text{vec}\left(\frac{\partial \Sigma}{\partial x}\Sigma^{-1}(S(\beta)-(T-p)\Sigma)+(S(\beta)-(T-p)\Sigma)\Sigma^{-1}\frac{\partial \Sigma}{\partial x}+(T-p)\frac{\partial \Sigma}{\partial x}\right) \\
&= -\left(\Sigma^{-1}\bigotimes\Sigma^{-1}\right)\left((S(\beta)-(T-p)\Sigma)\Sigma^{-1}\bigotimes I_n\right)\frac{\partial \text{vec}\,(\Sigma)}{\partial x} \\
&\quad -\left(\Sigma^{-1}\bigotimes\Sigma^{-1}\right)\left(I_n\bigotimes(S(\beta)-(T-p)\Sigma)\Sigma^{-1}\right)\frac{\partial \text{vec}\,(\Sigma)}{\partial x} \\
&\quad -(T-p)\left(\Sigma^{-1}\bigotimes\Sigma^{-1}\right)\frac{\partial \text{vec}\,(\Sigma)}{\partial x} \\
&= -\left(\Sigma^{-1}(S(\beta)-(T-p)\Sigma)\Sigma^{-1}\bigotimes\Sigma^{-1}\right)\frac{\partial \text{vec}\,(\Sigma)}{\partial x} \\
&\quad -\left(\Sigma^{-1}\bigotimes\Sigma^{-1}(S(\beta)-(T-p)\Sigma)\Sigma^{-1}\right)\frac{\partial \text{vec}\,(\Sigma)}{\partial x} \\
&\quad -(T-p)\left(\Sigma^{-1}\bigotimes\Sigma^{-1}\right)\frac{\partial \text{vec}\,(\Sigma)}{\partial x} \\
&= (T-p)\left(\Sigma^{-1}\bigotimes\Sigma^{-1}\right)\frac{\partial \text{vec}\,(\Sigma)}{\partial x} \\
&\quad -\left[\left(\Sigma^{-1}S(\beta)\Sigma^{-1}\bigotimes\Sigma^{-1}\right)+\left(\Sigma^{-1}\bigotimes\Sigma^{-1}S(\beta)\Sigma^{-1}\right)\right]\frac{\partial \text{vec}\,(\Sigma)}{\partial x},
\end{aligned}$$

and as such,

$$\begin{aligned}
&\frac{\partial \text{vech}\,(\Sigma^{-1}(S(\beta)-(T-p)\Sigma)\Sigma^{-1})}{\partial x} \\
&= D_n^{+}\frac{\partial \text{vec}\,(\Sigma^{-1}(S(\beta)-(T-p)\Sigma)\Sigma^{-1})}{\partial x} \\
&= (T-p)D_n^{+}\left(\Sigma^{-1}\bigotimes\Sigma^{-1}\right)D_n\frac{\partial \text{vech}\,(\Sigma)}{\partial x}
\end{aligned}$$

$$- D_n^+ \left[\left(\Sigma^{-1} S(\beta) \Sigma^{-1} \bigotimes \Sigma^{-1} \right) + \left(\Sigma^{-1} \bigotimes \Sigma^{-1} S(\beta) \Sigma^{-1} \right) \right] D_n \frac{\partial \text{vech}\,(\Sigma)}{\partial x}.$$

Therefore,

$$\frac{\partial^2 l(\theta)}{\partial \gamma \partial \gamma'} = \frac{1}{2} \frac{\partial \text{vech}\,(\Sigma^{-1}(S(\beta) - (T-p)\Sigma)\Sigma^{-1})}{\partial \text{vech}\,(\Sigma)'}$$

$$= \frac{T-p}{2} D_n^+ \left(\Sigma^{-1} \bigotimes \Sigma^{-1} \right) D_n - D_n^+ \left[\left(\Sigma^{-1} S(\beta) \Sigma^{-1} \bigotimes \Sigma^{-1} \right) + \left(\Sigma^{-1} \bigotimes \Sigma^{-1} S(\beta) \Sigma^{-1} \right) \right] D_n$$

$$= \frac{T-p}{2} D_n^+ \left(\Sigma^{-1} \bigotimes \Sigma^{-1} \right) \left[I_{n^2} - \left(\hat{\Sigma}_T \Sigma^{-1} \bigotimes I_n \right) - \left(I_n \bigotimes \hat{\Sigma}_T \Sigma^{-1} \right) \right] D_n.$$

Similarly, for any $A \in \mathbb{R}^{(np+1)\times n}$ such that $v = \text{vec}\,(A)$,

$$\begin{aligned}
\frac{\partial\,(\Sigma^{-1} \otimes I_{np+1})\, v}{\partial x} &= \left(\frac{\partial \Sigma^{-1}}{\partial x} \bigotimes I_{np+1} \right) v \\
&= - \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial x} \Sigma^{-1} \bigotimes I_{np+1} \right) v \\
&= -\text{vec} \left(A \Sigma^{-1} \frac{\partial \Sigma}{\partial x} \Sigma^{-1} \right) \\
&= - \left(\Sigma^{-1} \bigotimes A \Sigma^{-1} \right) \frac{\partial \text{vec}\,(\Sigma)}{\partial x} \\
&= - \left(\Sigma^{-1} \bigotimes A \Sigma^{-1} \right) D_n \frac{\partial \text{vech}\,(\Sigma)}{\partial x},
\end{aligned}$$

which implies that

$$\frac{\partial\,(\Sigma^{-1} \otimes I_{np+1})\, v}{\partial \text{vech}\,(\Sigma)'} = - \left(\Sigma^{-1} \bigotimes A \Sigma^{-1} \right) D_n.$$

Therefore,

$$\frac{\partial l(\theta)}{\partial \beta \partial \gamma'} = - \left(\Sigma^{-1} \bigotimes \left[\sum_{t=p+1}^{T} X_t \left(Y_t' - X_t' \Pi \right) \right] \right) D_n.$$

Putting the two results together,

$$H(\theta) := \frac{\partial l(\theta)}{d\theta d\theta'} = \begin{pmatrix} \frac{\partial l(\theta)}{d\beta d\beta'} & \frac{\partial l(\theta)}{d\beta d\gamma'} \\ \frac{\partial l(\theta)}{d\gamma d\beta'} & \frac{\partial l(\theta)}{d\gamma d\gamma'} \end{pmatrix}$$

$$= \begin{pmatrix} - \left(\Sigma^{-1} \otimes \sum_{t=p+1}^{T} X_t X_t' \right) & - \left(\Sigma^{-1} \otimes \left[\sum_{t=p+1}^{T} X_t \left(Y_t' - X_t' \Pi \right) \right] \right) D_n \\ -D_n' \left(\Sigma^{-1} \otimes \left[\sum_{t=p+1}^{T} \left(Y_t - \Pi' X_t \right) X_t' \right] \right) & \frac{T-p}{2} D_n^+ \left(\Sigma^{-1} \otimes \Sigma^{-1} \right) \left[I_{n^2} - \left(\hat{\Sigma}_T \Sigma^{-1} \otimes I_n \right) - \left(I_n \otimes \hat{\Sigma}_T \Sigma^{-1} \right) \right] D_n \end{pmatrix}$$

For any consistent estimator $\tilde{\theta}_T = (\tilde{\beta}_T', \tilde{\gamma}_T')'$ of θ_0, denoting

$$\tilde{\beta}_T = \text{vec}\left(\tilde{\Pi}_T\right) \quad \text{and} \quad \tilde{\gamma}_T = \text{vech}\left(\tilde{\Sigma}_T\right),$$

note that

$$\frac{1}{T}\sum_{t=p+1}^{T} X_t \left(Y_t - \tilde{\Pi}_T' X_t\right)' = (\Pi_0 - \tilde{\Pi}_T)' \frac{1}{T}\sum_{t=p+1}^{T} X_t X_t' + \frac{1}{T}\sum_{t=p+1}^{T} X_t \varepsilon_t' \xrightarrow{p} O$$

and

$$D_n^+ (\tilde{\Sigma}_T^{-1} \bigotimes \tilde{\Sigma}_T^{-1}) \left[I_{n^2} - (\hat{\Sigma}_T \tilde{\Sigma}_T^{-1} \bigotimes I_n) - (I_n \bigotimes \hat{\Sigma}_T \tilde{\Sigma}_T^{-1})\right] D_n$$
$$\xrightarrow{p} D_n^+ (\Sigma_0^{-1} \bigotimes \Sigma_0^{-1}) \cdot O \cdot D_n = O,$$

so we have

$$\frac{1}{T} H(\tilde{\theta}_T) \xrightarrow{p} H_0 := \begin{pmatrix} -\Sigma_0^{-1} \otimes Q & O \\ O & -\frac{1}{2} D_n^+ \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1}\right) D_n \end{pmatrix}.$$

Therefore,

$$\left(-\frac{1}{T} H(\tilde{\theta}_T)\right)^{-1} \xrightarrow{p} -H_0^{-1} := \begin{pmatrix} \Sigma_0 \otimes Q^{-1} & O \\ O & 2D_n^+ \left(\Sigma_0 \otimes \Sigma_0\right) D_n \end{pmatrix}$$

for any consistent estimator $\tilde{\theta}_T$ of θ_0. Note that, when the errors are i.i.d. normal, H_0 and I_0 are almost equal; this can be seen as the VAR version of the information matrix inequality.

2.3.4 Asymptotic Variance of $\hat{\theta}_T$

Using the values of H_0 and I_0 we derived above, the asymptotic distribution of $\hat{\theta}_T$ can be expressed as

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{d} N\left[\mathbf{0}, H_0^{-1} I_0 H_0'^{-1}\right].$$

In the special case that the errors are i.i.d. normal, we obtain the special case

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{d} N\left[\mathbf{0}, \underbrace{\begin{pmatrix} \Sigma_0 \bigotimes Q^{-1} & O \\ O & 2D_n^+ \left(\Sigma_0 \bigotimes \Sigma_0\right) D_n^{+\prime} \end{pmatrix}}_{I_0^{-1}}\right].$$

This is exactly the result on Quasi MLEs; given consistent estimators of H_0 and I_0, we can obtain a consistent estimator of the asymptotic variance of $\hat{\theta}_T$.

It remains to procure consistent estimators of H_0 and I_0. In the general case, this can be done by making use of the finiteness of the fourth moments of X_t, which follows from the fact that ε_t has finite fourth moments and X_t is a linear process with absolutely summable coefficients and innovation process ε_t.

We can naturally consider the following estimators of H_0:

$$\hat{H}_T^{-1} = -\begin{pmatrix} \hat{\Sigma}_T \bigotimes \left(\frac{1}{T}\sum_{t=p+1}^T X_t X_t'\right)^{-1} & O \\ O & 2D_n^+ \left(\hat{\Sigma}_T \bigotimes \hat{\Sigma}_T\right) D_n \end{pmatrix}.$$

Since I_0 is the asymptotic variance of the score function $s(\theta_0)$, which is itself a partial sum that can be written as

$$s(\theta_0) = \sum_{t=p+1}^T \begin{pmatrix} \text{vec}\left(X_t (Y_t - \Pi_0' X_t)' \Sigma_0^{-1}\right) \\ \frac{1}{2}\text{vech}\left(\Sigma_0^{-1}(Y_t - \Pi_0' X_t)(Y_t - \Pi_0' X_t)' \Sigma_0^{-1} - (T-p)\Sigma_0^{-1}\right) \end{pmatrix}$$

$$= \sum_{t=p+1}^T \begin{pmatrix} \text{vec}\left(X_t \varepsilon_t' \cdot \Sigma_0^{-1}\right) \\ \frac{1}{2}\text{vech}\left(\Sigma_0^{-1}\varepsilon_t \varepsilon_t' \Sigma_0^{-1} - \Sigma_0^{-1}\right) \end{pmatrix},$$

we can heuristically view $\left\{\begin{pmatrix} \text{vec}\left(X_t \varepsilon_t' \cdot \Sigma_0^{-1}\right) \\ \frac{1}{2}\text{vech}\left(\Sigma_0^{-1}\varepsilon_t \varepsilon_t' \Sigma_0^{-1} - \Sigma_0^{-1}\right) \end{pmatrix}\right\}_{t\in\mathbb{Z}}$ as a sequence with mean 0 and variance

$$I_0 = \mathbb{E}\left[\begin{pmatrix} \text{vec}\left(X_t \varepsilon_t' \cdot \Sigma_0^{-1}\right) \\ \frac{1}{2}\text{vech}\left(\Sigma_0^{-1}\varepsilon_t \varepsilon_t' \Sigma_0^{-1} - \Sigma_0^{-1}\right) \end{pmatrix}\begin{pmatrix} \text{vec}\left(X_t \varepsilon_t' \cdot \Sigma_0^{-1}\right) \\ \frac{1}{2}\text{vech}\left(\Sigma_0^{-1}\varepsilon_t \varepsilon_t' \Sigma_0^{-1} - \Sigma_0^{-1}\right) \end{pmatrix}'\right]$$

for any $t \in \mathbb{Z}$. Thus, an intuitive estimator for I_0 is

$$\hat{I}_T = \frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} \mathrm{vec}\left(X_t\hat{\varepsilon}_t'\hat{\Sigma}_T^{-1}\right) \\ \frac{1}{2}\mathrm{vech}\left(\hat{\Sigma}_T^{-1}\hat{\varepsilon}_t\hat{\varepsilon}_t'\hat{\Sigma}_T^{-1} - \hat{\Sigma}_T^{-1}\right) \end{pmatrix} \begin{pmatrix} \mathrm{vec}\left(X_t\hat{\varepsilon}_t'\hat{\Sigma}_T^{-1}\right) \\ \frac{1}{2}\mathrm{vech}\left(\hat{\Sigma}_T^{-1}\hat{\varepsilon}_t\hat{\varepsilon}_t'\hat{\Sigma}_T^{-1} - \hat{\Sigma}_T^{-1}\right) \end{pmatrix}',$$

where $\hat{\epsilon}_t$ is the tth residual. The following result demonstrates that $\hat{I}_T$ is consistent for I_0:

Theorem (Consistent Estimation of Asymptotic Variance)
Under our assumptions, the estimator $\hat{I}_T$ defined above is consistent for the asymptotic variance I_0 of the score function.

Proof) $\hat{I}_T$ can be rewritten as a matrix quadratic form as follows:

$$\hat{I}_T = \begin{pmatrix} \hat{\Sigma}_T^{-1} \otimes I_{np+1} & O \\ O & \frac{1}{2}D_n^+\left(\hat{\Sigma}_T^{-1}\otimes\hat{\Sigma}_T^{-1}\right) \end{pmatrix} \left[\frac{1}{T}\sum_{t=p+1}^{T} \begin{pmatrix} \mathrm{vec}\left(X_t\hat{\epsilon}_t'\right) \\ \mathrm{vec}\left(\hat{\epsilon}_t\hat{\epsilon}_t' - \hat{\Sigma}_T\right) \end{pmatrix} \begin{pmatrix} \mathrm{vec}\left(X_t\hat{\epsilon}_t'\right) \\ \mathrm{vec}\left(\hat{\epsilon}_t\hat{\epsilon}_t' - \hat{\Sigma}_T\right) \end{pmatrix}'\right]$$
$$\times \begin{pmatrix} \hat{\Sigma}_T^{-1} \otimes I_{np+1} & O \\ O & \frac{1}{2}D_n^+\left(\hat{\Sigma}_T^{-1}\otimes\hat{\Sigma}_T^{-1}\right) \end{pmatrix}'.$$

It follows from the consistency of $\hat{\Sigma}_T$ that

$$\begin{pmatrix} \hat{\Sigma}_T^{-1} \otimes I_{np+1} & O \\ O & \frac{1}{2}D_n^+\left(\hat{\Sigma}_T^{-1}\otimes\hat{\Sigma}_T^{-1}\right) \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \Sigma_0^{-1} \otimes I_{np+1} & O \\ O & \frac{1}{2}D_n^+\left(\Sigma_0^{-1}\otimes\Sigma_0^{-1}\right) \end{pmatrix}.$$

Furthermore, we know from the asymptotic results derived above that

$$\frac{1}{T}\sum_{t=p+1}^{T} \begin{pmatrix} \mathrm{vec}\left(X_t\varepsilon_t'\right) \\ \mathrm{vec}\left(\varepsilon_t\varepsilon_t' - \Sigma_0\right) \end{pmatrix} \begin{pmatrix} \mathrm{vec}\left(X_t\varepsilon_t'\right) \\ \mathrm{vec}\left(\varepsilon_t\varepsilon_t' - \Sigma_0\right) \end{pmatrix}' \xrightarrow{p} \begin{pmatrix} \Sigma_0 \otimes Q & (I_n \otimes \bar{\mu})\kappa_3 \\ \kappa_3'(I_n \otimes \bar{\mu}') & \kappa_4 - \Sigma_0 \otimes \Sigma_0 \end{pmatrix}.$$

If we can show that

$$\frac{1}{T}\sum_{t=p+1}^{T} \begin{pmatrix} \mathrm{vec}\left(X_t\hat{\epsilon}_t'\right) \\ \mathrm{vec}\left(\hat{\epsilon}_t\hat{\epsilon}_t' - \hat{\Sigma}_T\right) \end{pmatrix} \begin{pmatrix} \mathrm{vec}\left(X_t\hat{\epsilon}_t'\right) \\ \mathrm{vec}\left(\hat{\epsilon}_t\hat{\epsilon}_t' - \hat{\Sigma}_T\right) \end{pmatrix}' - \frac{1}{T}\sum_{t=p+1}^{T} \begin{pmatrix} \mathrm{vec}\left(X_t\varepsilon_t'\right) \\ \mathrm{vec}\left(\varepsilon_t\varepsilon_t' - \Sigma_0\right) \end{pmatrix} \begin{pmatrix} \mathrm{vec}\left(X_t\varepsilon_t'\right) \\ \mathrm{vec}\left(\varepsilon_t\varepsilon_t' - \Sigma_0\right) \end{pmatrix}' \xrightarrow{p} O,$$

or, equivalently,

$$\frac{1}{T}\sum_{t=p+1}^{T}\left[s_t(\hat{\theta}_T)s_t(\hat{\theta}_T)' - s_t(\theta_0)s_t(\theta_0)'\right] \xrightarrow{p} O,$$

then the claim of the theorem will follow.

We focus on each of the four block matrices that comprise the sum on the left hand side above. First, we show that

$$\frac{1}{T}\sum_{t=p+1}^{T}\left(\mathrm{vec}\left(X_t\hat{\varepsilon}_t'\right)\mathrm{vec}\left(X_t\hat{\varepsilon}_t'\right)' - \mathrm{vec}\left(X_t\varepsilon_t'\right)\mathrm{vec}\left(X_t\varepsilon_t'\right)'\right) \xrightarrow{p} O.$$

Note that

$$\begin{aligned}
&\left\|\mathrm{vec}\left(X_t\hat{\varepsilon}_t'\right)\mathrm{vec}\left(X_t\hat{\varepsilon}_t'\right)' - \mathrm{vec}\left(X_t\varepsilon_t'\right)\mathrm{vec}\left(X_t\varepsilon_t'\right)'\right\| \\
&\qquad \le \left|\mathrm{vec}\left(X_t(\hat{\varepsilon}_t-\varepsilon_t)'\right)\right|^2 + 2\left|\mathrm{vec}\left(X_t(\hat{\varepsilon}_t-\varepsilon_t)'\right)\right|\cdot\left|\mathrm{vec}\left(X_t\varepsilon_t'\right)\right| \\
&\qquad\qquad = \|X_t(\hat{\varepsilon}_t-\varepsilon_t)\|^2 + 2\|X_t(\hat{\varepsilon}_t-\varepsilon_t)\|\cdot\left\|X_t\varepsilon_t'\right\|
\end{aligned}$$

for any $t\in\mathbb{Z}$, where the last equality follows because, for any $A\in\mathbb{R}^{m\times k}$,

$$|\mathrm{vec}\,(A)|^2 = \sum_{i=1}^{m}\sum_{j=1}^{k}|A_{ij}|^2 = \|A\|^2.$$

Since

$$\hat{\varepsilon}_t - \varepsilon_t = (\Pi_0' - \hat{\Pi}_T')X_t,$$

we can see that, for any $t\in\mathbb{Z}$,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=p+1}^{T}\left\|\mathrm{vec}\left(X_t\hat{\varepsilon}_t'\right)\mathrm{vec}\left(X_t\hat{\varepsilon}_t'\right)' - \mathrm{vec}\left(X_t\varepsilon_t'\right)\mathrm{vec}\left(X_t\varepsilon_t'\right)'\right\| \\
&\qquad \le \left(\frac{1}{T}\sum_{t=p+1}^{T}|X_t|^4\right)\left\|\Pi_0-\hat{\Pi}_T\right\|^2 + 2\left(\frac{1}{T}\sum_{t=p+1}^{T}|X_t|^3|\varepsilon_t|\right)\left\|\Pi_0-\hat{\Pi}_T\right\|.
\end{aligned}$$

Since $\{X_t\}_{t\in\mathbb{Z}}$ is $L^{4+2\eta}$-bounded, the sequence $\left\{\frac{1}{T}\sum_{t=p+1}^{T}|X_t|^4\right\}_{T\in N_+}$ is $L^{1+\eta/2}$-bounded, indicating that it is uniformly integrable. Uniform integrability implies boundedness in probability, so by definition

$$\frac{1}{T}\sum_{t=p+1}^{T}|X_t|^4 = O_p(1).$$

Likewise, since X_t and ε_t are independent for any $t\in\mathbb{Z}$, and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is $L^{4+2\eta}$-bounded, the sequence $\left\{\frac{1}{T}\sum_{t=p+1}^{T}|X_t|^3|\varepsilon_t|\right\}_{T\in N_+}$ is also $L^{1+\eta/2}$-bounded, uniformly integrable and satisfies

$$\frac{1}{T}\sum_{t=p+1}^{T}|X_t|^3|\varepsilon_t| = O_p(1).$$

Since $\Pi_0 - \hat{\Pi}_T = o_p(1)$ by the consistency of $\hat{\theta}_T$, it follows that

$$\frac{1}{T}\sum_{t=p+1}^{T}\left\|\operatorname{vec}\left(X_t\hat{\varepsilon}_t'\right)\operatorname{vec}\left(X_t\hat{\varepsilon}_t'\right)' - \operatorname{vec}\left(X_t\varepsilon_t'\right)\operatorname{vec}\left(X_t\varepsilon_t'\right)'\right\| = o_p(1).$$

We now focus our attention on the second (and, by symmetry, third) block

$$\frac{1}{T}\sum_{t=p+1}^{T}\left(\operatorname{vec}\left(X_t\hat{\varepsilon}_t'\right)\operatorname{vec}\left(\hat{\varepsilon}_t\hat{\varepsilon}_t' - \hat{\Sigma}_T\right)' - \operatorname{vec}\left(X_t\varepsilon_t'\right)\operatorname{vec}\left(\varepsilon_t\varepsilon_t' - \Sigma_0\right)'\right).$$

As above, for any $t \in \mathbb{Z}$ we can see that

$$\begin{aligned}
&\left\|\operatorname{vec}\left(X_t\hat{\varepsilon}_t'\right)\operatorname{vec}\left(\hat{\varepsilon}_t\hat{\varepsilon}_t' - \hat{\Sigma}_T\right)' - \operatorname{vec}\left(X_t\varepsilon_t'\right)\operatorname{vec}\left(\varepsilon_t\varepsilon_t' - \Sigma_0\right)'\right\| \\
&\quad\le \left|\operatorname{vec}\left(X_t(\hat{\varepsilon}_t - \varepsilon_t)'\right)\right| \cdot \left|\operatorname{vec}\left((\hat{\varepsilon}_t\hat{\varepsilon}_t' - \varepsilon_t\varepsilon_t') + (\Sigma_0 - \hat{\Sigma}_T)\right)\right| \\
&\qquad + \left|\operatorname{vec}\left(X_t(\hat{\varepsilon}_t - \varepsilon_t)'\right)\right| \cdot \left|\operatorname{vec}\left(\varepsilon_t\varepsilon_t' - \Sigma_0\right)\right| \\
&\qquad + \left|\operatorname{vec}\left((\hat{\varepsilon}_t\hat{\varepsilon}_t' - \varepsilon_t\varepsilon_t') + (\Sigma_0 - \hat{\Sigma}_T)\right)\right| \cdot \left|\operatorname{vec}\left(X_t\varepsilon_t'\right)\right| \\
&\quad= \left\|X_t(\hat{\varepsilon}_t - \varepsilon_t)'\right\| \cdot \left\|(\hat{\varepsilon}_t\hat{\varepsilon}_t' - \varepsilon_t\varepsilon_t') + (\Sigma_0 - \hat{\Sigma}_T)\right\| \\
&\qquad + \left\|X_t(\hat{\varepsilon}_t - \varepsilon_t)'\right\| \cdot \left\|\varepsilon_t\varepsilon_t' - \Sigma_0\right\| \\
&\qquad + \left\|(\hat{\varepsilon}_t\hat{\varepsilon}_t' - \varepsilon_t\varepsilon_t') + (\Sigma_0 - \hat{\Sigma}_T)\right\| \cdot \left\|X_t\varepsilon_t'\right\| \\
&\quad\le |X_t|^2 \cdot \left\|\Pi_0 - \hat{\Pi}_T\right\| \cdot \left(|\hat{\varepsilon}_t - \varepsilon_t|^2 + 2|\hat{\varepsilon}_t - \varepsilon_t| \cdot |\varepsilon_t| + \left\|\Sigma_0 - \hat{\Sigma}_T\right\|\right) \\
&\qquad + |X_t|^2 \cdot \left\|\Pi_0 - \hat{\Pi}_T\right\| \cdot \left(|\varepsilon_t|^2 + \|\Sigma_0\|\right) \\
&\qquad + |X_t||\varepsilon_t| \cdot \left(|\hat{\varepsilon}_t - \varepsilon_t|^2 + 2|\hat{\varepsilon}_t - \varepsilon_t| \cdot |\varepsilon_t| + \left\|\Sigma_0 - \hat{\Sigma}_T\right\|\right) \\
&\quad\le |X_t|^2 \cdot \left\|\Pi_0 - \hat{\Pi}_T\right\| \cdot \left(|X_t|^2 \cdot \left\|\Pi_0 - \hat{\Pi}_T\right\|^2 + 2|X_t||\varepsilon_t| \cdot \left\|\Pi_0 - \hat{\Pi}_T\right\| + \left\|\Sigma_0 - \hat{\Sigma}_T\right\|\right) \\
&\qquad + |X_t|^2 \cdot \left\|\Pi_0 - \hat{\Pi}_T\right\| \cdot \left(|\varepsilon_t|^2 + \|\Sigma_0\|\right) \\
&\qquad + |X_t||\varepsilon_t| \cdot \left(|X_t|^2 \cdot \left\|\Pi_0 - \hat{\Pi}_T\right\|^2 + 2|X_t||\varepsilon_t| \cdot \left\|\Pi_0 - \hat{\Pi}_T\right\| + \left\|\Sigma_0 - \hat{\Sigma}_T\right\|\right).
\end{aligned}$$

As above,

$$\frac{1}{T}\sum_{t=p+1}^{T}\left\|\operatorname{vec}\left(X_t\hat{\varepsilon}_t'\right)\operatorname{vec}\left(\hat{\varepsilon}_t\hat{\varepsilon}_t' - \hat{\Sigma}_T\right)' - \operatorname{vec}\left(X_t\varepsilon_t'\right)\operatorname{vec}\left(\varepsilon_t\varepsilon_t' - \Sigma_0\right)'\right\|$$

ends up being majorized by a sum of $O_p(1)o_p(1) = o_p(1)$ terms, so that it is itself $o_p(1)$.

Finally, we deal with the final block

$$\frac{1}{T}\sum_{t=p+1}^{T}\left(\mathrm{vec}\left(\hat\varepsilon_t\hat\varepsilon_t'-\hat\Sigma_T\right)\mathrm{vec}\left(\hat\varepsilon_t\hat\varepsilon_t'-\hat\Sigma_T\right)'-\mathrm{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma_0\right)\mathrm{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma_0\right)'\right).$$

Analogously with the first block, for any $t\in\mathbb{Z}$

$$\begin{aligned}
&\left\|\mathrm{vec}\left(\hat\varepsilon_t\hat\varepsilon_t'-\hat\Sigma_T\right)\mathrm{vec}\left(\hat\varepsilon_t\hat\varepsilon_t'-\hat\Sigma_T\right)'-\mathrm{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma_0\right)\mathrm{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma_0\right)'\right\|\\
&\quad\le\left\|\left(\hat\varepsilon_t\hat\varepsilon_t'-\varepsilon_t\varepsilon_t'\right)+\left(\Sigma_0-\hat\Sigma_T\right)\right\|^2+2\left\|\left(\hat\varepsilon_t\hat\varepsilon_t'-\varepsilon_t\varepsilon_t'\right)+\left(\Sigma_0-\hat\Sigma_T\right)\right\|\cdot\left\|\varepsilon_t\varepsilon_t'-\Sigma_0\right\|\\
&\quad\le\left(|\hat\varepsilon_t-\varepsilon_t|^2+\left\|\Sigma_0-\hat\Sigma_T\right\|\right)^2+2\left(|\hat\varepsilon_t-\varepsilon_t|^2+\left\|\Sigma_0-\hat\Sigma_T\right\|\right)\left(|\varepsilon_t|^2+\|\Sigma_0\|\right)\\
&\le\left(|X_t|^2\left\|\Pi_0-\hat\Pi_T\right\|^2+\left\|\Sigma_0-\hat\Sigma_T\right\|\right)^2+2\left(|X_t|^2\left\|\Pi_0-\hat\Pi_T\right\|^2+\left\|\Sigma_0-\hat\Sigma_T\right\|\right)\left(|\varepsilon_t|^2+\|\Sigma_0\|\right).
\end{aligned}$$

Again,

$$\frac{1}{T}\sum_{t=p+1}^{T}\left\|\mathrm{vec}\left(\hat\varepsilon_t\hat\varepsilon_t'-\hat\Sigma_T\right)\mathrm{vec}\left(\hat\varepsilon_t\hat\varepsilon_t'-\hat\Sigma_T\right)'-\mathrm{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma_0\right)\mathrm{vec}\left(\varepsilon_t\varepsilon_t'-\Sigma_0\right)'\right\|$$

is majorized by a sum of $O_p(1)o_p(1)=o_p(1)$ terms, so that it is itself $o_p(1)$. This completes the proof.

Q.E.D.

Summarizing the results above, defining

$$\hat H_T^{-1}=-\begin{pmatrix}\hat\Sigma_T\otimes\left(\frac{1}{T}\sum_{t=p+1}^{T}X_tX_t'\right)^{-1} & O\\ O & 2D_n^{+}\left(\hat\Sigma_T\otimes\hat\Sigma_T\right)D_n\end{pmatrix},$$

$$\hat I_T=\frac{1}{T}\sum_{t=p+1}^{T}\begin{pmatrix}\mathrm{vec}\left(X_t\hat\varepsilon_t'\hat\Sigma_T^{-1}\right)\\ \frac{1}{2}\mathrm{vech}\left(\hat\Sigma_T^{-1}\hat\varepsilon_t\hat\varepsilon_t'\hat\Sigma_T^{-1}-\hat\Sigma_T^{-1}\right)\end{pmatrix}\begin{pmatrix}\mathrm{vec}\left(X_t\hat\varepsilon_t'\hat\Sigma_T^{-1}\right)\\ \frac{1}{2}\mathrm{vech}\left(\hat\Sigma_T^{-1}\hat\varepsilon_t\hat\varepsilon_t'\hat\Sigma_T^{-1}-\hat\Sigma_T^{-1}\right)\end{pmatrix},$$

a consistent estimator of the asymptotic variance of $\hat\theta_T$ can be constructed as

$$\hat V_T=\hat H_T^{-1}\hat I_T\hat H_T'^{-1}.$$

2.4 Asymptotics of Structural VARs

So far, we have dealt with the theory of reduced-form VAR models. In practice, we often turn to structural VARs to make causal inferences and disentangle the influence of one shock from another.

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional time series. We say that it follows a structural VAR(p) process if there exist $n\times n$ matrices $B_0, B_1, \cdots, B_p$ and an n-dimensional nonrandom vector α such that

$$B_0 Y_t = \alpha + B_1 Y_{t-1} + \cdots + B_p Y_{t-p} + \varepsilon_t$$

for any $t\in\mathbb{Z}$, where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is an n-dimensional white noise process with variance I_n. This means that the componenets of the error process are uncorrelated; often, we assume that they are independent. Note that the variances of the shocks are normalized to 1; an equivalent normalization is putting the diagonal elements of B_0 equal to 1, but we choose to normalize the variances of the shocks instead to make interpreting the impulse responses easier. The elements of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are referred to as structural errors.

Assuming that B_0 is non-singular, pre-multiplying both sides of the above equation by B_0^{-1} yields the reduced-form version of the model,

$$\begin{aligned} Y_t &= B_0^{-1}\alpha + (B_0^{-1}B_1)Y_{t-1} + \cdots + (B_0^{-1}B_p)Y_{t-p} + B_0^{-1}\varepsilon_t \\ &= \delta + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + u_t, \end{aligned}$$

where $\{u_t\}_{t\in\mathbb{Z}}$ is an n-dimensional white noise process with positive definite variance equal to $\Sigma = B_0^{-1}B_0'^{-1}$. Under the reduced-form model, the componenets of the errors are no longer independent. To distinguish them from the structural errors, we call the elements of $\{u_t\}_{t\in\mathbb{Z}}$ reduced-form errors.

Suppose that $\{Y_t\}_{t\in\mathbb{Z}}$ is square integrable with bounded second moments and that the companion matrix $F\in\mathbb{R}^{np\times np}$ defined as

$$F = \begin{pmatrix} \Phi_1 & \cdots & \Phi_{p-1} & \Phi_p \\ I_n & \cdots & O & O \\ \vdots & \cdots & \vdots & \vdots \\ O & \cdots & I_n & O \end{pmatrix}$$

has eigenvalues within the unit circle. Then, $\{Y_t\}_{t\in\mathbb{Z}}$ is weakly stationary with a one-summable causal linear process representation

$$Y_t = \mu + \sum_{j=0}^{\infty} \Psi_j \cdot u_{t-j} = \mu + \Psi(L)u_t$$

for any $t \in \mathbb{Z}$. Since $u_t = B_0^{-1}\varepsilon_t$ for any $t \in \mathbb{Z}$ by definition, defining

$$\Theta_j = \Psi_j \cdot B_0^{-1}$$

for any $j \in \mathbb{N}$, $\{\Theta_j\}_{j\in\mathbb{Z}}$ is a one-summable sequence of $n \times n$ matrices such that

$$Y_t = \mu + \sum_{j=0}^{\infty} \Theta_j \cdot \varepsilon_{t-j} = \mu + \Theta(L)\varepsilon_t$$

for any $t \in \mathbb{Z}$. This indicates that any Y_t is a function of current and past structural errors, where the function is time-invariant.

The h-period impulse response of the ith dependent variable, Y_{it}, to a one standard deviation shock to the jth structural error, ε_{jt}, can now be defined as

$$IRF_{t+h,ij} = \frac{\partial Y_{i,t+h}}{\partial \varepsilon_{jt}} = \Theta_{h,ij},$$

due to the independence of the components of ε_t and its independence with the structural errors of all other periods. We collect the h-period ahead impulse responses for all dependent variables with respect to each structural shock in the $n \times n$ matrix

$$IRF_{t+h} = \frac{\partial Y_{t+h}}{\partial \varepsilon_t'} = \Theta_h.$$

An estimate of this matrix requires knowledge of Ψ_h and B_0^{-1}. Since Ψ_h is just the $n \times n$ matrix in the (1,1) position of F^h, which itself is comprised of the reduced-form mean reversion coefficients $\Phi_1, \cdots, \Phi_p$, it can be consistently estimated by estimating the reduced-form model via QMLE, as we did in the previous section.

The problem arises when attempting to recover B_0^{-1}. As is, B_0^{-1} contains n^2 entries that we must estimate. However, in the course of estimating the reduced-form model, we estimate $\Sigma = B_0^{-1} B_0'^{-1}$, which only has $\frac{n(n+1)}{2}$ free parameters due to being a symmetric matrix. Thus, the order condition for identification is not satisfied; we require $\frac{n(n-1)}{2}$ more restrictions at the very least. In what follows, we study some popular methods to impose this restriction, and, for a select few cases, derive the asymptotic distribution of the impulse responses using the delta method.

2.4.1 Recursive Identification

Perhaps the most popular and certaintly the simplest method of identifying B_0^{-1} is recursive identification, which imposes the constraint that B_0^{-1} is lower triangular with non-zero diagonal elements. Since B_0^{-1} dictates the contemporaneous effects of the structural shocks on the dependent variables, recursive identification implies that the variables comprising Y_t are ordered so that, for any $1 \leq j \leq n$, the structural error ε_{jt} only contemporaneously affects the dependent variables Y_{it} for $i \geq j$. A famous example of this identification scheme in practice is the monetary VAR found in Stock and Watson (2001).

Formally, this allows us to identify B_0^{-1} as the Cholesky factor of the positive definite matrix Σ. Thus, our estimate of B_0^{-1} is simply the Cholesky factor of the MLE of Σ. Recall that the reduced-form parameters are

$$\Pi = \begin{pmatrix} \delta' \\ \Phi_1' \\ \vdots \\ \Phi_p' \end{pmatrix} \quad \text{and} \quad \Sigma.$$

Let $\theta = (\text{vec}(\Pi)', \text{vech}(\Sigma)')' \in \Theta = \mathbb{R}^{n(np+1)} \times \mathcal{A}$ be the reduced form parameters collected in vector form. We define the reduced-form impulse response $\Psi_h : \mathbb{R}^{n(np+1)} \to \mathbb{R}^{n \times n}$ as

$$\Psi_h(\beta) = J \cdot \begin{pmatrix} \Phi_1 & \cdots & \Phi_{p-1} & \Phi_p \\ I_n & \cdots & O & O \\ \vdots & \cdots & \vdots & \vdots \\ O & \cdots & I_n & O \end{pmatrix}^h \cdot J'$$

for any $\beta = \text{vec}(\Pi) \in \mathbb{R}^{n(np+1)}$, where $J = (I_n, O, \cdots, O) \in \mathbb{R}^{n \times np}$. Then, the h-period ahead impulse response function $IRF_h^{RI} : \Theta \to \mathbb{R}^{n \times n}$ under recursive identification is defined as

$$IRF_h^{RI}(\theta) = \Psi_h(\beta) \cdot \text{chol}\left(\text{vech}^{-1}(\gamma)\right)$$

for any $\theta = (\beta', \gamma')' \in \Theta$. To apply the delta method later on, we first obtain the derivative of $\text{vec}\left(IRF_h^{RI}\right)$ on the open set $\Theta = \mathbb{R}^{n(np+1)} \times \mathcal{A}$.

Choose any $\theta = (\beta', \gamma')' = (\text{vec}(\Pi)', \text{vech}(\Sigma)')' \in \Theta$. We first derive the derivative of $\text{vec}\left(IRF_h^{RI}\right)$ at θ with respect to $\beta = \text{vec}(\Pi)$. To do so, it suffices to derive

$$\frac{\partial \text{vec}(\Psi_h(\beta))}{\partial \beta'},$$

the derivative of $\text{vec}(\Psi_h)$ at β. This derivative proves useful under other identification methods as well, so we present it as a lemma:

Lemma (Derivative of Reduced-Form Impulse Response)
For any $h \geq 0$,

$$\frac{\partial \text{vec}\,(\Psi_h(\beta))}{\partial \beta'} = \begin{pmatrix} O_{n^2 \times n} & \sum_{i=0}^{h-1} \left(J(F^{h-i-1})' \bigotimes \Psi_i(\beta) \right) \end{pmatrix} \cdot K_{np+1,n}$$

for any $\beta = \text{vec}\left(\begin{pmatrix} \alpha & \Phi_1 & \cdots & \Phi_p \end{pmatrix}' \right) = \mathbb{R}^{n(np+1)}$, where $F \in \mathbb{R}^{np \times np}$ is the companion matrix

$$F = \begin{pmatrix} \Phi_1 & \cdots & \Phi_{p-1} & \Phi_p \\ I_n & \cdots & O & O \\ \vdots & \cdots & \vdots & \vdots \\ O & \cdots & I_n & O \end{pmatrix}$$

and $J = (I_n, O_{n\times n}, \cdots, O_{n\times n}) \in \mathbb{R}^{n \times np}$.

Proof) Note initially that

$$\begin{aligned} \frac{\partial \Psi_h(\beta)}{\partial x} &= J \cdot \left(\frac{\partial F}{\partial x} F^{h-1} + F \frac{\partial F^{h-1}}{\partial x} \right) J' \\ &= \cdots = J \cdot \left(\sum_{i=0}^{h-1} F^i \frac{\partial F}{\partial x} F^{h-i-1} \right) J' = \sum_{i=0}^{h-1} J F^i \frac{\partial F}{\partial x} F^{h-i-1} J'. \end{aligned}$$

It follows that

$$\frac{\partial \text{vec}\,(\Psi_h(\beta)))}{\partial x} = \sum_{i=0}^{h-1} \left(J(F^{h-i-1})' \bigotimes J F^i \right) \cdot \frac{\partial \text{vec}\,(F)}{\partial x}.$$

Letting $\Pi'_{2:e}$ collect the last np columns of $\Pi' = (\delta, \Phi_1, \cdots, \Phi_p)$, that is,

$$\Pi'_{2:e} = \begin{pmatrix} \Phi_1 & \cdots & \Phi_p \end{pmatrix},$$

we can see that

$$\frac{\partial \text{vec}\,(F)}{\partial \text{vec}\,(\Pi'_{2:e})'} = I_{np} \bigotimes J'.$$

It follows that

$$\begin{aligned} \frac{\partial \text{vec}\,(\Psi_h(\beta)))}{\partial \text{vec}\,(\Pi'_{2:e})'} &= \sum_{i=0}^{h-1} \left(J(F^{h-i-1})' \bigotimes J F^i \right) \cdot (I_{np} \bigotimes J') \\ &= \sum_{i=0}^{h-1} \left(J(F^{h-i-1})' \bigotimes J F^i J' \right) \\ &= \sum_{i=0}^{h-1} \left(J(F^{h-i-1})' \bigotimes \Psi_i(\beta) \right) \end{aligned}$$

and since the elements of δ do not appear in the formula for $\Psi_h(\beta)$, we can see that

$$\begin{aligned}
\frac{\partial \text{vec}\left(\Psi_h(\beta)\right))}{\partial \text{vec}\left(\Pi'\right)'} &= \Big(O_{n^2 \times n} \quad \textstyle\sum_{i=0}^{h-1} \Big(J(F^{h-i-1})' \otimes \Psi_i(\beta)\Big)\Big) \\
&= \sum_{i=0}^{h-1} \Big(O_{n\times 1} \quad J(F^{h-i-1})'\Big) \bigotimes \Psi_i(\beta)
\end{aligned}$$

Since $K_{n,np+1} \cdot \text{vec}\left(\Pi'\right) = \text{vec}\left(\Pi\right)$ by definition of the commutation matrix, we finally have the derivative

$$\begin{aligned}
\frac{\partial \text{vec}\left(\Psi_h(\beta)\right))}{\partial \beta'} &= \frac{\partial \text{vec}\left(\Psi_h(\beta)\right))}{\partial \text{vec}\left(\Pi'\right)'} \cdot K'_{n,np+1} \\
&= \left[\sum_{i=0}^{h-1} \Big(O_{n\times 1} \quad J(F^{h-i-1})'\Big) \bigotimes \Psi_i(\beta)\right] K_{np+1,n}.
\end{aligned}$$

Q.E.D.

Using the fact that

$$\text{vec}\left(IRF_h^{RI}(\theta)\right) = \left(\text{chol}\left(\text{vech}^{-1}\left(\gamma\right)\right)' \bigotimes I_n\right) \cdot \text{vec}\left(\Psi_h(\beta)\right),$$

we can see that

$$\begin{aligned}
\frac{\partial \text{vec}\left(IRF_h^{RI}(\theta)\right)}{\partial \beta'} &= \Big(\text{chol}\left(\Sigma\right)' \bigotimes I_n\Big) \cdot \frac{\partial \text{vec}\left(\Psi_h(\beta)\right))}{\partial \beta'} \\
&= \Big(\text{chol}\left(\Sigma\right)' \bigotimes I_n\Big) \cdot \left[\sum_{i=0}^{h-1} \Big(O_{n\times 1} \quad J(F^{h-i-1})'\Big) \bigotimes \Psi_i(\beta)\right] K_{np+1,n} \\
&= \left[\sum_{i=0}^{h-1} \Big(O_{n\times 1} \quad \text{chol}\left(\Sigma\right)' \cdot J(F^{h-i-1})'\Big) \bigotimes \Psi_i(\beta)\right] K_{np+1,n}.
\end{aligned}$$

We now move onto the derivative of $\text{vec}\left(IRF_h^{RI}\right)$ with respect to $\gamma = \text{vech}(\Sigma)$. The differentiation is slightly more involved than the earlier case, so we start by introducing the derivative of the function $\text{chol}: PS^{n\times n} \to \mathbb{R}^{n\times n}$. Since $PS^{n\times n}$ is an open set with respect to the metric induced by the trace norm, we do not have to worry about boundary problems when dealing with differentiation. Define the $n^2 \times \frac{n(n+1)}{2}$ matrix L_n as the matrix satisfying

$$L_n \cdot \text{vech}(A) = \text{vec}(A)$$

for any lower triangular matrix $A \in \mathbb{R}^{n\times n}$. We can construct L_n be taking the $n^2 \times n^2$ identity matrix and removing the $n(i-1)+j$th columns for any $1 \le i, j \le n$ such that $j > i$.

In addition, since each column of L_n contains exactly one element equal to 1, $L_n' L_n = I_{n(n+1)/2}$, making L_n an orthogonal matrix. By definition, for any lower triangular $A \in \mathbb{R}^{n\times n}$,

$$L_n' \cdot \text{vec}(A) = L_n' L_n \cdot \text{vech}(A) = \text{vech}(A).$$

This shows us that L_n' serves a similar function as D_n^+ for lower trianguar matrices.

Choose any $A \in PS^{n\times n}$ and note that

$$A = \text{chol}(A)\,\text{chol}(A)'.$$

As such,

$$\frac{\partial A}{\partial x} = \text{chol}(A)\left(\frac{\partial \text{chol}(A)}{\partial x}\right)' + \left(\frac{\partial \text{chol}(A)}{\partial x}\right)\cdot \text{chol}(A)'.$$

Therefore,

$$\begin{aligned}
\frac{\partial \text{vech}(A)}{\partial x} &= D_n^+ \cdot \frac{\partial \text{vec}(A)}{\partial x} \\
&= D_n^+ (I_n \bigotimes \text{chol}(A)) \cdot \frac{\partial \text{vec}\left(\text{chol}(A)'\right)}{\partial x} + D_n^+ (\text{chol}(A) \bigotimes I_n) \cdot \frac{\partial \text{vec}(\text{chol}(A))}{\partial x} \\
&= D_n^+ \left[(I_n \bigotimes \text{chol}(A)) K_n + (\text{chol}(A) \bigotimes I_n)\right] \cdot \frac{\partial \text{vec}(\text{chol}(A))}{\partial x} \\
&= D_n^+ (K_n + I_{n^2})(\text{chol}(A) \bigotimes I_n) \cdot \frac{\partial \text{vec}(\text{chol}(A))}{\partial x} \\
&= D_n^+ (K_n + I_{n^2})(\text{chol}(A) \bigotimes I_n) L_n \cdot \frac{\partial \text{vech}(\text{chol}(A))}{\partial x} \\
&= 2 D_n^+ (D_n D_n^+)(\text{chol}(A) \bigotimes I_n) L_n \cdot \frac{\partial \text{vech}(\text{chol}(A))}{\partial x} \\
&= 2 D_n^+ (\text{chol}(A) \bigotimes I_n) L_n \cdot \frac{\partial \text{vech}(\text{chol}(A))}{\partial x},
\end{aligned}$$

where we used the fact that $K_n \text{vec}(\text{chol}(A)) = \text{vec}\left(\text{chol}(A)'\right)$ by definition of the commutation matrix and $\frac{1}{2}(I_{n^2}+K_n)=D_nD_n^+$. Therefore,

$$\frac{\partial \text{vech}\,(A)}{\partial \text{vech}\,(\text{chol}\,(A))}=2D_n^+(\text{chol}\,(A)\bigotimes I_n)L_n,$$

and as such,

$$\frac{\partial \text{vech}\,(\text{chol}\,(A))}{\partial \text{vech}\,(A)}=\left(2D_n^+(\text{chol}\,(A)\bigotimes I_n)L_n\right)^{-1}.$$

Now we move onto the differentiation of vec $\left(IRF_h^{RI}(\theta)\right)$. Since

$$\text{vec}\left(IRF_h^{RI}(\theta)\right)=\left(I_n\bigotimes\Psi_h(\beta)\right)\cdot\text{vec}\,(\text{chol}\,(\Sigma))=\left(I_n\bigotimes\Psi_h(\beta)\right)L_n\cdot\text{vech}\,(\text{chol}\,(\Sigma))\,,$$

by the result derived above we can see that

$$\begin{aligned}\frac{\partial \text{vec}\left(IRF_h^{RI}(\theta)\right)}{\partial\gamma'}=&\frac{\partial \text{vec}\left(IRF_h^{RI}(\theta)\right)}{\partial \text{vech}\,(\Sigma)'}\\ &=\left(I_n\bigotimes\Psi_h(\beta)\right)L_n\cdot\frac{\partial \text{vech}\,(\text{chol}\,(\Sigma))}{\partial \text{vech}\,(\Sigma)'}\\ &=\left(I_n\bigotimes\Psi_h(\beta)\right)L_n\left[2D_n^+(\text{chol}\,(\Sigma)\bigotimes I_n)L_n\right]^{-1}.\end{aligned}$$

Putting the two results together, we can see that

$$\frac{\partial \text{vec}\left(IRF_h^{RI}(\theta)\right)}{\partial\theta'}=\left(\frac{\partial \text{vec}\left(IRF_h^{RI}(\theta)\right)}{\partial\beta'}\quad\frac{\partial \text{vec}\left(IRF_h^{RI}(\theta)\right)}{\partial\gamma'}\right)$$

$$=\left(\left[\textstyle\sum_{i=0}^{h-1}\left(O_{n\times1}\quad\text{chol}\,(\Sigma)'\cdot J(F^{h-i-1})'\right)\otimes\Psi_i(\beta)\right]K_{np+1,n}\quad\left(I_n\otimes\Psi_h(\beta)\right)L_n\left[2D_n^+(\text{chol}\,(\Sigma)\otimes I_n)L_n\right]^{-1}\right).$$

The Delta method can now be applied to study the asymptotic properties of the h-period ahead impulse response under recursive identification; we state this as a theorem.

Theorem (Asymptotic Normality of IRF under Recursive Identification)
Maintain assumptions A1 to A3, and retain the notations of the previous section. Let $\hat{\theta}_T$ be the QMLE of θ_0. Then,

$$\sqrt{T}\left[\text{vec}\left(IRF_h^{RI}(\hat{\theta}_T)\right) - \text{vec}\left(IRF_h^{RI}(\theta_0)\right)\right]$$

$$\xrightarrow{d} N\left[\mathbf{0}, \left(\frac{\partial \text{vec}\left(IRF_h^{RI}(\theta_0)\right)}{\partial \theta'}\right) H_0^{-1} I_0 H_0'^{-1} \left(\frac{\partial \text{vec}\left(IRF_h^{RI}(\theta_0)\right)}{\partial \theta'}\right)'\right],$$

where H_0, I_0 are as defined above.

Proof) This follows from the Delta method, together with the preceding result that, under our assumptions, the QMLE of θ is asymptotically normal with asymptotic distribution

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{d} N\left[\mathbf{0}, H_0^{-1} I_0 H_0'^{-1}\right].$$

Here,

$$H_0^{-1} = \begin{pmatrix} \Sigma_0 \bigotimes Q^{-1} & O \\ O & 2D_n^+(\Sigma_0 \bigotimes \Sigma_0) D_n \end{pmatrix}$$

and I_0 is the information matrix.

Q.E.D.

Using consistency results concerning $\hat{\beta}_T = \text{vec}\left(\hat{\Pi}_T\right)$, $\hat{\Sigma}_T$ and $\hat{I}_T$, we can easily construct a consistent estimator for the asymptotic variance of the estimaed h-step ahead impulse response under recursive identification, $IRF_h^{RI}(\hat{\theta}_T)$.

2.5 Likelihood Ratio Tests under Linear Restrictions

Here, we study likelihood ratio tests for linear restrictions on the coefficient parameters β. We do not discuss tests of linear restrictions on the covariance terms for two reasons. First, the analysis of LR test statistics for linear restrictions on covariance terms is very complicated, in contrast to the simplicity of Wald tests; thus, when testing linear restrictions on covariance terms, we opt for Wald tests. Second, we will later be interested in sequentially tests to find the correct lag order, which can be formulated in terms of linear restrictions on the coefficient parameters.

Recall that we derived the (quasi) log likelihood function for a reduced form VAR(p) model as

$$l(\theta) = -\frac{n(T-p)}{2}\log(2\pi) - \frac{T-p}{2}\log|\Sigma| - \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\cdot\sum_{t=p+1}^{T}(Y_t - \Pi' X_t)(Y_t - \Pi' X_t)'\right)$$

for any $\theta \in \Theta = \mathbb{R}^{n(np+1)} \times \mathcal{A}$. Since the QMLE of Π and Σ are

$$\hat{\Pi}_T = \left(\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\left(\sum_{t=p+1}^{T} X_t Y_t'\right)$$

and

$$\hat{\Sigma}_T = \frac{1}{T-p}\sum_{t=p+1}^{T}(Y_t - \hat{\Pi}_T' X_t)(Y_t - \hat{\Pi}_T' X_t)',$$

the maximized log likelihood function is given by

$$\begin{aligned} l(\hat{\theta}_T) &= -\frac{n(T-p)}{2}\log(2\pi) - \frac{T-p}{2}\log\left|\hat{\Sigma}_T\right| - \frac{1}{2}\operatorname{tr}\left(\hat{\Sigma}^{-1}\cdot\sum_{t=p+1}^{T}(Y_t - \hat{\Pi}_T' X_t)(Y_t - \hat{\Pi}_T' X_t)'\right) \\ &= -\frac{n(T-p)}{2}(\log(2\pi)+1) - \frac{T-p}{2}\log\left|\hat{\Sigma}_T\right|. \end{aligned}$$

In other words, $\log\left|\hat{\Sigma}_T\right|$ can be used as an indicator of the fit of the model, since the mean maximized log likelihood, $\frac{1}{T-p}\hat{l}(p)$, is a linear transformation of $\log\left|\hat{\Sigma}_T\right|$.

2.5.1 Feasible Restricted Estimators of θ

We consider linear restrictions of the form $R\beta = q$, where $R \in \mathbb{R}^{r \times n(np+1)}$ is a matrix of full rank r, and $q \in \mathbb{R}^r$. We want to find the Gaussian QMLE of θ udner the restriction $R\beta = q$. To do so, for every outcome we must solve the constrained optimization problem

$$\begin{aligned} \max_{\theta \in \Theta} \quad & -\frac{T-p}{2}\log|\Sigma| - \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1} \cdot \sum_{t=p+1}^{T}(Y_t - \Pi' X_t)(Y_t - \Pi' X_t)'\right) \\ \text{subject to} \quad & R\beta = q. \end{aligned}$$

Let $\tilde{\theta}_T = (\tilde{\beta}_T', \tilde{\gamma}_T')'$ be a solution to the above problem, where

$$\tilde{\beta}_T = \operatorname{vec}\left(\tilde{\Pi}_T\right) \quad \text{and} \quad \tilde{\gamma}_T = \operatorname{vech}\left(\tilde{\Sigma}_T\right).$$

Fixing an outcome $\omega \in \Omega$, the Lagrangian for constrained maximization is given as

$$\begin{aligned} \mathcal{L} &= -\frac{T-p}{2}\log|\Sigma| - \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1} \cdot \sum_{t=p+1}^{T}(Y_t - \Pi' X_t)(Y_t - \Pi' X_t)'\right) + \lambda(q - R\beta) \\ &= -\frac{T-p}{2}\log|\Sigma| - \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1} \cdot \sum_{t=p+1}^{T}(Y_t - (I_n \bigotimes X_t')\beta)(Y_t - (I_n \bigotimes X_t')\beta)'\right) + \lambda'(q - R\beta). \end{aligned}$$

The first order conditions for maximization are

$$\begin{aligned} \sum_{t=p+1}^{T}(I_n \bigotimes X_t)\tilde{\Sigma}_T^{-1}(Y_t - (I_n \bigotimes X_t')\tilde{\beta}_T) - R'\lambda = \mathbf{0} \\ \frac{T-p}{2}\operatorname{vech}\left(\tilde{\Sigma}_T^{-1}\right) + \frac{1}{2}\operatorname{vech}\left(\tilde{\Sigma}_T^{-1} S(\tilde{\beta}_T)\tilde{\Sigma}_T^{-1}\right) = \mathbf{0}, \end{aligned}$$

where the function $S(\cdot)$ is defined as above; that is, as

$$S(\beta) = \sum_{t=p+1}^{T}(Y_t - (I_n \bigotimes X_t')\beta)(Y_t - (I_n \bigotimes X_t')\beta)'$$

for any $\beta \in \mathbb{R}^{n(np+1)}$.

Since there is no restriction imposed on the covariance, $\tilde{\Sigma}_T$ is given identically as in the unrestricted case, except that the unrestricted estimator $\hat{\beta}_T$ of β is used in $\hat{\Sigma}_T$, while the restricted estimator $\tilde{\beta}_T$ is used in $\tilde{\Sigma}_T$:

$$\tilde{\Sigma}_T = \frac{1}{T-p}S(\tilde{\beta}_T) = \frac{1}{T-p}\sum_{t=p+1}^{T}(Y_t - (I_n \bigotimes X_t')\tilde{\beta}_T)(Y_t - (I_n \bigotimes X_t')\tilde{\beta}_T)'.$$

Turning our attention to the first order condition for the coefficient parameters, since

$$(I_n \bigotimes X_t)\tilde{\Sigma}_T^{-1}(Y_t - (I_n \bigotimes X_t')\tilde{\beta}_T) = (\tilde{\Sigma}_T^{-1} \bigotimes X_t)(Y_t - (I_n \bigotimes X_t')\tilde{\beta}_T),$$

we can see that

$$\sum_{t=p+1}^{T} (\tilde{\Sigma}_T^{-1} \bigotimes X_t)(Y_t - (I_n \bigotimes X_t')\tilde{\beta}_T) = R'\lambda.$$

Sovlingn for $\tilde{\beta}_T$ now yields

$$\begin{aligned}
\tilde{\beta}_T &= \left[\tilde{\Sigma}_T \bigotimes \left(\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\right] \left[\sum_{t=p+1}^{T} (\tilde{\Sigma}_T^{-1} \bigotimes X_t) Y_t - R'\lambda\right] \\
&= \left[I_n \bigotimes \left(\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\right] \sum_{t=p+1}^{T} (I_n \bigotimes X_t) Y_t - \left[\tilde{\Sigma}_T \bigotimes \left(\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\right] R'\lambda \\
&= \text{vec}\left(\left(\sum_{t=p+1}^{T} X_t X_t'\right)^{-1} \sum_{t=p+1}^{T} X_t Y_t'\right) - \left[\tilde{\Sigma}_T \bigotimes \left(\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\right] R'\lambda \\
&= \hat{\beta}_T - \left[\tilde{\Sigma}_T \bigotimes \left(\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\right] R'\lambda.
\end{aligned}$$

Since R is of full rank and $\tilde{\Sigma}_T \otimes \left(\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}$ is nonsingular, we can see that

$$R\left[\tilde{\Sigma}_T \bigotimes \left(\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\right] R'$$

is a nonsingular $r \times r$ matrix. $\tilde{\beta}_T$ must also satisfy the constraint $R\tilde{\beta}_T = q$, so pre-multiplying both sides of the above equation by R allows us to express the Lagrange multiplier λ as

$$\lambda = \left(R\left[\tilde{\Sigma}_T \bigotimes \left(\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\right] R'\right)^{-1} (R\hat{\beta}_T - q).$$

This holds for any outcome $\omega \in \Omega$, so the preservation of measurability under continuous mappings implies that the Lagrange multiplier itself is a random vector; to emphasize this, we write $\tilde{\lambda}_T$.

Given the above formulation for the Lagrange multiplier, the restricted estimator of β is finally given as

$$\tilde{\beta}_T = \hat{\beta}_T - \left[\tilde{\Sigma}_T \bigotimes \left(\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\right] R' \left(R\left[\tilde{\Sigma}_T \bigotimes \left(\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\right] R'\right)^{-1} (R\hat{\beta}_T - q).$$

A clear problem is that $\tilde{\beta}_T$ and $\tilde{\Sigma}_T$ are simultaneously determined by the equations

$$\tilde{\beta}_T = \hat{\beta}_T - \left[\tilde{\Sigma}_T \bigotimes \left(\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\right] R' \left(R\left[\tilde{\Sigma}_T \bigotimes \left(\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\right] R'\right)^{-1} (R\hat{\beta}_T - q)$$

$$\tilde{\Sigma}_T = \frac{1}{T-p} \sum_{t=p+1}^{T} (Y_t - (I_n \bigotimes X_t')\tilde{\beta}_T)(Y_t - (I_n \bigotimes X_t')\tilde{\beta}_T)'.$$

We therefore use the feasible restricted estimators

$$\tilde{\beta}_T^{FR} = \hat{\beta}_T - \left[\hat{\Sigma}_T \bigotimes \left(\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\right] R' \left(R\left[\hat{\Sigma}_T \bigotimes \left(\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\right] R'\right)^{-1} (R\hat{\beta}_T - q)$$

$$\tilde{\Sigma}_T^{FR} = \frac{1}{T-p} \sum_{t=p+1}^{T} (Y_t - (I_n \bigotimes X_t')\tilde{\beta}_T^{FR})(Y_t - (I_n \bigotimes X_t')\tilde{\beta}_T^{FR})',$$

where $\hat{\Sigma}_T$ is the unrestricted QMLE of Σ. Note that, $\tilde{\beta}_T^{FR}$ continues to satisfy the linear restrictions $R\beta = q$.

2.5.2 Asymptotic Properties of Restricted Estimators

The following theorem shows us that, if the restrictions are true, then $\tilde{\beta}_T^{FR}$ and $\tilde{\Sigma}_T^{FR}$ are consistent and that the difference between the FR and unrestricted QMLE coefficient estimators is asymptotically stable:

Theorem (Asymptotic Properties of FR Estimators)
Maintain assumptions A1 to A3 made in the previous section. If the linear restrictions $R\beta = q$ hold for the true parameter β_0, then

$$\tilde{\beta}_T^{FR} \xrightarrow{p} \beta_0,$$

$$\tilde{\Sigma}_T^{FR} \xrightarrow{p} \Sigma_0,$$

$$\sqrt{T}\left(\tilde{\beta}_T^{FR} - \hat{\beta}_T\right) \xrightarrow{d} N\left[\mathbf{0}, \left(\Sigma_0 \bigotimes Q^{-1}\right) R' \left[R\left(\Sigma_0 \bigotimes Q^{-1}\right) R'\right]^{-1} R\left(\Sigma_0 \bigotimes Q^{-1}\right)\right]$$

$$\tilde{\Sigma}_T^{FR} - \hat{\Sigma}_T = O_p(T^{-1}).$$

Proof) Under assumptions A1 to A3, the unrestricted QMLEs are consistent and asymptotically normal:

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{d} N\left[\mathbf{0}, \begin{pmatrix} \Sigma_0 \otimes Q^{-1} & \left(I_n \otimes Q^{-1}\bar{\mu}\right)\kappa_3 D_n^{+\prime} \\ D_n^{+}\kappa_3'\left(I_n \otimes \bar{\mu}' Q^{-1}\right) & D_n^{+}\left(\kappa_4 - \Sigma_0 \otimes \Sigma_0\right) D_n^{+\prime} \end{pmatrix}\right].$$

Suppose that $R\beta_0 = q$. Then,

$$\tilde{\beta}_T^{FR} = \hat{\beta}_T - \left[\hat{\Sigma}_T \bigotimes \left(\frac{1}{T}\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\right] R' \left(R\left[\hat{\Sigma}_T \bigotimes \left(\frac{1}{T}\sum_{t=p+1}^{T} X_t X_t'\right)^{-1}\right] R'\right)^{-1} (R\hat{\beta}_T - q)$$

$$\xrightarrow{p} \beta_0 - \left[\Sigma_0 \bigotimes Q^{-1}\right] R' \left(R\left[\Sigma_0 \bigotimes Q^{-1}\right] R'\right)^{-1} (R\beta_0 - q) = \beta_0$$

and, letting $\tilde{\beta}_T^{FR} = \text{vec}\left(\tilde{\Pi}_T^{FR}\right)$,

$$\begin{aligned}
\frac{T}{T-p}\tilde{\Sigma}_T^{FR} &= \frac{1}{T}\sum_{t=p+1}^{T} (Y_t - \tilde{\Pi}_T^{FR\prime} X_t)(Y_t - \tilde{\Pi}_T^{FR\prime} X_t)' \\
&= \frac{1}{T}\sum_{t=p+1}^{T} \left[\left(\Pi_0 - \tilde{\Pi}_T^{FR}\right)' X_t + \varepsilon_t\right]\left[\left(\Pi_0 - \tilde{\Pi}_T^{FR}\right)' X_t + \varepsilon_t\right]' \\
&= \left(\Pi_0 - \tilde{\Pi}_T^{FR}\right)' \left(\frac{1}{T}\sum_{t=p+1}^{T} X_t X_t'\right)\left(\Pi_0 - \tilde{\Pi}_T^{FR}\right) + \left(\Pi_0 - \tilde{\Pi}_T^{FR}\right)' \left(\frac{1}{T}\sum_{t=p+1}^{T} X_t \varepsilon_t'\right)
\end{aligned}$$

$$+\left(\frac{1}{T}\sum_{t=p+1}^{T} X_t\varepsilon_t'\right)'\left(\Pi_0-\tilde{\Pi}_T^{FR}\right)+\frac{1}{T}\sum_{t=p+1}^{T}\varepsilon_t\varepsilon_t'$$

$$\xrightarrow{p}\Sigma_0$$

by the consistency of $\tilde{\Pi}_T^{FR}$ for Π_0 and the variance ergodicity of ε_t. This establishes consistency.

As for asymptotic normality, note that

$$\begin{aligned}\tilde{\beta}_T^{FR}-\hat{\beta}_T &= -\left[\hat{\Sigma}_T\bigotimes\left(\sum_{t=p+1}^{T} X_tX_t'\right)^{-1}\right]R'\left(R\left[\hat{\Sigma}_T\bigotimes\left(\sum_{t=p+1}^{T} X_tX_t'\right)^{-1}\right]R'\right)^{-1}(R\hat{\beta}_T-q)\\ &= -\left[\hat{\Sigma}_T\bigotimes\left(\sum_{t=p+1}^{T} X_tX_t'\right)^{-1}\right]R'\left(R\left[\hat{\Sigma}_T\bigotimes\left(\sum_{t=p+1}^{T} X_tX_t'\right)^{-1}\right]R'\right)^{-1}R(\hat{\beta}_T-\beta_0),\end{aligned}$$

where we used the fact that $R\beta_0=q$. By the consistency of $\hat{\theta}_T$,

$$\begin{aligned}&\left[\hat{\Sigma}_T\bigotimes\left(\frac{1}{T}\sum_{t=p+1}^{T} X_tX_t'\right)^{-1}\right]R'\left(R\left[\hat{\Sigma}_T\bigotimes\left(\frac{1}{T}\sum_{t=p+1}^{T} X_tX_t'\right)^{-1}\right]R'\right)^{-1}R\\ &\qquad\xrightarrow{p}\left(\Sigma_0\bigotimes Q^{-1}\right)R'\left[R\left(\Sigma_0\bigotimes Q^{-1}\right)R'\right]^{-1}R.\end{aligned}$$

Furthermore, by the asymptotic normality result above,

$$\sqrt{T}\left(\hat{\beta}_T-\beta_0\right)\xrightarrow{d}N\left[\mathbf{0},\Sigma_0\bigotimes Q^{-1}\right].$$

By Slutsky's theorem,

$$\begin{aligned}\sqrt{T}\left(\tilde{\beta}_T^{FR}-\hat{\beta}_T\right)&\xrightarrow{d}\left(\Sigma_0\bigotimes Q^{-1}\right)R'\left[R\left(\Sigma_0\bigotimes Q^{-1}\right)R'\right]^{-1}R\times N\left[\mathbf{0},\Sigma_0\bigotimes Q^{-1}\right]\\ &=N\left[\mathbf{0},\left(\Sigma_0\bigotimes Q^{-1}\right)R'\left[R\left(\Sigma_0\bigotimes Q^{-1}\right)R'\right]^{-1}R\left(\Sigma_0\bigotimes Q^{-1}\right)\right].\end{aligned}$$

Similarly,

$$\begin{aligned}\tilde{\Sigma}_T^{FR}&=\frac{1}{T-p}\sum_{t=p+1}^{T}(Y_t-\tilde{\Pi}_T^{FR\prime}X_t)(Y_t-\tilde{\Pi}_T^{FR\prime}X_t)'\\ &=\frac{1}{T-p}\sum_{t=p+1}^{T}\left[Y_t-\hat{\Pi}_T'X_t+(\hat{\Pi}_T-\tilde{\Pi}_T^{FR})'X_t\right]\end{aligned}$$

$$\times \left[Y_t - \hat{\Pi}_T' X_t + (\hat{\Pi}_T - \tilde{\Pi}_T^{FR})' X_t\right]'$$

$$= \hat{\Sigma}_T + (\hat{\Pi}_T - \tilde{\Pi}_T^{FR})' \left(\frac{1}{T-p} \sum_{t=p+1}^{T} X_t X_t' \right) (\hat{\Pi}_T - \tilde{\Pi}_T^{FR})$$

$$+ (\hat{\Pi}_T - \tilde{\Pi}_T^{FR})' \left(\frac{1}{T-p} \sum_{t=p+1}^{T} X_t (Y_t - \hat{\Pi}_T' X_t)' \right) + \left(\frac{1}{T-p} \sum_{t=p+1}^{T} X_t (Y_t - \hat{\Pi}_T' X_t)' \right)' (\hat{\Pi}_T - \tilde{\Pi}_T^{FR})$$

$$= \hat{\Sigma}_T + (\hat{\Pi}_T - \tilde{\Pi}_T^{FR})' \left(\frac{1}{T-p} \sum_{t=p+1}^{T} X_t X_t' \right) (\hat{\Pi}_T - \tilde{\Pi}_T^{FR}),$$

where we used the fact that $\sum_{t=p+1}^{T} X_t (Y_t - \hat{\Pi}_T' X_t)' = \mathbf{0}$ by design. We saw above that $\sqrt{T}(\tilde{\beta}_T^{FR} - \hat{\beta}_T) = O_p(1)$, or

$$\tilde{\beta}_T^{FR} - \hat{\beta}_T = O_p(T^{-1/2}).$$

It follows that

$$\hat{\Pi}_T - \tilde{\Pi}_T^{FR} = O_p(T^{-1/2})$$

as well, and therefore,

$$\begin{aligned} \tilde{\Sigma}_T^{FR} - \hat{\Sigma}_T &= (\hat{\Pi}_T - \tilde{\Pi}_T^{FR})' \left(\frac{1}{T-p} \sum_{t=p+1}^{T} X_t X_t' \right) (\hat{\Pi}_T - \tilde{\Pi}_T^{FR}) \\ &= O_p(T^{-1/2}) O_p(1) O_p(T^{-1/2}) \\ &= O_p(T^{-1}). \end{aligned}$$

Q.E.D.

The above result shows us that the difference between the variance matrices converge at a faster rate than the coefficients.

2.5.3 Asymptotic Distribution of LR Test Statistic

Denoting

$$\tilde{\theta}_T^{FR} = \begin{pmatrix} \tilde{\beta}_T^{FR} \\ \text{vech}\left(\tilde{\Sigma}_T^{FR}\right) \end{pmatrix},$$

the value of the quasi log likelihood evaluated at the feasible restricted estimators is

$$l(\tilde{\theta}_T^{FR}) = -\frac{n(T-p)}{2}(\log(2\pi)+1) - \frac{T-p}{2}\log\left(\left|\tilde{\Sigma}_T^{FR}\right|\right),$$

and the difference between the unrestricted maximum log likelihood and restricted maximum log likelihood becomes

$$l(\hat{\theta}_T) - l(\tilde{\theta}_T^{FR}) = \frac{T-p}{2}\left[\log\left|\tilde{\Sigma}_T^{FR}\right| - \log\left|\hat{\Sigma}_T\right|\right].$$

The LR test statistic is now given as

$$\hat{LR}_T := 2\cdot\left(l(\hat{\theta}_T) - l(\tilde{\theta}_T^{FR})\right) = (T-p)\left[\log\left|\tilde{\Sigma}_T^{FR}\right| - \log\left|\hat{\Sigma}_T\right|\right].$$

In light of the result above and the continuous mapping theorem, it stands to reason that the LR statistic is $O_p(1)$. The next theorem shows us that this is indeed the case, and that the asymptotic distribution is a chi-squared distributed with r degrees of freedom under the null:

Theorem (Asymptotic Distribution of LR Statistic)
Maintain assumptions A1 to A3 made in the previous section. Suppose that the null hypothesis $H_0 : R\beta = q$ is true. Then,

$$\begin{aligned} \hat{LR}_T &= (\tilde{\beta}_T^{FR} - \hat{\beta}_T)' \left(\hat{\Sigma}_T^{-1} \bigotimes \frac{1}{T}\sum_{t=p+1}^{T} X_t X_t'\right)(\tilde{\beta}_T^{FR} - \hat{\beta}_T) + o_p(1) \\ &\xrightarrow{d} \chi_r^2. \end{aligned}$$

Proof) The asymptotic properties of the LR test statistic is best studied by employing a Taylor expansion of the log likelihood function. By the stochastic version of the second-order Taylor expansion, there exists a random vector $\bar{\theta}_T$ that is a convex combination of $\hat{\theta}_T$ and $\tilde{\theta}_T^{FR}$ such that

$$\begin{aligned} l(\tilde{\theta}_T^{FR}) &= l(\hat{\theta}_T) + \frac{\partial l(\hat{\theta}_T)}{\partial \theta'}(\tilde{\theta}_T^{FR} - \hat{\theta}_T) + \frac{1}{2}(\tilde{\theta}_T^{FR} - \hat{\theta}_T)'\frac{\partial^2 l(\bar{\theta}_T)}{\partial\theta\partial\theta'}(\tilde{\theta}_T^{FR} - \hat{\theta}_T) \\ &= l(\hat{\theta}_T) + \frac{1}{2}(\tilde{\theta}_T^{FR} - \hat{\theta}_T)' \cdot H(\bar{\theta}_T)\cdot(\tilde{\theta}_T^{FR} - \hat{\theta}_T), \end{aligned}$$

since $\frac{\partial l(\hat{\theta}_T)}{\partial\theta} = \mathbf{0}$ by the definition of the unrestricted QMLE $\hat{\theta}_T$. Since both $\hat{\theta}_T$ and $\tilde{\theta}_T^{FR}$ are consistent for θ_0, and $\bar{\theta}_T$ is a convex combination of the two estimators, $\bar{\theta}_T$ is also consistent for θ_0. We showed above that this implies $H(\bar{\theta}_T) \xrightarrow{p} H_0$, where

$$H_0 = \begin{pmatrix} -\Sigma_0^{-1} \bigotimes Q & O \\ O & -\frac{1}{2} D_n^+ (\Sigma_0^{-1} \bigotimes \Sigma_0^{-1}) D_n \end{pmatrix}.$$

We can write

$$\begin{aligned} l(\tilde{\theta}_T^{FR}) - l(\hat{\theta}_T) &= \frac{1}{2} (\tilde{\theta}_T^{FR} - \hat{\theta}_T)' \cdot H(\bar{\theta}_T) \cdot (\tilde{\theta}_T^{FR} - \hat{\theta}_T) \\ &= \frac{1}{2} \left(\sqrt{T} (\tilde{\beta}_T^{FR} - \hat{\beta}_T)' \quad \sqrt{T} (\tilde{\gamma}_T^{FR} - \hat{\gamma}_T)' \right) \cdot \left[\frac{1}{T} H(\bar{\theta}_T) \right] \begin{pmatrix} \sqrt{T} (\tilde{\beta}_T^{FR} - \hat{\beta}_T) \\ \sqrt{T} (\tilde{\gamma}_T^{FR} - \hat{\gamma}_T) \end{pmatrix}. \end{aligned}$$

The previous theorem tells us that $\sqrt{T}(\tilde{\beta}_T^{FR} - \hat{\beta}_T) = O_p(1)$ but that $\tilde{\Sigma}_T^{FR} - \hat{\Sigma}_T = O_p(T^{-1})$ and thus $\sqrt{T}(\tilde{\gamma}_T^{FR} - \hat{\gamma}_T) = o_p(1)$. Therefore,

$$\hat{LR}_T = 2 \left(l(\hat{\theta}_T) - l(\tilde{\theta}_T^{FR}) \right) = \left[\sqrt{T} (\tilde{\beta}_T^{FR} - \hat{\beta}_T) \right]' \left(\Sigma_0^{-1} \bigotimes Q \right) \left[\sqrt{T} (\tilde{\beta}_T^{FR} - \hat{\beta}_T) \right] + o_p(1),$$

where $-\Sigma_0^{-1} \bigotimes Q$ is the matrix in the (1,1) block of H_0. Due to the consistency of $\hat{\Sigma}_T^{-1} \bigotimes \frac{1}{T} \sum_{t=p+1}^T X_t X_t'$ for $\Sigma_0^{-1} \bigotimes Q$, it follows that

$$\hat{LR}_T = (\tilde{\beta}_T^{FR} - \hat{\beta}_T)' \left(\hat{\Sigma}_T^{-1} \bigotimes \frac{1}{T} \sum_{t=p+1}^T X_t X_t' \right) (\tilde{\beta}_T^{FR} - \hat{\beta}_T) + o_p(1).$$

It is now easy to derive the asymptotic distribution of the LR statistic. Recall that

$$\sqrt{T} (\tilde{\beta}_T^{FR} - \hat{\beta}_T) \xrightarrow{d} (\Sigma_0 \bigotimes Q^{-1}) R' L \times Z,$$

where $L \in \mathbb{R}^{r \times r}$ is the Cholesky factor of

$$\left[R \left(\Sigma_0 \bigotimes Q^{-1} \right) R' \right]^{-1}$$

and $Z = (Z_1, \cdots, Z_r)$ is an r-dimensional standard normally distributed random variable. It follows from the continuous mapping theorem and Slutsky's theorem that

$$\begin{aligned} \hat{LR}_T &\xrightarrow{d} Z' \left[L' R (\Sigma_0 \bigotimes Q^{-1}) \left(\Sigma_0^{-1} \bigotimes Q \right) (\Sigma_0 \bigotimes Q^{-1}) R' L \right] Z \\ &= Z' \left[L' R (\Sigma_0 \bigotimes Q^{-1}) R' L \right] Z. \end{aligned}$$

Since $(LL')^{-1} = R(\Sigma_0 \bigotimes Q^{-1}) R'$, it follows that $L' R (\Sigma_0 \bigotimes Q^{-1}) R' L = I_r$, and we have

$$\hat{LR}_T \xrightarrow{d} Z' Z = \sum_{i=1}^r Z_i^2.$$

The random variable on the right hand side is the sum of r squared independent standard normally distributed random variables, so by definition it follows a chi-squared distribution with r degrees of freedom.

Q.E.D.

2.6 Lag Length Selection

So far, we have taken the lag order p of a VAR model as given. In reality, it is often the case that p is unknown, and we must estimate the lag order. To obtain a rudimentary criterion with which to compare models with different lag orders, we return once again to the (quasi) log likelihood for a VAR(p) process.

Throughout, we will assume that the maximum lag order is given by $k > 0$, and that we search between 0 and k for the true lag order. In order to use the same number of observations in any case, we assume that observations from time $k+1$ to T are used. For notational convenience, in what follows we denote

$$\Pi(p) = \begin{pmatrix} \delta' \\ \Phi_1' \\ \vdots \\ \Phi_p' \end{pmatrix} \quad \text{and} \quad \Pi(p:q) = \begin{pmatrix} \Phi_{p+1}' \\ \vdots \\ \Phi_q' \end{pmatrix} \quad \text{for any } p < q \leq k$$

$$X_t(p) = \begin{pmatrix} 1 \\ Y_t \\ \vdots \\ Y_{t-p} \end{pmatrix} \quad \text{and} \quad X_t(p:q) = \begin{pmatrix} Y_{t-p-1} \\ \vdots \\ Y_{t-q} \end{pmatrix} \quad \text{for any } p < q \leq k$$

$$\hat{\Pi}_T(p) = \left(\sum_{t=k+1}^{T} X_t(p) X_t(p)' \right)^{-1} \sum_{t=k+1}^{T} X_t(p) Y_t'$$

$$\hat{\Sigma}_T(p) = \frac{1}{T-k} \sum_{t=k+1}^{T} \left[Y_t - \hat{\Pi}_T(p)' X_t(p) \right] \left[Y_t - \hat{\Pi}_T(p)' X_t(p) \right]'.$$

In other words, in a model with $p \leq k$ lags, $X_t(p)$ is the regressor used for QML estimation, $\hat{\Pi}_T(p)$ the QMLE of $\Pi = (\delta, \Phi_1, \cdots, \Phi_p)'$, and $\hat{\Sigma}_T(p)$ the QMLE of Σ. In a model with $p \leq k$ lags, the paramter space is denoted by

$$\Theta(p) = \mathbb{R}^{n(np+1)} \times \mathcal{A}.$$

The quasi log likelihood for a model with $p \leq k$ lags is given as

$$l(\theta; p) = -\frac{n(T-k)}{2} \log(2\pi) - \frac{T-k}{2} \log|\Sigma| - \frac{1}{2} \operatorname{tr} \left(\Sigma^{-1} \cdot \sum_{t=k+1}^{T} (Y_t - \Pi' X_t(p))(Y_t - \Pi' X_t(p))' \right)$$

for any $\theta \in \Theta$, and the maximized log likelihood is

$$\hat{l}_T(p) = -\frac{n(T-k)}{2} (\log(2\pi) + 1) - \frac{T-k}{2} \log \left| \hat{\Sigma}_T(p) \right|.$$

2.6.1 Sequential LR Testing

We saw above that $\log\left|\hat{\Sigma}_T(p)\right|$ can serve as an indicator of the fit of a model with p lags. Thus, an intuitively appealing way to search for the true lag order is to conduct sequential LR tests.

Specifically, suppose we wish to test the null hypothesis of $p-1$ lags against p lags. This can be viewed as testing whether $\Phi_p = O$ in the VAR(p) model

$$Y_t = \delta + \Phi_1 Y_{t-1} + \cdots + \Phi_{p-1} Y_{t-p+1} + \Phi_p Y_{t-p} + \varepsilon_t.$$

The unrestricted estimators of Π and Σ are, of course, given by $\hat{\Pi}_T(p)$ and $\hat{\Sigma}_T(p)$.

Defining

$$R = \underbrace{\begin{pmatrix} O_{n^2\times(n^2(p-1)+n)} & I_{n^2} \end{pmatrix}}_{J} K_{np+1,n} \in \mathbb{R}^{n^2\times n(np+1)},$$

R is a matrix of full rank n^2 such that

$$R\beta = R\text{vec}(\Pi) = J\cdot\text{vec}(\Pi') = \begin{pmatrix} O_{n^2\times(n^2(p-1)+n)} & I_{n^2} \end{pmatrix} \text{vec}\begin{pmatrix} \delta & \Phi_1 & \cdots & \Phi_{p-1} & \Phi_p \end{pmatrix} = \text{vec}(\Phi_p).$$

Thus, the null $H_0 : \Phi_p = O$ can be expressed equivalently as

$$H_0 : R\beta = \mathbf{0}.$$

Note that the restricted QMLEs $\tilde{\theta}_T$ of θ solve the problem

$$\max_{\theta\in\Theta(p)} \quad -\frac{T-k}{2}\log|\Sigma| - \frac{1}{2}\text{tr}\left(\Sigma^{-1}\cdot\sum_{t=k+1}^{T}(Y_t - \Pi(p)'X_t(p))(Y_t - \Pi(p)'X_t(p))\right)$$
$$\text{subject to} \quad \Phi_p = O.$$

Substituting the constraint into the objective function, the maximization problem can be rewritten as the unconstrained problem

$$\max_{\theta\in\Theta(p-1)} \quad -\frac{T-k}{2}\log|\Sigma| - \frac{1}{2}\text{tr}\left(\Sigma^{-1}\cdot\sum_{t=k+1}^{T}(Y_t - \Pi(p-1)'X_t(p-1))(Y_t - \Pi(p-1)'X_t(p-1))\right).$$

This indicates that $\tilde{\beta}_T$, the restricted QMLE of β, is simply the QMLEs of β in a VAR(p-1) model together with Φ_p equal to O. Similarly, $\tilde{\Sigma}_T$, the restricted QMLE of Σ, can also put equal to the covariance estimator in a VAR(p-1) model:

$$\tilde{\Sigma}_T = \frac{1}{T-p}\sum_{t=p+1}^{T}(Y_t - \hat{\Pi}_T(p-1)'X_t(p-1))(Y_t - \hat{\Pi}_T(p-1)'X_t(p-1))' = \hat{\Sigma}_T(p-1).$$

Since an estimate of Σ no longer appears in the equation for $\tilde{\beta}_T$, in this case the restricted estimators are identical to the FR estimators. Therefore, the asymptotic theory developed above continues to hold, and we can see that

$$\hat{LR}_T = (T-k)\left(\log\left|\hat{\Sigma}_T(p-1)\right| - \log\left|\hat{\Sigma}_T(p)\right|\right) \xrightarrow{d} \chi^2_{n^2}.$$

2.6.2 Information Criteria

Often, the lag order of a VAR model is chosen by minimizing an information criterion. Generally, information criteria are given as

$$IC(p) = \log\left|\hat{\Sigma}_T(p)\right| + p \cdot \frac{c_T}{T}$$

for lag orders $0 \le p \le k$, where c_T is a penalty term that is often deterministic functions of the sample size T. The first part of the criterion, $\log\left|\hat{\Sigma}_T(p)\right|$, is easy to understand; it represents the negative log-likelihood, so that the lower it is, the better the model fit. However, relying on only the negative log-likelihood may induce overfitting, since the higher the lag order, the better the model fit and thus the higher the log-likelihood. Therefore, a penalty term $p \cdot \frac{c_T}{T}$ is introduced in order to penalize lag orders that are too high without significantly improving the log-likelihood.

Given an information criterion, the optimal lag order is chosen as the value $\hat{p}_T$ that minimizes $IC(p)$ over $\{0, \cdots, k\}$. Letting $p_0 \in \{0, \cdots, k\}$ be the true lag order, we say that $IC(p)$ consistently estimates the true lag order if

$$\hat{p}_T \xrightarrow{p} p_0,$$

which is equivalent in this case to

$$\lim_{T\to\infty} \mathbb{P}(\hat{p}_T = p_0) = 1.$$

Since $\hat{p}_T$ is chosen as the minimizer of $IC(p)$ over $\{0, \cdots, k\}$,

$$\{\hat{p}_T = p_0\} = \bigcap_{p \neq p_0} \{IC(p) > IC(p_0)\}.$$

Therfore, another equivalent characterization of consistency is

$$\lim_{T\to\infty} \mathbb{P}(IC(p) > IC(p_0)) = 1$$

for any $0 \le p \le k$ such that $p \neq p_0$.

We can furnish a sufficient condition for the consistency of information criteria. First, a lemma:

Lemma (Simultaneous Diagonlization of Positive Semidefinite Matrices)
Suppose $A, B \in \mathbb{R}^{n\times n}$ are positive semidefinite matrices, and that A is positive definite. Then, there exists a nonsingular $P \in \mathbb{R}^{n\times n}$ and a diagonal matrix $D \in \mathbb{R}^{n\times n}$ with non-negative diagonal entries such that

$$A = PP' \quad \text{and} \quad B = PDP'.$$

Proof) Let $A = \Lambda_1 D_1 \Lambda_1'$ be the eigendecomposition of A, where Λ_1 is an orthogonal matrix and D_1 is a diagonal matrix collecting the eigenvalues of A. Since A is positive definite, the diagonal entries of D_1 are all positive, meaning that $D_1^{\frac{1}{2}}$ is well defined and nonsingular. Define

$$M = \Lambda_1 D_1^{\frac{1}{2}}.$$

Defining

$$C = M^{-1} B M'^{-1},$$

since B is positive semidefinite, so is C. Therefore, C has eigendecomposition

$$C = \Lambda_2 D \Lambda_2',$$

where the diagonal elements of D are non-negative. It follows that

$$B = MCM' = \Lambda_1 D_1^{\frac{1}{2}} \Lambda_2 \cdot D \cdot \Lambda_2' D_1^{\frac{1}{2}} \Lambda_1'.$$

Define

$$P = \Lambda_1 D_1^{\frac{1}{2}} \Lambda_2.$$

Since all three matrices that comprise P are invertible, so is P. By definition,

$$B = PDP',$$

where D is a diagonal matrix with non-negative entries. Finally,

$$PP' = \Lambda_1 D_1^{\frac{1}{2}} \Lambda_2 \Lambda_2' D_1^{\frac{1}{2}} \Lambda_1' = \Lambda_1 D_1 \Lambda_1' = A,$$

where we used the fact that Λ_2 is an orthogonal matrix.

Q.E.D.

Theorem (Consistency of Information Criteria)
Maintain assumptions A1 to A3 made in the previous section. Let $IC(p)$ be an information criterion given by

$$IC(p) = \log\left|\hat{\Sigma}_T(p)\right| + p \cdot \frac{c_T}{T}$$

for any $0 \le p \le k$, and $\hat{p}_T$ the lag order chosen by $IC(p)$. $\hat{p}_T$ is consistent for the true lag order p_0 if

$$c_T \to +\infty \quad \text{and} \quad \frac{c_T}{T} \to 0$$

as $T \to \infty$.

Proof) The theorem is proven by showing that

$$\mathbb{P}\left(IC(p) > IC(p_0)\right) \to 1$$

as $T \to \infty$ for any $p < p_0$ since $\frac{c_T}{T} \to 0$ and for any $p > p_0$ since $c_T \to +\infty$. We first study the case $p < p_0$.

Step 1: $p < p_0$

Suppose that $p < p_0$. Recall that

$$\hat{\Pi}_T(p) = \left(\sum_{t=k+1}^{T} X_t(p)X_t(p)'\right)^{-1} \sum_{t=k+1}^{T} X_t(p)Y_t$$
$$\hat{\Sigma}_T(p) = \frac{1}{T-k}\sum_{t=k+1}^{T} (Y_t - \hat{\Pi}_T(p)'X_t(p))(Y_t - \hat{\Pi}_T(p)'X_t(p))',$$

and likewise for p_0. From the asymptotic properties of the QMLE estimators under the correctly specified VAR model, $\hat{\Pi}_T(p_0)$ and $\hat{\Sigma}_T(p_0)$ are consistent for the true parameters $\Pi_0 = \Pi(p_0)$ and Σ_0. Meanwhile, note that

$$Y_t = \Pi(p_0)'X_t(p_0) + \varepsilon_t = \Pi(p)'X_t(p) + \Pi(p:p_0)'X_t(p:p_0) + \varepsilon_t$$

for any $t \in \mathbb{Z}$, so that

$$\begin{aligned}\hat{\Pi}_T(p) &= \left(\sum_{t=k+1}^{T} X_t(p)X_t(p)'\right)^{-1} \sum_{t=k+1}^{T} X_t(p)(\Pi(p)'X_t(p) + \Pi(p:p_0)'X_t(p:p_0) + \varepsilon_t)' \\ &= \Pi(p) + \left(\frac{1}{T}\sum_{t=k+1}^{T} X_t(p)X_t(p)'\right)^{-1}\left(\frac{1}{T}\sum_{t=k+1}^{T} X_t(p)X_t(p:p_0)'\right)\Pi(p:p_0)\end{aligned}$$

$$+\left(\frac{1}{T}\sum_{t=k+1}^{T}X_t(p)X_t(p)'\right)^{-1}\left(\frac{1}{T}\sum_{t=k+1}^{T}X_t(p)\varepsilon_t'\right).$$

By assumption,

$$\frac{1}{T}\sum_{t=k+1}^{T}X_t(p_0)X_t(p_0)' \xrightarrow{p} Q = \begin{pmatrix} Q(p) & Q_{12} \\ Q_{21} & Q(p:p_0) \end{pmatrix},$$

where Q and by extension $Q(p), Q(p:p_0)$ are symmetric positive definite matrices. We can thus see that

$$\frac{1}{T}\sum_{t=k+1}^{T}X_t(p)X_t(p)' \xrightarrow{p} Q(p) \in \mathbb{R}^{(np+1)\times(np+1)}$$

$$\frac{1}{T}\sum_{t=k+1}^{T}X_t(p)X_t(p:p_0)' \xrightarrow{p} Q_{12} \in \mathbb{R}^{(np+1)\times n(p_0-p)}.$$

Also as shown in a theorem above, under our assumptions $\frac{1}{T}\sum_{t=k+1}^{T}X_t(p)\varepsilon_t' = o_p(1)$. Therefore, by the continuous mapping theorem,

$$\hat{\Pi}_T(p) \xrightarrow{p} \Pi(p) + Q(p)^{-1}Q_{12}\cdot\Pi(p:p_0).$$

Defining

$$\tilde{\Pi}_T(p) = \begin{pmatrix} \hat{\Pi}_T(p) \\ O_{n(p_0-p)\times n} \end{pmatrix},$$

we can see that

$$\hat{\Sigma}_T(p) = \frac{1}{T-k}\sum_{t=k+1}^{T}\left[Y_t - \tilde{\Pi}_T(p)'X_t(p_0)\right]\left[Y_t - \tilde{\Pi}_T(p)'X_t(p_0)\right]'.$$

Expanding this expression as usual yields

$$\begin{aligned}\hat{\Sigma}_T(p) = &\frac{1}{T-k}\sum_{t=k+1}^{T}\varepsilon_t\varepsilon_t' \\ &+(\Pi(p_0)-\tilde{\Pi}_T(p))'\left(\frac{1}{T-k}\sum_{t=k+1}^{T}X_t(p_0)X_t(p_0)'\right)(\Pi(p_0)-\tilde{\Pi}_T(p)) \\ &+(\Pi(p_0)-\tilde{\Pi}_T(p))'\left(\frac{1}{T-k}\sum_{t=k+1}^{T}X_t(p_0)\varepsilon_t'\right)+\left(\frac{1}{T-k}\sum_{t=k+1}^{T}X_t(p_0)\varepsilon_t'\right)'(\Pi(p_0)-\tilde{\Pi}_T(p)).\end{aligned}$$

Since

$$\tilde{\Pi}_T(p) \xrightarrow{p} \begin{pmatrix} \Pi(p) + Q(p)^{-1}Q_{12}\cdot\Pi(p:p_0) \\ O_{n(p_0-p)\times n} \end{pmatrix},$$

$\tilde{\Pi}_T(p)$ is $O_p(1)$, so that the last two terms above converge in probability to 0. Meanwhile, we can see that

$$\begin{aligned}
&(\Pi(p_0)-\tilde{\Pi}_T(p))'\left(\frac{1}{T-k}\sum_{t=k+1}^{T}X_t(p_0)X_t(p_0)'\right)(\Pi(p_0)-\tilde{\Pi}_T(p)) \\
&\xrightarrow{p}\begin{pmatrix}-Q(p)^{-1}Q_{12}\cdot\Pi(p:p_0)\\ \Pi(p:p_0)\end{pmatrix}' Q\begin{pmatrix}-Q(p)^{-1}Q_{12}\cdot\Pi(p:p_0)\\ \Pi(p:p_0)\end{pmatrix} \\
&=\Pi(p:p_0)'\cdot\begin{pmatrix}-Q_{21}'Q(p)^{-1} & I_{n(p_0-p)}\end{pmatrix}Q\begin{pmatrix}-Q(p)^{-1}Q_{12}\\ I_{n(p_0-p)}\end{pmatrix}\cdot\Pi(p:p_0) \\
&=\Pi(p:p_0)'\left[Q(p:p_0)-Q_{21}Q(p)^{-1}Q_{12}\right]\Pi(p:p_0):=\tilde{Q}.
\end{aligned}$$

Since $Q(p:p_0)-Q_{21}Q(p)^{-1}Q_{12}$ is the Schur complement of Q, a positive definite matrix, it is also positive definite. Therefore, $\tilde{Q}$ is an $n(p_0-p)\times n(p_0-p)$ positive semidefinite matrix with at least one positive eigenvalue (otherwise, $\Pi(p:p_0)$ must be equal to 0, which contradicts the fact that p_0 is the true lag order). Together, we can see that

$$\hat{\Sigma}_T(p)\xrightarrow{p}\Sigma_0+\tilde{Q}.$$

The log-likelihood ratio of a model with p lags and one with p_0 lags is given by

$$\log\left|\hat{\Sigma}_T(p)\right|-\log\left|\hat{\Sigma}_T(p_0)\right|.$$

By the consistency results derived above,

$$\log\left|\hat{\Sigma}_T(p)\right|-\log\left|\hat{\Sigma}_T(p_0)\right|\xrightarrow{p}\log\left|\Sigma_0+\tilde{Q}\right|-\log|\Sigma_0|.$$

Since Σ_0 is positive definite and $\tilde{Q}$ is positive semidefinite, the preceding lemma tells us that there exist a nonsingular $P\in\mathbb{R}^{n\times n}$ and a diagonal matrix $D\in\mathbb{R}^{n\times n}$ with non-negative diagonal elements such that

$$\Sigma_0=PP'\quad\text{and}\quad\tilde{Q}=PDP'.$$

Note that at least one element of D is positive, since otherwise $\tilde{Q}=O$, a contradiction. We can now see that

$$\begin{aligned}
\log\left|\Sigma_0+\tilde{Q}\right|-\log|\Sigma_0| &= \log|P(D+I_n)P'|-\log|PP'| \\
&=\log\left(|P|^2\cdot|D+I_n|\right)-\log\left(|P|^2\right)=\log|D+I_n|>0,
\end{aligned}$$

where the last inequality follows because $|D+I_n|>1$. Therefore,

$$\log\left|\hat{\Sigma}_T(p)\right|-\log\left|\hat{\Sigma}_T(p_0)\right|\xrightarrow{p}\log|D+I_n|>0.$$

Since $\frac{c_T}{T} \to 0$ as $T \to \infty$ by assumption, this tells us that

$$IC(p) - IC(p_0) = \log\left|\hat{\Sigma}_T(p)\right| - \log\left|\hat{\Sigma}_T(p_0)\right| + (p - p_0) \cdot \frac{c_T}{T} \xrightarrow{p} \log|D + I_n| > 0.$$

By definition,

$$\lim_{T\to\infty} \mathbb{P}\left(\left|(IC(p) - IC(p_0)) - \log|D + I_n|\right| < \epsilon\right) = 1$$

for any $\epsilon > 0$. Putting $\epsilon = \log|D + I_n| > 0$ yields

$$\lim_{T\to\infty} \mathbb{P}\left(IC(p) - IC(p_0) > 0\right) = 1.$$

Step 2: $p > p_0$

Now suppose that $p_0 < p \leq k$. In this case, the true model is a restricted version of the model with p lags; that is, the solutions to the constrained maximization problem

$$\begin{aligned} &\max_{\theta \in \Theta(p)} \quad -\frac{n(T-k)}{2}\log|\Sigma| - \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1} \cdot \sum_{t=k+1}^{T} (Y_t - \Pi' X_t(p))(Y_t - \Pi' X_t(p))'\right) \\ &\text{subject to} \quad \Phi_{p_0+1} = \cdots = \Phi_p = O \end{aligned}$$

are

$$\tilde{\Pi}_T = \begin{pmatrix} \hat{\Pi}_T(p_0) \\ O_{n(p-p_0)\times n} \end{pmatrix} \quad \text{and} \quad \hat{\Sigma}_T(p_0).$$

Here, no covariance terms enter into the solution for β, so that the restricted estimators are identical to the feasible restricted estimators. Therefore, as we derived for the LR test statistic computed using FR estimators,

$$\hat{LR}_T = T \cdot \left(\log\left|\hat{\Sigma}_T(p_0)\right| - \log\left|\hat{\Sigma}_T(p)\right|\right) \xrightarrow{d} \chi^2_{n^2(p-p_0)}.$$

In other words,

$$\log\left|\hat{\Sigma}_T(p_0)\right| - \log\left|\hat{\Sigma}_T(p)\right| = O_p(T^{-1}),$$

which implies that $\log\left|\hat{\Sigma}_T(p_0)\right| - \log\left|\hat{\Sigma}_T(p)\right| = o_p(1)$.

The difference between the information criteria for lags p and p_0 is given as

$$IC(p) - IC(p_0) = -\left(\log\left|\hat{\Sigma}_T(p_0)\right| - \log\left|\hat{\Sigma}_T(p)\right|\right) + (p - p_0)\frac{c_T}{T}.$$

By the definition of boundedness in probability, for any $\epsilon > 0$ there exists an $M > 0$ such that

$$\mathbb{P}((p - p_0) \cdot c_T - T(IC(p) - IC(p_0)) > M) = \mathbb{P}\left(T\left(\log\left|\hat{\Sigma}_T(p_0)\right| - \log\left|\hat{\Sigma}_T(p)\right|\right) > M\right) < \epsilon$$

for large enough T. By assumption, $c_T \to +\infty$ as $T \to \infty$, and $p - p_0 > 0$, so for large enough T we have $(p - p_0)c_T > M + 1$ and as such

$$\mathbb{P}(T(IC(p) - IC(p_0)) \leq 0) \leq \mathbb{P}(T(IC(p) - IC(p_0)) < (p - p_0) \cdot c_T - M - 1) < \epsilon$$

for any T that is large enough. Thus,

$$\limsup_{T\to\infty} \mathbb{P}(T(IC(p) - IC(p_0)) \leq 0) \leq \epsilon,$$

and since this holds for any $\epsilon > 0$,

$$\lim_{T\to\infty} \mathbb{P}(T(IC(p) - IC(p_0)) \leq 0) = 0,$$

or equivalently,

$$\lim_{T\to\infty} \mathbb{P}(T(IC(p) - IC(p_0)) > 0) = 1.$$

Finally, for any $T \geq k + 1$,

$$\{T(IC(p) - IC(p_0)) > 0\} = \{IC(p) - IC(p_0) > 0\},$$

so that

$$\lim_{T\to\infty} \mathbb{P}(IC(p) - IC(p_0) > 0) = 1$$

as well.

Q.E.D.

We have thus seen that the condition $\frac{c_T}{T} \to 0$ is required to ensure that lag orders smaller than p_0 are not chosen; in effect, it does not penalize lag orders that are smaller than p_0. On the other hand, the condition $c_T \to +\infty$ is required to preculde lag orders greater than p_0; heuristically, because lag orders greater than p_0 actually yield higher log likelihoods than p_0, they must be penalized heavily.

Examples of information criteria include:

AIC: **Akaike Information Criterion**

This is perhaps the most widely used information criterion, and is given as

$$AIC(p) = \log\left|\hat{\Sigma}_T(p)\right| + np(np+1)\cdot\frac{2}{T},$$

where $np(np+1)$ is the number of freely estimated coefficient parameters $(n(np+1))$ multiplied by the lag length p.

Despite its renown, it is inconsistent for lag orders higher than the true lag length p_0, since $c_T = 2$ for any $T \in N_+$ and therefore $\frac{c_T}{T} = O_p(T^{-1})$ in this case. It follows that $AIC(p) - AIC(p_0)$ is the sum of two $O_p(T^{-1})$ terms for any $p > p_0$, so that $AIC(p) - AIC(p_0) > 0$ does not hold in the limit with probability 1.

BIC: **Bayesian Information Criterion**

This is a widely used consistent information criterion, and is given as

$$BIC(p) = \log\left|\hat{\Sigma}_T(p)\right| + np(np+1)\cdot\frac{\log(T)}{T}.$$

Clearly, $\frac{\log(T)}{T} \to 0$ but $\log(T) \to +\infty$ as $T \to \infty$, so that it is consistent in light of the preceding theorem.

HQ: **Hannan-Quinn Information Criterion**

This is an information criterion that is strongly consistent, or in other words, $\hat{p}_T \overset{a.s.}{\to} p_0$ (we omit the proof):

$$HQ(p) = \log\left|\hat{\Sigma}_T(p)\right| + np(np+1)\cdot\frac{2\log(\log(T))}{T}.$$

The penalty term also satisfies the conditions of the preceding theorem.

Spectral Analysis

In this chapter we study the frequency domain representation of time series through the spectral representation theorem. In addition, we introduce the spectral density and investigate some consistent non-parametric estimators of the spectral density.

3.1 The Spectral Density

So far, we have presented time series as a doubly infinite sequence $\{Y_t\}_{t\in\mathbb{Z}}$ of random vectors taking values in euclidean space. We can instead express this time series in terms of its frequencies, that is, weighted averages of sinusoidal waves of the form $t \mapsto \exp(ir\cdot)$ for some $r \in \mathbb{R}$. The spectral representation theorem, the proof of which is the main objective of this section, shows that every weakly stationary and well-behaved time series possesses both time domain and frequency domain representations.

First, we define and study the properties of the spectral density of time series. Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional mean zero weakly stationary time series with autocovariance function $\Gamma : \mathbb{Z} \to \mathbb{R}^n$. If the autocovariances of $\{Y_t\}_{t\in\mathbb{Z}}$ are absolutely summable, then we can define the function $f : (-\pi, \pi] \to \mathbb{C}^n$ as

$$f(w) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \Gamma(\tau) \exp(-i\tau w)$$

for any $w \in (-\pi, \pi]$; note that this series converges to an $n \times n$ complex matrix in this case because it is absolutely convergent under absolutely summable autocovariances:

$$\sum_{\tau=-\infty}^{\infty} \|\Gamma(\tau) \exp(-i\tau w)\| = \sum_{\tau=-\infty}^{\infty} \|\Gamma(\tau)\| < +\infty.$$

Note that

$$\operatorname{tr}(f(w)) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \operatorname{tr}(\Gamma(\tau)) \exp(-irw)$$

for any $w \in (-\pi, \pi]$. Since $\Gamma(-\tau) = \Gamma(\tau)'$ for any $\tau \in \mathbb{Z}$, $\operatorname{tr}(\Gamma(\tau)) = \operatorname{tr}(\Gamma(-\tau))$, and we can see that

$$\operatorname{tr}(f(w)) = \frac{1}{2\pi} \left[\operatorname{tr}(\Gamma(0)) + \sum_{\tau=1}^{\infty} \operatorname{tr}(\Gamma(\tau)) (\exp(-i\tau w) + \exp(i\tau w)) \right]$$

$$= \frac{1}{2\pi}\left[\operatorname{tr}(\Gamma(0)) + 2\sum_{\tau=1}^{\infty}\operatorname{tr}(\Gamma(\tau))\cos(\tau w)\right],$$

so that $\operatorname{tr}(f)$ is a real-valued function. We can in fact show a stronger result, namely that $\operatorname{tr}(f)$ is non-negative everywhere on $(-\pi, \pi]$.

Lemma Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional weakly stationary time series with absolutely summable autocovariance function $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$. Letting $f : (-\pi, \pi] \to \mathbb{C}^{n\times n}$ be the spectral density of $\{Y_t\}_{t\in\mathbb{Z}}$, $\operatorname{tr}(f(w)) \geq 0$ for any $w \in (-\pi, \pi]$. Furthermore, $f(0)$ is a positive semidefinite matrix.

Proof) For any $T \in N_+$, define $f_T : (-\pi, \pi] \to \mathbb{C}^{n\times n}$ as

$$\begin{aligned} f_T(w) &= \frac{1}{2\pi T}\sum_{s=1}^{T}\sum_{r=1}^{T}\Gamma(s-r)\exp(-isw)\exp(irw) \\ &= \frac{1}{2\pi}\sum_{\tau=-T+1}^{T-1}\left(1 - \frac{|\tau|}{T}\right)\Gamma(\tau)\exp(-i\tau w) \end{aligned}$$

for any $w \in (-\pi, \pi]$. Fixing $w \in (-\pi, \pi]$,

$$f(w) - f_T(w) = \frac{1}{2\pi}\sum_{|\tau|<T}\frac{|\tau|}{T}\Gamma(\tau)\exp(-i\tau w) + \frac{1}{2\pi}\sum_{|\tau|\geq T}\Gamma(\tau)\exp(-i\tau w).$$

so we have

$$\|f(w) - f_T(w)\| \leq \frac{1}{2\pi}\sum_{|\tau|\geq T}\|\Gamma(\tau)\| + \frac{1}{2\pi}\sum_{|\tau|<T}\frac{|\tau|}{T}\cdot\|\Gamma(\tau)\|.$$

Choose some $\epsilon > 0$. The absolute summability of the autocovariances tells us that there exists some $N \in N_+$ such that

$$\sum_{|\tau|\geq T}\|\Gamma(\tau)\| < \frac{\epsilon}{3}.$$

for any $T \geq N$. For such T, we also have

$$\begin{aligned} \sum_{|\tau|<T}\frac{|\tau|}{T}\|\Gamma(\tau)\| &\leq \sum_{|\tau|<N}\frac{|\tau|}{T}\|\Gamma(\tau)\| + \sum_{N\leq|\tau|<T}\|\Gamma(\tau)\| \\ &\leq \frac{N}{T}\sum_{|\tau|<N}\|\Gamma(\tau)\| + \sum_{|\tau|\geq N}\|\Gamma(\tau)\| \\ &< \frac{N}{T}\sum_{\tau=-N+1}^{N-1}\|\Gamma(\tau)\| + \frac{\epsilon}{3}. \end{aligned}$$

The first term here goes to 0 as $T \to \infty$, so there exists some $N_0 \geq N$ such that

$$\frac{N}{T} \sum_{\tau=-N+1}^{N-1} \|\Gamma(\tau)\| < \frac{\epsilon}{3}$$

for any $T \geq N_0$. Therefore, for any $T \geq N_0$, since $T \geq N$ as well, we have

$$\begin{aligned} \|f(w) - f_T(w)\| &\leq \frac{1}{2\pi} \sum_{|\tau| \geq T} \|\Gamma(\tau)\| + \frac{1}{2\pi} \left(\frac{N}{T} \sum_{\tau=-N+1}^{N-1} \|\Gamma(\tau)\| + \frac{\epsilon}{3} \right) \\ &\leq \frac{1}{2\pi} \frac{\epsilon}{3} + \frac{1}{2\pi} \left(\frac{\epsilon}{3} + \frac{\epsilon}{3} \right) = \frac{1}{2\pi} \epsilon < \epsilon. \end{aligned}$$

Such an $N_0 \in N_+$ exists for any $\epsilon > 0$, so it follows that

$$\lim_{T \to \infty} f_T(w) = f(w).$$

For any $w \in (-\pi, \pi]$ and $T \in N_+$, we can show that $\mathrm{tr}\left(f_T(w)\right) \geq 0$. To this end, note that we can write

$$\begin{aligned} \mathrm{tr}\left(f_T(w)\right) &= \frac{1}{2\pi T} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathrm{tr}(\Gamma(s-r)) \exp(-isw) \exp(irw) \\ &= \frac{1}{2\pi T} \begin{pmatrix} \exp(-iw) & \cdots & \exp(-iTw) \end{pmatrix} \begin{pmatrix} \mathrm{tr}(\Gamma(0)) & \cdots & \mathrm{tr}(\Gamma(T-1)) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(\Gamma(T-1)) & \cdots & \mathrm{tr}(\Gamma(0)) \end{pmatrix} \begin{pmatrix} \exp(iw) \\ \vdots \\ \exp(iTw) \end{pmatrix}. \end{aligned}$$

Defining

$$\beta = \begin{pmatrix} \exp(iw) \\ \vdots \\ \exp(iTw) \end{pmatrix},$$

we can rewrite $\mathrm{tr}(f_T(w))$ as

$$\begin{aligned} \mathrm{tr}(f_T(w)) &= \frac{1}{2\pi T} \overline{\beta}' \mathbb{E} \left[\begin{pmatrix} Y_1' Y_1 & \cdots & Y_1' Y_T \\ \vdots & \ddots & \vdots \\ Y_T' Y_1 & \cdots & Y_T' Y_T \end{pmatrix} \right] \beta \\ &= \frac{1}{2\pi T} \mathbb{E} \left[\overline{\beta}' \begin{pmatrix} Y_1' \\ \vdots \\ Y_T' \end{pmatrix} \begin{pmatrix} Y_1 & \cdots & Y_T \end{pmatrix} \beta \right] = \frac{1}{2\pi T} \mathbb{E} \left[\overline{Z_T}' Z_T \right] = \frac{1}{2\pi T} \mathbb{E} |Z_T|^2 \geq 0, \end{aligned}$$

where we define the n-dimensional random vector Z_T as

$$Z_T = \begin{pmatrix} Y_1 & \cdots & Y_T \end{pmatrix} \beta.$$

Similarly, we can show that the $n \times n$ real valued matrix $f_T(0)$ is positive semidefinite. Choose any $\alpha \in \mathbb{R}^n$, and define $y_t = \alpha' Y_t$ for any $t \in \mathbb{Z}$. Then, $\{y_t\}_{t\in\mathbb{Z}}$ is a univariate mean zero weakly stationary process with autocovariance function $\gamma : \mathbb{Z} \to \mathbb{R}$ defined as

$$\gamma(\tau) = \mathbb{E}\left[y_t y_{t-\tau}\right] = \alpha' \mathbb{E}\left[Y_t Y_{t-\tau}'\right] \alpha = \alpha' \Gamma(\tau) \alpha$$

for any $\tau \in \mathbb{Z}$. Then,

$$\begin{aligned}
\alpha' f_T(0) \alpha &= \frac{1}{2\pi T} \sum_{s=1}^{T} \sum_{t=1}^{T} \left(\alpha' \Gamma(s-r) \alpha\right) \\
&= \frac{1}{2\pi T} \iota_T' \begin{pmatrix} \gamma(0) & \cdots & \gamma(T-1) \\ \vdots & \ddots & \vdots \\ \gamma(T-1) & \cdots & \gamma(0) \end{pmatrix} \iota_T \\
&= \frac{1}{2\pi T} \mathbb{E}\left[\iota_T' \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix} \begin{pmatrix} y_1 & \cdots & y_T \end{pmatrix} \iota_T \right] = \frac{1}{2\pi T} \mathbb{E}\left[z_T^2\right] \geq 0,
\end{aligned}$$

where we define $z_T = \begin{pmatrix} y_1 & \cdots & y_T \end{pmatrix} \iota_T = \sum_{t=1}^{T} y_t$. This holds for any $\alpha \in \mathbb{R}^n$, so by definition $f_T(0)$ is positive semidefinite.

We showed that $f_T \to f$ pointwise on $(-\pi, \pi]$, that $\text{tr}(f_T)$ is non-negative valued for any $T \in N_+$, and that $f_T(0)$ is positive semidefinite for any $T \in N_+$. Therefore, $\text{tr}(f)$ is non-negative valued on $(-\pi, \pi]$ and $f(0)$ is positive semidefinite by the continuity of the trace operation and (ordered) eigenvalues of real symmetric matrices.

Q.E.D.

Heuristically, $f(w)$ represents the contribution that the sinusoid of frequency w makes to the variance, or power, of the stationary time series $\{Y_t\}_{t\in\mathbb{Z}}$; this is formalized below.

Theorem (Relationship between Spectral Density and Autocovariance Function) Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional weakly stationary time series with absolutely summable autocovariance function $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$. Letting $f : (-\pi,\pi] \to \mathbb{C}^{n\times n}$ be the spectral density of $\{Y_t\}_{t\in\mathbb{Z}}$,

$$\Gamma(\tau) = \int_{-\pi}^{\pi} \exp(i\tau w) f(w) dw$$

for any $\tau \in \mathbb{Z}$.

Proof) Choose any $\tau \in \mathbb{Z}$. Note that, for any $T \in N_+$,

$$\left\| \sum_{s=-T}^{T} \Gamma(s) \exp(i(\tau - s)w) \right\| \leq \sum_{s=-\infty}^{\infty} \|\Gamma(s)\| < +\infty$$

for any $w \in (-\pi,\pi]$ by absolute summability. Defining the function $g_T : (-\pi,\pi] \to \mathbb{C}^{n\times n}$ and $g : (-\pi,\pi] \to [0,+\infty)$ as

$$g_T(w) = \sum_{s=-T}^{T} \Gamma(s) \exp(i(\tau - s)w)$$

and

$$g(w) = \sum_{s=-\infty}^{\infty} \|\Gamma(s)\|$$

for any $w \in (-\pi,\pi]$,

$$\int_{-\pi}^{\pi} g(w) dw = 2\pi \cdot \left(\sum_{s=-\infty}^{\infty} \|\Gamma(s)\| \right) < +\infty.$$

Since $\{g_T\}_{T\in N_+}$ is a sequence of complex matrix valued functions with limit

$$\sum_{s=-\infty}^{\infty} \Gamma(s) \exp(i(\tau - s)w)$$

such that $\|g_T\| \leq g$ for any $T \in N_+$, and g is a non-negative function that is integrable on $(-\pi,\pi]$ with respect to the Lebesgue measure, so by the DCT,

$$\begin{aligned}
2\pi \cdot \int_{-\pi}^{\pi} \exp(i\tau w) f(w) dw &= \int_{-\pi}^{\pi} \left(\sum_{s=-\infty}^{\infty} \Gamma(s) \exp(i(\tau - s)w) \right) dw \\
&= \lim_{T\to\infty} \int_{-\pi}^{\pi} \left(\sum_{s=-T}^{T} \Gamma(s) \exp(i(\tau - s)w) \right) dw \\
&= \lim_{T\to\infty} \sum_{s=-T}^{T} \left(\int_{-\pi}^{\pi} \Gamma(s) \exp(i(\tau - s)w) dw \right).
\end{aligned}$$

For any $s \neq \tau$,

$$\begin{aligned}\int_{-\pi}^{\pi} \Gamma(s)\exp(i(\tau - s)w)dw &= \Gamma(s) \cdot \int_{-\pi}^{\pi} \exp(i(\tau - s)w)dw \\ &= \Gamma(s) \cdot \frac{1}{i(\tau - s)}\left(\exp(i(\tau - s)\pi) - \exp(-i(\tau - s)\pi)\right) = O,\end{aligned}$$

because $r \mapsto \exp(ir)$ is a function with period 2π, while

$$\int_{-\pi}^{\pi} \Gamma(s)\exp(i(\tau - s)w)dw = 2\pi \cdot \Gamma(s)$$

if $s = \tau$. Therefore,

$$\int_{-\pi}^{\pi} \exp(i\tau w)f(w)dw = \Gamma(\tau),$$

as desired.

Q.E.D.

This result tells us that, for $\tau = 0$,

$$\Gamma(0) = \int_{-\pi}^{\pi} f(w)dw,$$

so that the variance $\Gamma(0)$ of $\{Y_t\}_{t\in\mathbb{Z}}$ is the sum of the spectral densities for frequencies between $-\pi$ to π. As such, we can interpret $f(w)$ as the contribution the time series of frequency w makes to $\Gamma(0)$. The spectral representation theorem, which is the topic of the next section, shows that any time series can indeed be decomposed into the weighted sum of sinusoids of various frequencies.

3.2 The Spectral Representation of Time Series

Given an n-dimensional time series $\{Y_t\}_{t\in\mathbb{Z}}$, our goal in this section is to express each Y_t as a stochastic integral of the mapping $r \mapsto \exp(itr)$ with respect to some stochastic process on the index space $(-\pi,\pi]$ with orthogonal increments. To this end, we first rigorously define the stochastic integral of deterministic functions; because we only deal with non-random integrands, the development of this particular theory of stochastic integration is much simpler than that of the Ito integral, for instance.

3.2.1 Orthogonal Increment Processes

Let $\{Z_t\}_{-\pi\le t\le\pi}$ be a stochastic process with index set $[-\pi,\pi]$ such that each Z_t is a random vector taking values in $\mathbb{C}^n$. Suppose that $\{Z_t\}_{-\pi\le t\le\pi}$ is in $L^2_n(\mathcal{H},\mathbb{P})$, that is,

$$\|Z_t\|_{n,2} = \left(\mathbb{E}|Z_t|^2\right)^{\frac{1}{2}} < +\infty$$

for any $-\pi \le t \le \pi$. In this case, we say that $\{Z_t\}_{-\pi\le t\le\pi}$ is a mean-zero process with orthogonal increments if

$$\mathbb{E}\left[Z_t\right] = \mathbf{0} \quad \text{for any } -\pi \le t \le \pi, \text{ and}$$

$$\langle Z_t - Z_s, Z_u - Z_w\rangle_{n,2} = \mathbb{E}\left[(Z_t - Z_s)'(Z_u - Z_w)\right] = 0 \quad \text{for any } -\pi \le w \le u \le s \le t \le \pi.$$

In light of the mean-zero assumption, the second condition is equivalent to requiring that the trace of the covariance of $Z_t - Z_s$ and $Z_u - Z_w$ is 0 if the index intervals do not overlap.

$\{Z_t\}_{-\pi\le t\le\pi}$ is said in addition to be right-continuous in mean square (or simply just right-continuous) if, for any $-\pi \le t < \pi$,

$$\|Z_{t+\delta} - Z_t\|_{n,2} = \left(\mathbb{E}|Z_{t+\delta} - Z_t|^2\right)^{\frac{1}{2}} \to 0$$

as $\delta \downarrow 0$. From here on, we will be working with stochastic processes $\{Z_t\}_{-\pi\le t\le\pi}$ that are **square integrable** (that is, is a process in $L^2_n(\mathcal{H},\mathbb{P})$), with **mean zero** and **orthogonal increments**, that are also **right-continuous**. When we refer to orthogonal increment processes on $[-\pi,\pi]$, we will be referring to a process $\{Z_t\}_{-\pi\le t\le\pi}$ that possesses all the above properties.

Brownian motion is a famous example of a square integrable, right-continuous and mean zero stochastic process with orthogonal (in its case independent) increments. In fact, Brownian motion also has the added condition that the increments are stationary, that is, the distribution of the increments only depend on the difference between the time indices. Furthermore, in the case of Brownian motion, the entire paths themselves are continuous, that is, each realization $t \mapsto Z_t(\omega)$ of the process is a continuous function, which implies the kind of right-continuity in mean square discussed above.

The right continuity property of an orthogonal increment process $\{Z_t\}_{-\pi\le t\le\pi}$ actually allows us to construct a finite measure on $\mathbb{R}$ whose value on any interval $(s,t] \subset [-\pi,\pi]$ coincides with the variance $\|Z_t - Z_s\|_{n,2}^2 = \mathbb{E}|Z_t - Z_s|^2$. This result is formalized below:

Lemma Let $\{Z_t\}_{-\pi \le t \le \pi}$ be an n-dimensional orthogonal increment process. Then, there exists an increasing, right-continuous and bounded function $F : \mathbb{R} \to \mathbb{R}$, a σ-algebra $\mathcal{L}_F$ on $\mathbb{R}$ and a finite measure λ_F on $(\mathbb{R}, \mathcal{L}_F)$ such that:

i) $\mathcal{L}_F$ contains every Borel set on $\mathbb{R}$, that is, $\mathcal{B}(\mathbb{R}) \subset \mathcal{L}_F$.

ii) **(Completeness)** $(\mathbb{R}, \mathcal{L}_F, \lambda_F)$ is a complete measure space.

iii) **(Regularity)** λ_F is a regular Borel measure, that is,

$$\begin{aligned} \lambda_F(A) &= \inf\{\lambda_F(V) \mid A \subset V,\ V \text{ is open}\} \\ &= \sup\{\lambda_F(K) \mid K \subset A\ K \text{ is compact}\} \end{aligned}$$

for any $A \in \mathcal{L}_F$.

iv) **(Approximation Property)** For any $A \in \mathcal{L}_F$ and $\epsilon > 0$, there exists an open set V and a closed set K such taht $K \subset A \subset V$ and

$$\lambda_F(V \setminus K) < \epsilon.$$

v) For any half-open interval $(s,t] \subset [-\pi,\pi]$,

$$\lambda_F((s,t]) = F(t) - F(s) = \|Z_t - Z_s\|_{n,2}^2.$$

vi) The entire mass of λ_F is concentrated on $(-\pi,\pi]$, that is,

$$\lambda_F((-\pi,\pi]^c) = 0.$$

Proof) Define $F : \mathbb{R} \to \mathbb{R}$ as

$$F(t) = \begin{cases} \|Z_\pi - Z_{-\pi}\|_{n,2}^2 & \text{if } t > \pi \\ \|Z_t - Z_{-\pi}\|_{n,2}^2 & \text{if } -\pi < t \le \pi \\ 0 & \text{if } t \le -\pi \end{cases}$$

for any $t \in \mathbb{R}$. Then, F is bounded above by $\mathbb{E}|Z_\pi - Z_{-\pi}|^2 < +\infty$, which is finite by the square integrability of $\{Z_t\}_{-\pi \le t \le \pi}$. It is clearly increasing because, for any $-\pi \le s \le$

$t \leq \pi$,

$$\begin{aligned} F(t) &= \|Z_t - Z_{-\pi}\|_{n,2}^2 = \|(Z_t - Z_s) + (Z_s - Z_{-\pi})\|_{n,2}^2 \\ &= \|Z_t - Z_s\|_{n,2}^2 + \|Z_s - Z_{-\pi}\|_{n,2}^2 + 2 \cdot Re\left(\langle Z_t - Z_s, Z_s - Z_{-\pi}\rangle_{n,2}\right) \\ &= \|Z_t - Z_s\|_{n,2}^2 + \|Z_s - Z_{-\pi}\|_{n,2}^2 = \|Z_t - Z_s\|_{n,2}^2 + F(s), \end{aligned}$$

where the fourth equality follows from the orthogonality of the increments $Z_t - Z_s$ and $Z_s - Z_{-\pi}$. It follows that

$$F(t) - F(s) = \|Z_t - Z_s\|_{n,2}^2 \geq 0,$$

and as such that F is an increasing function on $\mathbb{R}$.

Finally, F is right-continuous because, for any $-\pi \leq t < \pi$,

$$\lim_{\delta \downarrow 0} (F(t+\delta) - F(t)) = \lim_{\delta \downarrow 0} \|Z_{t+\delta} - Z_t\|_{n,2}^2 = 0$$

by the right-continuty property of $\{Z_t\}_{-\pi \leq t \leq \pi}$.

Given an increasing and right continuous function $F : \mathbb{R} \to \mathbb{R}$, the theorem on the construction of the Lebesgue-Stieltjes measure on $\mathbb{R}$ found in the probability theory text shows us that there exist an σ-algebra $\mathcal{L}_F$ on $\mathbb{R}$ and a measure λ_F on $(\mathbb{R}, \mathcal{L}_F)$ such that properties i) to iv) are satisfied, and

$$\lambda_F((s,t]) = F(t) - F(s)$$

for any half-open interval $(s,t] \subset \mathbb{R}$. λ_F must be finite because

$$\lambda_F(\mathbb{R}) = \lim_{t \nearrow +\infty} F(t) \leq \mathbb{E}|Z_\pi - Z_{-\pi}|^2 < +\infty,$$

where the first equality follows from sequential continuity.

Choose any half-open interval $(s,t] \subset [-\pi, \pi]$. Then, by what we showed above,

$$\lambda_F((s,t]) = F(t) - F(s) = \|Z_t - Z_s\|_{n,2}^2.$$

Finally, note that the sequence $\{A_k\}_{k \in N_+}$ of sets defined as $A_k = (-\pi - k, -\pi]$ for any $k \in N_+$ is an increasing sequence of Borel sets on $\mathbb{R}$ such that $\bigcup_k A_k = (-\infty, \pi]$. For any $k \in N_+$,

$$\lambda_F(A_k) = F(-\pi) - F(-\pi - k) = 0,$$

so by sequential continuity

$$\lambda_F((-\infty, \pi]) = \lim_{k \to \infty} \lambda_F(A_k) = 0.$$

Likewise, let $\{B_k\}_{k\in N_+}$ be defined as $B_k = (\pi, \pi + k]$ for any $k \in N_+$. Then, $\{B_k\}_{k\in N_+}$ is an increasing sequence of Borel sets on $\mathbb{R}$ with union $(\pi, +\infty)$ such that

$$\lambda_F(B_k) = F(\pi + k) - F(\pi) = 0$$

for any $k \in N_+$. Sequential continuity again tells us that

$$\lambda_F((\pi, +\infty)) = 0,$$

so that

$$\lambda_F((-\pi, \pi]^c) = \lambda_F((-\infty, \pi]) + \lambda_F((\pi, +\infty)) = 0.$$

This completes the proof.

Q.E.D.

The function F and measure λ_F constructed above are called the distribution function and distribution associated with the orthogonal increment process $\{Z_t\}_{-\pi\leq t\leq\pi}$. Similarly, we refer to $\mathcal{L}_F$ as the σ-algebra associated with $\{Z_t\}_{-\pi\leq t\leq\pi}$. The term "distribution" is a bit of a misnomer, since the total mass of λ_F is not necessarily equal to 1, but we overlook this abuse in notation for the time being.

3.2.2 Stochastic Integration of Elementary Functions

Let $\{Z_t\}_{-\pi\leq t\leq\pi}$ be an n-dimensional orthogonal increment process with associated distribution function $F:\mathbb{R}\to\mathbb{R}$ and distribution μ_F on $(\mathbb{R},\mathcal{L}_F)$, where the σ-algebra $\mathcal{L}_F$ associated with this process contains $\mathcal{B}(\mathbb{R})$.

Let $\mathcal{D}$ be the collection of all complex functions on $\mathbb{R}$ taking values in $\mathbb{C}$ that can be written as

$$f=\sum_{i=0}^{k} r_i\cdot I_{(\lambda_i,\lambda_{i+1}]}$$

for some partition $-\pi=\lambda_0<\cdots<\lambda_{k+1}=\pi$ of $[-\pi,\pi]$ and $r_0,\cdots,r_k\in\mathbb{C}$. In other words, $\mathcal{D}$ collects complex functions on $\mathbb{R}$ taking finitely many values, which are equal to 0 at $-\pi$, and have support $[-\pi,\pi]$. These kind of functions are called elementary functions on $[-\pi,\pi]$, and they are Borel measurable and thus $\mathcal{L}_F$-measurable because they are left-continuous. Furthermore,

$$\begin{aligned}\int_{-\infty}^{\infty}|f|^2d\mu_F=\int_{-\pi}^{\pi}|f|^2d\mu_F&=\sum_{i=0}^{k}|r_i|^2\cdot\mu_F\left((\lambda_i,\lambda_{i+1}]\right)\\&=\sum_{i=0}^{k}|r_i|^2\left(F(\lambda_{i+1})-F(\lambda_i)\right)<+\infty\end{aligned}$$

by the definition of μ_F as the distribution associated with $\{Z_t\}_{-\pi\leq t\leq\pi}$, so it follows that $f\in L^2(\mathcal{L}_F,\mu_F)$. Thus, $\mathcal{D}$ is a subset of the inner product space $L^2(\mathcal{L}_F,\mu_F)$ over the complex field. Denote the inner product on $L^2(\mathcal{L}_F,\mu_F)$ by $\langle\cdot,\cdot\rangle_F$, and the L^2-norm induced by this inner product as $\|\cdot\|_F$.

Note that $\mathcal{D}$ is a linear subspace of $L^2(\mathcal{L}_F,\mu_F)$; the zero function is trivially included in $\mathcal{D}$, and for any $a\in\mathbb{C}$ and $f,g\in\mathcal{D}$, since $af+g$ takes finitely many values, equals 0 at $-\pi$, and has support $[-\pi,\pi]$, it must be the case that $af+g\in\mathcal{D}$.

We define the stochastic integral of the elementary function f with respect to $\{Z_t\}_{-\pi\leq t\leq\pi}$ as the n-dimensional complex valued random variable

$$I(f):=\sum_{i=0}^{k} r_i\cdot\left(Z_{\lambda_{i+1}}-Z_{\lambda_i}\right).$$

We also denote $I(f)$ by

$$\int_{-\pi}^{\pi}f(\lambda)dZ(\lambda).$$

$I(f)$ is well-defined for any two representations of an elementary function $f\in\mathcal{D}$ by the same line of reasoning used to show that the integral of a non-negative simple function is well-defined[1].

Since $I(f)$ is square integrable by the square integrability of $\{Z_t\}_{-\pi\leq t\leq\pi}$, we can view the stochastic integration operation I as a mapping from the vector space $\mathcal{D}$ over the complex field

[1]For the sake of completeness, we present the formal argument in this footnote. Let f,g be two elementary

into the vector space $L_n^2(\mathcal{H},\mathbb{P})$ over the complex field.

We can establish the following properties of the mapping $I:\mathcal{D}\to L_n^2(\mathcal{H},\mathbb{P})$:

Theorem (Properties of the Stochastic Integral for Elementary Functions)
Let $\{Z_t\}_{-\pi\le t\le\pi}$ be an n-dimensional orthogonal increment process with associated distribution function F, distribution μ_F, and σ-algebra $\mathcal{L}_F$. Let the operation $I:\mathcal{D}\to L_n^2(\mathcal{H},\mathbb{P})$ denote stochastic integration with respect to the orthogonal increment process $\{Z_t\}_{-\pi\le t\le\pi}$. Then, the following hold true:

i) **(Preservation of Inner Products)**
For any $f,g\in\mathcal{D}$,

$$\langle I(f),I(g)\rangle_{n,2}=\langle f,g\rangle_F.$$

In particular,

$$\mathbb{E}\left|\int_{-\pi}^{\pi} f(\lambda)dZ(\lambda)\right|^2=\int_{-\pi}^{\pi}|f(\lambda)|^2d\mu_F(\lambda).$$

ii) **(Linearity)**
$I:\mathcal{D}\to L_n^2(\mathcal{H},\mathbb{P})$ is a linear transformation: for any $a\in\mathbb{C}$ and $f,g\in\mathcal{D}$,

$$I(af+g)=a\cdot I(f)+I(g).$$

iii) For any $f\in\mathcal{D}$, $\mathbb{E}[I(f)]=\mathbf{0}$.

---

functions on $[-\pi,\pi]$ with representations

$$f=\sum_{i=0}^{l} r_i\cdot I_{(\lambda_i^f,\lambda_{i+1}^f]}\quad\text{and}\quad g=\sum_{i=0}^{m} s_i\cdot I_{(\lambda_i^g,\lambda_{i+1}^g]}$$

for partitions $-\pi=\lambda_0^f<\cdots<\lambda_{l+1}^f=\pi$ and $-\pi=\lambda_0^g<\cdots<\lambda_{m+1}^g=\pi$ of $[-\pi,\pi]$. Letting $-\pi=\lambda_0<\cdots<\lambda_{k+1}=\pi$ be the common refinement (for a definition, consult chapter 6 of PMA) of the two partitions above, we can express

$$f=\sum_{i=0}^{k}\tilde{r}_i\cdot I_{(\lambda_i,\lambda_{i+1}]}\quad\text{and}\quad g=\sum_{i=0}^{k}\tilde{s}_i\cdot I_{(\lambda_i,\lambda_{i+1}]}.$$

Suppose that $f\le g$. Then, $\tilde{r}_i\le\tilde{s}_i$ for any $0\le i\le k$, so it follows that

$$I(f)=\sum_{i=0}^{k}\tilde{r}_i\left(Z_{\lambda_{i+1}}-Z_{\lambda_i}\right)\le\sum_{i=0}^{k}\tilde{s}_i\left(Z_{\lambda_{i+1}}-Z_{\lambda_i}\right)=I(g)$$

everywhere on the sample space. Thus, if $f=g$, that is, if $\sum_{i=0}^{l} r_i\cdot I_{(\lambda_i^f,\lambda_{i+1}^f]}$ and $\sum_{i=0}^{m} s_i\cdot I_{(\lambda_i^g,\lambda_{i+1}^g]}$ are two different elementary function representations of the same function in $\mathcal{D}$, then

$$I(f)=I(g);$$

the stochastic integral of the function is invariant to the elementary function representation of functions in $\mathcal{D}$.

Proof) Let $f, g \in \mathcal{D}$. Then, there exists a partition $-\pi = \lambda_0 < \cdots < \lambda_{k+1} = \pi$ of $[-\pi, \pi]$ and complex numbers $r_0, \cdots, r_k \in \mathbb{C}$ and $s_0, \cdots, s_k \in \mathbb{C}$ such that[2]

$$f = \sum_{i=0}^{k} r_i \cdot I_{(\lambda_i, \lambda_{i+1}]} \quad \text{and} \quad g = \sum_{i=0}^{k} s_i \cdot I_{(\lambda_i, \lambda_{i+1}]}.$$

Note that

$$f\overline{g} = \sum_{i=0}^{k} r_i \overline{s_i} \cdot I_{(\lambda_i, \lambda_{i+1}]}.$$

By the definition of the stochastic integral of f and g with respect to the n-dimensional orthogonal increment process $\{Z_t\}_{-\pi \le t \le \pi}$,

$$I(f) = \sum_{i=0}^{k} r_i \cdot \left(Z_{\lambda_{i+1}} - Z_{\lambda_i}\right)$$
$$I(g) = \sum_{i=0}^{k} s_i \cdot \left(Z_{\lambda_{i+1}} - Z_{\lambda_i}\right).$$

It follows from the linearity properties of the inner product that

$$\langle I(f), I(g)\rangle_{n,2} = \sum_{i=0}^{k} \sum_{j=0}^{k} r_i \overline{s_j} \cdot \langle Z_{\lambda_{i+1}} - Z_{\lambda_i}, Z_{\lambda_{j+1}} - Z_{\lambda_j}\rangle_{n,2}.$$

Due to the orthogonality of the increments of $\{Z_t\}_{-\pi \le t \le \pi}$,

$$\langle Z_{\lambda_{i+1}} - Z_{\lambda_i}, Z_{\lambda_{j+1}} - Z_{\lambda_j}\rangle_{n,2} = 0$$

if $i \neq j$, so that

$$\langle I(f), I(g)\rangle_{n,2} = \sum_{i=0}^{k} r_i \overline{s_i} \cdot \left\| Z_{\lambda_{i+1}} - Z_{\lambda_i} \right\|_{n,2}^2.$$

From the preceding lemma, we know that

$$\left\| Z_{\lambda_{i+1}} - Z_{\lambda_i} \right\|_{n,2}^2 = F(\lambda_{i+1}) - F(\lambda_i) = \mu_F((\lambda_i, \lambda_{i+1}])$$

for each $0 \le i \le k$, so

$$\langle I(f), I(g)\rangle_{n,2} = \sum_{i=0}^{k} r_i \overline{s_i} \cdot \mu_F((\lambda_i, \lambda_{i+1}])$$
$$= \int_{-\pi}^{\pi} \left(\sum_{i=0}^{k} r_i \overline{s_i} \cdot I_{(\lambda_i, \lambda_{i+1}]} \right) d\mu_F$$

[2] For the existence of this common partition, consult the preceding footnote.

$$= \int_{-\pi}^{\pi} (f\overline{g})\, d\mu_F.$$

The last integral is precisely the inner product of f and g on the space $L^2(\mathcal{L}_F, \mu_F)$, so our first result has been proven.

Now choose $a \in \mathbb{C}$. Using the same representation for f, g as above, we can see that

$$af + g = \sum_{i=0}^{k} (ar_i + s_i) \cdot I_{(\lambda_i, \lambda_{i+1}]},$$

so we have

$$\begin{aligned} I(af+g) &= \sum_{i=0}^{k} (ar_i + s_i) \cdot (Z_{\lambda_{i+1}} - Z_{\lambda_i}) \\ &= a \cdot \sum_{i=0}^{k} r_i \cdot (Z_{\lambda_{i+1}} - Z_{\lambda_i}) + \sum_{i=0}^{k} s_i \cdot (Z_{\lambda_{i+1}} - Z_{\lambda_i}) \\ &= a \cdot I(f) + I(g). \end{aligned}$$

This establishes the linearity of the mapping I.

Finally, let $f \in \mathcal{D}$ have the representation

$$f = \sum_{i=0}^{k} r_i \cdot I_{(\lambda_i, \lambda_{i+1}]}$$

for some $r_0, \cdots, r_k \in \mathbb{C}$ and partition $-\pi = \lambda_0 < \cdots < \lambda_{k+1} = \pi$ of $[-\pi, \pi]$. Then,

$$I(f) = \sum_{i=0}^{k} r_i \left(Z_{\lambda_{i+1}} - Z_{\lambda_i} \right),$$

and since $\{Z_t\}_{-\pi \le t \le \pi}$ is a mean-zero process,

$$\mathbb{E}[I(f)] = \mathbf{0}$$

as well.

Q.E.D.

3.2.3 Stochastic Integration of Square Integrable Functions

Let $\overline{\mathcal{D}}$ be the closure of $\mathcal{D}$ with respect to the metric induced by the L^2-norm on $L^2(\mathcal{L}_F, \mu_F)$. Consider an arbitrary function $f \in \overline{\mathcal{D}}$. Then, there exists a sequence $\{f_k\}_{k \in N_+}$ of functions in $\mathcal{D}$ such that

$$\lim_{k \to \infty} \|f_k - f\|_F = 0.$$

Note now that the sequence $\{I(f_k)\}_{k \in N_+}$ of stochastic integrals in $L^2_n(\mathcal{H}, \mathbb{P})$ is Cauchy in the metric induced by $\langle \cdot, \cdot \rangle_{n,2}$; for any $k, m \in N_+$,

$$\begin{aligned}
\|I(f_k) - I(f_m)\|^2_{n,2} &= \|I(f_k - f_m)\|^2_{n,2} \\
&= \langle I(f_k - f_m), I(f_k - f_m) \rangle_{n,2} \\
&= \langle f_k - f_m, f_k - f_m \rangle_F = \|f_k - f_m\|^2_F
\end{aligned}$$

by the linearity of the mapping $I : \mathcal{D} \to L^2_n(\mathcal{H}, \mathbb{P})$ and the fact that it preserves inner prodcuts across $\mathcal{D} \subset L^2(\mathcal{L}_F, \mu_F)$ and $L^2_n(\mathcal{H}, \mathbb{P})$. Since $\{f_k\}_{k \in N_+}$ is convergent in L^2, it is also Cauchy in L^2, so that

$$\lim_{k,m \to \infty} \|I(f_k) - I(f_m)\|_{n,2} = \lim_{k,m \to \infty} \|f_k - f_m\|_F = 0.$$

Therefore, $\{I(f_k)\}_{k \in N_+}$ is Cauchy in L^2 as well, and since $L^2_n(\mathcal{H}, \mathbb{P})$ is a Hilbert space under the inner product $\langle \cdot, \cdot \rangle_{n,2}$, it follows that $\{I(f_k)\}_{k \in N_+}$ converges to some unique (up to almost sure equivalence) random vector in $L^2_n(\mathcal{H}, \mathbb{P})$ in mean square. We choose to denote this limit by $\tilde{I}(f)$, that is,

$$\tilde{I}(f) := \underset{k \to \infty}{\text{m.s.lim}}\, I(f_k).$$

We can verify that $\tilde{I}(f)$ is well-defined in the following two respects:

- **$\tilde{I}(f)$ is invariant to the choice of convergent sequence**
  Suppose that $\{g_k\}_{k \in N_+}$ and $\{f_k\}_{k \in N_+}$ are two sequences in $\mathcal{D}$ converging to $f \in \overline{\mathcal{D}}$ in L^2. Denote

  $$I_g = \underset{k \to \infty}{\text{m.s.lim}}\, I(g_k) \quad \text{and} \quad I_f = \underset{k \to \infty}{\text{m.s.lim}}\, I(f_k).$$

  Then, for any $k \in N_+$,

  $$\|I_g - I_f\|_{n,2} \le \|I_g - I(g_k)\|_{n,2} + \|I(g_k) - I(f_k)\|_{n,2} + \|I(f_k) - I_f\|_{n,2}.$$

  By assumption, the first and third terms go to 0 as $k \to \infty$. As for the second term,

  $$\begin{aligned}
  \|I(g_k) - I(f_k)\|_{n,2} &= \|I(g_k - f_k)\|_{n,2} = \|g_k - f_k\|_F \\
  &\le \|g_k - f\|_F + \|f - f_k\|_F,
  \end{aligned}$$

where the first equality follows from the linearity of the mapping $I:\mathcal{D}\to L_n^2(\mathcal{H},\mathbb{P})$, the second from the fact that it preserves inner prodcuts, and the last inequality is Minkowski's inequality. Both terms on the right hand side go to 0 as $k\to\infty$, so $\|I(g_k)-I(f_k)\|_{n,2}$ also goes to 0 as $k\to\infty$. It follows that

$$\|I_g - I_f\|_{n,2} = 0,$$

so that $I_g = I_f = \tilde{I}(f)$ almost surely. This shows us that $\tilde{I}(f)$ does not depend on the choice of sequence in $\mathcal{D}$ that converges to f.

- **$\tilde{I}(f)$ is $I(f)$ for elementary functions**
  Suppose $f\in\mathcal{D}$. Then, $\{f_k\}_{k\in N_+}$ defined as $f_k = f$ for any $k\in N_+$ is a sequence in $\mathcal{D}$ converging to f in L^2, so

$$\tilde{I}(f) = \underset{k\to\infty}{\text{m.s.lim}}\, I(f_k) = I(f).$$

  Therefore, $\tilde{I}(f)$ is precisely the stochastic integral of f with respect to $\{Z_t\}_{-\pi\leq t\pi}$ if f is an elementary function.

The two remarks above allow us to define, for any $f\in\overline{\mathcal{D}}$, the stochastic integral of f with respect to $\{Z_t\}_{-\pi\leq t\leq\pi}$ as

$$I(f) = \underset{k\to\infty}{\text{m.s.lim}}\, I(f_k),$$

where $\{f_k\}_{k\in N_+}$ is any sequence in $\mathcal{D}$ converging to f in L^2.

Now it remains to see which functions in $L^2(\mathcal{L}_F,\mu_F)$ are included in $\overline{\mathcal{D}}$, which is a subset of $L^2(\mathcal{L}_F,\mu_F)$. Fortunately, it turns out the space $L^2(\mathcal{L}_F,\mu_F)$ is exactly $\overline{\mathcal{D}}$; $\mathcal{D}$ is dense in $L^2(\mathcal{L}_F,\mu_F)$ in the L^2-sense thanks to the fact that μ_F is concentrated on $(-\pi,\pi]$, so we can define the stochastic integral for every function in $L^2(\mathcal{L}_F,\mu_F)$. The formal result is presented below:

Lemma **($\mathcal{D}$ is dense in the set of all Square Integrable Functions)**
Let $\{Z_t\}_{-\pi\leq t\leq\pi}$ be an n-dimensional orthogonal increment process with associated distribution function F, distribution μ_F, and σ-algebra $\mathcal{L}_F$. Denot by $\mathcal{D}$ the collection of elementary functions defined above. Then, for any $f\in L^2(\mathcal{L}_F,\mu_F)$ and $\epsilon>0$, there exists a $g\in\mathcal{D}$ such that

$$\|f-g\|_F < \epsilon.$$

Proof) Let $C_c(\mathbb{R},\mathbb{C})$ be the set of all continuous complex valued functions on $\mathbb{R}$ with compact support. Every function in $C_c(\mathbb{R},\mathbb{C})$ is continuous and continuous functions are $\mathcal{L}_F$-measurable, so $C_c(\mathbb{R},\mathbb{C})$ is a collection of $\mathcal{L}_F$-measurable functions. In addition, for any

$f \in C_c(\mathbb{R},\mathbb{C})$, letting $K = \overline{\{f \neq 0\}}$ be the compact support of f, f is bounded on K by the extreme value theorem. Letting $M > 0$ be this bound, we can now see that

$$\int_{\mathbb{R}} |f|^2 d\mu_F = \int_K |f|^2 d\mu_F \leq M^2 \cdot \mu_F(K) < +\infty,$$

where the final inequality follows because μ_F is a finite measure. Thus, $f \in L^2(\mathcal{L}_F, \mu_F)$, so that $C_c(\mathbb{R},\mathbb{C}) \subset L^2(\mathcal{L}_F, \mu_F)$.

The proof wil proceed in two steps. First, we will show that $\mathcal{D}$ is dense in $C_c(\mathbb{R},\mathbb{C})$ in the L^2-norm. Afterward, we show that $C_c(\mathbb{R},\mathbb{C})$ is dense in $L^2(\mathcal{L}_F, \mu_F)$ in the L^2-norm, at which point the proof will be complete.

Step 1: $\mathcal{D}$ is dense in $C_c(\mathbb{R},\mathbb{C})$

Choose any $f \in C_c(\mathbb{R},\mathbb{C})$, and let $\epsilon > 0$. Since f is a complex valued continuous function with compact support, it is uniformly continuous[3] on $\mathbb{R}$, and as such there exists a $\delta > 0$ such that

$$|f(x) - f(y)| < \frac{\epsilon}{1 + \sqrt{\mu_F((-\pi,\pi])}}$$

for any $x, y \in \mathbb{R}$ such that $|x - y| < \delta$. Choosing $k \in N_+$ so that $\frac{2\pi}{k} < \delta$, define

$$\lambda_j = -\pi + \frac{2\pi}{k+1} \cdot j$$

for any $0 \leq j \leq k+1$. Then, $-\pi = \lambda_0 < \lambda_1 < \cdots < \lambda_{k+1} = \pi$ is a partition of $[-\pi,\pi]$ such that $\lambda_{i+1} - \lambda_i = \frac{2\pi}{k+1} < \delta$ for any $0 \leq i \leq k$. Now define the function $g : \mathbb{R} \to \mathbb{C}$ as

$$g = \sum_{i=0}^{k} f(\lambda_i) \cdot I_{(\lambda_i, \lambda_{i+1}]}.$$

Clearly, $g \in \mathcal{D}$, and for any $x \in (-\pi,\pi]$, letting $\lambda_i < x \leq \lambda_{i+1}$ for some $0 \leq i \leq k$,

$$|f(x) - g(x)| = |f(x) - f(\lambda_i)| < \frac{\epsilon}{1 + \sqrt{\mu_F((-\pi,\pi])}}$$

because $|x - \lambda_i| \leq \lambda_{i+1} - \lambda_i < \delta$. Therefore,

$$|f - g|^2 \cdot I_{(-\pi,\pi]} < \left(\frac{\epsilon}{1 + \sqrt{\mu_F((-\pi,\pi])}} \right)^2$$

[3]For a formal proof, consult the theorem in chapter 4 of the measure theory text that constructs the Lebesgue measure using the Riemann integral and the Riesz representation theorem.

and since $\mu_F((-\pi,\pi]^c)=0$, we can see that

$$\|f-g\|_F=\left(\int_{(-\pi,\pi]}|f-g|^2 d\mu_F\right)^{\frac{1}{2}}\leq \epsilon\cdot\frac{\sqrt{\mu_F((-\pi,\pi])}}{1+\sqrt{\mu_F((-\pi,\pi])}}<\epsilon.$$

Such a $g\in\mathcal{D}$ exists for any $\epsilon>0$, so $\mathcal{D}$ is dense in $C_c(\mathbb{R},\mathbb{C})$.

Step 2: $C_c(\mathbb{R},\mathbb{C})$ is dense in $L^2(\mathcal{L}_F,\mu_F)$

Choose some $A\in\mathcal{L}_F$ and $\epsilon>0$. Focusing on the regulariy property of μ_F, since

$$\begin{aligned}\mu_F(A)&=\inf\{\mu_F(V)\mid A\subset V,\ V\text{ is open}\}\\&=\sup\{\mu_F(K)\mid K\subset A,\ K\text{ is compact}\}\end{aligned}$$

and $\mu_F(A)<+\infty$, by the definitions of the infimum and supremum there exist an open set V and a compact set K such that $K\subset A\subset V$ and

$$\mu_F(V)<\mu_F(A)+\frac{\epsilon}{2},\quad \mu_F(A)-\frac{\epsilon}{2}<\mu_F(K).$$

Putting these results together, we have

$$\mu_F(V\setminus K)=\mu_F(V)-\mu_F(K)<\mu_F(A)+\frac{\epsilon}{2}+\frac{\epsilon}{2}-\mu_F(A)=\epsilon.$$

By Urysohn's lemma for locally compact Hausdorff spaces, since $\mathbb{R}$ is locally compact Hausdorff, $K\subset V$, V is open, and K is compact, there exists a function $g\in C_c(\mathbb{R},\mathbb{C})$ such that

$$g(x)\in\begin{cases}\{1\} & \text{if } x\in K\\ [0,1] & \text{if } x\in V\setminus K\\ \{0\} & \text{if } x\notin V\end{cases}$$

for any $x\in\mathbb{R}$. In other words, $I_K\leq g\leq I_V$. If $x\in K$, then

$$|I_A(x)-g(x)|=0$$

because $x\in A$ and $g(x)=1$; if $x\notin V$, then

$$|I_A(x)-g(x)|=0$$

because $x\notin A$ and $g(x)=0$; if $x\in V\setminus K$, then

$$|I_A(x)-g(x)|\leq 2$$

since both I_A and g are bounded above by 1. It follows that

$$\|I_A - g\|_F^2 = \int_{\mathbb{R}} |I_A - g|^2 d\mu_F = \int_{V\setminus K} |I_A - g|^2 d\mu_F$$
$$\leq 4 \cdot \mu_F(V \setminus K) < 4\epsilon.$$

Thus, we can find a continuous compactly supported function g on $\mathbb{R}$ that approximates I_A arbitrarily closely in mean square.

Let f be a non-negative $\mathcal{L}_F$-measurable simple function; then, there exist $a_1, \cdots, a_k \in [0, +\infty)$ and disjoint $A_1, \cdots, A_k \in \mathcal{L}_F$ such that

$$f = \sum_{i=1}^{k} a_i \cdot I_{A_i}.$$

Choose any $\epsilon > 0$. For each A_i, we saw above that there exists a $g_i \in C_c(\mathbb{R}, \mathbb{C})$ such that

$$\|I_{A_i} - g_i\|_F \leq \frac{\epsilon}{k(a_i + 1)}.$$

Defining

$$g = \sum_{i=1}^{k} a_i g_i \in C_c(\mathbb{R}, \mathbb{C}),$$

we can now see that, by Minkowski's inequality,

$$\|f - g\|_F \leq \sum_{i=1}^{k} a_i \cdot \|I_{A_i} - g\|_F < \epsilon \cdot \frac{1}{k}\left(\sum_{i=1}^{k} \frac{a_i}{a_i + 1}\right) < \epsilon.$$

It follows that any non-negative $\mathcal{L}_F$-measurable simple function can be arbitrarily closely approximated in mean square by a continuous compactly supported funciton.

Now let f be an arbitrary non-negative function in $L^2(\mathcal{L}_F, \mu_F)$. Then, there exists a sequence $\{f_k\}_{k \in N_+}$ of simple non-negative $\mathcal{L}_F$-measurable function that increases pointwise to f. $\{|f - f_k|^2\}_{k \in N_+}$ is a sequence of $\mathcal{L}_F$-measurable functions such that $|f - f_k|^2 \leq 4|f|^2$, where $4|f|^2$ is μ_F-integrable due to the assumption that $f \in L^2(\mathcal{L}_F, \mu_F)$, and which converges pointwise to 0. Therefore, by the DCT,

$$\lim_{k \to \infty} \int_{\mathbb{R}} |f - f_k|^2 d\mu_F = 0,$$

or equiavlently, $\|f - f_k\|_F \to 0$ as $k \to \infty$. For any $\epsilon > 0$, there thus exists a $k \in N_+$ such that

$$\|f_k - f\|_F < \frac{\epsilon}{2},$$

and since this f_k is a simple non-negative $\mathcal{L}_F$-measurable function, by the preceding result there exists a $g \in C_c(\mathbb{R},\mathbb{C})$ such that

$$\|f_k - g\|_F < \frac{\epsilon}{2}.$$

Putting these two results together, we can conclude that $\|f-g\|_F < \epsilon$; a non-negative square integrable function f can be arbitrarily closely approximated in mean square by a continuous compactly supported function.

Moving onto real-valued functions, suppose $f \in L^2(\mathcal{L}_F, \mu_F)$ is real valued. Then, its positive and negative parts f^+ and f^- are also μ_F-square integrable $\mathcal{L}_F$-measurable functions. For any $\epsilon > 0$, by the preceding result, there exist $g_1, g_2 \in C_c(\mathbb{R},\mathbb{C})$ such that

$$\left\|f^+ - g_1\right\|_F, \|f^- - g_2\|_F < \frac{\epsilon}{2}.$$

It follows that, defining $g = g_1 - g_2 \in C_c(\mathbb{R},\mathbb{C})$,

$$\|f-g\|_F = \left\|(f^+ - g_1) - (f^- - g_2)\right\|_F \leq \left\|f^+ - g_1\right\|_F + \|f^- - g_2\|_F < \epsilon.$$

Finally, let $f \in L^2(\mathcal{L}_F, \mu_F)$ in general. Then, its real and imaginary parts $Re(f)$ and $Im(f)$ are real-valued functions in $L^2(\mathcal{L}_F, \mu_F)$, and a process similar to the proof for real valued functions in $L^2(\mathcal{L}_F, \mu_F)$ shows that, for any $\epsilon > 0$, there exists a $g \in C_c(\mathbb{R},\mathbb{C})$ such that

$$\|f-g\|_F < \epsilon.$$

Q.E.D.

With the above lemma, we can formally claim that, for any $f \in L^2(\mathcal{L}_F, \mu_F)$, there exists an almost surely unique stochastic integral

$$I(f) = \int_{-\pi}^{\pi} f(\lambda) dZ(\lambda)$$

of f with respect to the orthogonal increment process $\{Z_t\}_{-\pi \leq t \leq \pi}$, defined as the L^2-limit of the sequence $\{I(f_k)\}_{k \in N_+} \subset L^2_n(\mathcal{H}, \mathbb{P})$, where $\{f_k\}_{k \in N_+}$ is a sequence of elementary functions converging in mean square to f. The domain of the mapping I as now been extended from $\mathcal{D}$ to $L^2(\mathcal{L}_F, \mu_F)$. The following are some properties of the mapping $I : L^2(\mathcal{L}_F, \mu_F) \to L^2_n(\mathcal{H}, \mathbb{P})$:

Theorem (Properties of the Stochastic Integral)
Let $\{Z_t\}_{-\pi\leq t\leq\pi}$ be an n-dimensional orthogonal increment process with associated distribution function F, distribution μ_F, and σ-algebra $\mathcal{L}_F$. Let the operation $I : L^2(\mathcal{L}_F,\mu_F) \to L_n^2(\mathcal{H},\mathbb{P})$ denote stochastic integation with respect to the orthogonal increment process $\{Z_t\}_{-\pi\leq t\leq\pi}$. Then, the following hold true:

i) **(Preservation of Inner Products)**
For any $f,g \in L^2(\mathcal{L}_F,\mu_F)$,

$$\langle I(f), I(g)\rangle_{n,2} = \langle f,g\rangle_F.$$

In particular,

$$\mathbb{E}\left|\int_{-\pi}^{\pi} f(\lambda)dZ(\lambda)\right|^2 = \int_{-\pi}^{\pi} |f(\lambda)|^2 d\mu_F(\lambda).$$

ii) **(Linearity)**
$I : L^2(\mathcal{L}_F,\mu_F) \to L_n^2(\mathcal{H},\mathbb{P})$ is a linear transformation: for any $a \in \mathbb{C}$ and $f,g \in L^2(\mathcal{L}_F,\mu_F)$,

$$I(af+g) = a\cdot I(f) + I(g).$$

iii) For any $f \in L^2(\mathcal{L}_F,\mu_F)$, $\mathbb{E}[I(f)] = \mathbf{0}$.

Proof) Let $f,g \in L^2(\mathcal{L}_F,\mu_F)$. Then, because $f,g\in\overline{\mathcal{D}}$, there exists sequences $\{f_k\}_{k\in N_+}$ and $\{g_k\}_{k\in N_+}$ of elementary functions that converge in mean square to f and g. For each $k\in N_+$, the properties of stochastic integration for elementary functions tell us that

$$\langle I(f_k), I(g_k)\rangle_{n,2} = \langle f_k, g_k\rangle_F.$$

By the definition of stochastic integrals,

$$\|I(f_k)-I(f)\|_{n,2} \to 0, \quad \|I(g_k)-I(g)\|_{n,2} \to 0$$

as $k\to\infty$. Note now that

$$\begin{aligned}\left|\langle I(f_k), I(g_k)\rangle_{n,2} - \langle I(f), I(g)\rangle_{n,2}\right| &= \left|\langle I(f_k)-I(f), I(g_k)-I(g)\rangle_{n,2} + \langle I(f_k)-I(f), I(g)\rangle_{n,2} + \langle I(f), I(g_k)-I(g)\rangle_{n,2}\right| \\ &\leq \|I(f_k)-I(f)\|_{n,2}\cdot\|I(g_k)-I(g)\|_{n,2} \\ &\quad + \|I(g)\|_{n,2}\cdot\|I(f_k)-I(f)\|_{n,2} + \|I(f)\|_{n,2}\cdot\|I(g_k)-I(g)\|_{n,2},\end{aligned}$$

where the last inequality follows from the Cauchy-Schwarz inequality. Taking $k\to\infty$ on both sides now yields

$$\lim_{k\to\infty}\langle I(f_k), I(g_k)\rangle_{n,2} = \langle I(f), I(g)\rangle_{n,2}.$$

Likewise,

$$\lim_{k\to\infty}\langle f_k, g_k\rangle_F = \langle f, g\rangle_F,$$

so we can see that

$$\langle I(f), I(g)\rangle_{n,2} = \langle f, g\rangle_F.$$

Now choose some $a \in \mathbb{C}$ and $f, g \in L^2(\mathcal{L}_F, \mu_F)$. Then, as above, there exists sequences $\{f_k\}_{k\in N_+}$ and $\{g_k\}_{k\in N_+}$ of elementary functions that converge in mean square to f and g, and by definition

$$\|I(f_k) - I(f)\|_{n,2} \to 0, \quad \|I(g_k) - I(g)\|_{n,2} \to 0.$$

By the linearity of stochastic integration for elementary functions,

$$I(af_k + g_k) = a \cdot I(f_k) + I(g_k)$$

for any $k \in N_+$. Since $\{af_k + g_k\}_{k\in N_+}$ is a sequence of elementary functions converging in mean square to $af + g \in L^2(\mathcal{L}_F, \mu_F)$, by definition

$$I(af_k + g_k) \xrightarrow{L^2} I(af + g).$$

Similarly, because $I(f_k) \xrightarrow{L^2} I(f)$ and $I(g_k) \xrightarrow{L^2} I(g)$, we can conclude that

$$a \cdot I(f_k) + I(g_k) \xrightarrow{L^2} a \cdot I(f) + I(g)$$

as well. By the almost sure uniqueness of L^2-limits, we can conclude that

$$I(af + g) = a \cdot I(f) + i(g)$$

almost surely.

Finally, let $f \in L^2(\mathcal{L}_F, \mu_F)$ and $\{f_k\}_{k\in N_+}$ a sequence of elementary functions converginig to f in mean square. Recall that $\mathbb{E}[I(f_k)] = \mathbf{0}$ for any $k \in N_+$. Thus,

$$|\mathbb{E}[I(f)]| = |\mathbb{E}[I(f) - I(f_k)]| \le \mathbb{E}|I(f) - I(f_k)| \le \left(\mathbb{E}|I(f) - I(f_k)|^2\right)^{\frac{1}{2}} = \|I(f) - I(f_k)\|_{n,2}$$

for any $k \in N_+$, where the second inequality is Hölder's. Taking $k \to \infty$ on both sides now tells us that $|\mathbb{E}[I(f)]| = 0$, that is, $\mathbb{E}[I(f)] = \mathbf{0}$.

Q.E.D.

3.2.4 Trigonometric Polynomials

Before moving onto the spectral representation theorem, we take a brief detour to prove that the space of trigonometric polynomials is dense in the space of all functions on $\mathbb{R}$ that are square integrable with respect to a finite measure on $\mathbb{R}$ that assigns all of its mass to $(-\pi, \pi]$. Note that this is precisely the type of measure that is naturally associated with orthogonal increment processes on $[-\pi, \pi]$, hence the usefulness of the result we are about to show.

We first review the Stone-Weierstrass theorem, which plays a central role in the exposition this section. A vector space V over a field F is said to be an algebra over the field F if it is equipped with a product operation $\times : V^2 \to V$ satisfying the following properties:

- **The Distributive Law**
  For any $x, y, z \in V$,

$$(x+y) \times z = x \times z + y \times z$$
$$z \times (x+y) = z \times x + z \times y$$

- **Compatibility with Scalars**
  For any $a, b \in F$ and $x, y \in V$,

$$(a \cdot x) \times (b \cdot y) = (ab) \cdot (x \times y),$$

  where $\cdot$ is the scalar multiplication operation.

Given any set E and a field F, the collection $\mathcal{F}$ of all functions $f : E \to F$ is an algebra over the field F under the product $\times : \mathcal{F}^2 \to \mathcal{F}$ defined as

$$(f \times g)(x) = f(x)g(x)$$

for any $x \in E$. In particular, if (E, τ) is a topological space and $F = \mathbb{R}$ or $\mathbb{C}$, then the space $C_b(E, F)$ of all bounded continuous functions from E into F is a subalgebra of $\mathcal{F}$ equipped with the same product operation; this can be easily seen since the product of bounded continuous functions is also bounded and continuous, and scalar products obey the distributivity and associativity properties. Recall that the space of bounded continuous functions can itself be considered a metric space under the supremum metric $d_{\mathcal{C}}$, which is the metric induced by the supremium norm $\|\cdot\|_{\mathcal{C}} : C_b(E, F) \to \mathbb{R}_+$ defined as

$$\|f\|_{\mathcal{C}} = \sup_{x \in E} |f(x)|.$$

for any $f \in C_b(E,F)$.

Let E be a set, F a field with additive identity 0_F, $\mathcal{F}$ the algebra of all functions from E to F, and $\mathcal{A}$ a subalgebra of $\mathcal{F}$. We say that $\mathcal{A}$:

- **Separates points on** E
  If, for any $x_1, x_2 \in E$ such that $x_1 \neq x_2$, there exists an $f \in \mathcal{A}$ such that $f(x_1) \neq f(x_2)$.

- **Vanishes at no point in** E
  If, for any $x \in E$, there exists an $f \in \mathcal{A}$ such that $f(x) \neq 0_F$.

- **Is Self-Adjoint**
  If $F = \mathbb{C}$ and, for any $f \in \mathcal{A}$, its conjugate $\overline{f}$ is also contained in $\mathcal{A}$.

The Stone-Weierstrass theorem can now be stated as follows:

Theorem (The Stone-Weierstrass Theorem)
Let (E,τ) be a topological space, $F = \mathbb{R}$ or $\mathbb{C}$, $C_b(E,F)$ the set of all continuous and bounded functions from E to F, and $d_{\mathcal{C}}$ the supremum metric on $C_b(E,F)$. Let $\mathcal{A}$ be a subalgebra of $C_b(E,F)$ over the field F. Then, the following hold true:

i) If $\mathcal{A}$ separates points on E, then (E,τ) is a Hausdorff space.

ii) **(Real Version of the Stone-Weierstrass Theorem)**
Suppose (E,τ) is a compact space and that $F = \mathbb{R}$. If $\mathcal{A}$ separates points on E and vanishes at no point in E, then $\mathcal{A}$ is uniformly dense in $C(E,F)$, that is, $C(E,F)$ is the closure of $\mathcal{A}$ under the metric $d_{\mathcal{C}}$.

iii) **(Complex Version of the Stone-Weierstrass Theorem)**
Suppose (E,τ) is a compact space and that $F = \mathbb{C}$. If $\mathcal{A}$ separates points on E, vanishes at no point in E, and is self-adjoint, then $\mathcal{A}$ is uniformly dense in $C(E,F)$, that is, $C(E,F)$ is the closure of $\mathcal{A}$ under the metric $d_{\mathcal{C}}$.

Proof) Consult chapter 6 of the probability theory text.

Q.E.D.

The complex version of the Stone-Weierstrass theorem can now be used to show the first part of our desired result. Let $\mathbb{T}$ be the unit circle on $\mathbb{C}$, that is,

$$\mathbb{T} = \{x \in \mathbb{C} \mid |x| = 1\}.$$

Clearly, $\mathbb{T}$ is compact under the euclidean metric on $\mathbb{C}$. Furthermore, functions on $\mathbb{T}$ are equivalent to periodic functions on $\mathbb{R}$ with period 2π in the sense below:

Lemma (Characterization of Periodic Functions)
Let $f : \mathbb{T} \to \mathbb{C}$ be a c ontinuous function on $\mathbb{T}$. Then, there exists a continuous periodic function $g : \mathbb{R} \to \mathbb{C}$ with period 2π such that

$$f(\exp(ix)) = g(x)$$

for any $x \in \mathbb{R}$.

Conversely, suppose $f : \mathbb{R} \to \mathbb{C}$ is a continuous periodic function with period 2π. Then, there exists a continuous function $g : \mathbb{T} \to \mathbb{C}$ such that

$$g(\exp(ix)) = f(x)$$

for any $x \in \mathbb{R}$.

Proof) First choose some continuous $f : \mathbb{T} \to \mathbb{C}$, and define the function $g : \mathbb{R} \to \mathbb{C}$ as

$$g(x) = f(\exp(ix))$$

for any $x \in \mathbb{R}$. g is well-defined because $\exp(ix) \in \mathbb{T}$ for any $x \in \mathbb{R}$, and it has period 2π because, for any $x \in \mathbb{R}$,

$$g(x + 2\pi) = f(\exp(i(x + 2\pi))) = f(\exp(ix)) = g(x)$$

by the periodicity of the mapping $x \mapsto \exp(ix)$. Furthermore, it is continuous because it is the composition of two continuous functions.

Conversely, let $f : \mathbb{R} \to \mathbb{C}$ be a continuous function with period 2π. Let $\theta : \mathbb{R} \to \mathbb{T}$ be defined as $\theta(x) = \exp(ix)$ for any $x \in \mathbb{R}$. By Euler's formula, the real and imaginary parts of θ are continuous, so θ is also itself a continuous function. Define $g : \mathbb{T} \to \mathbb{C}$ as

$$g(z) \in f\left(\theta^{-1}(\{z\})\right)$$

for any $z \in \mathbb{T}$, where $\theta^{-1}(\{z\})$ is the inverse image of the singleton $\{z\}$, and $f\left(\theta^{-1}(\{z\})\right)$ is the image of the set $\theta^{-1}(\{z\})$. g is a well-defined function because the periodicity of f ensures that $f\left(\theta^{-1}(\{z\})\right)$ contains one and only one element. It is also easy to see that the relationship

$$g(\exp(ix)) = g(\theta(x)) = f(x)$$

holds for any $x \in \mathbb{R}$.

It remains to show that g is continuous on $\mathbb{T}$. To this end, let V be a subset of $\mathbb{C}$; we initially note that

$$g^{-1}(V) = \theta(f^{-1}(V)).$$

To rigorously establish this result, first choose some $z \in g^{-1}(V)$. Then, since $g(z) \in V$ and $f(\theta^{-1}(\{z\})) = \{g(z)\}$, it follows that

$$\theta^{-1}(\{z\}) \subset f^{-1}(V).$$

By definition, there exists a $x \in \theta^{-1}(\{z\})$ such that $\theta(x) = z$, and since this x is contained in $f^{-1}(V)$, we have

$$z \in \theta(f^{-1}(V)),$$

which implies that $g^{-1}(V) \subset \theta(f^{-1}(V))$.

Conversely, choose some $z \in \theta(f^{-1}(V))$. Then, there exists some $x \in f^{-1}(V)$ such that $z = \theta(x)$, which shows us that $x \in \theta^{-1}(\{z\})$. By definition, $g(z) = f(x) \in V$, so it follows that $z \in g^{-1}(V)$. This establishes the reverse inclusion, and we are able to conclude that

$$g^{-1}(V) = \theta(f^{-1}(V)).$$

To establish the continuity of g on $\mathbb{T}$, it suffices then to show that

$$\theta(f^{-1}(V))$$

is an open subset of $\mathbb{T}$ for any open subset V of $\mathbb{C}$. Since the set of all open intervals on $\mathbb{R}$ forms a base generating the standard topology on $\mathbb{R}$, we can express U as the arbitrary union of open intervals (a,b). The proof will therefore be over if we can show that the image

$$\theta((a,b))$$

is an open subset of $\mathbb{T}$ for any open interval (a,b) on $\mathbb{R}$.

Choose any such interval (a,b); note that $\theta((a,b))$ forms an arc on the unit circle. Let

$z \in \theta((a,b))$. Then, there exists an $x \in (a,b)$ such that $z = \theta(x)$, and using this we can choose a $0 < \delta < \pi$ such that

$$(x-\delta, x+\delta) \subset (a,b).$$

We will now search for an $\epsilon > 0$ such that

$$B(z,\epsilon) \cap \mathbb{T} \subset \theta((x-\delta, x+\delta)).$$

For any $y \in \mathbb{R}$, note that

$$\begin{aligned}
|\theta(x)-\theta(y)|^2 &= |\exp(ix)-\exp(iy)|^2 \\
&= (\cos(x)-\cos(y))^2 + (\sin(x)-\sin(y))^2 \\
&= 2 - 2(\cos(x)\cos(y)+\sin(x)\sin(y)) \\
&= 2(1-\cos(x-y)) \\
&= 2\left(1-\left(1-2\sin^2\left(\frac{x-y}{2}\right)\right)\right) \\
&= 4\sin^2\left(\frac{x-y}{2}\right).
\end{aligned}$$

Therefore, for any $0 < \epsilon < 4$, if $w \in B(z,\epsilon) \cap \mathbb{T}$, then there exists some $y \in \mathbb{R}$ such that $w = \theta(y)$, where the y is chosen so that $|x-y| \leq \pi$. This y then satisfies

$$4\sin^2\left(\frac{x-y}{2}\right) = |\theta(x)-\theta(y)|^2 < \epsilon^2,$$

which implies that

$$|x-y| < \arcsin\left(\frac{\epsilon}{2}\right),$$

where we can employ the inverse sine function because $-\frac{\pi}{2} \leq \frac{x-y}{2} \leq \frac{\pi}{2}$. It follows that, if we set

$$\epsilon = 2\sin\left(\frac{\delta}{2}\right) > 0,$$

then since $0 < \frac{\delta}{2} < \frac{\pi}{2}$, we have

$$|x-y| < \delta$$

and therefore

$$w = \theta(y) \in \theta((x-\delta, x+\delta)).$$

We have just shown that

$$B\left[z, 2\sin\left(\frac{\delta}{2}\right)\right] \cap \mathbb{T} \subset \theta((x-\delta, x+\delta)) \subset \theta((a,b)).$$

Therefore, any point z in $\theta((a,b))$ has a neighborhood open in $\mathbb{T}$ that is once again contained in $\theta((a,b))$. This shows us that $\theta((a,b))$ is open in $\mathbb{T}$, and we can now conclude that g is a continuous function.

Q.E.D.

The main functions of interest in this section are trigonometric polynomials. A trigonometric polynomial is a function $P : \mathbb{R} \to \mathbb{C}$ of the form

$$P(x) = a_0 + \sum_{t=1}^{k} (a_t \cdot \cos(tx) + b_t \cdot \sin(tx))$$

for any $x \in \mathbb{R}$, where $a_0, \cdots, a_k, b_1, \cdots, b_k \in \mathbb{C}$. By Euler's formula, we can formulate the cosine and sine functions in terms of the complex exponential as follows:

$$\cos(tx) = \frac{1}{2}(\exp(itx) + \exp(-itx))$$
$$\sin(tx) = \frac{1}{2i}(\exp(itx) - \exp(-itx))$$

for any $t \in \mathbb{Z}$ and $x \in \mathbb{R}$. Therefore, the trigonometric polynomial above can be written in terms of the complex exponential as

$$\begin{aligned} P(x) &= a_0 + \sum_{t=1}^{k} \left[\frac{a_t}{2} \cdot (\exp(itx) + \exp(-itx)) + \frac{b_t}{2i} \cdot (\exp(itx) - \exp(-itx))\right] \\ &= a_0 + \sum_{t=1}^{k} \frac{a_t - ib_t}{2} \cdot \exp(itx) + \sum_{t=1}^{k} \frac{a_t + ib_t}{2} \cdot \exp(-itx) \\ &= \sum_{t=-k}^{k} c_t \cdot \exp(itx) \end{aligned}$$

for any $x \in \mathbb{R}$, where

$$c_t = \begin{cases} \frac{a_t - ib_t}{2} & \text{if } 1 \leq t \leq k \\ a_0 & \text{if } t = 0 \\ \frac{a_t + ib_t}{2} & \text{if } -k \leq t \leq -1 \end{cases}.$$

Euler's formula also tells us that we can write any finite partial linear combination of complex exponentials as a trigonometric polynomial, so from here on we write trigonometric polynomials as a doubly finite linear combination of complex exponentials, which is more convenient for our purposes.

The preceding characterization of 2π-periodic continuous functions, combined with the Stone-Weierstrass theorem, produces the following result central to the ubiquity of trigonometric polynomials in mathematics:

Lemma (Density of Trigonometric Polynomials for 2π-Periodic Functions)
Let $\mathcal{T}$ be the space of all trigonometric polynomials on $\mathbb{R}$. Then, $\mathcal{T}$ is uniformly dense in the space $C_{2\pi}(\mathbb{R}, \mathbb{C})$ of all complex periodic and continuous functions on $\mathbb{R}$ with period 2π.

Proof) Define $\mathcal{A}$ as the set of all polynomials from the unit circle $\mathbb{T}$ to the complex field $\mathbb{C}$. $\mathcal{A}$ is clearly an algebra over the complex field, since the linear combination of polynomials, as well as the product of polynomials are still polynomials. $\mathcal{A}$ also separates points on $\mathbb{T}$; for any distinct $z_1, z_2 \in \mathbb{T}$, the simple linear function $P \in \mathcal{A}$ defined as $P(x) = x$ for any $x \in \mathbb{T}$ separates z_1 and z_2. For any $z \in \mathbb{T}$, the polynomial $P \in \mathcal{A}$ defined as $P(x) = 1$ for any $x \in \mathbb{T}$ satisfies $P(z) = 1 \neq 0$, and therefore $\mathcal{A}$ vanishes at no point in $\mathbb{T}$. Finally, for any polynomial $P \in \mathcal{A}$ defined as

$$P(z) = \sum_{t=0}^{k} a_t \cdot z^t$$

for any $z \in \mathbb{T}$, its conjugate $\overline{P}$ is defined as

$$\overline{P}(z) = \sum_{t=0}^{k} \overline{a_t} \cdot z^t$$

for any $z \in \mathbb{T}$ and is thus also a complex-valued polynomial on $\mathbb{T}$, meaning that it is contained in $\mathcal{A}$. In other words, $\mathcal{A}$ is also a self-adjoint algebra.

The properties of $\mathcal{A}$ shown above, together with the compactness of $\mathbb{T}$, allow us to use the complex version of the Stone-Weierstrass theorem; we can conclude that $\mathcal{A}$ is uniformly dense in the space $C(\mathbb{T}, \mathbb{C})$ of all complex continuous functions on $\mathbb{T}$.

Now choose any function $f \in C_{2\pi}(\mathbb{R}, \mathbb{C})$. The preceding lemma tells us that there exists a $g \in C(\mathbb{T}, \mathbb{C})$ such that

$$f(x) = g(\exp(ix))$$

for any $x \in \mathbb{R}$. Since $\mathcal{A}$ is uniformly dense in $\mathbb{T}$, there exists a sequence $\{P_n\}_{n \in N_+}$ of polynomials on $\mathbb{T}$, where each P_n is defined as

$$P_n(z) = \sum_{t=0}^{k_n} a_t^{(n)} \cdot z^t$$

for any $z \in \mathbb{T}$, such that

$$\lim_{n\to\infty} \sup_{z\in\mathbb{T}} |P_n(z) - g(z)| = 0.$$

Define the sequence $\{T_n\}_{n\in N_+}$ as

$$T_n(x) = P_n(\exp(ix)) = \sum_{t=0}^{k_n} a_t^{(n)} \cdot \exp(itx)$$

for any $x \in \mathbb{R}$ and $n \in N_+$; $\{T_n\}_{n\in N_+}$ is, by definition, a sequence of trigonometric polynomials on $\mathbb{R}$. Since $g(\exp(ix)) = f(x)$ for any $x \in \mathbb{R}$, it follows that, for any $n \in N_+$,

$$\sup_{z\in\mathbb{T}} |P_n(z) - g(z)| = \sup_{x\in\mathbb{R}} |T_n(x) - f(x)|.$$

Therefore,

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |T_n(x) - f(x)|,$$

and since this holds for any continuous 2π-periodic function $f : \mathbb{R} \to \mathbb{C}$, the space $\mathcal{T}$ of trigonometric polynomials on $\mathbb{R}$ is uniformly dense in $C_{2\pi}(\mathbb{R}, \mathbb{C})$.

Q.E.D.

We can now show that trigonometric polynomials are dense in certain kinds of L^2-spaces.

Lemma (Density of Trigonometric Polynomials for Square Integrable Functions) Let $\mathcal{T}$ be the space of all trigonometric polynomials on $\mathbb{R}$. Let $\mathcal{E}$ be a σ-algebra on $\mathbb{R}$ that contains the Borel σ-algebra $\mathcal{B}(\mathbb{R})$, and μ a finite measure on $(\mathbb{R}, \mathcal{E})$ concentrated on $(-\pi, \pi]$, that is,

$$\mu((-\pi, \pi]^c) = 0.$$

Then, $\mathcal{T}$ is dense in $L^2(\mathcal{E}, \mu)$ in the mean-square sense, that is, $L^2(\mathcal{E}, \mu)$ is the closure of $\mathcal{T}$ with respect to the L^2-norm $\|\cdot\|_2$ on $L^2(\mathcal{E}, \mu)$.

Proof) We first show that $\mathcal{T}$ is dense in the space of all complex continuous functions f on $\mathbb{R}$ such that $f(-\pi) = f(\pi)$, in the mean-square sense. Let $f : \mathbb{R} \to \mathbb{C}$ be a continuous function such that $f(-\pi) = f(\pi)$, and choose any $\epsilon > 0$. Define $\tilde{f}$ as the restriction of f to $(-\pi, \pi]$, and construct $g : \mathbb{R} \to \mathbb{C}$ by connecting an infinite number of $\tilde{f}$ side by side. Then, g is a continuous periodic function on $\mathbb{R}$ with period 2π. By the previous result, there exists a trigonometric polynomial $P \in \mathcal{T}$ such that

$$\sup_{x\in\mathbb{R}} |P(x) - g(x)| < \frac{\epsilon}{1+\sqrt{\mu((-\pi,\pi])}}.$$

By implication, we have

$$\begin{aligned} \sup_{x\in(-\pi,\pi]} |P(x) - f(x)| &= \sup_{x\in(-\pi,\pi]} |P(x) - g(x)| \\ &\le \sup_{x\in\mathbb{R}} |P(x) - g(x)| < \frac{\epsilon}{1+\sqrt{\mu((-\pi,\pi])}}, \end{aligned}$$

since f and g agree on $(-\pi,\pi]$, and since the measure μ is concentrated on $(-\pi,\pi]$,

$$\begin{aligned} \|f-P\|_2 &= \left(\int_{\mathbb{R}} |f(x) - P(x)|^2 dx\right)^{\frac{1}{2}} \\ &= \left(\int_{(-\pi,\pi]} |f(x) - P(x)|^2 dx\right)^{\frac{1}{2}} \le \epsilon \cdot \frac{\sqrt{\mu((-\pi,\pi])}}{1+\sqrt{\mu((-\pi,\pi])}} < \epsilon. \end{aligned}$$

Now we show that the sapce of all complex continuous functions f on $\mathbb{R}$ such that $f(-\pi) = f(\pi)$ is dense in the space $C(\mathbb{R},\mathbb{C})$ of continuous complex functions on $\mathbb{R}$ in the mean-square sense. Choose any $f \in C(\mathbb{R},\mathbb{C})$ and $\epsilon > 0$. Since f is continuous on the compact set $[-\pi,\pi]$, by the Weierstrass theorem it is bounded on this set; let $M > 1$ be an upper bound of $|f(x)|$ for $x \in [-\pi,\pi]$. Define

$$K = (\pi^2+1)^2 M^2 > 0,$$

and choose $\delta \in (0,1)$ so that

$$\mu((-\pi,-\pi+\delta)) + \mu((\pi-\delta,\pi]) < \frac{\epsilon}{2K};$$

such a δ exists due to sequential continuity and the finiteness of μ. Having chosen this δ, since

$$\lim_{h\downarrow 0} \left(1+\left(\pi^2\right)^{2h} - 2(2\pi\delta-\delta^2)^h\right) = 0,$$

there exists an $h \in (0,1)$ such that

$$\left(1+\left(\pi^2\right)^{2h} - 2(2\pi\delta-\delta^2)^h\right) < \frac{\epsilon}{2M^2 \cdot \mu((-\pi,\pi])}.$$

Now define $g : \mathbb{R} \to \mathbb{C}$ as

$$g(x) = f(x)\left(x^2-\pi^2\right)^h$$

for any $x \in \mathbb{R}$. We can clearly see that g is a continuous function on $\mathbb{R}$ such that

$f(-\pi) = f(\pi) = 0$, with upper bound

$$\sup_{x\in[-\pi,\pi]} |g(x)| \leq \left(\sup_{x\in[-\pi,\pi]} |f(x)|\right)\left(\pi^2\right)^h \leq M \cdot \pi^2$$

on $[-\pi, \pi]$. It follows that

$$\sup_{x\in[-\pi,\pi]} |f(x) - g(x)|^2 \leq M^2\left(\pi^2 + 1\right)^2 = K,$$

and since

$$\sup_{x\in[-\pi+\delta,\pi-\delta]} (\pi^2 - x^2)^h = \left(\pi^2\right)^h$$
$$\inf_{x\in[-\pi+\delta,\pi-\delta]} (\pi^2 - x^2)^h = \left(\pi^2 - (\pi-\delta)^2\right)^h = \left(2\pi\delta - \delta^2\right)^h,$$

we can see that

$$\begin{aligned}\sup_{x\in[-\pi+\delta,\pi-\delta]} |f(x) - g(x)|^2 &= \sup_{x\in[-\pi+\delta,\pi-\delta]} \left(1 - (\pi^2 - x^2)^h\right)^2 |f(x)|^2 \\ &= M^2\left[\sup_{x\in[-\pi+\delta,\pi-\delta]}\left(1 + \left(\pi^2 - x^2\right)^{2h} - 2\left(\pi^2 - x^2\right)^h\right)\right] \\ &= M^2\left(1 + \left(\pi^2\right)^{2h} - 2\left(2\pi\delta - \delta^2\right)^h\right) < \frac{\epsilon}{2 \cdot \mu((-\pi,\pi])}\end{aligned}$$

by our choice of $\delta > 0$ and $h > 0$. We can now see that

$$\begin{aligned}\|f - g\|_2^2 &= \int_{\mathbb{R}} |f(x) - g(x)|^2 d\mu(x) \\ &= \int_{(-\pi,\pi]} |f(x) - g(x)|^2 d\mu(x) \\ &= \int_{(-\pi,-\pi+\delta)} |f(x) - g(x)|^2 d\mu(x) + \int_{(\pi-\delta,\pi]} |f(x) - g(x)|^2 d\mu(x) \\ &\quad + \int_{[-\pi+\delta,\pi-\delta]} |f(x) - g(x)|^2 d\mu(x) \\ &\leq K\left(\mu((-\pi,-\pi+\delta)) + \mu((\pi-\delta,\pi])\right) \\ &\quad + \left(\sup_{x\in[-\pi+\delta,\pi-\delta]} |f(x) - g(x)|^2\right) \cdot \mu([-\pi+\delta,\pi-\delta]) \\ &\leq \frac{\epsilon}{2} + \frac{\epsilon}{2 \cdot \mu((-\pi,\pi])}\mu([-\pi+\delta,\pi-\delta]) \leq \epsilon.\end{aligned}$$

Putting these two results together, we can see that, for any continuous function $f : \mathbb{R} \to$

$\mathbb{C}$ and $\epsilon > 0$, there exists a trigonometric polynomial $P \in \mathcal{T}$ such that

$$\|f - P\|_2 < \epsilon.$$

In other words, $\mathcal{T}$ is dense in $C(\mathbb{R},\mathbb{C})$ in the mean square sense. Since we already saw in the previous section that the collection $C_c(\mathbb{R},\mathbb{C})$ of all continuous compactly supported functions on $\mathbb{R}$ is dense in $L^2(\mathcal{E},\mu)$ in the mean-square sense, it now follows that $\mathcal{T}$ is dense in $L^2(\mathcal{E},\mu)$ in the mean-square sense.

Q.E.D.

3.2.5 The Spectral Representation Theorem

Now we return to the main objective of this section, namely furnishing a representation of some time series as the stochastic integral of sinusoidal functions. Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional mean zero weakly stationary process with absolutely summable autocovariances. Letting $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$ be the autocovariance function of $\{Y_t\}_{t\in\mathbb{Z}}$, we defined the spectral density $f : (-\pi,\pi] \to \mathbb{C}^{n\times n}$ of $\{Y_t\}_{t\in\mathbb{Z}}$ as

$$f(w) = \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty}\Gamma(\tau)\exp(-i\tau w)$$

for any $w \in (-\pi,\pi]$. The spectral distribution $F : \mathbb{R} \to \mathbb{R}$ of $\{Y_t\}_{t\in\mathbb{Z}}$ is defined using f as

$$F(x) = \begin{cases} \operatorname{tr}(\Gamma(0)) & \text{if } x > \pi \\ \int_{-\pi}^{x} \operatorname{tr}(f(w))\,dw & \text{if } -\pi < x \le \pi \\ 0 & \text{if } -\pi \le x \end{cases}$$

for any $x \in \mathbb{R}$. F is continuous at π and $-\pi$ because $\Gamma(0) = \int_{-\pi}^{\pi} f(w)dw$, and it is differentiable on $(-\pi,\pi)$ because $\operatorname{tr}(f)$ is continuous on $(-\pi,\pi)$. Therefore, F is continuous on $\mathbb{R}$, bounded above by $\operatorname{tr}(\Gamma(0))$, and increasing because $\operatorname{tr}(f)$ is a non-negative function on $(-\pi,\pi]$. This indicates, in light of the construction of the Lebesgue-Stieltjes measure, that there exists a σ-algebra $\mathcal{L}_F$ on $\mathbb{R}$ and a finite measure μ_F on $(\mathbb{R},\mathcal{L}_F)$ satisfying the following properties:

i) $\mathcal{L}_F$ contains every Borel set on $\mathbb{R}$, that is, $\mathcal{B}(\mathbb{R}) \subset \mathcal{L}_F$.

ii) **(Completeness)** $(\mathbb{R},\mathcal{L}_F,\mu_F)$ is a complete measure space.

iii) **(Regularity)** μ_F is a regular Borel measure, that is,

$$\begin{aligned}\mu_F(A) &= \inf\{\mu_F(V) \mid A \subset V,\ V \text{ is open}\} \\ &= \sup\{\mu_F(K) \mid K \subset A\ K \text{ is compact}\}\end{aligned}$$

for any $A \in \mathcal{L}_F$.

iv) **(Approximation Property)** For any $A \in \mathcal{L}_F$ and $\epsilon > 0$, there exists an open set V and a closed set K such taht $K \subset A \subset V$ and

$$\mu_F(V \setminus K) < \epsilon.$$

v) For any half-open interval $(s,t] \subset [-\pi,\pi]$,

$$\mu_F((s,t]) = F(t) - F(s) = \int_s^t \mathrm{tr}(f(w))dw.$$

vi) The entire mass of μ_F is concentrated on $(-\pi,\pi]$, that is,

$$\mu_F((-\pi,\pi]^c) = 0.$$

The last property follows in a manner similarly to how we showed the distribution associated with an orthogonal increment process is concentrated on $(-\pi,\pi]$. We call μ_F the spectral distribution of $\{Y_t\}_{t\in\mathbb{Z}}$, and $\mathcal{L}_F$ the associated σ-algebra.

Suppose we extend the definition of f so that

$$f(-\pi) = f(\pi), \quad f(w) = O \text{ for any } w \notin [-\pi,\pi].$$

Then, for any half-open interval $(s,t] \subset \mathbb{R}$, we have

$$\mu_F((s,t]) = \int_s^t \mathrm{tr}(f(w))dw,$$

and since the set of half-open intervals generates the Borel σ-algebra on $\mathbb{R}$, this makes $\mathrm{tr}(f)$ the Radon-Nikodym derivative of μ_F with respect to the Lebesgue measure on $\mathbb{R}$. It follows that

$$\int_{-\pi}^{\pi} g d\mu_F = \int_{-\pi}^{\pi} (g(w)\,\mathrm{tr}(f(w)))\,dw$$

for any $\mathcal{L}_F$-measurable and μ_F-integrable complex valued function g.

Our goal is to first and foremost construct an n-dimensional orthogonal increment process $\{Z_t\}_{-\pi\le t\le\pi}$ with respect to which each Y_t can be expressed as a stochastic integral. To this end, we first consider the linear subspaces

$$W_2 \subset L_n^2(\mathcal{H},\mathbb{P}), \quad \text{and} \quad W_F \subset L^2(\mathcal{L}_F,\mu_F)$$

defined as

$$W_2 = \mathrm{span}\{Y_t \mid t \in \mathbb{Z}\}, \quad \text{and} \quad \mathcal{T} = \mathrm{span}\{w \mapsto \exp(itw) \mid t \in \mathbb{Z}\}.$$

Note that $\mathcal{T}$ is simply the space of all trigonometric polynomials; similarly, for any $X \in W_2$, we can write

$$X = \sum_{t=-N}^{N} a_t \cdot Y_t$$

for some $N \in N_+$ and $N \in N_+$.

Define the mapping $T : W_2 \to \mathcal{T}$ as

$$T\left(\sum_{t=-N}^{N} a_t \cdot Y_t\right) = \sum_{t=-N}^{N} a_t \cdot \exp(it\cdot),$$

for any $\sum_{t=-N}^{N} a_t \cdot Y_t \in W_2$, where $\exp(it\cdot)$ represents the complex exponential $x \mapsto \exp(itx)$. The following is our first result concerning the operation T:

Lemma Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional mean-zero weakly stationary time series with absolutely summable autocovariances, and denote the autocovariance function of $\{Y_t\}_{t\in\mathbb{Z}}$ by $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$. Let f be the spectral density, F the spectral distribution, and $(\mathbb{R}, \mathcal{L}_F, \mu_F)$ the spectral measure space associated with the process $\{Y_t\}_{t\in\mathbb{Z}}$.

Define the linear subspaces $W_2 \subset L_n^2(\mathcal{H}, \mathbb{P})$ and $\mathcal{T} \subset L^2(\mathcal{L}_F, \mu_F)$ as above, and let $T : W_2 \to \mathcal{T}$ be the operation introduced above. Then, T is an inner product space isomorphism from $(W_2, \langle\cdot,\cdot\rangle_{n,2})$ onto $(\mathcal{T}, \langle\cdot,\cdot\rangle_F)$.

Proof) We first establish a key result. Choose any

$$\sum_{t=-N}^{N} a_t \cdot Y_t,\ \sum_{t=-N}^{N} b_t \cdot Y_t \in W_2.$$

Then, we have

$$\left\| T\left(\sum_{t=-N}^{N} a_t \cdot Y_t\right) - T\left(\sum_{t=-N}^{N} b_t \cdot Y_t\right)\right\|_F^2 = \left\|\sum_{t=-N}^{N} a_t \cdot \exp(it\cdot) - \sum_{t=-N}^{N} b_t \cdot \exp(it\cdot)\right\|_F^2$$

$$= \int_{\mathbb{R}} \left|\sum_{t=-N}^{N} (a_t - b_t)\exp(itw)\right|^2 d\mu_F(w)$$

$$= \sum_{|t|\le N}\sum_{|s|\le N} (a_t - b_t)\overline{(a_s - b_s)} \cdot \int_{-\pi}^{\pi} \exp(i(t-s)w)d\mu_F(w)$$

$$= \sum_{|t|\le N}\sum_{|s|\le N} (a_t - b_t)\overline{(a_s - b_s)} \cdot \int_{-\pi}^{\pi} \exp(i(t-s)w)\,\mathrm{tr}\,(f(w))\,dw$$

$$= \sum_{|t|\le N}\sum_{|s|\le N} (a_t - b_t)\overline{(a_s - b_s)} \sum_{\tau=-\infty}^{\infty} \mathrm{tr}\,(\Gamma(\tau)) \cdot \int_{-\pi}^{\pi} \exp(i((t-s)-\tau)w)dw$$

$$= \sum_{|t|\leq N}\sum_{|s|\leq N}(a_t - b_t)\overline{(a_s - b_s)}\,\mathrm{tr}\left(\Gamma(t-s)\right)$$

$$= \sum_{|t|\leq N}\sum_{|s|\leq N}(a_t - b_t)\overline{(a_s - b_s)}\cdot \mathbb{E}\left[Y_t' Y_s\right]$$

$$= \mathbb{E}\left|\sum_{|t|\leq N}(a_t - b_t)\,Y_t\right|^2 = \left\|\sum_{t=-N}^{N} a_t\cdot Y_t - \sum_{t=-N}^{N} b_t\cdot Y_t\right\|_{n,2}^2.$$

As such,

$$\left\|T\left(\sum_{t=-N}^{N} a_t\cdot Y_t\right) - T\left(\sum_{t=-N}^{N} b_t\cdot Y_t\right)\right\|_F = \left\|\sum_{t=-N}^{N} a_t\cdot Y_t - \sum_{t=-N}^{N} b_t\cdot Y_t\right\|_{n,2};$$

if $\sum_{t=-N}^{N} a_t\cdot Y_t$ and $\sum_{t=-N}^{N} b_t\cdot Y_t$ are two representations of the same random vector $X\in W_2$, then the right hand side is 0, so that

$$T\left(\sum_{t=-N}^{N} a_t\cdot Y_t\right) = T\left(\sum_{t=-N}^{N} b_t\cdot Y_t\right) = T(X)$$

almost surely. This shows us that T is a well-defined operation.

Another way to express the above equation is that, for any $X, Z\in W_2$, we have

$$\|T(X) - T(Z)\|_F = \|X - Z\|_{n,2}.$$

Putting $Z = \mathbf{0}$, we can see that

$$\|T(X)\|_F = \|X\|_{n,2},$$

so that $T(X) = 0$ if and only if $X = \mathbf{0}$. This shows us that the operation T is injective. In addition, for any $X, Z\in W_2$ and $c\in\mathbb{C}$ such that

$$X = \sum_{t=-N}^{N} a_t\cdot Y_t,\; Z = \sum_{t=-N}^{N} b_t\cdot Y_t,$$

since

$$cX + Z = \sum_{t=-N}^{N}(c\cdot a_t + b_t)\cdot Y_t,$$

we have

$$T(cX+Z) = \sum_{t=-N}^{N}(c\cdot a_t + b_t)\cdot \exp(it\cdot)$$

$$= c\left(\sum_{t=-N}^{N} a_t \cdot \exp(it\cdot)\right) + \left(\sum_{t=-N}^{N} b_t \cdot \exp(it\cdot)\right) = c \cdot T(X) + T(Z),$$

which demonstrates that T is a linear transformation. This, together with the injectivity of T, shows us that T is a vector space isomorphism from W_2 onto $\mathcal{T}$.

It remains to show that T preserves inner products. This is seen below; for any

$$\sum_{t=-N}^{N} a_t \cdot Y_t, \sum_{t=-N}^{N} b_t \cdot Y_t \in W_2.$$

we have

$$\begin{aligned}
\left\langle T\left(\sum_{|t|\leq N} a_t \cdot Y_t\right), T\left(\sum_{|t|\leq N} b_t \cdot Y_t\right)\right\rangle_F &= \left\langle \sum_{|t|\leq N} a_t \cdot \exp(it\cdot), \sum_{|t|\leq N} b_t \cdot \exp(it\cdot)\right\rangle_F \\
&= \sum_{|t|\leq N}\sum_{|s|\leq N} a_t\overline{b_s} \cdot \int_{-\pi}^{\pi} \exp(i(t-s)w)d\mu_F(w) \\
&= \sum_{|t|\leq N}\sum_{|s|\leq N} a_t\overline{b_s} \cdot \int_{-\pi}^{\pi} \exp(i(t-s)w)\operatorname{tr}(f(w))\,dw \\
&= \sum_{|t|\leq N}\sum_{|s|\leq N} a_t\overline{b_s} \sum_{\tau=-\infty}^{\infty} \operatorname{tr}(\Gamma(\tau)) \cdot \int_{-\pi}^{\pi} \exp(i((t-s)-\tau)w)dw \\
&= \sum_{|t|\leq N}\sum_{|s|\leq N} a_t\overline{b_s}\operatorname{tr}(\Gamma(t-s)) \\
&= \sum_{|t|\leq N}\sum_{|s|\leq N} a_t\overline{b_s} \cdot \mathbb{E}\left[Y_t' Y_s\right] \\
&= \mathbb{E}\left[\left(\sum_{t=-N}^{N} a_t \cdot Y_t\right)' \overline{\left(\sum_{t=-N}^{N} b_t \cdot Y_t\right)}\right] \\
&= \left\langle \sum_{t=-N}^{N} a_t \cdot Y_t, \sum_{t=-N}^{N} b_t \cdot Y_t\right\rangle_{n,2}.
\end{aligned}$$

Therefore, T is an inner product space isomorphism from W_2 onto $\mathcal{T}$.

Q.E.D.

Now, we want to extend the domain of T to $\overline{W_2}$ and its target space to $\overline{\mathcal{T}}$, where $\overline{W_2}$ is the L^2-closure of W_2 and $\overline{\mathcal{T}}$ the L^2-closure of W_F. Note that, since $\mathcal{T}$ is the set of all trigonometric polynomials, $\mathcal{L}_F$ is a σ-algebra on $\mathbb{R}$ containing $\mathcal{B}(\mathbb{R})$, and μ_F is a finite measure on $(\mathbb{R}, \mathcal{L}_F)$ concentrated on $(-\pi, \pi]$, by the result shown in the previous section the L^2-closure $\overline{\mathcal{T}}$ is precisely the space $L^2(\mathcal{L}_F, \mu_F)$.

T is extended to $\overline{W_2}$ in almost the same manner as the stochastic integral. For any $X \in \overline{W_2}$, there exists a sequence $\{X_k\}_{k \in N_+} \in W_2$ such that

$$\lim_{n \to \infty} \|X_k - X\|_{n,2} = 0.$$

Since the operation T on W_2 preserves norms, we can see that

$$\|T(X_k) - T(X_m)\|_F = \|X_k - X_m\|_{n,2}$$

for any $k, m \in N_+$. Since the right hand side goes to 0 as $n, m \to \infty$ (all convergent sequences are Cauchy), so does the left hand side; this tells us that $\{T(X_k)\}_{k \in N_+} \subset L^2(\mathcal{L}_F, \mu_F)$ is Cauchy in L^2, and by the completeness of L^2-spaces as Hilbert spaces, it follows that this sequence converges to some quantity in $L^2(\mathcal{L}_F, \mu_F) = \overline{W_F}$. We then define

$$\tilde{T}(X) = \underset{k \to \infty}{m.s.lim} T(X_k).$$

As during the construction of stochastic integrals, we must verify the following to see that $\tilde{T}(X)$ is well-defined:

- **$\tilde{T}(X)$ is invariant to the choice of convergent sequence**
  Suppose that $\{X_k\}_{k \in N_+}$ and $\{Z_k\}_{k \in N_+}$ are two sequences in W_2 converging to $X \in \overline{W_2}$ in L^2. Denote

  $$T_X = \underset{k \to \infty}{\text{m.s.lim}}\, T(X_k) \quad \text{and} \quad T_Z = \underset{k \to \infty}{\text{m.s.lim}}\, T(Z_k).$$

  Then, for any $k \in N_+$,

  $$\|T_X - T_Z\|_F \leq \|T_X - T(X_k)\|_F + \|T(X_k) - T(Z_k)\|_F + \|T(Z_k) - T_Z\|_F.$$

  By assumption, the first and third terms go to 0 as $k \to \infty$. As for the second term,

  $$\begin{aligned} \|T(X_k) - T(Z_k)\|_F &= \|X_k - Z_k\|_{n,2} \\ &\leq \|X_k - X\|_{n,2} + \|X - Z_k\|_{n,2}. \end{aligned}$$

  Both terms on the right hand side go to 0 as $k \to \infty$, so $\|T(X_k) - T(Z_k)\|_F$ also goes to 0 as $k \to \infty$. It follows that

  $$\|T_X - T_Z\|_F = 0,$$

so that $T_X = T_Z = \tilde{T}(X)$ almost surely. This shows us that $\tilde{T}(X)$ does not depend on the choice of sequence in W_2 that converges to X.

- **$\tilde{T}(X)$ is $T(X)$ for $X \in W_2$**
  Suppose $X \in W_2$. Then, $\{X_k\}_{k\in N_+}$ defined as $X_k = X$ for any $k \in N_+$ is a sequence in W_2 converging to X in L^2, so

$$\tilde{T}(X) = \underset{k\to\infty}{\text{m.s.lim}}\, T(X_k) = T(X).$$

  Therefore, $\tilde{T}(X)$ is precisely $T(X)$ if $X \in W_2$.

As in the case of stochastic integration, these two remarks allow us to define the operation $T : \overline{W_2} \to L^2(\mathcal{L}_F, \mu_F)$ as

$$T(X) = \underset{k\to\infty}{\text{m.s.lim}}\, T(X_k)$$

for any $X \in \overline{W_2}$ and sequence $\{X_k\}_{k\in N_+}$ converging to X in L^2.

The extended operation $T : \overline{W_2} \to L^2(\mathcal{L}_F, \mu_F)$ is actually an inner product isomorphsim from $\overline{W_2}$ onto $L^2(\mathcal{L}_F, \mu_F)$, as we show below:

Lemma Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional mean-zero weakly stationary time series with absolutely summable autocovariances, and denote the autocovariance function of $\{Y_t\}_{t\in\mathbb{Z}}$ by $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$. Let f be the spectral density, F the spectral distribution, and $(\mathbb{R}, \mathcal{L}_F, \mu_F)$ the spectral measure space associated with the process $\{Y_t\}_{t\in\mathbb{Z}}$.

Define the linear subspace $W_2 \subset L^2_n(\mathcal{H}, \mathbb{P})$ as above, and let $T : \overline{W_2} \to L^2(\mathcal{L}_F, \mu_F)$ be the operation introduced above. Then, T is an inner product space isomorphism from $(\overline{W_2}, \langle\cdot,\cdot\rangle_{n,2})$ onto $L^2(\mathcal{L}_F, \mu_F)$.

Proof) Let $c \in \mathbb{C}$, $X, Z \in \overline{W_2}$, and $\{X_k\}_{k\in N_+}$, $\{Z_k\}_{k\in N_+}$ sequences in W_2 converging to X and Z in L^2. Then, since $\{cX_k + Z_k\}_{k\in N_+}$ is a sequence in W_2 converging to $cX + Z$ in L^2,

$$\begin{aligned} T(cX+Z) &= \underset{k\to\infty}{\text{m.s.lim}}\, T(cX_k + Z_k) \\ &= c\left(\underset{k\to\infty}{\text{m.s.lim}}\, T(X_k)\right) + \underset{k\to\infty}{\text{m.s.lim}}\, T(Z_k) \\ &= cT(X) + T(Z), \end{aligned}$$

where the second equality follows from the linearity of T on W_2. Therefore, T is a linear transformation on $\overline{W_2}$.

In addition, we can see that, because inner products are continuous functions,

$$\begin{aligned}\langle T(X), T(Z)\rangle_F &= \lim_{k\to\infty}\langle T(X_k), T(Z_k)\rangle_F \\ &= \lim_{k\to\infty}\langle X_k, Z_k\rangle_{n,2} = \langle X, Z\rangle_{n,2};\end{aligned}$$

the second equality follows because T preserves inner products on W_2. Therefore, T preserves inner products on $\overline{W_2}$, and as a special case, when $X = Z$,

$$\|T(X)\|_F = \|X\|_{n,2}.$$

This shows us that the operation T is injective, and as such T is an inner product space isomorphism from $\overline{W_2}$ onto $L^2(\mathcal{L}_F, \mu_F)$.

Q.E.D.

The inner product space isomorphism $T : \overline{W_2} \to L^2(\mathcal{L}_F, \mu_F)$ defined above can be used to construct the orthogonal increment process of interest.

Lemma Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional mean-zero weakly stationary time series with absolutely summable autocovariances, and denote the autocovariance function of $\{Y_t\}_{t\in\mathbb{Z}}$ by $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$. Let f be the spectral density, F the spectral distribution, and $(\mathbb{R}, \mathcal{L}_F, \mu_F)$ the spectral measure space associated with the process $\{Y_t\}_{t\in\mathbb{Z}}$.

Define the linear subspace $W_2 \subset L_n^2(\mathcal{H}, \mathbb{P})$ as above, and let $T : \overline{W_2} \to L^2(\mathcal{L}_F, \mu_F)$ be the operation introduced above. Letting $T^{-1} : L^2(\mathcal{L}_F, \mu_F) \to \overline{W_2}$ be the inverse of T, define the process $\{Z_\lambda\}_{-\pi\le\lambda\le\pi}$ as

$$Z_\lambda = T^{-1}\left(I_{(-\pi,\lambda]}\right)$$

for any $-\pi \le \lambda \le \pi$. Then, $\{Z_\lambda\}_{-\pi\le\lambda\le\pi}$ is a mean-zero, square integrable and right continuous orthogonal increment process, and F is precisely the distribution function associated with $\{Z_\lambda\}_{-\pi\le\lambda\le\pi}$.

Proof) $\{Z_\lambda\}_{-\pi\le\lambda\le\pi}$ is clearly square integrable, since each Z_λ takes values in $\overline{W_2} \subset L_n^2(\mathcal{H}, \mathbb{P})$. In addition, since Z_λ is an element of $\overline{W_2}$, there exists a sequence $\{X_k\}_{k\in N_+}$ in W_2 that converges in mean square to Z_λ. Each X_k has mean zero because it is the linear combination of random vectors with mean zero. Since

$$|\mathbb{E}[Z_\lambda]| = |\mathbb{E}[Z_\lambda] - \mathbb{E}[X_k]| \le \mathbb{E}|Z_\lambda - X_k| \le \|Z_\lambda - X_k\|_{n,2}$$

for any $k \in N_+$ by Hölder's inequality, taking $k \to \infty$ on both sides shows us that

$$|\mathbb{E}[Z_\lambda]| = 0,$$

or that Z_λ has mean zero.

For any $-\pi \leq w \leq u \leq s \leq t \leq \pi$,

$$\begin{aligned}\langle Z_t - Z_s,\, Z_u - Z_w\rangle_{n,2} &= \langle T(Z_t) - T(Z_s),\, T(Z_u) - T(Z_w)\rangle_F \\ &= \langle I_{(s,t]},\, I_{(w,u]}\rangle_F = \sqrt{\mu_F((w,u] \cap (s,t])},\end{aligned}$$

where the first equality follows because the operation T is linear and preserves inner products. Since $(w,u] \cap (s,t] = \emptyset$, it follows that

$$\langle Z_t - Z_s,\, Z_u - Z_w\rangle_{n,2} = 0,$$

so that $\{Z_\lambda\}_{-\pi \leq \lambda \leq \pi}$ has orthogonal increments.

Finally, for any $-\pi < \lambda \leq \pi$, note that

$$\begin{aligned}\mathbb{E}|Z_\lambda - Z_{-\pi}|^2 &= \|Z_\lambda - Z_{-\pi}\|_{n,2}^2 \\ &= \|T(Z_\lambda) - T(Z_{-\pi})\|_F^2 = \left\|I_{(-\pi,\lambda]}\right\|_F^2 \\ &= \mu_F((-\pi,\lambda]) = F(\lambda).\end{aligned}$$

$F(\lambda) = 0$ for any $\lambda \leq \pi$ and $F(\lambda) = F(\pi)$ for any $\lambda > \pi$, so it follows that F is precisely the distribution function associated with $\{Z_\lambda\}_{-\pi \leq \lambda \leq \pi}$. By implication,

$$\|Z_{\lambda+\delta} - Z_\lambda\|_{n,2} = \sqrt{F(\lambda+\delta) - F(\lambda)}$$

for any $-\pi \leq \lambda < \pi$ and sufficiently small $\delta > 0$, so by the continuity of F,

$$\lim_{\delta \downarrow 0} \|Z_{\lambda+\delta} - Z_\lambda\|_{n,2} = 0.$$

This demonstrates that $\{Z_\lambda\}_{-\pi \leq \lambda \leq \pi}$ is a process that is right continuous (in mean square).

Q.E.D.

Finally, we can show that each Y_t can be expressed as the stochastic integral of $\exp(it\cdot)$ with respect to the orthogonal increment process constructed above. This is the formal statement of the spectral representation theorem.

Theorem (Spectral Representation Theorem)
Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional mean-zero weakly stationary time series with absolutely summable autocovariances, and denote the autocovariance function of $\{Y_t\}_{t\in\mathbb{Z}}$ by $\Gamma:\mathbb{Z}\to\mathbb{R}^{n\times n}$. Let f be the spectral density, F the spectral distribution, and $(\mathbb{R},\mathcal{L}_F,\mu_F)$ the spectral measure space associated with the process $\{Y_t\}_{t\in\mathbb{Z}}$.

Then, there exists a mean-zero, square integrable and right continuous (in mean square) orthogonal increment process $\{Z_\lambda\}_{-\pi\le\lambda\le\pi}$ such that

$$\mathbb{E}|Z_\lambda - Z_{-\pi}|^2 = F(\lambda)$$

for any $-\pi\le\lambda\le\pi$ and

$$Y_t = \int_{-\pi}^{\pi}\exp(it\lambda)dZ(\lambda)$$

for any $t\in\mathbb{Z}$.

Proof) Let $W_2\subset L_n^2(\mathcal{H},\mathbb{P})$ be the linear subspace and $T:\overline{W_2}\to L^2(\mathcal{L}_F,\mu_F)$ the inner product space isomorphism defined above. Furthermore, let $\{Z_\lambda\}_{-\pi\le\lambda\le\pi}$ be the orthogonal increment process constructed in the preceding lemma as

$$Z_\lambda = T^{-1}\left(I_{(-\pi,\lambda]}\right)$$

for any $-\pi\le\lambda\le\pi$.

Let the operation $I:L^2(\mathcal{L}_F,\mu_F)\to L_n^2(\mathcal{H},\mathbb{P})$ denote stochastic integration with respect to the process $\{Z_\lambda\}_{-\pi\le\lambda\le\pi}$, that is,

$$I(f) = \int_{-\pi}^{\pi} f(\lambda)dZ(\lambda)$$

for any $f\in L^2(\mathcal{L}_F,\mu_F)$. Recall that I is a linear transformation that preserves inner products. We must show that $I=T^{-1}$ on $L^2(\mathcal{L}_F,\mu_F$; then,

$$Y_t = T^{-1}(\exp(it\cdot)) = I(\exp(it\cdot)) = \int_{-\pi}^{\pi}\exp(it\lambda)dZ(\lambda)$$

for any $t\in\mathbb{Z}$, as we desired.

To this end, choose some elementary function $f\in\mathcal{D}\subset L^2(\mathcal{L}_F,\mu_F)$, and represent it as

$$f=\sum_{i=0}^{k} r_i\cdot I_{(\lambda_i,\lambda_{i+1}]},$$

where $-\pi = \lambda_0 < \cdots < \lambda_{k+1} = \pi$ is a partition of $[-\pi, \pi]$. Then,

$$\begin{aligned}
T^{-1}(f) &= \sum_{i=0}^{k} r_i \cdot T^{-1}\left(I_{(\lambda_i, \lambda_{i+1}]}\right) \\
&= \sum_{i=0}^{k} r_i \cdot \left[T^{-1}\left(I_{(-\pi, \lambda_{i+1}]}\right) - T\left(I_{(-\pi, \lambda_i]}\right)\right] \\
&= \sum_{i=0}^{k} r_i \cdot (Z_{\lambda_{i+1}} - Z_{\lambda_i}) = I(f),
\end{aligned}$$

so that $I = T^{-1}$ on the space $\mathcal{D}$ of all elementary functions on $[-\pi, \pi]$.

Now choose some $f \in L^2(\mathcal{L}_F, \mu_F)$. Since $\mathcal{D}$ is dense in $L^2(\mathcal{L}_F, \mu_F)$ in the mean square sense, there exists a sequence $\{f_k\}_{k \in N_+}$ of elementary functions that converges to f in L^2. By definition,

$$I(f) = \underset{k\to\infty}{\text{m.s.lim}}\, I(f_k) = \underset{k\to\infty}{\text{m.s.lim}}\, T^{-1}(f_k).$$

Note that, for any $k \in N_+$,

$$\left\|T^{-1}(f_k) - T^{-1}(f)\right\|_{n,2} = \left\|T(T^{-1}(f_k)) - T(T^{-1}(f))\right\|_F = \|f_k - f\|_F$$

by the linearity and inner product perserving property of T. Thus, taking $k \to \infty$ on both sides shows that

$$T^{-1}(f) = \underset{k\to\infty}{\text{m.s.lim}}\, T^{-1}(f_k),$$

and by the almost sure uniqueness of L^2-limits,

$$I(f) = T^{-1}(f).$$

In other words, $I = T^{-1}$ on $L^2(\mathcal{L}_F, \mu_F)$, which completes the proof.

Q.E.D.

What the spectral representation tells us is that any time series $\{Y_t\}_{t\in\mathbb{Z}}$ can be expressed as the weighted sum of periodic functions of frequencies ranging from $-\pi$ to π. Here, Z_λ can be interpreted as the weight assigned to the periodic function $t \mapsto \exp(it\lambda)$. Use Euler's formula to see that

$$\exp(it\lambda) = \cos(t\lambda) + i\sin(t\lambda).$$

$\exp(it\lambda)$, as a function of t, is a periodic function with period $\frac{2\pi}{\lambda}$, since

$$\cos(t\lambda) = \cos(t\lambda + 2\pi) = \cos\left(\lambda\left(t + \frac{2\pi}{\lambda}\right)\right).$$

In other words, this function repeats every $\frac{2\pi}{\lambda}$ time periods. An equivalent way to say this is to say that the function has frequency $\frac{\lambda}{2\pi}$; the frequency of a periodic function is the average number of times the function is expected to repeat its behavior in a single time period. The lower the frequency, the less frequently we observe the same behavior.

Therefore, Z_λ is the weight assigned to a wavelength of frequency $\frac{\lambda}{2\pi}$. The larger λ, the more the associated wavelength recurs in a given length of time, meaning that it represents cyclical behavior. In contrast, a wavelegnth with a lower frequency takes much longer to recur, so that it can be interpreted as representing the trending behavior of a time series. Thus, the spectral representation theorem, by allowing us to decompose a time series into its higher and lower frequency components, allows us to extract and study separately its trend and cycle components.

3.3 Time Invariant Linear Filters

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional, mean-zero, square integrable and weakly stationary time series with absolutely summable autocovariance function $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$. An absolutely summable sequence $\{\Psi_j\}_{j\in\mathbb{Z}}$ of $n \times n$ matrices is referred to as a time-invariant linear filter (TLF), and the process $\{X_t\}_{t\in\mathbb{Z}}$ defined as

$$X_t = \Psi(L)Y_t = \sum_{j=-\infty}^{\infty} \Psi_j \cdot Y_{t-j}$$

for any $t \in \mathbb{Z}$ is said to be obtained from $\{Y_t\}_{t\in\mathbb{Z}}$ via the filter $\Psi = \{\Psi_j\}_{j\in\mathbb{Z}}$. Recall that $\{X_t\}_{t\in\mathbb{Z}}$ is itself a mean-zero weakly stationary process and absolutely summable autocovariances. The filter Ψ is said to be causal if $\Psi_j = O$ for any $j < 0$.

TLFs arise often in time series analysis; for instance, the h-period moving average process $\{X_t\}_{t\in\mathbb{Z}}$ defined as

$$X_t = \frac{1}{2h} \sum_{j=-h}^{h} Y_{t-j}$$

for any $t \in \mathbb{Z}$ is obtained from $\{Y_t\}_{t\in\mathbb{Z}}$ via the filter $H = \{H_j\}_{j\in\mathbb{Z}}$ defined as

$$H_j = \begin{cases} \frac{1}{2h} I_n & \text{if } |j| \leq h \\ O & \text{if } |j| > h \end{cases}.$$

Similarly, the first difference process $\{\Delta Y_t\}_{t\in\mathbb{Z}}$ defined as

$$\Delta Y_t = Y_t - Y_{t-1}$$

for any $t \in \mathbb{Z}$ is obtained from $\{Y_t\}_{t\in\mathbb{Z}}$ via the causal filter $D = \{D_j\}_{j\in\mathbb{Z}}$ defined as

$$D_j = \begin{cases} I_n & \text{if } j = 0 \\ -I_n & \text{if } j = 1 \\ O & \text{otherwise} \end{cases}.$$

However, one should exercise caution when employing TLFs because they have the effect of eliminating wavelengths of certain frequencies from the original time series; this may leave us only with the trend or cycle components of the original series. This is formally articulated in the following theorem:

Theorem (Transformations in Spectrum via TLF)
Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a univariate mean-zero weakly stationary time series with absolutely summable autocovariances. Let f be the spectral density of $\{Y_t\}_{t\in\mathbb{Z}}$, F its spectral distribution, and $(\mathbb{R}, \mathcal{L}_F, \mu_F)$ the associated measure space. Denote the L^2-norm on $L^2(\mathcal{L}_F, \mu_F)$ by $\|\cdot\|_F$. Finally, let

$$Y_t = \int_{-\pi}^{\pi} \exp(it\lambda) dZ(\lambda)$$

the spectral representation of $\{Y_t\}_{t\in\mathbb{Z}}$, where F is the distribution function associated with the orthogonal increment process $\{Z_\lambda\}_{-\pi\le\lambda\le\pi}$.

Let $\Psi = \{\Psi_j\}_{j\in\mathbb{Z}}$ be a TLF, and let $\Psi(z)$ be the associated polynomial defined as

$$\Psi(z) = \sum_{j=-\infty}^{\infty} \Psi_j \cdot z^j$$

for any $z \in \mathbb{C}$.

Suppose we obtain the univariate mean-zero weakly stationary time series $\{X_t\}_{t\in\mathbb{Z}}$ with absolutely summable autocovariances from $\{Y_t\}_{t\in\mathbb{Z}}$ via the TLF Ψ. Then, the spectral density $f_X : (-\pi, \pi] \to \mathbb{C}$ of $\{X_t\}_{t\in\mathbb{Z}}$ is given as

$$f_X(w) = \left|\Psi(e^{-iw})\right|^2 f(w)$$

for any $w \in (-\pi, \pi]$, and $\{X_t\}_{t\in\mathbb{Z}}$ has spectral representation

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda} \Psi(e^{-i\lambda}) dZ(\lambda)$$

for any $t \in \mathbb{Z}$.

Proof) We first show the result for the spectral density. Letting $\Gamma : \mathbb{Z} \to \mathbb{R}$ and $G : \mathbb{Z} \to \mathbb{R}$ be the autocovariance functions of $\{Y_t\}_{t\in\mathbb{Z}}$ and $\{X_t\}_{t\in\mathbb{Z}}$, recall that

$$G(\tau) = \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} \Psi_j \cdot \Gamma(\tau+k-j)\cdot \Psi_k$$

for any $\tau \in \mathbb{Z}$. By definition, for any $w \in (-\pi, \pi]$ we now have

$$\begin{aligned}
f_X(w) &= \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty} G(\tau)\exp(-i\tau w) \\
&= \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} [\Psi_j \cdot \Gamma(\tau+k-j)\cdot \Psi_k]\exp(-i\tau w) \\
&= \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} \Psi_j\Psi_k\left[\frac{1}{2\pi}\sum_{s=-\infty}^{\infty}\Gamma(s)\exp(-i(s+j-k)w)\right] \\
&= \left(\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\Psi_j\Psi_k\exp(i(k-j)w)\right) f(w) \\
&= \left(\sum_{j=-\infty}^{\infty}\Psi_j\cdot\exp(-ijw)\right)^2 f(w) = \left|\Psi(e^{-iw})\right|^2 f(w).
\end{aligned}$$

Now we move onto the spectral representation. For any $t \in \mathbb{Z}$ and $N \in N_+$, by the linearity of stochastic integration,

$$\begin{aligned}
\sum_{|j|\le N}\Psi_j\cdot Y_{t-j} &= \sum_{|j|\le N}\Psi_j\cdot\int_{-\pi}^{\pi}\exp(i(t-j)\lambda)dZ(\lambda) \\
&= \int_{-\pi}^{\pi}\exp(it\lambda)\left(\sum_{|j|\le N}\Psi_j\exp(-ij\lambda)\right)dZ(\lambda).
\end{aligned}$$

The sequence

$$\Big\{\sum_{|j|\le N}\Psi_j\cdot Y_{t-j}\Big\}_{N\in N_+}$$

converges in mean square and almost surely to X_t. Furthermore, defining

$$X_{N,t} = \int_{-\pi}^{\pi}\exp(it\lambda)\left(\sum_{|j|\le N}\Psi_j\exp(-ij\lambda)\right)dZ(\lambda)$$

for any $N \in N_+$,

$$\left\|X_{N,t} - \int_{-\pi}^{\pi}e^{it\lambda}\Psi(e^{-i\lambda})dZ(\lambda)\right\|_2^2 = \left\|\exp(it\cdot)\left(\sum_{|j|\le N}\Psi_j\exp(-ij\cdot) - \Psi(\exp(-i\cdot))\right)\right\|_F^2$$

$$= \int_{\mathbb{R}} \left| e^{it\lambda} \left(\sum_{|j| \le N} \Psi_j e^{-ij\lambda} - \Psi(e^{-i\lambda}) \right) \right|^2 d\mu_F(\lambda)$$

$$= \int_{\mathbb{R}} \left| \sum_{|j| \le N} \Psi_j e^{-ij\lambda} - \Psi(e^{-i\lambda}) \right|^2 d\mu_F(\lambda),$$

where the first equality follows from the fact that stochastic integration perserves inner products.

$$\left\{ \left| \sum_{|j| \le N} \Psi_j \exp(-ij\cdot) - \Psi(\exp(-i\cdot)) \right|^2 \right\}_{N \in N_+}$$

is a sequence of continuous functions on $\mathbb{R}$ that converges pointwise to 0 and which is dominated by the μ_F-integrable function $3|\Psi(\exp(-i\cdot))|$, so by the DCT,

$$\lim_{N \to \infty} \left\| X_{N,t} - \int_{-\pi}^{\pi} e^{it\lambda} \Psi(e^{-i\lambda}) dZ(\lambda) \right\|_{n,2}^2$$
$$= \lim_{N \to \infty} \int_{\mathbb{R}} \left| \sum_{|j| \le N} \Psi_j e^{-ij\lambda} - \Psi(e^{-i\lambda}) \right|^2 d\mu_F(\lambda) = 0.$$

It follows from the almost sure uniqueness of L^2-limits that

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda} \Psi(e^{-i\lambda}) dZ(\lambda).$$

Q.E.D.

The mapping $w \mapsto \Psi(e^{-iw})$ is called the gain function, and $w \mapsto \left|\Psi(e^{-iw})\right|^2$ the squared gain function. For example, the gain function of the h-period moving average filter is

$$H(e^{-iw}) = \frac{1}{2h} \sum_{j=-h}^{h} e^{-ijw} = \frac{1}{h} \left(\sum_{j=1}^{h} \cos(jw) + \frac{1}{2} \right),$$

and that of the first difference filter is

$$D(e^{-iw}) = 1 - e^{-iw}.$$

The above theorem basically tells us that wavelengths of frequencies for which the gain function equals 0 are erased from the transformed process $\{X_t\}_{t \in \mathbb{Z}}$.

Unit Root Asymptotics

Here we introduce the mathematics needed to study the asymptotics of non-stationary and cointegrated processes. As in the previous section, we take $(\Omega, \mathcal{H}, \mathbb{P})$ as our underlying probability space.

4.1 The FCLT and its Extensions

The FCLT is a generalization of the CLT studied above. We first re-state some results concerning continuous function spaces and weak convergence on such spaces.

4.1.1 Continuous Function Spaces

For any topological space (E, τ) and $F = \mathbb{R}^n$ or $\mathbb{C}$, the space $\mathcal{C}_b(E, F)$ collects every bounded and continuous function mapping E into F; the boundedness condition can be omitted if E is compact due to the extreme value theorem. The supremum norm on $\mathcal{C}_b(E, F)$ is defined as

$$\|f\|_{\mathcal{C}} = \sup_{x \in E} |f(x)|$$

for any $f \in \mathcal{C}_b(E, F)$, and the supremum metric d on $\mathcal{C}_b(E, F)$ as

$$d(f, g) = \|f - g\|_{\mathcal{C}}$$

for any $f, g \in \mathcal{C}_b(E, F)$. We can show that $(\mathcal{C}_b(E, F), d)$ is a complete metric space.

Using the Stone-Weierstrass theorem, it is also possible to show that, if (E, ρ) is a compact metric space, then $(\mathcal{C}(E, F), d)$ defines a separable metric space. Thus, $(\mathcal{C}(E, F), d)$ is a Polish space (a complete and separable metric space) if (E, ρ) is a compact metric space.

Let $\mathcal{B}_{\mathcal{C}}(E, F)$ be the Borel σ-algebra generated by the metric topology induced by d. Defining the set of all finite-dimensional distributions as

$$\mathcal{C}_f = \{\pi^{-1}_{t_1, \cdots, t_k}(A) \mid t_1, \cdots, t_k \in E, A \in \mathcal{B}(F^k)\},$$

where $\pi_{t_1,\cdots,t_k} : \mathcal{C}(E,F) \to F^k$ is the projection function defined as

$$\pi_{t_1,\cdots,t_k} \circ f = (f(t_1),\cdots,f(t_k))$$

for any $f \in \mathcal{C}(E,F)$, $\mathcal{C}_f$ is a π-system that generates $\mathcal{B}_{\mathcal{C}}(E,F)$.

4.1.2 Random Functions

A random function X is a random variable that takes values in the measurable space $(\mathcal{C}(E,F),\mathcal{B}_{\mathcal{C}}(E,F))$. To any X there corresponds a stochastic process $\{X_t\}_{t\in E}$ with continuous paths taking values in $(F,\mathcal{B}(F))$ defined as

$$X_t = \pi_t \circ X$$

for any $t \in E$. Conversely, for any stochastic process $\{X_t\}_{t\in E}$ with continuous paths taking values in $(F,\mathcal{B}(F))$, we can define a corresponding random function X by letting $X(\omega)$ be the continuous mapping

$$t \mapsto X_t(\omega)$$

for any $\omega \in \Omega$. The random function X and the stochastic process $\{X_t\}_{t\ inE}$ are in this case said to correspond to one another.

We are mostly interested in the collection of continuous functions defined on the compact metric space $[0,1]$ equipped with the euclidean metric. The properties mentioned above all apply to the metric space $(\mathcal{C}([0,1],\mathbb{R}^n),d)$, where d is the supremum metric, and to the measurable space $(\mathcal{C}([0,1],\mathbb{R}^n),\mathcal{B}_{\mathcal{C}}([0,1],\mathbb{R}^n))$, where $\mathcal{B}_{\mathcal{C}}([0,1],\mathbb{R}^n)$ is the Borel σ-algebra generated by the metric topology induced by d.

4.1.3 The FCLT

Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be an i.i.d. sequence of n-dimensional random vectors with mean 0, positive definite covariance matrix Σ, and finite fourth moments. For any $T \in N_+$, we can define the stochastic process $\{X_T(r)\}_{r\in[0,1]}$ with continuous paths as

$$X_T(r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}\varepsilon_t + \frac{1}{\sqrt{T}}(Tr - \lfloor Tr\rfloor)\varepsilon_{\lfloor Tr\rfloor+1}$$

for any $r \in [0,1]$. Let the random function corresponding to $\{X_T(r)\}_{r\in[0,1]}$ be denoted X^T, and μ_T its distribution.

The Functional Central Limit Theorem (FCLT) tells us that the sequence $\{X^T\}_{T\in N_+}$ of random functions converges weakly to the n-dimensional Brownian function B^n with covariance matrix Σ, that is,

$$B^n = \Sigma^{\frac{1}{2}} W^n$$

for the standard n-dimensional Wiener function W^n and $\Sigma^{\frac{1}{2}}$ the Cholesky factor of Σ. The stochastic process $\{B^n(r)\}_{r\in[0,1]}$ corresponding to B^n is the n-dimensional Brownian motion with covariance matrix Σ, and

$$B^n(r) = \Sigma^{\frac{1}{2}} W^n(r)$$

for any $r \in [0,1]$, where $\{W^n(r)\}_{r\in[0,1]}$ is the standard n-dimensional Wiener process on $[0,1]$.

By the continuous mapping theorem, for any $0 \leq r_1 < \cdots < r_k \leq 1$,

$$(X_T(r_1), \cdots, X_T(r_k)) \xrightarrow{d} (B^n(r_1), \cdots, B^n(r_k))$$

as $T \to \infty$, which holds because the projection $\pi_{r_1,\cdots,r_k}$ is a uniformly continuous function from $\mathcal{C}([0,1], \mathbb{R}^n)$ to $\mathbb{R}^{nk}$.

The FCLT implies the Lindeberg-Levy CLT, since the FCLT implies

$$X_T(1) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t \xrightarrow{d} B^n(1) \sim N(\mathbf{0}, \Sigma).$$

4.1.4 Linear Processes and the BN Decomposition

We often find it necessary to extend the FCLT beyond i.i.d. processes. A natural class of time series to which to apply the FCLT is the class of linear processes. Recall that, given a white noise process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ with positive definite covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$, and a square summable $\{\Psi_j\}_{j\in\mathbb{N}}$, that is,

$$\sum_{j=0}^{\infty} \operatorname{tr}\left(\Psi_j \Sigma \Psi_j'\right) < +\infty,$$

we can define the (causal) zero mean linear process $\{Y_t\}_{t\in\mathbb{Z}}$ as

$$Y_t = \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j}$$

for any $t \in \mathbb{Z}$, where the limit is taken in L^2. If $\{\Psi_j\}_{j\in\mathbb{N}}$ is absolutely summable instead of square summable, we showed above that the convergence can be extended to almost sure convergence as well.

A stronger result than square summability and even absolute summability is one-summability;

$\{\Psi_j\}_{j\in\mathbb{N}}$ is said to be one-summable if

$$\sum_{j=0}^{\infty} j \cdot \|\Psi_j\| < +\infty.$$

This implies that $\{\Psi_j\}_{j\in\mathbb{N}}$ is absolutely summable, or that

$$\sum_{j=0}^{\infty} \|\Psi_j\| < +\infty,$$

and both these conditions imply square summability. In these cases,

$$\sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j}$$

can be viwed as both the L^2 and the almost sure limit of the corresponding partial sum process.

The following theorem allows us to decompose the partial sum process of a linear process into a pure trend component and a stationary component.

Theorem (The Beveridge-Nelson Decomposition)
Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional zero-mean linear process with underlying white noise process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ with covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ and one-summable filter $\{\Psi_j\}_{j\in\mathbb{N}}$. Then, defining

$$\alpha_j = -\sum_{h=j+1}^{\infty} \Psi_h$$

for any $j \in \mathbb{N}$, $\{\alpha_j\}_{j\in\mathbb{N}}$ is absolutely summable, and there exists an almost sure set $\Omega_0 \in \mathcal{H}$ such that, for any $t \in \mathbb{Z}$,

$$Y_t = \Psi(1) \cdot \varepsilon_t + \eta_t - \eta_{t-1}$$

on Ω_0, where $\Psi(1)$ is defined as

$$\Psi(1) = \sum_{j=0}^{\infty} \Psi_j \in \mathbb{R}^{n\times n}$$

and $\{\eta_t\}_{t\in\mathbb{Z}}$ is a zero-mean weakly stationary process such that

$$\eta_t = \alpha(L)\varepsilon_t = \sum_{j=0}^{\infty} \alpha_j \cdot \varepsilon_{t-j}$$

for any $t \in \mathbb{Z}$.

Consequently, for any $T > 0$,

$$\sum_{t=1}^{T} Y_t = \Psi(1) \cdot \sum_{t=1}^{T} \varepsilon_t + \eta_T - \eta_0$$

on Ω_0.

Proof) We first prove the claim that $\{\alpha_j\}_{j\in\mathbb{N}}$ is absolutely summable. Clearly, the process itself is well-defined because of the absolute summability of $\{\Psi_j\}_{j\in\mathbb{N}}$. Now note that, for any $m, k \in N_+$,

$$\sum_{j=0}^{\infty} \|\alpha_j\| \leq \sum_{j=0}^{\infty} \sum_{h=j+1}^{\infty} \|\Psi_h\| = \sum_{j=1}^{\infty} j \cdot \|\Psi_j\| < +\infty,$$

where the terms in the series can be rearranged due to the absolute summability of $\{\Psi_j\}_{j\in\mathbb{N}}$ and the last inequality follows from one-summability. Thsi shows that $\{\alpha_j\}_{j\in\mathbb{N}}$ is absolutely summable.

By the absolute summability of $\{\Psi_j\}_{j\in\mathbb{N}}$, Y_t is the almost sure limit of the sequence

$$\left\{\sum_{j=0}^{m} \Psi_j \cdot \varepsilon_{t-j}\right\}_{m\in N_+} = \{Y_{t,m}\}_{m\in N_+}$$

for any $t \in \mathbb{Z}$. Likewise, the absolute summability of $\{\alpha_j\}_{j\in\mathbb{N}}$ ensures that each η_t is the almost sure limit of the sequence

$$\left\{\sum_{j=0}^{m} \alpha_j \cdot \varepsilon_{t-j}\right\}_{m\in N_+} = \{\eta_{t,m}\}_{m\in N_+}.$$

$\mathbb{Z}$ is countable, so we can define the almost sure set $\Omega_0 \in \mathcal{H}$ on which every $\{Y_{t,m}\}_{m\in N_+}$ and $\{\eta_{t,m}\}_{m\in N_+}$ converges absolutely.
Now choose any $t \in \mathbb{Z}$. It follows that, for any $\omega \in \Omega_0$,

$$\begin{aligned}
Y_t(\omega) &= \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j}(\omega) \\
&= (\Psi(1) + \alpha_0) \cdot \varepsilon_t(\omega) + \sum_{j=1}^{\infty} (\alpha_j - \alpha_{j-1}) \cdot \varepsilon_{t-j}(\omega) \\
&= (\Psi(1) + \alpha_0) \cdot \varepsilon_t(\omega) + \sum_{j=1}^{\infty} (\alpha_j \cdot \varepsilon_{t-j}(\omega) - \alpha_{j-1} \cdot \varepsilon_{t-j}(\omega)) \\
&= \Psi(1) \cdot \varepsilon_t(\omega) + \alpha_0 \cdot \varepsilon_t(\omega) + \sum_{j=1}^{\infty} \alpha_j \cdot \varepsilon_{t-j}(\omega) - \sum_{j=0}^{\infty} \alpha_j \cdot \varepsilon_{t-j-1}(\omega) \\
&= \Psi(1) \cdot \varepsilon_t(\omega) + \eta_t(\omega) - \eta_{t-1}(\omega),
\end{aligned}$$

where the additive operations above all hold because the series involved are all absolutely convergent. Therefore, on Ω_0, for any $t \in \mathbb{Z}$ we have

$$Y_t = \Psi(1) \cdot \varepsilon_t + \eta_t - \eta_{t-1}.$$

It follows that, for any $T > 0$,

$$\sum_{t=1}^{T} Y_t = \sum_{t=1}^{T} (\Psi(1) \cdot \varepsilon_t + \eta_t - \eta_{t-1}) = \Psi(1) \cdot \sum_{t=1}^{T} \varepsilon_t + \eta_T - \eta_0$$

on Ω_0.
Q.E.D.

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a zero-mean linear process as the one above, and assume that the process $\{S_t\}_{t\in\mathbb{N}}$ is defined as

$$S_t = \sum_{s=1}^{t} Y_s + S_0$$

for any $t \in N_+$, so that

$$S_t = S_{t-1} + Y_t$$

for any $t \in N_+$. $\{S_t\}_{t\in\mathbb{N}}$ thus looks like a random walk process, but has potentially serially correlated errors.
Similarly, define the pure random walk process $\{\tau_t\}_{t\in\mathbb{N}}$ as $\tau_0 = 0$ and

$$\tau_t = \sum_{s=1}^{t} \varepsilon_s$$

for any $t \in N_+$, so that

$$\tau_t = \tau_{t-1} + \varepsilon_t$$

again, but this time with WN errors.
Then, the BN Decomposition of S_t is

$$\begin{aligned} S_t = \sum_{s=1}^{t} Y_s + S_0 &= \Psi(1) \cdot \sum_{s=1}^{t} \varepsilon_s + \eta_t - \eta_0 + S_0 \\ &= \Psi(1) \cdot \tau_t + \eta_t + (S_0 - \eta_0) \end{aligned}$$

almost surely for any $t \in N_+$. Thus, the BN decomposition allows us to decompose S_t into a trend component $\Psi(1) \cdot \tau_t$, a stationary component η_t, and a component consisting of initial values $S_0 - \eta_0$. This will come in handy later on when defining I(1) processes.

4.1.5 The CLT for Linear Processes

The BN Decomposition allows us to establish the CLT for zero mean linear processes when the underlying white noise process is an MDS with finite fourth moments.

Theorem (CLT for Linear Processes)
Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be an n-dimensional MDS with finite fourth moments and common positive definite covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$. Suppose $\{\Psi_j\}_{j\in\mathbb{N}}$ is a one-summable sequence of $n\times n$ matrices and define the linear process $\{Y_t\}_{t\in\mathbb{Z}}$ as

$$Y_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}.$$

for any $t \in \mathbb{Z}$. Then,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Y_t \xrightarrow{d} N\left[\mathbf{0}, \Psi(1)\Sigma\Psi(1)'\right],$$

where $\Psi(1) = \sum_{j=0}^{\infty} \Psi_j$.

Proof) Define $\{\alpha_j\}_{j\in\mathbb{N}} \subset \mathbb{R}^{n\times n}$ and $\{\eta_t\}_{t\in\mathbb{Z}}$ as in the BN decomposition. By that theorem, there exists an almost sure set $\Omega_0 \in \mathcal{H}$ such that, for any $T \in N_+$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Y_t = \Psi(1) \cdot \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t + \frac{1}{\sqrt{T}}(\eta_T - \eta_0)$$

on Ω_0.
Letting $\gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$ be the autocovariance function of the weakly stationary process $\{\eta_t\}_{t\in\mathbb{Z}}$,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{\sqrt{T}}(\eta_T - \eta_0)\right| > \delta\right) &\le \frac{1}{\delta^2}\mathbb{E}\left|\frac{1}{\sqrt{T}}(\eta_T - \eta_0)\right|^2 \\ &= \frac{1}{\delta^2 T}\mathbb{E}\left[\eta_T'\eta_T + \eta_0'\eta_0 - 2\eta_T'\eta_0\right] \\ &= \frac{2}{\delta^2 T}\left[\mathrm{tr}(\gamma(0)) - \mathrm{tr}(\gamma(T))\right]\end{aligned}$$

for any $T \in N_+$. Since $\gamma(T) \to O$ as $T \to \infty$ due to the absolute summability of the autocovariances $\gamma(\cdot)$ (which follows from the absolute summability of the coefficient matrices $\{\alpha_j\}_{j\in\mathbb{N}}$), taking $T \to \infty$ on both sides tells us that

$$\lim_{T\to\infty} \mathbb{P}\left(\left|\frac{1}{\sqrt{T}}(\eta_T - \eta_0)\right| > \delta\right) = 0.$$

This holds for any $\delta > 0$, so

$$\frac{1}{\sqrt{T}}(\eta_T - \eta_0) \xrightarrow{p} \mathbf{0}$$

by definition.
On the other hand,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t \xrightarrow{d} N(\mathbf{0},\Sigma)$$

by the Martingale Difference CLT. It follows from Slutsky's theorem that

$$\Psi(1)\cdot\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t+\frac{1}{\sqrt{T}}(\eta_T-\eta_0) \xrightarrow{d} N(\mathbf{0},\Psi(1)\Sigma\Psi(1)').$$

Finally,

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Y_t-\left(\Psi(1)\cdot\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t+\frac{1}{\sqrt{T}}(\eta_T-\eta_0)\right)\right|>\delta\right)\leq\mathbb{P}\left(\Omega^c\right)=0$$

for any $\delta>0$, so

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Y_t-\left(\Psi(1)\cdot\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t+\frac{1}{\sqrt{T}}(\eta_T-\eta_0)\right)\xrightarrow{p}\mathbf{0},$$

and by Slutsky's theorem once more,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Y_t \xrightarrow{d} N(\mathbf{0},\Psi(1)\Sigma\Psi(1)').$$

Q.E.D.

Here, the positive semidefinite covariance matrix

$$\Sigma_Y=\Psi(1)\Sigma\Psi(1)'$$

is called the long run variance. This is clearly different from the ordinary variance

$$\Gamma(0)=\sum_{j=0}^{\infty}\Psi_j\Sigma\Psi_j'$$

of the process $\{Y_t\}_{t\in\mathbb{Z}}$.

4.1.6 Extending the FCLT to Linear Processes

With the BN decomposition, we can extend the FCLT in a manner thata allows the underlying process to be a linear process. Since linear processes are generally autocorrelated, this means that the FCLT can be formulated even for partial sums of some autocorrelated processes. The formal statement is as follows:

Theorem (The Extended FCLT)
Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be an n-dimensional i.i.d. white noise process with positive definite covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ and finite fourth moments, and $\{\Psi_j\}_{j\in\mathbb{N}}$ a one-summable sequence of $n\times n$ matrices. Let $\{u_t\}_{t\in\mathbb{Z}}$ be the mean zero linear process defined as

$$u_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}$$

for any $t \in \mathbb{Z}$.

For any $T \in N_+$, define the n-dimensional stochastic process $\{X_T(r)\}_{r\in[0,1]}$ with continuous paths as

$$X_T(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} u_t + \frac{1}{\sqrt{T}}(Tr - \lfloor Tr \rfloor) u_{\lfloor Tr \rfloor + 1}$$

for any $r \in [0,1]$. Then, letting X^T be the random function in $\mathcal{C}([0,1],\mathbb{R}^n)$ corresponding to $\{X_T(r)\}_{r\in[0,1]}$,

$$X^T \xrightarrow{d} B^n$$

as $T \to \infty$, where B^n is the n-dimensional Wiener function with covariance matrix $\Sigma_u = \Psi(1)\Sigma\Psi(1)'$. By implication, for any $0 \le r_1 < \cdots < r_k \le 1$,

$$(X_T(r_1), \cdots, X_T(r_k)) \xrightarrow{d} (B^n(r_1), \cdots, B^n(r_k))$$

as $T \to \infty$.

Proof) By the Beveridge-Nelson decomposition, there exists an almost sure set $\Omega_0 \in \mathcal{H}$ such that, for any $T \in N_+$ and $r \in [0,1]$,

$$\sum_{t=1}^{\lfloor Tr \rfloor} u_t = \Psi(1) \cdot \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t + \eta_{\lfloor Tr \rfloor} - \eta_0$$

and

$$u_{\lfloor Tr \rfloor + 1} = \Psi(1) \cdot \varepsilon_{\lfloor Tr \rfloor + 1} + \eta_{\lfloor Tr \rfloor + 1} - \eta_{\lfloor Tr \rfloor}$$

on Ω_0, where

$$\alpha_j = -\sum_{h=j+1}^{\infty} \Psi_h$$

for any $j \in \mathbb{N}$ and $\{\eta_t\}_{t\in\mathbb{Z}}$ is the weakly stationary process defined as

$$\eta_t = \sum_{j=0}^{\infty} \alpha_j \cdot \varepsilon_{t-j}$$

for any $t \in \mathbb{Z}$. We now decompose each random function X^T in a convenient way.

Defining $\{V_T(r)\}_{r\in[0,1]}$ as

$$V_T(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t + \frac{1}{\sqrt{T}}(Tr - \lfloor Tr \rfloor)\varepsilon_{\lfloor Tr \rfloor + 1}$$

for any $r \in [0,1]$ and $\{A_T(r)\}_{r\in[0,1]}$ as

$$A_T(r) = \frac{1}{\sqrt{T}}\left(\eta_{\lfloor Tr \rfloor} - \eta_0\right) + \frac{1}{\sqrt{T}}(Tr - \lfloor Tr \rfloor)(\eta_{\lfloor Tr \rfloor + 1} - \eta_{\lfloor Tr \rfloor})$$

for any $r \in [0,1]$, both $\{V_T(r)\}_{r\in[0,1]}$ and $\{A_T(r)\}_{r\in[0,1]}$ have continuous paths, so that there exist random functions V^T and A^T in $\mathcal{C}([0,1],\mathbb{R}^n)$ corresponding to these processes.

We can see that, for any $T \in N_+$ and $r \in [0,1]$,

$$\begin{aligned}
X_T(r) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} u_t + \frac{1}{\sqrt{T}}(Tr - \lfloor Tr \rfloor)u_{\lfloor Tr \rfloor + 1} \\
&= \frac{1}{\sqrt{T}}\left(\Psi(1) \cdot \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t + \eta_{\lfloor Tr \rfloor} - \eta_0\right) + \frac{1}{\sqrt{T}}(Tr - \lfloor Tr \rfloor)\left(\Psi(1) \cdot \varepsilon_{\lfloor Tr \rfloor + 1} + \eta_{\lfloor Tr \rfloor + 1} - \eta_{\lfloor Tr \rfloor}\right) \\
&= \Psi(1) \cdot \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t + \frac{1}{\sqrt{T}}(Tr - \lfloor Tr \rfloor)\varepsilon_{\lfloor Tr \rfloor + 1}\right] \\
&\quad + \left[\frac{1}{\sqrt{T}}\left(\eta_{\lfloor Tr \rfloor} - \eta_0\right) + \frac{1}{\sqrt{T}}(Tr - \lfloor Tr \rfloor)(\eta_{\lfloor Tr \rfloor + 1} - \eta_{\lfloor Tr \rfloor})\right] \\
&= \Psi(1) \cdot V_T(r) + A_T(r)
\end{aligned}$$

on Ω_0. It follows that

$$X^T = \Psi(1) \cdot V^T + A^T$$

on Ω_0 for any $T \in N_+$. We study the limiting behavior of each term:

i) **The First Term** V^T

By the FCLT, we know that

$$V^T \xrightarrow{d} \Sigma^{\frac{1}{2}} W^n,$$

where W^n is the n-dimensional standard Wiener function on $[0,1]$ and $\Sigma^{\frac{1}{2}}$ is the Cholesky factor of Σ.

ii) **The Second Term** A^T

Meanwhile, for any $T \in N_+$ and $r \in [0,1]$,

$$|A_T(r)| \leq \frac{1}{\sqrt{T}} \left(2 \cdot \left|\eta_{\lfloor Tr \rfloor}\right| + \left|\eta_{\lfloor Tr \rfloor + 1}\right| + |\eta_0|\right)$$

because $Tr - \lfloor Tr \rfloor \leq 1$. We now have

$$\left\|A^T\right\|_{\mathcal{C}} = \sup_{r \in [0,1]} |A_T(r)| \leq \frac{1}{\sqrt{T}} 4 \cdot \max_{0 \leq t \leq T+1} |\eta_t|$$

and as such, for any $\delta > 0$,

$$\mathbb{P}\left(\left\|A^T\right\|_{\mathcal{C}} > \delta\right) \leq \mathbb{P}\left(\frac{1}{\sqrt{T}} \max_{0 \leq t \leq T+1} |\eta_t| > \frac{\delta}{4}\right)$$

Note that

$$\left\{\frac{1}{\sqrt{T}} \max_{0 \leq t \leq T+1} |\eta_t| > \frac{\delta}{4}\right\} = \bigcup_{t=0}^{T+1} \left\{\frac{1}{\sqrt{T}} |\eta_t| > \frac{\delta}{4}\right\},$$

so that, by finite suabdditivity and the generalized Markov inequality,

$$\begin{aligned} \mathbb{P}\left(\frac{1}{\sqrt{T}} \max_{0 \leq t \leq T+1} |\eta_t| > \frac{\delta}{4}\right) &\leq \sum_{t=0}^{T+1} \mathbb{P}\left(\frac{1}{\sqrt{T}} |\eta_t| > \frac{\delta}{4}\right) \\ &\leq \left(\frac{\delta}{4}\right)^{-4} \frac{1}{T^2} \sum_{t=0}^{T+1} \mathbb{E}\left[|\eta_t|^4\right]. \end{aligned}$$

Since $\{\eta_t\}_{t \in \mathbb{Z}}$ has finite fourth moments due to the finiteness of the fourth moments of the underyling WN process $\{\varepsilon_t\}_{t \in \mathbb{Z}}$, letting

$$\mathbb{E}|\eta_t|^4 = \mu_4 < +\infty$$

for any $t \in \mathbb{Z}$, we can see that

$$\begin{aligned} \mathbb{P}\left(\left\|A^T\right\|_{\mathcal{C}} > \delta\right) &\leq \left(\frac{\delta}{4}\right)^{-4} \frac{1}{T^2} \sum_{t=0}^{T+1} \mathbb{E}\left[|\eta_t|^4\right] \\ &\leq \mu_4 \cdot \left(\frac{\delta}{4}\right)^{-4} \frac{T+2}{T^2}. \end{aligned}$$

Therefore,

$$\lim_{T\to\infty} \mathbb{P}\left(\left\|A^T\right\|_{\mathcal{C}} > \delta\right) = 0,$$

and because this holds for any $\delta > 0$,

$$A^T \xrightarrow{p} \mathbf{0}_{\mathcal{C}}$$

as $T \to \infty$, where $\mathbf{0}_{\mathcal{C}}$ is the zero function on $[0,1]$.

By Slutsky's theorem, it now follows that

$$\Psi(1) \cdot V^T + A^T \xrightarrow{d} \Psi(1)\Sigma^{\frac{1}{2}} W^n,$$

and because $\Omega_0 \subset \{X^T = \Psi(1) \cdot V^T + A^T\}$,

$$\mathbb{P}\left(\left\|X^T - (\Psi(1) \cdot V^T + A^T)\right\|_{\mathcal{C}} > \delta\right) \leq \mathbb{P}(\Omega_0^c) = 0$$

for any $\delta > 0$; this trivially implies that

$$X^T - (\Psi(1) \cdot V^T + A^T) \xrightarrow{p} \mathbf{0}_{\mathcal{C}},$$

and by Slutsky's theorem again,

$$X^T \xrightarrow{d} \Psi(1)\Sigma^{\frac{1}{2}} W^n.$$

Here, $\Psi(1)\Sigma^{\frac{1}{2}} W^n$ is an n-dimensional Wiener function with covariance matrix $\Psi(1)\Sigma\Psi(1)' = \Sigma_u$, which is the result we desired.
Q.E.D.

4.2 The Limit of Functions of Trending Processes

In this section, we use the results of the preceding section to derive asymptotic results pertaining to trending processes.

4.2.1 Continuous Functions on $\mathcal{C}([0,1],\mathbb{R}^n)$

To set the stage, we first note that the following are continuous functions:

- **The Integral of a Continuous Function**
  Define the function $g : \mathcal{C}([0,1],\mathbb{R}^n) \to \mathbb{R}^n$ as

$$g(f) = \int_0^1 f(r) d\mu(r)$$

for any $f \in \mathcal{C}([0,1],\mathbb{R}^n)$, where μ is any finite measure on $[0,1]$; μ is most often taken to be the Lebesgue measure on $[0,1]$. The integral is well-defined because f, being continuous, is measurable and it is also bounded by the extreme value theorem, which, in light of the finiteness of μ, means that f is μ-integrable. Thus, the usual integral arithmetic operatinos apply.
For any $f, h \in \mathcal{C}([0,1],\mathbb{R}^n)$,

$$\begin{aligned}|g(f) - g(h)| = \left| \int_0^1 (f(r) - h(r)) d\mu(r) \right| &\leq \int_0^1 |f(r) - h(r)| d\mu(r) \\ &\leq \sup_{r \in [0,1]} |f(r) - h(r)| \cdot \mu([0,1]) = \mu([0,1]) \cdot \|f - h\|_{\mathcal{C}},\end{aligned}$$

and because $\mu([0,1]) < +\infty$, this shows us that g is Lipschitz continuous on $\mathcal{C}([0,1],\mathbb{R}^n)$.

When μ is the Lebesgue measure on $[0,1]$, we can discuss a stronger form of continuity. Consider the product space $[0,1] \times \mathcal{C}([0,1],\mathbb{R}^n)$ given the product metric ρ of the euclidean metric on $\mathbb{R}$ and the supremum metric on $\mathcal{C}([0,1],\mathbb{R}^n)$, which is defined as

$$\rho((r,f),(s,g)) = \max(|r-s|, \|f-g\|_{\mathcal{C}})$$

for any $(r,f),(s,g) \in [0,1] \times \mathcal{C}([0,1],\mathbb{R}^n)$.
Define $G : [0,1] \times \mathcal{C}([0,1],\mathbb{R}^n) \to \mathbb{R}^n$ as

$$G(r,f) = \int_0^r f(x) dx$$

for any $(r,f) \in [0,1] \times \mathcal{C}([0,1],\mathbb{R}^n)$. To see that G is continuous, choose any $\epsilon > 0$ and $(r,f) \in [0,1] \times \mathcal{C}([0,1],\mathbb{R}^n)$. For any $(s,g) \in \mathcal{C}([0,1],\mathbb{R}^n)$ such that

$$\|f-g\|_{\mathcal{C}}, |r-s| \leq \rho((r,f),(s,g)) < \frac{\varepsilon}{2 \cdot \max(\|f\|_{\mathcal{C}}, 1)},$$

we have

$$
\begin{aligned}
|G(r,f)-G(s,g)| &= \left|\int_0^r f(x)dx - \int_0^s g(x)dx\right| \\
&\le \left|\int_0^r f(x)dx - \int_0^s f(x)dx\right| + \left|\int_0^s f(x)dx - \int_0^s g(x)dx\right| \\
&\le \int_{\min(r,s)}^{\max(r,s)} |f(x)|dx + \int_0^s |f(x)-g(x)|dx \\
&\le \|f\|_{\mathcal{C}} \cdot |r-s| + \|f-g\| \cdot s \\
&\le \|f\|_{\mathcal{C}} \cdot |r-s| + \|f-g\| \\
&< \|f\|_{\mathcal{C}} \cdot \frac{\epsilon}{2\cdot \|f\|_{\mathcal{C}}} + \frac{\epsilon}{2} \le \epsilon.
\end{aligned}
$$

By definition, G is continuous at (r,f), and because this point was chosen arbitrarily, G is continuous on the entire product space.
Since the pair of any two measurable random variables on metric spaces is also measurable with respect to the Borel σ-algebra associated with the product metric on the product space, this result is sufficient for us to apply the continuous mapping theorem.

- **More Integrals**

While we only considered the integral of a function itself in the above discussion, we can in fact define a wider variety of functions from $\mathcal{C}([0,1],\mathbb{R}^n)$ into arbitrary metric spaces defined via integration.

For any finite measure μ on $[0,1]$, define $g_2 : \mathcal{C}([0,1],\mathbb{R}^n) \to \mathbb{R}^{n\times n}$ as

$$
g_2(f) = \int_0^1 f(r)f(r)'d\mu(r)
$$

for any $f \in \mathcal{C}([0,1],\mathbb{R}^n)$. The integral is once again well defined because of the boundedness and continuity of f, as well as the finiteness of μ.
To see that g_2 is continuous, choose any $f,h \in \mathcal{C}([0,1],\mathbb{R}^n)$ and note that

$$
\begin{aligned}
\|g_2(f)-g_2(h)\| &\le \sum_{i=1}^n\sum_{j=1}^n \left|\int_0^1 \left(f_i(r)f_j(r) - h_i(r)h_j(r)\right)d\mu(r)\right| \\
&\le \sum_{i=1}^n\sum_{j=1}^n \int_0^1 |f_i(r)f_j(r) - h_i(r)h_j(r)|d\mu(r) \\
&\le \mu([0,1]) \cdot \sum_{i=1}^n\sum_{j=1}^n \|f_if_j - h_ih_j\|_{\mathcal{C}}.
\end{aligned}
$$

Each $f_if_j - h_ih_j$ is bounded above by linear combinations of $\|f\|_{\mathcal{C}}$ and $\|f-h\|_{\mathcal{C}}$, so g_2 is continuous on $\mathcal{C}([0,1],\mathbb{R}^n)$.

Now define $\bar{g} : \mathcal{C}([0,1],\mathbb{R}^n) \to \mathbb{R}^n$ as

$$\bar{g}(f) = \int_0^1 r \cdot f(r) d\mu(r)$$

for any $f \in \mathcal{C}([0,1],\mathbb{R}^n)$. Then, for any $f,h \in \mathcal{C}([0,1],\mathbb{R}^n)$,

$$\begin{aligned} |\bar{g}(f) - \bar{g}(h)| &= \left| \int_0^1 r \cdot (f(r) - g(r)) d\mu(r) \right| \leq \int_0^1 r \cdot |f(r) - g(r)| d\mu(r) \\ &\leq \|f - g\|_{\mathcal{C}} \cdot \int_0^1 r d\mu(r) \leq \|f - g\|_{\mathcal{C}} \cdot \mu([0,1]), \end{aligned}$$

so $\bar{g}$ is Lipschitz continuous on $\mathcal{C}([0,1],\mathbb{R}^n)$.

- **Joint Continuity of Projections**
  For any $0 \leq r_1 < \cdots < r_k \leq 1$, we saw that the projection $\pi_{r_1,\cdots,r_k}$ is uniformly continuous on $\mathcal{C}([0,1],\mathbb{R}^n)$.
  Choosing any $r \in [0,1]$, we can also view the projection $\pi_r(f)$ of $f \in \mathcal{C}([0,1],\mathbb{R}^n)$ as a function of both r and f. Reflecting this change in perspective, define $\pi : [0,1] \times \mathcal{C}([0,1],\mathbb{R}^n) \to \mathbb{R}^n$ as

$$\pi(r,f) = \pi_r(f)$$

  for any $(r,f) \in [0,1] \times \mathcal{C}([0,1],\mathbb{R}^n)$. Define the product metric ρ as above. We can now show that π is continuous.

  To this end, chose any $(r,f) \in [0,1] \times \mathcal{C}([0,1],\mathbb{R}^n)$. Then, for any $\epsilon > 0$, by the uniform continuity of f there exists a $\delta_1 > 0$ such that

$$|f(x) - f(y)| < \frac{\epsilon}{2}$$

  for any $x,y \in [0,1]$ such that $|x - y| < \delta_1$. Now define

$$\delta = \min\left(\delta_1, \frac{\epsilon}{2}\right) > 0.$$

  It follows that, for any $(s,g) \in [0,1] \times \mathcal{C}([0,1],\mathbb{R}^n)$ such that

$$|r - s|, \|f - g\|_{\mathcal{C}} \leq \rho((r,f),(s,g)) < \delta,$$

  since $\|f - g\|_{\mathcal{C}} < \frac{\epsilon}{2}$ and $|r - s| < \delta_1$ implies

$$|f(r) - f(s)| < \frac{\epsilon}{2},$$

we have

$$
\begin{aligned}
|\pi(r,f)-\pi(s,g)| = |f(r)-g(s)| &\leq |f(r)-f(s)|+|f(s)-g(s)| \\
&< \frac{\epsilon}{2}+\|f-g\|_{\mathcal{C}} \leq \frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon.
\end{aligned}
$$

By definition, π is continuous. Since the pair of any two measurable random variables on metric spaces is also measurable with respect to the Borel σ-algebra associated with the product metric on the product space, this result is sufficient for us to apply the continuous mapping theorem.

4.2.2 Convergence to Stochastic Integrals

So far, given an n-dimensional linear process $\{u_t\}_{t\in\mathbb{Z}}$, our main convergence result has been that of the random function X_T, defined as the random function corresponding to the continuous stochastic process $\{X_T(r)\}_{r\in[0,1]}$ defined as

$$X_T(r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor} u_t + \frac{1}{\sqrt{T}}(Tr - \lfloor Tr\rfloor)u_{\lfloor Tr\rfloor+1}$$

for any $r \in [0,1]$. As we will soon see, however, we require not only the convergence of X_T to an n-dimensional Brownian motion, but the convergence of a partial sum process to a stochastic integral as well. Furthermore, this convergence should hold jointly for X_T and the partial sum process. We thus show here that this sort of joint convergence holds true.

First, we state some of the results to be used to derive this result. The two main ones are as follows:

Theorem (Skorokhod's Reprsentation Theorem)
Let (E,d) be a complete and separable metric space, τ the metric topology induced by d, and $\mathcal{E}$ the Borel σ-algebra on E generated by τ. Let $\{\mu_T\}_{T\in N_+}$ be a sequence of probability measures on $(E,\mathcal{E})$ weakly converging to a probability measure μ on $(E,\mathcal{E})$.
There exists a probability space and random variables $\{X_T\}_{T\in N_+}$, X defined on that probability space, take values in $(E,\mathcal{E})$, and satisfy the following:

i) X_T has distribution μ_T for any $T \in N_+$ and X has distribution μ.

ii) X_T converges almost surely to X.

The above theorem allows us to move back and forth between the weak convergence of probability measures and the almost sure convergence of specific random variables.

We also require the following result, which we state without proof (refer to the text on the convergence theory for a proof):

Theorem (Egorov's Theorem)
Let (E,d) be a separable metric space, τ the metric topology induced by d and $\mathcal{E}$ the Borel σ-algebra on E generated by τ. Suppose $\{X_T\}_{T\in N_+}$ is a sequence of random variables on $(E,\mathcal{E})$ that converges almost surely to the random variable X.
Then, for any $\varepsilon > 0$ there exists a measurable set $\Omega_0 \in \mathcal{H}$ such that $\mathbb{P}(\Omega_0) < \varepsilon$ and $\{X_T\}_{T\in N_+}$ converges to X uniformly on $\Omega\setminus\Omega_0$.

This theorem furnishes sufficient conditions for pointwise convergence of a sequence of random variables taking values in a separable metric space to converge uniformly. It turns out that the sequence converges uniformly except on a set whose measure can be made arbitrarily small.

Both of the above theorems are proved in a separate text exclusively concerning the convergence of random variables and probability measures.

We are now ready to prove our joint convergence result. The proof below is based on Chen and Wei (1988), where we have imposed stronger assumptions to make use of convergence results in $\mathcal{C}([0,1],\mathbb{R}^n)$.

Theorem (Joint Convergence to Stochastic Integrals for IID Processes)
Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be an n-dimensional i.i.d. process with mean 0, positive definite covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ and finite fourth moments. In addition, define $\{V_t\}_{t\in\mathbb{N}}$ as the partial sum process of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ such that $V_0 = 0$ and

$$V_t = \sum_{s=1}^{t} \varepsilon_s + V_0$$

for any $t \in N_+$. For any $T \in N_+$, define the stochastic process $\{X_T(r)\}_{r\in[0,1]}$ as

$$X_T(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t + \frac{1}{\sqrt{T}}(Tr - \lfloor Tr \rfloor)\varepsilon_{\lfloor Tr \rfloor+1}$$

for any $r \in [0,1]$, and let X^T be the random function taking values in $\mathcal{C}([0,1],\mathbb{R}^n)$ corresponding to $\{X_T(r)\}_{r\in[0,1]}$. Defining

$$\mathcal{V}_T^h = \frac{1}{T} \sum_{t=1}^{T-h} V_{t-1}\varepsilon_t'$$

for any $h \in \mathbb{N}$, for any $p \in N_+$, X^T and $\mathcal{V}_T^0, \cdots, \mathcal{V}_T^p$ jointly converge in distribution to their limits:

$$X^T \overset{d}{\to} \Sigma^{\frac{1}{2}} W^n$$

$$(\mathcal{V}_T^0, \cdots, \mathcal{V}_T^p) \overset{d}{\to} \iota_{p+1}' \bigotimes \left[\Sigma^{\frac{1}{2}} \int_0^1 W^n(r) dW^n(r)' \Sigma^{\frac{1}{2}\prime} \right]$$

jointly, where $\{W^n(r)\}_{r\in[0,1]}$ is the n-dimensional Wiener process, where ι_{p+1} is a $p+1$-dimensional vector of ones.

Proof) To avoid confusion, we denote by $\mathcal{W}^n$ the standard n-dimensional Wiener function.

We know from the FCLT that $X^T \overset{d}{\to} \Sigma^{\frac{1}{2}} \mathcal{W}^n$. Because $(\mathcal{C}([0,1],\mathbb{R}^n), d)$, where d is the supremum metric, is a complete and separable metric space, by Skorokhod's representation theorem there exists a probability space $(\Omega_0, \mathcal{H}_0, \mathbb{P}_0)$ and random functions $\{Y^T\}_{T\in N_+}$, $\mathcal{W}_0^n$ on Ω_0 taking values in the measurable space $(\mathcal{C}([0,1],\mathbb{R}^n), \mathcal{B}_{\mathcal{C}}([0,1],\mathbb{R}^n))$ such that

– $Y^T \sim X^T$ for any $T \in N_+$,

– W_0^n is an n-dimensional standard Wiener function, and

– $Y^T \overset{a.s.}{\to} \Sigma^{\frac{1}{2}} \mathcal{W}_0^n$.

as $T \to \infty$. For any $T \in N_+$, let $\{Y_T(r)\}_{r\in[0,1]}$ be the stochastic process corresponding to the random function Y^T.

Now we fix some $0 \le h \le p$. Note that

$$\frac{1}{T}\sum_{t=1}^{T-h} V_{t-1}\varepsilon_t' = \sum_{t=0}^{T-h-1} X_T\left(\frac{t}{T}\right)\left[X_T\left(\frac{t+1}{T}\right) - X_T\left(\frac{t}{T}\right)\right]',$$

a function of X_T. By implication, defining

$$Z_T^h = \sum_{t=0}^{T-h-1} Y_T\left(\frac{t}{T}\right)\left[Y_T\left(\frac{t+1}{T}\right) - Y_T\left(\frac{t}{T}\right)\right]',$$

we have

$$(Y^T, Z_T) \sim \left(X^T, \frac{1}{T}\sum_{t=1}^{T-h} V_{t-1}\varepsilon_t'\right).$$

We will now show that (Y^T, Z_T^h) converges in probability to the limit in the claim of the theorem.
For notational simplicity, we denote $\mathcal{W}_0^n$ by $\mathcal{W}^n$, and the expectation with respect to $\mathbb{P}_0$ by $\mathbb{E}[\cdot]$.

The separability of $(\mathcal{C}([0,1],\mathbb{R}^n), d)$ allows us to use Egorov's theorem, which tells us that, for any $\epsilon > 0$, there exists a set $\Omega_\epsilon \in \mathcal{H}_0$ such that

$$\mathbb{P}_0(\Omega_\epsilon) \ge 1-\epsilon, \quad \text{and}$$
$$\delta_T = \sup_{\omega\in\Omega_\epsilon} \left\| Y^T(\omega) - \Sigma^{\frac{1}{2}}\mathcal{W}^n(\omega)\right\|_{\mathcal{C}} \to 0 \quad \text{as } T\to\infty.$$

We can choose a subsequence $\{S_T\}_{T\in N_+}$ of N_+ such that $S_T \le T-h-1$ for any $T \ge h$ and

$$S_T\delta_T \to 0 \quad \text{and} \quad \frac{S_T}{T} \to 0 \quad \text{as } T\to\infty.$$

In other words, S_T must converge to $+\infty$ at a slower rate than δ_T and $\frac{1}{T}$ converge to 0. One possible choice for S_T could be

$$S_T = \lfloor \min\left(\frac{1}{\sqrt{\delta_T}}, \sqrt{T-h-1}\right)\rfloor.$$

Once $\{S_T\}_{T\in N_+}$ has been established, for any $T \in N_+$ choose a partition $\left\{\frac{t_0}{T}, \cdots, \frac{t_{S_T}}{T}\right\}$

of $\left[0, \frac{T-h-1}{T}\right]$, where $t_0 = 0$, $t_{S_T} = T-h-1$ and $t_1, \cdots, t_{S_T-1} \in N_+$ are chosen so that

$$\lim_{T\to\infty} \frac{1}{T}\left(\max_{1\leq i\leq S_T} |t_i - t_{i-1}|\right) = 0.$$

One possible choice is to set

$$t_i = \lfloor \frac{T-h-2}{S_T-1} \rfloor \cdot i \quad \forall 1 \leq i \leq S_T - 1.$$

The natural numbers $t_0, \cdots, t_{S_T}$ represent a non-overlapping partition of the interval $[0, T-h-1]$, since $t_i - t_{i-1} > 0$ for any $1 \leq i \leq S_T - 1$ for large enough T (this is the case because $\frac{S_T}{T} \to 0$ as $T \to \infty$) and

$$t_{S_T-1} = \lfloor \frac{T-h-2}{S_T-1} \rfloor \cdot (S_T - 1) \leq \frac{T-h-2}{S_T-1} \cdot (S_T-1) = T-h-2 < T-h-1,$$

so that $t_{S_T} - t_{S_T-1} > 0$. Finally,

$$\frac{1}{T}\lfloor \frac{T-h-2}{S_T-1} \rfloor \leq \frac{T-h-1}{T} \cdot \frac{1}{S_T-1} \to 0$$

as $T \to \infty$ because $S_T \to +\infty$ as $T \to \infty$ and, letting $m = \lfloor \frac{T-h-1}{S_T-1} \rfloor$,

$$\frac{T-h-1-m\cdot(S_T-1)}{T} < \frac{S_T}{T} \to 0$$

as $T \to \infty$, where the inequality is justified because

$$m \leq \frac{T-h-2}{S_T-1} < m+1$$

and thus

$$T-h-1-m\cdot(S_T-1) < T-h-1+(S_T-1) - \frac{T-h-2}{S_T-1} \cdot (S_T-1) = S_T$$

Therefore,

$$\lim_{T\to\infty} \frac{1}{T}\left(\max_{1\leq i\leq S_T} |t_i - t_{i-1}|\right) = 0$$

under our choice of $t_0, \cdots, t_{S_T}$.

We now proceed in steps:

1) **Step 1:**
We first show that

$$Z_T^h = \sum_{i=1}^{S_T} Y_T\left(\frac{t_{i-1}}{T}\right)\left[Y_T\left(\frac{t_i}{T}\right) - Y_T\left(\frac{t_{i-1}}{T}\right)\right]' + o_p(1).$$

To this end, note that, because $\{S_T\}_{T\in N_+}$ is a subsequence of N_+,

$$\begin{aligned} J_T &= Z_T^h - \sum_{i=1}^{S_T} Y_T\left(\frac{t_{i-1}}{T}\right)\left[Y_T\left(\frac{t_i}{T}\right) - Y_T\left(\frac{t_{i-1}}{T}\right)\right]' \\ &= \sum_{i=1}^{S_T}\sum_{t=t_{i-1}}^{t_i-1}\left[Y_T\left(\frac{t}{T}\right) - Y_T\left(\frac{t_{i-1}}{T}\right)\right]\left[Y_T\left(\frac{t+1}{T}\right) - Y_T\left(\frac{t}{T}\right)\right]'. \end{aligned}$$

Since $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is an i.i.d. sequence, and

$$\begin{aligned} Y_T\left(\frac{t}{T}\right) - Y_T\left(\frac{t_{i-1}}{T}\right) &\sim \frac{1}{\sqrt{T}}\sum_{s=t_{i-1}+1}^{t}\varepsilon_s \\ Y_T\left(\frac{t+1}{T}\right) - Y_T\left(\frac{t}{T}\right) &\sim \frac{1}{\sqrt{T}}\varepsilon_{t+1}, \end{aligned}$$

for any $1 \le i \le S_T$ and $t_{i-1} \le t$, we can see that, for any $1 \le j,k \le n$,

$$\begin{aligned} \mathbb{E}|J_{T,jk}|^2 &= \sum_{i=1}^{S_T}\sum_{t=t_{i-1}}^{t_i-1}\mathbb{E}\left[\left(Y_{jT}\left(\frac{t}{T}\right) - Y_{jT}\left(\frac{t_{i-1}}{T}\right)\right)^2\right]\cdot\mathbb{E}\left[\left(Y_{kT}\left(\frac{t+1}{T}\right) - Y_{kT}\left(\frac{t}{T}\right)\right)^2\right] \\ &= \Sigma_{jj}\Sigma_{kk}\frac{1}{T^2}\sum_{i=1}^{S_T}\sum_{t=t_{i-1}}^{t_i-1}(t - t_{i-1}) \\ &= \Sigma_{jj}\Sigma_{kk}\frac{1}{T^2}\sum_{i=1}^{S_T}\frac{(t_i - 1 - t_{i-1})(t_i - t_{i-1})}{2} \\ &\le \frac{1}{2}\Sigma_{jj}\Sigma_{kk}\frac{1}{T^2}\sum_{i=1}^{S_T}(t_i - t_{i-1})^2 \\ &\le \frac{1}{2}\Sigma_{jj}\Sigma_{kk}\left(\frac{1}{T}\sum_{i=1}^{S_T}(t_i - t_{i-1})\right)\cdot\left(\max_{1\le i\le S_T}\left|\frac{t_i}{T} - \frac{t_{i-1}}{T}\right|\right) \\ &= \frac{1}{2}\Sigma_{jj}\Sigma_{kk}\cdot\left(\max_{1\le i\le S_T}\left|\frac{t_i}{T} - \frac{t_{i-1}}{T}\right|\right), \end{aligned}$$

where the first equality follows from the independence of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, and the second from the fact that

$$\mathbb{E}\left[\left(Y_{jT}\left(\frac{t}{T}\right) - Y_{jT}\left(\frac{t_{i-1}}{T}\right)\right)^2\right] = \frac{1}{T}\sum_{s=t_{i-1}+1}^{t}\mathbb{E}\left[\varepsilon_{j,s}^2\right] = \frac{(t - t_{i-1})\Sigma_{jj}}{T} \quad \text{and}$$

$$\mathbb{E}\left[\left(Y_{kT}\left(\frac{t+1}{T}\right) - Y_{kT}\left(\frac{t}{T}\right)\right)^2\right] = \frac{1}{T}\mathbb{E}\left[\varepsilon_{k,t+1}^2\right] = \frac{1}{T}\Sigma_{kk}.$$

By assumption, the right hand side of the above chain of inequalities goes to 0 as $T \to \infty$, and because

$$\mathbb{E}\|J_T\|^2 \leq \sum_{j=1}^{n}\sum_{k=1}^{n}\mathbb{E}|J_{T,jk}|^2,$$

we have $J_T \xrightarrow{p} O$ as $T \to \infty$. Therefore,

$$Z_T^h = \sum_{i=1}^{S_T} Y_T\left(\frac{t_{i-1}}{T}\right)\left[Y_T\left(\frac{t_i}{T}\right) - Y_T\left(\frac{t_{i-1}}{T}\right)\right]' + o_p(1),$$

where the $o_p(1)$ term represents J_T.

2) **Step 2:**

Now we show that

$$Z_T^h \cdot I_{\Omega_\epsilon} = I_{\Omega_\epsilon} \cdot \Sigma^{\frac{1}{2}} \sum_{i=1}^{S_T} W^n\left(\frac{t_{i-1}}{T}\right)\left[Y_T\left(\frac{t_i}{T}\right) - Y_T\left(\frac{t_{i-1}}{T}\right)\right]' + o_p(1).$$

The above result shows us that we need only prove

$$D_T = I_{\Omega_\epsilon} \cdot \sum_{i=1}^{S_T}\left[Y_T\left(\frac{t_{i-1}}{T}\right) - \Sigma^{\frac{1}{2}} W^n\left(\frac{t_{i-1}}{T}\right)\right]\left[Y_T\left(\frac{t_i}{T}\right) - Y_T\left(\frac{t_{i-1}}{T}\right)\right]' = o_p(1).$$

Since

$$\sup_{r\in[0,1]}\left|(Y_T(r))(\omega) - \Sigma^{\frac{1}{2}}(W^n(r))(\omega)\right| \leq \sup_{\omega'\in\Omega_\epsilon}\left\|Y^T(\omega') - \Sigma^{\frac{1}{2}}\mathcal{W}^n(\omega')\right\|_{\mathcal{C}} = \delta_T$$

for any $\omega \in \Omega_\epsilon$, for any $1 \leq j,k \leq n$ we have

$$\begin{aligned}
|D_{T,jk}|^2 &= I_{\Omega_\epsilon} \cdot \left[\sum_{i=1}^{S_T}\left(Y_{jT}\left(\frac{t_{i-1}}{T}\right) - \Sigma^{\frac{1}{2}} W_j\left(\frac{t_{i-1}}{T}\right)\right)\left(Y_{kT}\left(\frac{t_i}{T}\right) - Y_{kT}\left(\frac{t_{i-1}}{T}\right)\right)\right]^2 \\
&\leq I_{\Omega_\epsilon} \cdot \delta_T^2 \cdot \left[\sum_{i=1}^{S_T}\left(Y_{kT}\left(\frac{t_i}{T}\right) - Y_{kT}\left(\frac{t_{i-1}}{T}\right)\right)\right]^2.
\end{aligned}$$

By the independence of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, we once again have

$$\begin{aligned}
\mathbb{E}|D_{T,jk}|^2 &\leq I_{\Omega_\epsilon} \cdot \delta_T^2 \cdot \sum_{i=1}^{S_T}\mathbb{E}\left[\left(Y_{kT}\left(\frac{t_i}{T}\right) - Y_{kT}\left(\frac{t_{i-1}}{T}\right)\right)^2\right] \\
&= I_{\Omega_\epsilon} \cdot \Sigma_{kk}\delta_T^2 \frac{1}{T}\sum_{i=1}^{S_T}(t_i - t_{i-1}) \\
&= I_{\Omega_\epsilon} \cdot \Sigma_{kk}\delta_T^2.
\end{aligned}$$

The term on the right hand side converges to 0 as $T \to \infty$ by design, and since this holds for any $1 \leq j,k \leq n$ and

$$\mathbb{E}\|D_T\|^2 \leq \sum_{j=1}^{n}\sum_{k=1}^{n}\mathbb{E}|D_{T,jk}|^2$$

we can see that $D_T \xrightarrow{p} O$.

3) **Step 3:**
We now replace the remaining Y^T terms with $\Sigma^{\frac{1}{2}}W^n$; specifically, we want to show that

$$Z_T^h \cdot I_{\Omega_\epsilon} = I_{\Omega_\epsilon} \cdot \Sigma^{\frac{1}{2}} \sum_{i=1}^{S_T} W^n\left(\frac{t_{i-1}}{T}\right)\left[W^n\left(\frac{t_i}{T}\right) - W^n\left(\frac{t_{i-1}}{T}\right)\right]' \Sigma^{\frac{1}{2}\prime} + o_p(1).$$

To this end, we utilize the following summation by parts formula:

$$\Sigma^{\frac{1}{2}} \sum_{i=1}^{S_T} W^n\left(\frac{t_{i-1}}{T}\right)\left[Y_T\left(\frac{t_i}{T}\right) - Y_T\left(\frac{t_{i-1}}{T}\right)\right]' + \Sigma^{\frac{1}{2}} \sum_{i=1}^{S_T}\left[W^n\left(\frac{t_i}{T}\right) - W^n\left(\frac{t_{i-1}}{T}\right)\right] Y_T\left(\frac{t_i}{T}\right)' = \Sigma^{\frac{1}{2}} W^n(1) Y_T(1)'.$$

Following the same process as in step 2 reveals that

$$I_{\Omega_\epsilon} \cdot \Sigma^{\frac{1}{2}} \sum_{i=1}^{S_T}\left[W^n\left(\frac{t_i}{T}\right) - W^n\left(\frac{t_{i-1}}{T}\right)\right] Y_T\left(\frac{t_i}{T}\right)' = I_{\Omega_\epsilon} \cdot \Sigma^{\frac{1}{2}} \sum_{i=1}^{S_T}\left[W^n\left(\frac{t_i}{T}\right) - W^n\left(\frac{t_{i-1}}{T}\right)\right] W^n\left(\frac{t_i}{T}\right)' \Sigma^{\frac{1}{2}\prime} + o_p(1).$$

In addition, $Y_T(1) \xrightarrow{d} \Sigma^{\frac{1}{2}} W^n(1)$ implies that $Y_T(1) - \Sigma^{\frac{1}{2}} W^n(1) = o_p(1)$, so

$$I_{\Omega_\epsilon} \cdot \Sigma^{\frac{1}{2}} \sum_{i=1}^{S_T} W^n\left(\frac{t_{i-1}}{T}\right)\left[Y_T\left(\frac{t_i}{T}\right) - Y_T\left(\frac{t_{i-1}}{T}\right)\right]' = I_{\Omega_\epsilon} \cdot \Sigma^{\frac{1}{2}} W^n(1) W^n(1)' \Sigma^{\frac{1}{2}\prime} - I_{\Omega_\epsilon} \cdot \Sigma^{\frac{1}{2}} \sum_{i=1}^{S_T}\left[W^n\left(\frac{t_i}{T}\right) - W^n\left(\frac{t_{i-1}}{T}\right)\right] W^n\left(\frac{t_i}{T}\right)' \Sigma^{\frac{1}{2}\prime} + o_p(1).$$

The same summation by parts formula reveals that

$$\Sigma^{\frac{1}{2}} W^n(1) W^n(1)' \Sigma^{\frac{1}{2}\prime} - \Sigma^{\frac{1}{2}} \sum_{i=1}^{S_T}\left[W^n\left(\frac{t_i}{T}\right) - \left(\frac{t_{i-1}}{T}\right)\right] W^n\left(\frac{t_i}{T}\right)' \Sigma^{\frac{1}{2}\prime} = \Sigma^{\frac{1}{2}} \sum_{i=1}^{S_T} W^n\left(\frac{t_{i-1}}{T}\right)\left[W^n\left(\frac{t_i}{T}\right) - W^n\left(\frac{t_{i-1}}{T}\right)\right]' \Sigma^{\frac{1}{2}\prime},$$

so we have

$$I_{\Omega_\epsilon}\cdot\Sigma^{\frac{1}{2}}\sum_{i=1}^{S_T}W^n\left(\frac{t_{i-1}}{T}\right)\left[Y_T\left(\frac{t_i}{T}\right)-Y_T\left(\frac{t_{i-1}}{T}\right)\right]'$$
$$=I_{\Omega_\epsilon}\cdot\Sigma^{\frac{1}{2}}\sum_{i=1}^{S_T}W^n\left(\frac{t_{i-1}}{T}\right)\left[W^n\left(\frac{t_i}{T}\right)-W^n\left(\frac{t_{i-1}}{T}\right)\right]'\Sigma^{\frac{1}{2}\prime}+o_p(1).$$

In light of the previous developments, it stands to reason that

$$Z_T^h\cdot I_{\Omega_\epsilon}=I_{\Omega_\epsilon}\cdot\Sigma^{\frac{1}{2}}\sum_{i=1}^{S_T}W^n\left(\frac{t_{i-1}}{T}\right)\left[W^n\left(\frac{t_i}{T}\right)-W^n\left(\frac{t_{i-1}}{T}\right)\right]'\Sigma^{\frac{1}{2}\prime}+o_p(1).$$

It now remains to show that the partial sum above converges to the stochastic integral $\int_0^1 W^n(r)dW^n(r)$. Choose any $1\le i,j\le n$, and let W_i and W_j represent the i and jth coordinates of W^n; W_i and W_j are independent random functions corresponding to the univariate Wiener process.
For any $T\in N_+$, define the elementary function $\phi_T:\Omega\times[0,1]\to\mathbb{R}$ as

$$\phi_T(\cdot,t)=\sum_{l=1}^{S_T}W_i\left(\frac{t_{l-1}}{T}\right)I_{\left[\frac{t_{l-1}}{T},\frac{t_l}{T}\right)}(t).$$

Since

$$\phi_T(\cdot,t)-W_i(t)=\sum_{l=1}^{S_T}\left(W_i\left(\frac{t_{l-1}}{T}\right)-W_i(t)\right)\cdot I_{\left[\frac{t_{l-1}}{T},\frac{t_l}{T}\right)}(t)-W_i(t)\cdot I_{\left[\frac{T-h-1}{T},1\right]}(t)$$

for any $t\in[0,1]$,

$$\begin{aligned}\mathbb{E}\left[\int_0^1|\phi_T(\cdot,t)-W_i(t)|^2dt\right]&=\int_0^1\mathbb{E}|\phi_T(\cdot,t)-W_i(t)|^2dt \qquad \text{(Fubini's theorem)}\\&=\sum_{l=1}^{S_T}\int_{t_{l-1}/T}^{t_l/T}\mathbb{E}\left|W_i\left(\frac{t_{l-1}}{T}\right)-W_i(t)\right|^2dt-\int_{\frac{T-h-1}{T}}^1\mathbb{E}|W_i(t)|^2dt\\&=\sum_{l=1}^{S_T}\int_{t_{l-1}/T}^{t_l/T}\left(t-\frac{t_{l-1}}{T}\right)dt-\int_{\frac{T-h-1}{T}}^1 tdt\\&=\frac{1}{T^2}\sum_{l=1}^{S_T}\left[\frac{1}{2}\left(t_l^2-t_{l-1}^2\right)-t_{l-1}(t_l-t_{l-1})\right]+\frac{1}{2}\left[\left(\frac{T-h-1}{T}\right)^2-1\right]\\&=\frac{1}{2T^2}\sum_{l=1}^{S_T}(t_l-t_{l-1})^2+\frac{1}{2}\left[\left(\frac{T-h-1}{T}\right)^2-1\right]\\&\le\frac{1}{2}\left(\max_{1\le i\le S_T}\left|\frac{t_i}{T}-\frac{t_{i-1}}{T}\right|\right)+\frac{1}{2}\left[\left(\frac{T-h-1}{T}\right)^2-1\right].\end{aligned}$$

The term on the right goes to 0 as $T \to \infty$, so we have

$$\lim_{T\to\infty} \mathbb{E}\left[\int_0^1 |\phi_T(\cdot,t) - W_i(t)|^2 dt\right] = 0,$$

that is, the sequence $\{\phi_T\}_{T\in N_+}$ of elementary functions on $\Omega \times [0,1]$ converges in $L^2(\mathbb{P}_0 \times \lambda)$ to the mapping $(\omega,t) \mapsto (W_i(t))(\omega)$, where λ is the Lebesgue measure on $[0,1]$. By definition, the stochastic integral of W_i with respect to W_j is the $L^2(\mathbb{P}_0)$ limit of

$$\int_0^1 \phi_T(\cdot,t)dW_j(t) = \sum_{l=1}^{S_T} W_i\left(\frac{t_{l-1}}{T}\right)\left[W_j\left(\frac{t_l}{T}\right) - W_j\left(\frac{t_{l-1}}{T}\right)\right].$$

This holds for any $1 \le i,j \le n$, so we have

$$\sum_{i=1}^{S_T} W^n\left(\frac{t_{i-1}}{T}\right)\left[W^n\left(\frac{t_i}{T}\right) - W^n\left(\frac{t_{i-1}}{T}\right)\right]' \xrightarrow{L^2} \int_0^1 W^n(r)dW^n(r)',$$

which implies that the convergence is in probability as well. Therefore,

$$Z_T^h \cdot I_{\Omega_\epsilon} = I_{\Omega_\epsilon} \cdot \Sigma^{\frac{1}{2}} \int_0^1 W^n(r)dW^n(r)'\Sigma^{\frac{1}{2}\prime} + o_p(1).$$

By implication, for any $\delta > 0$,

$$\mathbb{P}_0\left(\left\| Z_T^h - \Sigma^{\frac{1}{2}} \int_0^1 W^n(r)dW^n(r)'\Sigma^{\frac{1}{2}\prime}\right\| > \delta\right)$$
$$\le \mathbb{P}_0\left(\left\| I_{\Omega_\epsilon} \cdot Z_T^h - I_{\Omega_\epsilon} \cdot \Sigma^{\frac{1}{2}} \int_0^1 W^n(r)dW^n(r)'\Sigma^{\frac{1}{2}\prime}\right\| > \delta\right) + \mathbb{P}_0\left(\Omega_\epsilon^c\right).$$

Taking $T \to \infty$ on both sides yields

$$\lim_{T\to\infty} \mathbb{P}_0\left(\left\| Z_T^h - \Sigma^{\frac{1}{2}} \int_0^1 W^n(r)dW^n(r)'\Sigma^{\frac{1}{2}\prime}\right\| > \delta\right) = \mathbb{P}_0\left(\Omega_\epsilon^c\right) < \epsilon,$$

and because this holds for any $\epsilon > 0$,

$$\lim_{T\to\infty} \mathbb{P}_0\left(\left\| Z_T^h - \Sigma^{\frac{1}{2}} \int_0^1 W^n(r)dW^n(r)'\Sigma^{\frac{1}{2}\prime}\right\| > \delta\right) = 0.$$

As such,

$$Z_T^h \xrightarrow{p} \Sigma^{\frac{1}{2}} \int_0^1 W^n(r)dW^n(r)'\Sigma^{\frac{1}{2}\prime},$$

and because almost sure convergence implies convergence in probability,

$$(Y^T, Z_T^h) \xrightarrow{p} \left(\Sigma^{\frac{1}{2}} W^n, \Sigma^{\frac{1}{2}} \int_0^1 W^n(r)dW^n(r)'\Sigma^{\frac{1}{2}\prime}\right).$$

This holds for any $0 \le h \le p$, so

$$(Y^T, Z_T^0, \cdots, Z_T^p) \xrightarrow{p} \left(\Sigma^{\frac{1}{2}} W^n, \quad \iota'_{p+1} \bigotimes \left[\Sigma^{\frac{1}{2}} \int_0^1 W^n(r) dW^n(r)' \Sigma^{\frac{1}{2}\prime} \right] \right).$$

Convergence in probability implies weak convergence, so the above convergence is in distribution as well. Finally, because

$$(Y^T, Z_T^0, \cdots, Z_T^p) \sim \left(X^T, \frac{1}{T} \sum_{t=1}^{T} V_{t-1} \varepsilon'_t, \cdots, \frac{1}{T} \sum_{t=1}^{T-h} V_{t-1} \varepsilon'_t \right)$$

for any $T \in N_+$, we can conclude that

$$\left(X^T, \frac{1}{T} \sum_{t=1}^{T} V_{t-1} \varepsilon'_t, \cdots, \frac{1}{T} \sum_{t=1}^{T-h} V_{t-1} \varepsilon'_t \right) \xrightarrow{d} \left(\Sigma^{\frac{1}{2}} W^n, \quad \iota'_{p+1} \bigotimes \left[\Sigma^{\frac{1}{2}} \int_0^1 W^n(r) dW^n(r)' \Sigma^{\frac{1}{2}\prime} \right] \right).$$

Q.E.D.

Now we show the main result, which generalizes the above theorem for the case when the underlying errors form a linear process. The proof follows the martingale approximation approach by Phillips (1988) almost verbatim, save for the use of the BN decomposition instead of a martingale approximation.

Theorem (Joint Convergence to Stochastic Integrals for Linear Processes)
Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be an n-dimensional mean zero i.i.d. process with positive definite covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ and finite fourth moments, and $\{\Psi_j\}_{j\in\mathbb{N}}$ a one-summable sequence of $n\times n$ matrices. Let $\{u_t\}_{t\in\mathbb{Z}}$ be the mean zero linear process defined as $u_t = \Psi(L)\varepsilon_t$ for any $t\in\mathbb{Z}$. Let $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$ be the autocovariance function of $\{u_t\}_{t\in\mathbb{Z}}$.
Define the process $\{S_t\}_{t\in\mathbb{N}}$ as $S_0 = 0$ and $S_t = \sum_{s=1}^{t} u_s$ for any $t \in N_+$. For any $T \in N_+$, define the stochastic processes $\{X_T(r)\}_{r\in[0,1]}$ and $\{V_T(r)\}_{r\in[0,1]}$ as

$$X_T(r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor} u_t + \frac{1}{\sqrt{T}}(Tr - \lfloor Tr\rfloor)u_{\lfloor Tr\rfloor+1}$$
$$V_T(r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor} \varepsilon_t + \frac{1}{\sqrt{T}}(Tr - \lfloor Tr\rfloor)\varepsilon_{\lfloor Tr\rfloor+1}$$

for any $r \in [0,1]$, and let X^T and V^T be the random functions taking values in $\mathcal{C}([0,1],\mathbb{R}^n)$ corresponding to $\{X_T(r)\}_{r\in[0,1]}$ and $\{V_T(r)\}_{r\in[0,1]}$. Defining

$$\mathcal{V}_T^h = \frac{1}{T}\sum_{t=1}^{T-h} S_{t-1}\varepsilon_t'$$
$$\mathcal{U}_T^h = \frac{1}{T}\sum_{t=1}^{T-h} S_{t-1}u_t'$$

for any $h \in \mathbb{N}$, for any $p \in N_+$

$$V^T \xrightarrow{d} \Sigma^{\frac{1}{2}} W^n$$

$$X^T \xrightarrow{d} \Psi(1)\Sigma^{\frac{1}{2}} W^n$$

$$(\mathcal{V}_T^0, \cdots, \mathcal{V}_T^p) \xrightarrow{d} \iota_{p+1}' \bigotimes \left[\Psi(1)\Sigma^{\frac{1}{2}}\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}\right]$$

$$(\mathcal{U}_T^0, \cdots, \mathcal{U}_T^p) \xrightarrow{d} \iota_{p+1}' \left[\Psi(1)\Sigma^{\frac{1}{2}}\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}\Psi(1)' + K\right]$$

jointly, where W^n is the n-dimensional Wiener function, $\{W^n(r)\}_{r\in[0,1]}$ the corresponding Wiener process, and $K = \sum_{j=1}^{\infty}\Gamma(j)'$.

Proof) We first define some notations and state preliminary results.

Recall from the BN decomposition that there exists an almost sure set $\Omega_0 \in \mathcal{H}$ such that

$$u_t = \Psi(1)\varepsilon_t + \eta_t - \eta_{t-1}$$
$$S_t = \sum_{s=1}^{t} u_s = \Psi(1)\left(\sum_{s=1}^{t} \varepsilon_s\right) + \eta_t - \eta_0$$

on Ω_0 for any $t \in \mathbb{N}_+$, where

- $\Psi(1) = \sum_{j=0}^{\infty} \Psi_j$,
- $\{\alpha_j\}_{j\in\mathbb{N}}$ is an absolutely summable sequence of $n \times n$ matrices such that

$$\alpha_j = -\sum_{h=j+1}^{\infty} \Psi_j \quad \text{for any } j \in \mathbb{N},$$

- $\{\eta_t\}_{t\in\mathbb{Z}}$ is an absolutely summable linear process such that $\eta_t = \sum_{j=0}^{\infty} \alpha_j \cdot \varepsilon_{t-j}$ for any $t \in N_+$.

Define the random walk τ_t as $\tau_0 = 0$ and

$$\tau_t = \sum_{s=1}^{t} \varepsilon_s + \tau_0$$

for any $t \in N_+$, so that we can express

$$S_t = \Psi(1)\tau_t + \eta_t - \eta_0$$

and

$$u_t = \Psi(1)\varepsilon_t + \eta_t - \eta_{t-1}$$

for any $t \in N_+$. Furthermore, it was shown in the proof of the extended CLT that we can write

$$X^T = \Psi(1)V^T + A^T$$

for some random function A_T taking values in $\mathcal{C}([0,1],\mathbb{R}^n)$ that converges to the zero function in probability.
By the stochastic integral convergence result we proved in the previous theorem,

$$V^T \xrightarrow{d} \Sigma^{\frac{1}{2}} W^n$$

$$\left(\frac{1}{T}\sum_{t=1}^{T} \tau_{t-1}\varepsilon_t, \cdots, \frac{1}{T}\sum_{t=1}^{T-p} \tau_{t-1}\varepsilon_t\right) \xrightarrow{d} \iota'_{p+1} \bigotimes \left[\Sigma^{\frac{1}{2}} \left(\int_0^1 W^n(r) dW^n(r)'\right) \Sigma^{\frac{1}{2}}\right]$$

jointly. We can easily see that, by the continuous mapping theorem, the convergence result

$$X^T \xrightarrow{d} \Psi(1)\Sigma^{\frac{1}{2}} W^n$$

occurs jointly with the above results. We are now ready to prove the remaining items.

We start with the easier of the two partial sum processes.
For any $0 \le h \le p$,

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T-h} S_{t-1}\varepsilon_t' &= \frac{1}{T}\sum_{t=1}^{T-h}\left(\Psi(1)\tau_{t-1} + \eta_{t-1} - \eta_0\right)\varepsilon_t' \\
&= \Psi(1)\left(\frac{1}{T}\sum_{t=1}^{T-h}\tau_{t-1}\varepsilon_t\right) + \frac{1}{T}\sum_{t=1}^{T-h}\left(\eta_{t-1} - \eta_0\right)\varepsilon_t'.
\end{aligned}$$

The first term converges in distribution to

$$\Psi(1)\Sigma^{\frac{1}{2}}\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}$$

by the previous theorem. Meanwhile, because the $\{\eta_t\}_{t\in\mathbb{Z}}$ is an absolutely summable causal linear process representation with an iid innovation process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ that has finite fourth moments, the earlier result on linear processes tells us that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\eta_{t-1}\varepsilon_t = O_p(1).$$

In addition,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\varepsilon_t = O_p(1)$$

by the usual Lindeberg-Levy CLT, so it follows that

$$\frac{1}{T}\sum_{t=1}^{T-h}\left(\eta_{t-1} - \eta_0\right)\varepsilon_t'$$

converges to 0. This holds for any $0 \le h \le p$, so using the result proven above,

$$X^T \xrightarrow{d} \Psi(1)\Sigma^{\frac{1}{2}}\cdot W^n$$

$$\left(\frac{1}{T}\sum_{t=1}^{T} S_{t-1}\varepsilon_t', \cdots, \frac{1}{T}\sum_{t=1}^{T-p} S_{t-1}\varepsilon_t'\right) \xrightarrow{d} \iota_{p+1}' \bigotimes \left[\Psi(1)\Sigma^{\frac{1}{2}}\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}\right]$$

jointly by the continuous mapping theorem and Slutsky's theorem.

The proof that the second partial sum process converges is slightly trickier. For any $0 \leq h \leq p$ we have

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T-h} S_{t-1}u_t' &= \frac{1}{T}\sum_{t=1}^{T-h}\left(\Psi(1)\tau_{t-1}+\eta_{t-1}-\eta_0\right)\left(\Psi(1)\varepsilon_t+\eta_t-\eta_{t-1}\right)' \\ &= \Psi(1)\left(\frac{1}{T}\sum_{t=1}^{T-h}\tau_{t-1}\varepsilon_t\right)\Psi(1)'+\Psi(1)\left(\frac{1}{T}\sum_{t=1}^{T-h}\tau_{t-1}\left(\eta_t-\eta_{t-1}\right)'\right) \\ &\quad+\left(\frac{1}{T}\sum_{t=1}^{T-h}\left(\eta_{t-1}-\eta_0\right)\varepsilon_t'\right)\Psi(1)'+\frac{1}{T}\sum_{t=1}^{T-h}\left(\eta_{t-1}-\eta_0\right)\left(\eta_t-\eta_{t-1}\right)'.\end{aligned}$$

We examine each term in turn.

The Last Term

Because the innovation process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ for $\{\eta_t\}_{t\in\mathbb{Z}}$ is i.i.d. and the linear filter $\{\alpha_j\}_{j\in\mathbb{N}}$ is absolutely summable, we can apply the WLLN for linear processes and conclude that

$$\frac{1}{T}\sum_{t=1}^{T-h}\left(\eta_{t-1}-\eta_0\right)\left(\eta_t-\eta_{t-1}\right)' \xrightarrow{p} G(1)'-G(0),$$

where $G:\mathbb{Z}\to\mathbb{R}^{n\times n}$ is the autocovariance function of η_t. Examining the limit $G(1)'-G(0)$ further, we can tell that

$$\begin{aligned}G(1)'-G(0) &= \sum_{j=1}^{\infty}\alpha_{j-1}\Sigma\alpha_j'-\sum_{j=0}^{\infty}\alpha_j\Sigma\alpha_j' \\ &= \sum_{j=1}^{\infty}\left(\alpha_{j-1}-\alpha_j\right)\Sigma\alpha_j'-\alpha_0\Sigma\alpha_0' \\ &= -\sum_{j=1}^{\infty}\Psi_j\Sigma\alpha_j'-\alpha_0\Sigma\alpha_0' \\ &= -\sum_{j=1}^{\infty}\Psi_j\Sigma\alpha_j'-\Sigma_u-\Psi_0\Sigma\Psi_0'+\Psi(1)\Sigma\Psi_0'+\Psi_0\Sigma\Psi(1)',\end{aligned}$$

since $\alpha_0=-\sum_{j=1}^{\infty}\Psi_j=\Psi_0-\Psi(1)$.

The Third Term

From our earlier result, we have

$$\frac{1}{T}\sum_{t=1}^{T-h}\left(\eta_{t-1}-\eta_0\right)\varepsilon_t'\Psi(1)' \xrightarrow{p} O.$$

The Second Term

As for the second term, we have

$$\begin{aligned}
\Psi(1)\left(\frac{1}{T}\sum_{t=1}^{T-h}\tau_{t-1}\left(\eta_t-\eta_{t-1}\right)'\right) &= \Psi(1)\frac{1}{T}\sum_{t=1}^{T-h}\tau_{t-1}\eta_t' - \Psi(1)\frac{1}{T}\sum_{t=1}^{T-h}\tau_{t-1}\eta_{t-1}' \\
&= \Psi(1)\frac{1}{T}\sum_{t=1}^{T}\tau_t\eta_t' - \Psi(1)\frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_t\eta_t' - \Psi(1)\frac{1}{T}\sum_{t=1}^{T-h}\tau_{t-1}\eta_{t-1}' \\
&= \Psi(1)\frac{1}{T}\tau_{T-h}\eta_{T-h}' - \Psi(1)\frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_t\eta_t'.
\end{aligned}$$

Since $\mathbb{E}|\tau_T|^2 = T\cdot\mathrm{tr}(\Sigma)$ for any $T\in N_+$,

$$\begin{aligned}
\mathbb{E}\left\|\frac{1}{T}\tau_{T-h}\eta_{T-h}'\right\| &\leq \frac{1}{T}\left(\mathbb{E}|\tau_{T-h}|^2\right)^{\frac{1}{2}}\mathrm{tr}\left(G(0)\right)^{\frac{1}{2}} \\
&= \mathrm{tr}(\Sigma)^{\frac{1}{2}}\,\mathrm{tr}(G(0))^{\frac{1}{2}}\,\frac{\sqrt{T-h}}{T},
\end{aligned}$$

so that $\frac{1}{T}\tau_{T-h}\eta_{T-h}' \xrightarrow{L^1} O$ and thus $\frac{1}{T}\tau_{T-h}\eta_{T-h}' = o_p(1)$. The sum of the product of η_t and ε_t can be expanded as

$$\frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_t\eta_t' = \frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_t(\eta_t-\alpha_0\varepsilon_t)' + \frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_t\varepsilon_t'\alpha_0'.$$

Because $\{\eta_t-\alpha_0\varepsilon_t\}_{t\in\mathbb{Z}}$ has the absolutely summable causal linear process representation

$$\eta_t-\alpha_0\varepsilon_t = \sum_{j=0}^{\infty}\alpha_{j+1}\cdot\varepsilon_{t-1-j}$$

for any $t\in\mathbb{Z}$, the result on linear processes tells us once again that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\varepsilon_t(\eta_t-\alpha_0\varepsilon_t)' = O_p(1).$$

By implication,

$$\frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_t(\eta_t-\alpha_0\varepsilon_t)' \xrightarrow{p} O.$$

As for the second term,

$$\frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_t\varepsilon_t'\alpha_0' \xrightarrow{p} \Sigma\alpha_0'$$

by the WLLN for iid processes. Considering the fact that $\alpha_0 = \Psi_0 - \Psi(1)$, we have

$$\Psi(1)\left(\frac{1}{T}\sum_{t=1}^{T-h}\tau_{t-1}\left(\eta_t-\eta_{t-1}\right)'\right) \xrightarrow{p} -\Psi(1)\Sigma\alpha_0' = \Psi(1)\Sigma\Psi(1)' - \Psi(1)\Sigma\Psi_0'.$$

The First Term

Finally, the first term has the limit

$$\Psi(1)\left(\frac{1}{T}\sum_{t=1}^{T-h}\tau_{t-1}\varepsilon_t\right)\Psi(1)' \xrightarrow{d} \Psi(1)\Sigma^{\frac{1}{2}}\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}\Psi(1)',$$

because we know the limit of the sum in the middle.

The above results hold for any $0 \le h \le p$, so

$$V^T \xrightarrow{d} \Sigma^{\frac{1}{2}} \cdot W^n$$

$$X^T \xrightarrow{d} \Psi(1)\Sigma^{\frac{1}{2}} \cdot W^n$$

$$\left(\frac{1}{T}\sum_{t=1}^{T}S_{t-1}\varepsilon_t', \cdots, \frac{1}{T}\sum_{t=1}^{T-p}S_{t-1}\varepsilon_t'\right) \xrightarrow{d} \iota_{p+1}' \bigotimes \left[\Psi(1)\Sigma^{\frac{1}{2}}\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}\right]$$

$$\left(\frac{1}{T}\sum_{t=1}^{T}S_{t-1}u_t', \cdots, \frac{1}{T}\sum_{t=1}^{T-p}S_{t-1}u_t'\right) \xrightarrow{d} \iota_{p+1}' \bigotimes \left[\Psi(1)\Sigma^{\frac{1}{2}}\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}\Psi(1)' + K\right]$$

jointly by the continuous mapping theorem and Slutsky's theorem, for some $K \in \mathbb{R}^{n\times n}$ by Slutsky's theorem and the joint convergence of V_T and $\frac{1}{T}\sum_{t=1}^{T}\tau_{t-1}\varepsilon_t'$.
The constant term K is given as

$$\begin{aligned} K &= \Psi(1)\Sigma\Psi(1)' - \Psi(1)\Sigma\Psi_0' - \sum_{j=1}^{\infty}\Psi_j\Sigma\alpha_j' - \Sigma_u - \Psi_0\Sigma\Psi_0' + \Psi(1)\Sigma\Psi_0' + \Psi_0\Sigma\Psi(1)' \\ &= -\sum_{j=1}^{\infty}\Psi_j\Sigma\alpha_j' + \Psi_0\Sigma\left(\Psi(1)-\Psi_0\right)' = -\sum_{j=0}^{\infty}\Psi_j\Sigma\alpha_j'. \end{aligned}$$

By the definition of each α_j,

$$-\sum_{j=0}^{\infty}\Psi_j\Sigma\alpha_j' = \sum_{j=0}^{\infty}\sum_{h=j+1}^{\infty}\Psi_j\Sigma\Psi_h' = \sum_{j=1}^{\infty}\sum_{h=j}^{\infty}\Psi_{h-j}\Sigma\Psi_h' = \sum_{j=1}^{\infty}\Gamma(j)'.$$

Q.E.D.

4.2.3 Main Limit Results

Theorem (Main Limit Results)
Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be an n-dimensional mean zero i.i.d. process with positive definite covariance matrix $\Sigma\in\mathbb{R}^{n\times n}$ and finite fourth moments, and $\{\Psi_j\}_{j\in\mathbb{N}}$ a one-summable sequence of $n\times n$ matrices. Let $\{u_t\}_{t\in\mathbb{Z}}$ be the mean zero linear process defined as

$$u_t=\sum_{j=0}^{\infty}\Psi_j\cdot\varepsilon_{t-j}$$

for any $t\in\mathbb{Z}$. Let $\Gamma:\mathbb{Z}\to\mathbb{R}^{n\times n}$ be the autocovariance function of $\{u_t\}_{t\in\mathbb{Z}}$.
Define the process $\{S_t\}_{t\in\mathbb{N}}$ as $S_0=0$ and

$$S_t=\sum_{s=1}^{t}u_s+S_0.$$

for any $t\in N_+$. Define $\Sigma_u=\Psi(1)\Sigma\Psi(1)'$, $\Sigma^{\frac{1}{2}}$ as the Cholesky factor of Σ, and $\Lambda=\Psi(1)\Sigma^{\frac{1}{2}}$. Let W^n denote the standard n-dimensional Wiener function, and $\{W^n(r)\}_{r\in[0,1]}$ the corresponding Brownian motion. For any $p\in N_+$, the following convergence results hold jointly:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t\xrightarrow{d}\Sigma^{\frac{1}{2}}W^n(1)$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t\xrightarrow{d}\Lambda\cdot W^n(1)$$

$$\frac{1}{T}\sum_{t=1}^{T}u_tu_{t-h}'\xrightarrow{p}\Gamma(h)\qquad\text{for any } h\geq 0$$

$$\frac{1}{T}\sum_{t=h+1}^{T}S_{t-1}\varepsilon_{t-h}'\xrightarrow{d}\Lambda\left(\int_0^1W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}+\Sigma\qquad\text{for any } 0\leq h\leq p$$

$$\frac{1}{T}\sum_{t=h+1}^{T}S_{t-1}u_{t-h}'\xrightarrow{d}\Lambda\left(\int_0^1W^n(r)dW^n(r)'\right)\Lambda'+\Sigma_u-\sum_{j=h}^{\infty}\Gamma(j)\qquad\text{for any } 0\leq h\leq p$$

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T}S_{t-1}\xrightarrow{d}\Lambda\cdot\int_0^1W^n(r)dr$$

$$\frac{1}{T^2}\sum_{t=1}^{T}S_{t-1}S_{t-1}'\xrightarrow{d}\Lambda\left(\int_0^1W^n(r)W^n(r)'dr\right)\Lambda'$$

$$\frac{1}{T^{3/2}}\sum_{t=h+1}^{T}t\cdot u_{t-h}\xrightarrow{d}\Lambda\cdot\int_0^1rdW^n(r)\qquad\text{for any } h\geq 0$$

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T}t\cdot\varepsilon_{t-1}\xrightarrow{d}\Sigma^{\frac{1}{2}}\cdot\int_0^1rdW^n(r)$$

$$\frac{1}{T^{5/2}}\sum_{t=1}^{T}t\cdot S_{t-1}\xrightarrow{d}\Lambda\cdot\int_0^1r\cdot W^n(r)dr$$

Proof) For any $T \in N_+$, define the n-dimensional stochastic processes $\{X_T(r)\}_{r\in[0,1]}$ and $\{V_T(r)\}_{r\in[0,1]}$ with continuous paths as

$$X_T(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} u_t + \frac{1}{\sqrt{T}} (Tr - \lfloor Tr \rfloor) u_{\lfloor Tr \rfloor + 1}$$
$$V_T(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t + \frac{1}{\sqrt{T}} (Tr - \lfloor Tr \rfloor) \varepsilon_{\lfloor Tr \rfloor + 1}$$

for any $r \in [0,1]$. Let X^T and V^T be the random functions in $\mathcal{C}([0,1], \mathbb{R}^n)$ corresponding to $\{X_T(r)\}_{r\in[0,1]}$ and $\{V_T(r)\}_{r\in[0,1]}$. Finally, defining

$$\mathcal{V}_T^h = \frac{1}{T} \sum_{t=1}^{T-h} S_{t-1} \varepsilon'_{t-h} \quad \text{and}$$
$$\mathcal{U}_T^h = \frac{1}{T} \sum_{t=1}^{T-h} S_{t-1} u'_{t-h}$$

for $0 \le h \le p$, the joint convergence results in the previous theorem tell us that

$$V^T \xrightarrow{d} \Sigma^{\frac{1}{2}} \cdot W^n$$

$$X^T \xrightarrow{d} \Lambda \cdot W^n$$

$$(\mathcal{V}_T^0, \cdots, \mathcal{V}_T^p) \xrightarrow{d} \iota'_{p+1} \bigotimes \left[\Lambda \left(\int_0^1 W^n(r) dW^n(r)' \right) \Sigma^{\frac{1}{2}\prime} \right]$$

$$(\mathcal{U}_T^0, \cdots, \mathcal{U}_T^p) \xrightarrow{d} \iota'_{p+1} \left[\Lambda \left(\int_0^1 W^n(r) dW^n(r)' \right) \Lambda' + K \right]$$

jointly, where

$$K = \sum_{j=1}^{\infty} \Gamma(j)'.$$

These results are now used to prove the results claimed in the theorem.

We proceed one by one:

i) For any $T \in N_+$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t = V_T(1) = \pi_1(V_T),$$

so by the continuity of projections,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t \xrightarrow{d} \pi_1(\Sigma^{\frac{1}{2}}W^n) = \Sigma^{\frac{1}{2}}W^n(1).$$

ii) The proof is almost identical to the above. For any $T \in N_+$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t = X_T(1) = \pi_1(X_T),$$

so by the continuity of projections,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t \xrightarrow{d} \pi_1(\Lambda \cdot W^n) = \Lambda \cdot W^n(1).$$

iii) This was proven in the lemma for the WLLN of Linear Processes.

iv) Choose any $0 \leq h \leq p$. Then,

$$\begin{aligned}
\frac{1}{T}\sum_{t=h+1}^{T} S_{t-1}\varepsilon_{t-h}' &= \frac{1}{T}\sum_{t=h+1}^{T}\left(S_{t-h-1} + \sum_{s=t-h}^{t}\varepsilon_s\right)\varepsilon_{t-h}' \\
&= \frac{1}{T}\sum_{t=1}^{T-h} S_{t-1}\varepsilon_t' + \frac{1}{T}\sum_{t=h+1}^{T}\sum_{s=t-h}^{t}(\varepsilon_s\varepsilon_{t-h}') \\
&= \mathcal{V}_T^h + \sum_{j=0}^{h}\left(\frac{1}{T}\sum_{t=j+1}^{T-h+j}\varepsilon_t\varepsilon_{t-j}'\right) \\
&\xrightarrow{d} \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime} + \Sigma,
\end{aligned}$$

where we used the fact that

$$\frac{1}{T}\sum_{t=j+1}^{T-h+j}\varepsilon_t\varepsilon_{t-j}' \xrightarrow{p} \mathbb{E}\left[\varepsilon_t\varepsilon_{t-j}'\right] = \begin{cases}\Sigma & \text{if } j = 0\\ O & \text{otherwise}\end{cases}$$

for any $j \in \mathbb{N}$. This holds jointly for any $0 \leq h \leq p$.

v) This result can be shown in a similar manner to the preceding result. Choose any $0 \leq h \leq p$. Then,

$$\begin{aligned}
\frac{1}{T}\sum_{t=h+1}^{T} S_{t-1}u_{t-h}' &= \frac{1}{T}\sum_{t=h+1}^{T}\left(S_{t-h-1} + \sum_{s=t-h}^{t-1} u_s\right)u_{t-h}' \\
&= \frac{1}{T}\sum_{t=1}^{T-h} S_{t-1}u_t' + \frac{1}{T}\sum_{t=h+1}^{T}\sum_{s=t-h}^{t-1}(u_s u_{t-h}') \\
&= \mathcal{V}_T^h + \sum_{j=0}^{h-1}\left(\frac{1}{T}\sum_{t=j+1}^{T-h+j} u_t u_{t-j}'\right) \\
&\xrightarrow{d} \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Lambda' + K + \sum_{j=0}^{j-1}\Gamma(j),
\end{aligned}$$

where this time we used the fact that

$$\frac{1}{T}\sum_{t=j+1}^{T-h+j} u_t u_{t-j}' \xrightarrow{p} \Gamma(j)$$

for any $j \in \mathbb{N}$, as shown in iii). Since $K = \sum_{j=1}^{\infty}\Gamma(j)'$, we can see that

$$\frac{1}{T}\sum_{t=h+1}^{T} S_{t-1}u_{t-h}' \xrightarrow{d} \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Lambda' + \sum_{j=1-h}^{\infty}\Gamma(j)'.$$

To obtain the alternative formulation of the nuisance term $\sum_{j=1-h}^{\infty}\Gamma(j)'$, note that, because $\{u_t\}_{t\in\mathbb{Z}}$ has an MA(∞) representation with coefficients $\{\Psi_j\}_{j\in\mathbb{N}}$,

$$\Sigma_u = \Psi(1)\Sigma\Psi(1)'$$

is 2π times the spectral density of u_t evaluated at 0, that is,

$$\Sigma_u = 2\pi \cdot f_{uu}(0).$$

By definition,

$$f_{uu}(z) = \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty}\exp(-i\tau z)\Gamma(\tau) = \frac{1}{2\pi}\left[\Gamma(0) + \sum_{\tau=1}^{\infty}\exp(-i\tau z)\left(\Gamma(\tau)+\Gamma(\tau)'\right)\right],$$

so we have

$$\Sigma_u = \Gamma(0) + \sum_{\tau=1}^{\infty}\left(\Gamma(\tau)+\Gamma(\tau)'\right).$$

It follows that

$$\sum_{j=1-h}^{\infty}\Gamma(j)' = \Sigma_u - \sum_{j=h}^{\infty}\Gamma(j).$$

This again holds jointly for any $0 \leq h \leq p$.

vi) Note that the Lebesgue integral of $X_T(r)$ with respect to r is given by

$$\begin{aligned}
\int_0^1 X_T(r)dr &= \sum_{t=1}^{T} \int_{[\frac{t-1}{T},\frac{t}{T})} X_T(r)dr \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\left(\frac{1}{T} \sum_{s=1}^{t-1} u_s \right) + \left(\int_{\frac{t-1}{T}}^{\frac{t}{T}} (Tr-(t-1))dr \right) \cdot u_t \right] \\
&= \frac{1}{T^{3/2}} \sum_{t=1}^{T} S_{t-1} + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[T\left(\frac{t^2}{2T^2} - \frac{(t-1)^2}{2T^2} \right) - \frac{t-1}{T} \right] u_t \\
&= \frac{1}{T^{3/2}} \sum_{t=1}^{T} S_{t-1} + \frac{1}{2} \cdot \frac{1}{T^{3/2}} \sum_{t=1}^{T} u_t.
\end{aligned}$$

By implication,

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} S_{t-1} = \int_0^1 X_T(r)dr - \frac{1}{2} \cdot \frac{1}{T^{3/2}} \sum_{t=1}^{T} u_t = g(X_T) - \frac{1}{2T} \cdot X_T(1).$$

It follows from the continuous mapping theorem and Slutsky's theorem that

$$\begin{aligned}
\frac{1}{T^{3/2}} \sum_{t=1}^{T} S_{t-1} &= g(X_T) - \frac{1}{2T} \cdot \pi_1(X_T) \\
&\xrightarrow{d} g(\Lambda \cdot W^n) = \Lambda \cdot \int_0^1 W^n(r)dr.
\end{aligned}$$

vii) As above,

$$\begin{aligned}
\int_0^1 X_T(r)X_T(r)'dr &= \sum_{t=1}^{T} \int_{[\frac{t-1}{T},\frac{t}{T})} X_T(r)X_T(r)'dr \\
&= \frac{1}{T} \sum_{t=1}^{T} \int_{[\frac{t-1}{T},\frac{t}{T})} \left(S_{t-1} + (Tr-(t-1))u_t\right)\left(S_{t-1} + (Tr-(t-1))u_t\right)' dr \\
&= \frac{1}{T} \sum_{t=1}^{T} \left[\frac{1}{T} S_{t-1}S_{t-1}' + S_{t-1} \cdot \left(\int_{\frac{t-1}{T}}^{\frac{t}{T}} (Tr-(t-1))dr \right) u_t' \right] \\
&\quad + \frac{1}{T} \sum_{t=1}^{T} \left[u_t \cdot \left(\int_{\frac{t-1}{T}}^{\frac{t}{T}} (Tr-(t-1))dr \right) S_{t-1}' + u_t \left(\int_{\frac{t-1}{T}}^{\frac{t}{T}} (Tr-(t-1))^2 dr \right) u_t' \right] \\
&= \frac{1}{T^2} \sum_{t=1}^{T} S_{t-1}S_{t-1}' + \frac{1}{2T^2} \sum_{t=1}^{T} (S_{t-1}u_t' + u_t S_{t-1}') + \frac{1}{3T^2} \sum_{t=1}^{T} u_t u_t'.
\end{aligned}$$

It follows that

$$\frac{1}{T^2}\sum_{t=1}^{T} S_{t-1}S_{t-1}' = \int_0^1 X_T(r)X_T(r)'dr - \frac{1}{2T}\left(\frac{1}{T}\sum_{t=1}^{T}(S_{t-1}u_t' + u_tS_{t-1}')\right) - \frac{1}{3T}\left(\frac{1}{T}\sum_{t=1}^{T}u_tu_t'\right)$$
$$= g_2(X_T) - \frac{1}{2T}\left(\frac{1}{T}\sum_{t=1}^{T}(S_{t-1}u_t' + u_tS_{t-1}')\right) - \frac{1}{3T}\left(\frac{1}{T}\sum_{t=1}^{T}u_tu_t'\right);$$

the last two terms converge in probability to 0 by the previous results, and by the continuous mapping theorem,

$$\frac{1}{T^2}\sum_{t=1}^{T} S_{t-1}S_{t-1}' \xrightarrow{d} \Lambda\cdot\left(\int_0^1 W^n(r)W^n(r)'dr\right)\cdot\Lambda.$$

viii) For any $h \geq 0$,

$$\frac{1}{\sqrt{T}}\sum_{t=h+1}^{T} u_{t-h} - \frac{1}{T^{3/2}}\sum_{t=h+1}^{T} t\cdot u_{t-h} = \frac{1}{T^{3/2}}\sum_{t=h+1}^{T}(T-t)u_{t-h}$$
$$= \frac{1}{T^{3/2}}(u_{T-h-1} + 2\cdot u_{T-h-2} + \cdots + (T-h-1)\cdot u_1)$$
$$= \frac{1}{T^{3/2}}\left(\sum_{t=1}^{T-h-1}u_t + \sum_{t=1}^{T-2}u_t + \cdots + \sum_{t=1}^{1}u_t\right)$$
$$= \frac{1}{T^{3/2}}\sum_{t=1}^{T-h}S_{t-1}.$$

From viii), we can deduce that

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T-h}S_{t-1} = G\left(\frac{T-h}{T}, X_T\right) - \frac{1}{2T}\pi\left(\frac{T-h}{T}, X_T\right),$$

and

$$\frac{1}{T^{3/2}}\sum_{t=1}^{h} t\cdot u_{t-h} \xrightarrow{p} \mathbf{0},$$

so we have

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T} t\cdot u_{t-h} = \frac{1}{T^{3/2}}\sum_{t=h+1}^{T} t\cdot u_{t-h} + \frac{1}{T^{3/2}}\sum_{t=1}^{h} t\cdot u_{t-h}$$
$$= \frac{1}{\sqrt{T}}\sum_{t=h+1}^{T} u_{t-h} - \frac{1}{T^{3/2}}\sum_{t=1}^{T-h}S_{t-1} + \frac{1}{T^{3/2}}\sum_{t=1}^{h} t\cdot u_{t-h}$$
$$= \sqrt{\frac{T-h}{T}}\pi\left(\frac{T-h}{T}, X_T\right) - G\left(\frac{T-h}{T}, X_T\right) + o_p(1)$$

$$\xrightarrow{d} \pi(1,W^n) - G(1,W^n) = \Lambda \cdot \left(W^n(1) - \int_0^1 W^n(r)dr\right).$$

By Ito's lemma, for any $1 \leq i \leq n$,

$$d(rW_i(r)) = W_i(r)dr + rdW_i(r),$$

so that

$$W_i(1) = \int_0^1 W_i(r)dr + \int_0^1 rdW_i(r).$$

Therefore,

$$W^n(1) - \int_0^1 W^n(r)dr = \int_0^1 rdW^n(r),$$

so the limiting distribution can be written as

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T} t \cdot u_{t-h} \xrightarrow{d} \Lambda \cdot \int_0^1 rdW^n(r).$$

ix) By the same process as the preceding result, we can see that

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\varepsilon_{t-1} - \frac{1}{T^{3/2}}\sum_{t=2}^{T} t \cdot \varepsilon_{t-1} = \frac{1}{T^{3/2}}\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\varepsilon_s\right).$$

Analogously to viii), we can deduce that

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T-1}\left(\sum_{s=1}^{t}\varepsilon_s\right) - \int_0^1 V_T(r)dr = o_p(1),$$

and by the continuous mapping theorem,

$$\int_0^1 V_T(r)dr \xrightarrow{d} \Sigma^{\frac{1}{2}} \int_0^1 W^n(r)dr.$$

Furthermore,

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\varepsilon_{t-1} = \sqrt{\frac{T-1}{T}} \cdot \pi\left(\frac{T-1}{T}, V_T\right),$$

so we can see that

$$
\begin{aligned}
\frac{1}{T^{3/2}}\sum_{t=1}^{T} t\cdot\varepsilon_{t-1} &= \frac{1}{T^{3/2}}\varepsilon_0 + \sqrt{\frac{T-1}{T}}\pi\left(\frac{T-1}{T}, V_T\right) - \frac{1}{T^{3/2}}\sum_{t=1}^{T-1}\left(\sum_{s=1}^{t}\varepsilon_s\right) \\
&\xrightarrow{d} \pi(1, \Sigma^{\frac{1}{2}}W^n) - \Sigma^{\frac{1}{2}}\int_0^1 W^n(r)dr = \Sigma^{\frac{1}{2}}\left(W^n - \int_0^1 W^n(r)dr\right).
\end{aligned}
$$

As in the preceding result,

$$
\frac{1}{T^{3/2}}\sum_{t=1}^{T} t\cdot\varepsilon_{t-1} \xrightarrow{d} \Sigma^{\frac{1}{2}}\cdot\int_0^1 r dW^n(r).
$$

x) Note that

$$
\begin{aligned}
\int_0^1 r\cdot X_T(r)dr &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\int_{[\frac{t-1}{T},\frac{t}{T})}\left(rS_{t-1} + (Tr^2 - (t-1)r)u_t\right)dr \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{1}{T^2}t\cdot S_{t-1} - \frac{1}{2T^2}S_{t-1}\right) + \frac{1}{6T^{5/2}}\sum_{t=1}^{T}(3t-1)u_t \\
&= \frac{1}{T^{5/2}}\sum_{t=1}^{T} t\cdot S_{t-1} - \frac{1}{2T^{5/2}}\sum_{t=1}^{T} S_{T-1} + \frac{1}{2T^{5/2}}\sum_{t=1}^{T} t\cdot u_t - \frac{1}{6T^{5/2}}\sum_{t=1}^{T} u_t.
\end{aligned}
$$

By implication,

$$
\begin{aligned}
\frac{1}{T^{5/2}}\sum_{t=1}^{T} t\cdot S_{t-1} &= \int_0^1 r\cdot X_T(r)dr + \frac{1}{2T^{5/2}}\sum_{t=1}^{T} S_{T-1} - \frac{1}{2T^{5/2}}\sum_{t=1}^{T} t\cdot u_t + \frac{1}{6T^{5/2}}\sum_{t=1}^{T} u_t \\
&= \bar{g}(X_T) + \frac{1}{2T^{5/2}}\sum_{t=1}^{T} S_{T-1} - \frac{1}{2T^{5/2}}\sum_{t=1}^{T} t\cdot u_t + \frac{1}{6T^{5/2}}\sum_{t=1}^{T} u_t \\
&\xrightarrow{p} \Lambda\cdot\left(\int_0^1 r\cdot W^n(r)dr\right)
\end{aligned}
$$

by the continuous mapping theorem and Slutsky's theorem.

To see that the convergence results hold jointly, note that each term can be expressed as the sum of, on the one hand, continuous functions of V^T, X^T and the partial sums $\mathcal{V}_T^0, \cdots, \mathcal{V}_T^p, \mathcal{U}_T^0, \cdots, \mathcal{U}_T^p$, and on the other, a term that converges to 0 in probability. Therefore, the continuous mapping theorem and Slutsky's theorem imply that the terms all converge to their weak limits jointly.

Q.E.D.

Items iv) and v) can be further simplified when $n=1$. In this case, denoting $\Sigma=\sigma^2>0$ and $\Sigma_u=\Psi(1)\Sigma\Psi(1)'=\sigma^2\Psi(1)^2$, the above theorem tells us that

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}S_{t-1}u_t &\xrightarrow{d}\frac{1}{2}\sigma^2\Psi(1)^2\int_0^1 W^1(r)dW^1(r)+\frac{\sigma^2\Psi(1)^2-\Gamma(0)}{2}\\ &=\frac{\Psi(1)^2}{2}\left(\sigma^2\int_0^1 W^1(r)dW^1(r)+\frac{\sigma^2\Psi(1)^2-\Gamma(0)}{\Psi(1)^2}\right),\end{aligned}$$

$$\frac{1}{T}\sum_{t=1}^{T}S_{t-1}u_{t-h}\xrightarrow{d}\frac{1}{2}\left[\sigma^2\Psi(1)^2W^1(1)^2+\Gamma(0)+2\sum_{j=1}^{h-1}\Gamma(j)\right].$$

Cointegration

Here, we investigate the properties of cointegrated time series using the limit results established above. We first define what is meant by I(1) and I(0) time series, and then define cointegration of I(1) processes. We derive the asymptotic properties of the Phillips-Ouliaris test for cointegration, and afterward move onto cointegrated VAR systems, or VECMs. There, we derive the Granger representation for such systems, and investigate methods, including Johansen's MLE approach and Ahn and Reinsel's least squares approach, to estimating VECMs.

As above, we let $(\Omega, \mathcal{H}, \mathbb{P})$ be our underlying probability space.

5.1 I(1) Processes and Cointegrated Time Series

5.1.1 I(0) Processes

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional time series. We say that $\{Y_t\}_{t\in\mathbb{Z}}$ is I(0) if it is a weakly stationary and causal linear process; that is, if

- There exist some $\mu \in \mathbb{R}^n$, an i.i.d. process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ and
- A one summable sequence $\{\Psi_j\}_{j\in\mathbb{N}}$ of $n \times n$ matrices such that
- For any $t \in \mathbb{Z}$,

$$Y_t = \mu + \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j},$$

  where the limit is both in L^2 and almost surely.

We limit the class of I(0) processes, which is often taken to be the class of weakly stationary processes in general, to the class of weakly stationary causal linear processes with i.i.d. innovations and one-summable coefficients, for analytical convenience. Note that any stationary VAR process with i.i.d. innovations satisfies the above conditions and is thus I(0), which demonstrates that our definition is not as restrictive as it first seems.

It is also important to note that the first difference of I(0) processes is again I(0). To see this, let the I(0) process $\{Y_t\}_{t\in\mathbb{Z}}$ be defined as

$$Y_t = \mu + \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j}$$

for any $t \in \mathbb{Z}$. Then, the first difference $\{\Delta Y_t\}_{t\in\mathbb{Z}}$ is given as

$$\begin{aligned}\Delta Y_t = Y_t - Y_{t-1} &= \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j} - \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-1-j} \\ &= \Psi_0 \varepsilon_t + \sum_{j=1}^{\infty} (\Psi_j - \Psi_{j-1}) \varepsilon_{t-j}\end{aligned}$$

for any $t \in \mathbb{Z}$, where the last equality is justified because the series $\sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j}$ converges absolutely with probability 1 for any $t \in \mathbb{Z}$. Defining $\Theta_0 = \Psi_0$ and $\Theta_j = \Psi_j - \Psi_{j-1}$ for any $j \in N_+$,

$$\begin{aligned}\sum_{j=0}^{\infty} j \cdot \|\Theta_j\| &\leq \sum_{j=0}^{\infty} j \cdot \|\Psi_j\| + \sum_{j=1}^{\infty} j \cdot \|\Psi_{j-1}\| \\ &= \sum_{j=0}^{\infty} j \cdot \|\Psi_j\| + \sum_{j=1}^{\infty} (j-1) \cdot \|\Psi_{j-1}\| + \sum_{j=1}^{\infty} \|\Psi_{j-1}\| \\ &= 2 \cdot \sum_{j=0}^{\infty} j \cdot \|\Psi_j\| + \sum_{j=0}^{\infty} \|\Psi_j\|.\end{aligned}$$

The rightmost term is finite due to the one-summability of $\{\Psi_j\}_{j\in\mathbb{N}}$, so it follows that $\{\Theta_j\}_{j\in\mathbb{N}}$ is also one-summable; it follows that the process $\{\Delta Y_t\}_{t\in\mathbb{Z}}$, written as

$$\Delta Y_t = \sum_{j=0}^{\infty} \Theta_j \cdot \varepsilon_{t-j} = \Theta(L)\varepsilon_t,$$

is an I(0) process with mean zero.

5.1.2 I(1) Processes

We say that the n-dimensional time series $\{Y_t\}_{t\in\mathbb{Z}}$ is I(d) for some $d \geq 1$ if its d difference process $\{(I_n - I_n L)^d Y_t\}_{t\in\mathbb{Z}}$ is I(0), or equivalently, a weakly stationary causal linear process with i.i.d. innovations and one-summable coefficients. Of special interest are I(1) processes, which are processes whose first difference process $\{\Delta Y_t\}_{t\in\mathbb{Z}}$ is I(0).

Suppose that $\{Y_t\}_{t\in\mathbb{Z}}$ is an I(1) process, and define $u_t = \Delta Y_t = Y_t - Y_{t-1}$ for any $t \in \mathbb{Z}$. Then, $\{u_t\}_{t\in\mathbb{Z}}$ is a weakly stationary causal linear process with mean $\delta \in \mathbb{R}^n$, i.i.d. innovation process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ and one-summable coefficients $\{\Psi_j\}_{j\in\mathbb{N}}$; by the Beveridge-Nelson decomposition, there exists a weakly stationary causal linear process $\{\eta_t\}_{t\in\mathbb{Z}}$ with mean zero such that

$$Y_t - Y_0 = \sum_{s=1}^{t} u_t = \delta t + \Psi(1) \cdot \sum_{s=1}^{t} \varepsilon_s + \eta_t - \eta_0,$$

or

$$Y_t = \underbrace{\delta t}_{\text{Deterministic Trend}} + \underbrace{\Psi(1) \cdot \sum_{s=1}^{t} \varepsilon_s}_{\text{Stochastic Trend}} + \underbrace{\eta_t}_{\text{Cycle}} + \underbrace{(Y_0 - \eta_0)}_{\text{Initial Values}}$$

for any $t \in \mathbb{N}$ with probability 1. We say that $\{Y_t\}_{t\in\mathbb{N}}$ is non-stationary if $\delta \neq \mathbf{0}$ or $\Psi(1) \neq O$; the latter condition is referred to as the "no MA unit root" condition. This definition is motivated by the fact that, if $\delta = \mathbf{0}$ and $\Psi(1) = O$, then for the inital value $Y_0 = \eta_0$, we have $Y_t = \eta_t$ for any $t \in \mathbb{N}$, which renders $\{Y_t\}_{t\in\mathbb{N}}$ weakly stationary.

5.1.3 Cointegrated Processes

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional I(1) process. $\{Y_t\}_{t\in\mathbb{N}}$ is said to be cointegrated if there exists a non-zero vector $\beta \in \mathbb{R}^n$ such that $\{\beta' Y_t\}_{t\in\mathbb{N}}$ is weakly stationary for some initial value Y_0. The vector β is called a cointegrating vector for $\{Y_t\}_{t\in\mathbb{N}}$.

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional I(1) process; we can carry over the Beveridge Nelson decomposition used in the previous section, so that

$$Y_t = \delta t + \Psi(1) \cdot \sum_{s=1}^{t} \varepsilon_s + \eta_t + (Y_0 - \eta_0)$$

for any $t \in \mathbb{N}$ with probability 1. This representation allows us to obtain a characterization of the space of all cointegration vectors, which we denote by V and refer to as the cointegrating space.

Suppose that $\beta \in \mathbb{R}^n$ is a cointegrating vector for $\{Y_t\}_{t\in\mathbb{N}}$. Then,

$$\beta' Y_t = \beta' \delta t + \beta' \Psi(1) \cdot \sum_{s=1}^{t} \varepsilon_s + \beta' \eta_t + (Y_0 - \eta_0)$$

defines a stationary process $\{\beta' Y_t\}_{t\in\mathbb{N}}$ for some choice of Y_0. If $\beta'\delta \neq 0$ or $\beta'\Psi(1) \neq \mathbf{0}'$, then for any choice of Y_0, $\beta' Y_t$ will possess either a deterministic time trend or a stochastic trend. Therefore, a necessary condition for cointegration under the cointegrating vector β is that

$$\beta'\delta = 0 \quad \text{and} \quad \Psi(1)'\beta = \mathbf{0}.$$

That is, $V \subset N_{\delta'} \cap N_{\Psi(1)'}$, where $N_{\delta'}$ and $N_{\Psi(1)'}$ are the null spaces of the linear transformations δ' and $\Psi(1)'$ defined on $\mathbb{R}^n$. Since the intersection of two vector spaces is also a vector space, $N_{\delta'} \cap N_{\Psi(1)'}$ is a linear subspace of $\mathbb{R}^n$.

On the other hand, suppose that $\beta \in N_{\delta'} \cap N_{\Psi(1)'}$ and β is non-zero. Then, $\beta'\delta = 0$ and

$\beta'\Psi(1) = \mathbf{0}'$, so that

$$\beta' Y_t = \beta' \eta_t + (Y_0 - \eta_0)$$

for any $t \in \mathbb{N}$, and given initial value $Y_0 = \eta_0$, $\beta' Y_t = \beta' \eta_t$ for any $t \in \mathbb{N}$. The weak stationarity of $\{\eta_t\}_{t\in\mathbb{Z}}$ implies that $\{\beta' Y_t\}_{t\in\mathbb{N}}$ is also weakly stationary, and as such β is a cointegrating vector for $\{Y_t\}_{t\in\mathbb{N}}$. We have thus shown that $(N_{\delta'} \cap N_{\Psi(1)'}) \setminus \{\mathbf{0}\} \subset V$, and putting the two results together, we can see that the space of all cointegrating vectors is characterized as

$$V = (N_{\delta'} \cap N_{\Psi(1)'}) \setminus \{\mathbf{0}\}.$$

This shows us that the cointegrating space V augmented by the zero vector $\mathbf{0}$ is a linear subspace of $\mathbb{R}^n$; for simplicity, denote the augmented cointegrating space by $\tilde{V} = V \cup \{\ \mathbf{0}\} = N_{\delta'} \cap N_{\Psi(1)'}$. The rank of $\tilde{V}$ is called the cointegrating rank, and any basis of $\tilde{V}$ is called a cointegrating basis for $\{Y_t\}_{t\in\mathbb{N}}$; since basis vectors must be non-zero, the cointegrating basis consists of cointegrating vectors.

Let q be the cointegrating rank and $\{\beta_1, \cdots, \beta_q\} \subset \tilde{V}$ a cointegrating basis. Define

$$B = \begin{pmatrix} \beta_1' \\ \vdots \\ \beta_q' \end{pmatrix}.$$

Because $\beta_i \in N_{\delta'} \cap N_{\Psi(1)'}$ for $1 \le i \le q$,

$$BY_t = B\eta_t$$

for any $t \in \mathbb{N}$ under the initial value $Y_0 = \eta_0$. Thus, $\{BY_t\}_{t\in\mathbb{N}}$ is a q-dimensional weakly stationary causal linear process.

Lemma (Cointegrating Rank of a Non-Stationary Cointegrated I(1) Process) Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional I(1) process. If $\{Y_t\}_{t\in\mathbb{N}}$ is non-stationary and cointegrated, then its cointegrating rank q cannot be 0 or n.

Proof) We continue to rely on the Beveridge-Nelson decomposition of Y_t, which tells us that

$$Y_t = \delta t + \Psi(1) \cdot \sum_{s=1}^{t} \varepsilon_s + \eta_t + (Y_0 - \eta_0)$$

for any $t \in \mathbb{N}$ with probability 1. Let V be the cointegrating space of $\{Y_t\}_{t\in\mathbb{N}}$, and $\tilde{V}$ its extension to a vector space. The cointegrating rank q is the dimension of $\tilde{V}$.

Suppose that $q = 0$. Then, $\tilde{V} = N_{\delta'} \cap N_{\Psi(1)'}$ contains only the zero vector, which implies that

$$V = (N_{\delta'} \cap N_{\Psi(1)'}) \setminus \{\mathbf{0}\} = \emptyset,$$

or that there are no cointegrating relationships among the elements of Y_t. This contradicts the fact that $\{Y_t\}_{t\in\mathbb{N}}$ is cointegrated, so we must have $q > 0$.

On the other hand, suppose that $q = n$. Then, $N_{\delta'} \cap N_{\Psi(1)'} = \mathbb{R}^n$, which implies that $N_{\delta'} = N_{\Psi(1)'} = \mathbb{R}^n$. Thus, $\delta = \mathbf{0}$ and $\Psi(1) = O$, which contradicts the non-stationarity of $\{Y_t\}_{t\in\mathbb{N}}$ ($\delta \neq \mathbf{0}$ or $\Psi(1) \neq O$). It follows that $q < n$.

Q.E.D.

Of particular interest is the case $\delta = \mathbf{0}$. In this case, the dimension of $\tilde{V} = N_{\Psi(1)'}$ is exactly the cointegrating rank. Denoting the cointegrating rank by r, by the dimension theorem we can see that

$$0 < rank(\Psi(1)) = n - r < n.$$

Since the rank of $\Psi(1)$ dictates the number of linearly independent and thus distinct random walks comprising the stochastic trend of Y_t, we call $n - r$ the number of common trends. Thus, the larger the cointegrating rank, or the more series are cointegrated, the more the dynamics of the n series are driven by a small number of common stochastic trends.

5.1.4 The Phillips-Ouliaris Test for Cointegration

Before studying a cointegrated system, we must first verify whether the system is cointegrated in the first place. The first of these types of tests was developed by Engle and Granger. Their test is based on the intuition that, if there is no cointegration between variables, then any linear combination of the variables is non-stationary. This indicates that the OLS residual from regressing one of the variables on the other ones must be non-stationary, and as such they propose running an ADF or PP type unit root test on the residuals. In this manner, the null of no cointegration is transformed into a null of a unit root in the residuals.

The Phillips-Ouliaris test extends the Engle-Granger test and derives the exact asymptotic distribution for the test statistic, which differs slightly from that of the ADF and PP tests. In addition, they implement a bias correction in the vein of the PP test to obtain a pivotal distribution for the test statistic. The exposition here follows that of Phillips and Ouliaris (1990).

The model is one in which there is at most one cointegrating relationship, where the variables have been ordered so that the last $n-1$ variables are not cointegrated. Formally, let $\{Y_t = (y_t, X_t')'\}_{t\in\mathbb{Z}}$ be an n-dimensional non-stationary I(1) process, where $\{X_t\}_{t\in\mathbb{Z}}$ is not cointegrated. Let $\{\Delta Y_t\}_{t\in\mathbb{Z}}$ be a weakly stationary causal linear process with one summable coefficients $\{\Psi_j\}_{j\in\mathbb{N}}$, mean $\delta \in \mathbb{R}^n$ and i.i.d. innovation process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ with positive definite variance $\Sigma \in \mathbb{R}^{n\times n}$ and finite fourth moments. By the Beveridge-Nelson decomposition, we have

$$Y_t = \delta t + \Psi(1)\cdot\sum_{s=1}^{t}\varepsilon_s + \eta_t + (Y_0 - \eta_0).$$

for any $t\in\mathbb{N}$ with probability 1, where $\{\eta_t\}_{t\in\mathbb{Z}}$ is an absolutely summable weakly stationary linear process. The augmented cointegrating space $\tilde{V}$ is given as

$$\tilde{V} = N_{\delta'} \cap N_{\Psi(1)'}.$$

Define the long run variance $\Sigma_u = \Psi(1)\Sigma\Psi(1)'$, and partition it as

$$\Sigma_u = \begin{pmatrix} \Sigma_{u,11} & \Sigma_{u,12} \\ \Sigma_{u,21} & \Sigma_{u,22} \end{pmatrix}$$

conformably with $(y_t, X_t')'$. We can now derive a convenient necessary and sufficient condition for cointegration using Σ_u:

Lemma (Characterization of Cointegration)
Let $\{Y_t = (y_t, X_t')'\}_{t\in\mathbb{Z}}$ be an n-dimensional non-stationary I(1) process, where $\{X_t\}_{t\in\mathbb{Z}}$ is not cointegrated. Retain the notations above, and let the long run variance Σ_u be defined and partitioned in the same way. Then, $\Sigma_{u,22}$ is positive definite.

Suppose, in addition, that $N_{\Psi(1)'} \subset N_{\delta'}$, so that the augmented cointegrating space $\tilde{V}$ equals the null space of $\Psi(1)'$. Then, $\{Y_t\}_{t\in\mathbb{Z}}$ is cointegrated if and only if

$$\Sigma_{u,1\cdot 2} = \Sigma_{u,11} - \Sigma_{u,12}\Sigma_{u,22}^{-1}\Sigma_{u,21}$$

is equal to 0.

Proof) We first show that $\Sigma_{u,22}$ is positive definite. Since $\{X_t\}_{t\in\mathbb{Z}}$ is not cointegrated, if

$$\begin{pmatrix} 0 & \beta' \end{pmatrix} \Psi(1) = \mathbf{0}'$$

for some $\beta \in \mathbb{R}^{n-1}$, then we must have $\beta = \mathbf{0}$; otherwise,

$$\beta' X_t = \beta' \eta_t^{(2)}$$

for any $t \in \mathbb{N}$ when the initial value $Y_0 = \eta_0$, where $\eta_t^{(2)}$ collects the last $n-1$ elements of η_t. this makes β a cointegrating vector for $\{X_t\}_{t\in\mathbb{N}}$, a contradiction.

Now choose some $\beta \in \mathbb{R}^{n-1}$; since Σ_u is positive semidefinite,

$$\beta'\Sigma_{u,22}\beta = \begin{pmatrix} 0 & \beta' \end{pmatrix} \Sigma_u \begin{pmatrix} 0 \\ \beta \end{pmatrix} \geq 0,$$

which tells us that $\Sigma_{u,22}$ is also positive semidefinite. To show that $\Sigma_{u,22}$ is positive definite, suppose $\beta'\Sigma_{u,22}\beta = 0$. Then,

$$0 = \begin{pmatrix} 0 & \beta' \end{pmatrix} \Sigma_u \begin{pmatrix} 0 \\ \beta \end{pmatrix} = \begin{pmatrix} 0 & \beta' \end{pmatrix} \Psi(1)\Sigma\Psi(1)' \begin{pmatrix} 0 \\ \beta \end{pmatrix}.$$

Since Σ is positive definite, this implies that

$$\Psi(1)' \begin{pmatrix} 0 \\ \beta \end{pmatrix} = \mathbf{0},$$

and we just showed above that this implies $\beta = \mathbf{0}$. Therefore, $\beta'\Sigma_{u,22}\beta > 0$ if and only if β is non-zero, which tells us that $\Sigma_{u,22}$ is positive definite.

Define $\Sigma_{u,1\cdot 2}$ as above. Then, since $\Sigma_{u,1\cdot 2}$ is the Schur complement of the positive semidefinite matrix Σ_u,

$$\det(\Sigma_u) = \det(\Sigma_{u,1\cdot 2}) \cdot \det(\Sigma_{u,22}) \geq 0,$$

and since $\det(\Sigma_{u,22}) > 0$, we have $\det(\Sigma_{u,1\cdot 2}) = \Sigma_{u,1\cdot 2} \geq 0$.

Suppose $\Sigma_{u,1\cdot 2} > 0$. Then, $\det(\Sigma_u) > 0$, and thus

$$\Sigma_u = \Psi(1)\Sigma\Psi(1)'$$

is positive definite. Σ_u and $\Psi(1)'$ share the same null space and are linear transformations on the same space $\mathbb{R}^n$, so it follows that they have the same rank; in other words, $\Psi(1)$ has full rank n. Otherwise, if $\Sigma_{u,1\cdot 2} = 0$, then Σ_u, and by extension $\Psi(1)$, has rank $n-1$.

Assume now that the augmented cointegrating space of $\{Y_t\}_{t\in\mathbb{Z}}$ is $\tilde{V} = N_{\Psi(1)'}$. If $\Sigma_{u,1\cdot 2} = 0$, then $N_{\Psi(1)'}$ is a subspace of dimension 1 and thus $\tilde{V}$ contains a non-zero vector, meaning $\{Y_t\}_{t\in\mathbb{Z}}$ is cointegrated. Otherwise, $\tilde{V} = \{\mathbf{0}\}$ and $\{Y_t\}_{t\in\mathbb{Z}}$ is not cointegrated.

Q.E.D.

In light of the above result, the null of no cointegration is equivalent to putting $\Sigma_{u,1\cdot 2} > 0$.

The Phillips-Ouliaris test considers the OLS residual $\hat{u}_t$ from regressing y_t on an intercept and X_t; letting $\hat{\mu}_T$ and $\hat{\beta}_T$ be the OLS estimators of the intercept and coefficient terms estimated using the sample up to time T,

$$\hat{u}_t = y_t - \hat{\mu}_T - \hat{\beta}_T' X_t$$

for any $1 \leq t \leq T$. We consider the asymptotic behavior of the Dickey-Fuller test statistic

$$\hat{\rho}_T = \frac{\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^{T} \hat{u}_{t-1}^2}.$$

Intuitively, if $\hat{\rho}_T$ is close to 1, then $\hat{u}_t$ is close to a unit root process and it is likely that y_t is cointegrated with X_t. Thus, we reject the null of no cointegration if $\hat{\rho}_T$ is significantly smaller than 1. To define what is meant by "significantly smaller", we derive the asymptotic distribution of $\hat{\rho}_T$ below:

Theorem (Asymptotic Distribution of Phillips-Ouliaris Test Statistic)
Let $\{Y_t = (y_t, X_t')'\}_{t\in\mathbb{Z}}$ be an n-dimensional non-stationary I(1) process, where $\{X_t\}_{t\in\mathbb{Z}}$ is not cointegrated. Retain the notations above, and assume that $\delta = \mathbf{0}$, so that there is no deterministic time trend. Suppose the autocovariance function of the mean zero weakly stationary linear process $\{\Delta Y_t = e_t\}_{t\in\mathbb{Z}}$ is given as $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$, and let $\lambda = \sum_{j=1}^{\infty} \Gamma(j)'$. Let $\{W^n(r) = (W_1^n(r), W_2^n(r)')'\}_{r\in[0,1]}$ be the conformably partitioned n-dimensional standard Wiener process on $[0,1]$.

Let $\{W^{n*}(r) = (W_1^{n*}(r), W_2^{n*}(r)')'\}_{r\in[0,1]}$ be the n-dimensional Brownian bridge defined as

$$W^{n*}(r) = W^n(r) - \int_0^1 W^n(s)ds$$

for any $r \in [0,1]$, and let $\{Q(r)\}_{r\in[0,1]}$ be the stochastic process defined as

$$Q(r) = W_1^{n*}(r) - \left(\int_0^1 W_1^{n*}(r)W_2^{n*}(r)'dr\right)\left(\int_0^1 W_2^{n*}(r)W_2^{n*}(r)'dr\right)^{-1} W_2^{n*}(r)$$

for any $r \in [0,1]$. Then, under the null of no cointegration, there exists an n-dimensional random vector η such tat

$$T(\hat{\rho}_T - 1) \xrightarrow{d} \frac{\int_0^1 Q(r)dQ(r)}{\int_0^1 Q(r)^2 dr} + \frac{1}{\int_0^1 Q(r)^2 dr} \cdot \frac{1}{2\Sigma_{u,1\cdot 2}} \eta' (\Sigma_u - \Gamma(0)) \eta.$$

Proof) By the previous lemma, $\Sigma_{u,22}$ is positive definite, and if $\{Y_t\}_{t\in\mathbb{Z}}$ is not cointegrated, $\Sigma_{u,1\cdot 2} > 0$. Denoting $\Delta Y_t = e_t = \Psi(L)\varepsilon_t$ for any $t \in \mathbb{Z}$, define the process $\{Z_t\}_{t\in\mathbb{N}}$ as

$$Z_t = Y_t - Y_0 = \sum_{s=1}^{t} e_s$$

for any $t \in \mathbb{N}$; note that $Z_0 = 0$. Letting $\{B^n(r) = (B_1^n(r), B_2^n(r)')'\}_{r\in[0,1]}$ be the n-dimensional Brownian motion with variance Σ_u, by the asymptotic results derived above we know that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_t \xrightarrow{d} B^n(1)$$

$$\frac{1}{T} \sum_{t=1}^{T} e_t e_t' \xrightarrow{p} \Gamma(0)$$

$$\frac{1}{T} \sum_{t=1}^{T} Z_{t-1} e_t' \xrightarrow{d} \int_0^1 B^n(r)dB^n(r)' + \underbrace{\Sigma_u - \sum_{j=0}^{\infty} \Gamma(j)}_{\lambda}$$

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} Z_{t-1} \xrightarrow{d} \int_0^1 B^n(r)dr$$

$$\frac{1}{T^2}\sum_{t=1}^{T} Z_{t-1}Z'_{t-1} \xrightarrow{d} \int_0^1 B^n(r)B^n(r)'dr.$$

Since the only difference between Z_t and Y_t is the presence of the initial value Y_0, the above results hold unchanged for Y_t in the place of Z_t. Furthermore, since $Y_t = Y_{t-1} + e_t$,

$$\begin{aligned}\frac{1}{T^{3/2}}\sum_{t=1}^{T} Y_t &= \frac{1}{T^{3/2}}\sum_{t=1}^{T} Y_{t-1} + \frac{1}{T^{3/2}}\sum_{t=1}^{T} e_t \\ &\xrightarrow{d} \int_0^1 B^n(r)dr \\ \frac{1}{T^2}\sum_{t=1}^{T} Y_tY'_t &= \frac{1}{T^2}\sum_{t=1}^{T} Y_{t-1}Y'_{t-1} + \frac{1}{T^2}\sum_{t=1}^{T} Y_{t-1}e'_t + \frac{1}{T^2}\sum_{t=1}^{T} e_tY'_{t-1} + \frac{1}{T^2}\sum_{t=1}^{T} e_te'_t \\ &\xrightarrow{d} \int_0^1 B^n(r)B^n(r)'dr.\end{aligned}$$

Note that Σ_u is nonsingular due to the assumption of no integration, which means $\int_0^1 B^n(r)B^n(r)'dr$ is positive definite-valued.

We now proceed in steps.

Step 1: Asymptotic Distribution of the OLS Estimators

Let $(\hat{\mu}_T, \hat{\beta}'_T)'$ be the OLS estimators from the regression of y_t on X_t. By definition,

$$\begin{pmatrix}\hat{\mu}_T \\ \hat{\beta}_T\end{pmatrix} = \begin{pmatrix} T & \sum_{t=1}^T X'_t \\ \sum_{t=1}^T X_t & \sum_{t=1}^T X_tX'_t\end{pmatrix}^{-1}\begin{pmatrix}\sum_{t=1}^T y_t \\ \sum_{t=1}^T X_ty_t\end{pmatrix},$$

and the residuals are given as

$$\hat{u}_t = y_t - \begin{pmatrix}1 & X'_t\end{pmatrix}\begin{pmatrix}\hat{\mu}_T \\ \hat{\beta}_T\end{pmatrix}$$

for any $1 \le t \le T$. We can scale the estimators above as

$$\begin{aligned}\begin{pmatrix}\sqrt{T}\hat{\mu}_T \\ T\cdot\hat{\beta}_T\end{pmatrix} &= \begin{pmatrix}\sqrt{T} & O \\ O & T\cdot I_{n-1}\end{pmatrix}\begin{pmatrix}\hat{\mu}_T \\ \hat{\beta}_T\end{pmatrix} \\ &= \begin{pmatrix} 1 & \frac{1}{T^{3/2}}\sum_{t=1}^T X'_t \\ \frac{1}{T^{3/2}}\sum_{t=1}^T X_t & \frac{1}{T^2}\sum_{t=1}^T X_tX'_t\end{pmatrix}^{-1}\begin{pmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^T y_t \\ \frac{1}{T}\sum_{t=1}^T X_ty_t\end{pmatrix}.\end{aligned}$$

Note how the last term diverges because the speed at which they converge is $T^{3/2}$ and

T^2. Thus, we divide both sides by T, which reveals that

$$\begin{pmatrix} \frac{1}{\sqrt{T}}\hat{\mu}_T \\ \hat{\beta}_T \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{T^{3/2}}\sum_{t=1}^T X_t' \\ \frac{1}{T^{3/2}}\sum_{t=1}^T X_t & \frac{1}{T^2}\sum_{t=1}^T X_t X_t' \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{T^{3/2}}\sum_{t=1}^T y_t \\ \frac{1}{T^2}\sum_{t=1}^T X_t y_t \end{pmatrix}$$

$$\xrightarrow{d} \begin{pmatrix} 1 & \int_0^1 B_2^n(r)'dr \\ \int_0^1 B_2^n(r)dr & \int_0^1 B_2^n(r)B_2^n(r)'dr \end{pmatrix}^{-1} \begin{pmatrix} \int_0^1 B^1(r)dr \\ \int_0^1 B_2^n(r)B_1^n(r)'dr \end{pmatrix}.$$

This shows us that $\hat{\beta}_T$ is $O_p(1)$, and that $\hat{\mu}_T$ diverges. In particular, using the formula for block matrix inversion, we can see that

$$\begin{pmatrix} \frac{1}{\sqrt{T}}\hat{\mu}_T \\ \hat{\beta}_T \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \int_0^1 B_1^n(r)dr - \left(\int_0^1 B_2^n(r)'dr\right)\left(\int_0^1 B_2^{n*}(r)B_2^{n*}(r)'dr\right)^{-1}\left(\int_0^1 B_2^{n*}(r)B_1^{n*}(r)dr\right) \\ \left(\int_0^1 B_2^{n*}(r)B_2^{n*}(r)'dr\right)^{-1}\left(\int_0^1 B_2^{n*}(r)B_1^{n*}(r)dr\right) \end{pmatrix},$$

where $\{B^{n*}(r) = (B_1^{n*}(r), B_2^{n*}(r)')'\}_{r\in[0,1]}$ is the Brownian bridge defined as

$$B^{n*}(r) = B^n(r) - \int_0^1 B^n(s)ds$$

for any $r \in [0,1]$. Note the similarities with the usual OLS intercept and slope estimators.

Step 2: Asymptotic Distribution of Test Statistic

Using the above result, we can derive the asymptotic distribution of the denominator of the test statistic $\hat{\rho}_T$ as follows:

$$\begin{aligned}
\frac{1}{T^2}D_T &= \frac{1}{T^2}\sum_{t=2}^T \hat{u}_{t-1}^2 \\
&= \frac{1}{T^2}\sum_{t=2}^T y_{t-1}^2 - 2\begin{pmatrix} \frac{1}{T^{3/2}}\sum_{t=2}^T y_{t-1} & \frac{1}{T^2}\sum_{t=2}^T y_{t-1}X_{t-1}' \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{T}}\hat{\mu}_T \\ \hat{\beta}_T \end{pmatrix} \\
&\quad + \begin{pmatrix} \frac{1}{\sqrt{T}}\hat{\mu}_T & \hat{\beta}_T' \end{pmatrix}\begin{pmatrix} \frac{T-1}{T} & \frac{1}{T^{3/2}}\sum_{t=2}^T X_{t-1}' \\ \frac{1}{T^{3/2}}\sum_{t=2}^T X_{t-1} & \frac{1}{T^2}\sum_{t=2}^T X_{t-1}X_{t-1}' \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{T}}\hat{\mu}_T \\ \hat{\beta}_T \end{pmatrix} \\
&\xrightarrow{d} \int_0^1 B_1^n(r)^2 dr \\
&\quad - \begin{pmatrix} \int_0^1 B_1^n(r)dr & \int_0^1 B_1^n(r)B_2^n(r)'dr \end{pmatrix} \\
&\quad \times \begin{pmatrix} 1 & \int_0^1 B_2^n(r)'dr \\ \int_0^1 B^n(r)dr & \int_0^1 B_2^n(r)B_2^n(r)'dr \end{pmatrix}^{-1}\begin{pmatrix} \int_0^1 B_1^n(r)dr \\ \int_0^1 B_2^n(r)B_1^n(r)'dr \end{pmatrix}
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^1 B_1^n(r)^2 dr - \left(\int_0^1 B_1^n(r)dr\right)^2 \\
&\quad + \left(\int_0^1 B_1^n(r)dr\right)\left(\int_0^1 B_2^n(r)'dr\right)\left(\int_0^1 B_2^{n*}(r)B_2^{n*}(r)'dr\right)^{-1}\left(\int_0^1 B_2^{n*}(r)B_1^{n*}(r)dr\right) \\
&\quad - \left(\int_0^1 B_1^n(r)B_2^n(r)'dr\right)\left(\int_0^1 B_2^{n*}(r)B_2^{n*}(r)'dr\right)^{-1}\left(\int_0^1 B_2^{n*}(r)B_1^{n*}(r)dr\right) \\
&= \int_0^1 B_1^{n*}(r)^2 dr - \left(\int_0^1 B_1^{n*}(r)B_2^{n*}(r)'dr\right)\left(\int_0^1 B_2^{n*}(r)B_2^{n*}(r)'dr\right)^{-1}\left(\int_0^1 B_2^{n*}(r)B_1^{n*}(r)dr\right).
\end{aligned}
$$

Defining the univariate stochastic process $\{J(r)\}_{r\in[0,1]}$ as

$$
J(r) = B_1^{n*}(r) - \left(\int_0^1 B_1^{n*}(s)B_2^{n*}(s)'ds\right)\left(\int_0^1 B_2^{n*}(s)B_2^{n*}(s)'dr\right)^{-1} B_2^{n*}(r)
$$

for any $r \in [0,1]$, the above result can be simply written as

$$
\frac{1}{T^2}D_T \xrightarrow{d} \int_0^1 J(r)^2 dr.
$$

Likewise, since

$$
\begin{aligned}
\hat{u}_t &= y_t - \hat{\mu}_T - \hat{\beta}_T' X_t \\
&= y_{t-1} - \hat{\mu}_T - \hat{\beta}_T' X_{t-1} + e_{1t} - \hat{\beta}_T' e_{2t} = \hat{u}_{t-1} + e_{1t} - \hat{\beta}_T' e_{2t},
\end{aligned}
$$

where $e_t = (e_{1t}, e_{2t}')'$ is the conformable partition of the first difference e_t, the numerator of the test statistic is given as

$$
\begin{aligned}
N_T &= D_T + \sum_{t=2}^T \left(e_{1t} - \hat{\beta}_T' e_{2t}\right)\hat{u}_{t-1} \\
&= D_T + \sum_{t=2}^T e_{1t}y_{t-1} - \hat{\beta}_T' \sum_{t=2}^T e_{2t}y_{t-1} \\
&\quad - \begin{pmatrix} \sum_{t=2}^T e_{1t} & \sum_{t=2}^T e_{1t}X_{t-1}' \end{pmatrix}\begin{pmatrix} \hat{\mu}_T \\ \hat{\beta}_T \end{pmatrix} + \hat{\beta}_T' \begin{pmatrix} \sum_{t=2}^T e_{2t} & \sum_{t=2}^T e_{2t}X_{t-1}' \end{pmatrix}\begin{pmatrix} \hat{\mu}_T \\ \hat{\beta}_T \end{pmatrix}.
\end{aligned}
$$

Denoting $N_T - D_T = \Delta_T$, we can write

$$
\begin{aligned}
\frac{1}{T}\Delta_T &= \frac{1}{T}\sum_{t=2}^T e_{1t}y_{t-1} - \hat{\beta}_T' \cdot \frac{1}{T}\sum_{t=2}^T e_{2t}y_{t-1} \\
&\quad - \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=2}^T e_{1t} & \frac{1}{T}\sum_{t=2}^T e_{1t}X_{t-1}' \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{T}}\hat{\mu}_T \\ \hat{\beta}_T \end{pmatrix} + \hat{\beta}_T' \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=2}^T e_{2t} & \frac{1}{T}\sum_{t=2}^T e_{2t}X_{t-1}' \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{T}}\hat{\mu}_T \\ \hat{\beta}_T \end{pmatrix}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=2}^{T} e_{1t}y_{t-1} - \left(\tfrac{1}{\sqrt{T}}\sum_{t=2}^{T} e_{1t} \quad \tfrac{1}{T}\sum_{t=2}^{T} e_{1t}X_{t-1}'\right)\begin{pmatrix}\frac{1}{\sqrt{T}}\hat{\mu}_T \\ \hat{\beta}_T\end{pmatrix} \\
&\qquad \xrightarrow{d} \int_0^1 B_1^n(r)dB_1^n(r) + \lambda_{11} \\
&- B_1(1)\cdot\int_0^1 B_1^n(r)dr + B_1(1)\cdot\left(\int_0^1 B_2^n(r)'dr\right)\left(\int_0^1 B_2^{n*}(r)B_2^{n*}(r)'dr\right)^{-1}\left(\int_0^1 B_2^{n*}(r)B_1^{n*}(r)dr\right) \\
&\quad - \left(\int_0^1 B_2(r)dB_1(r) + \lambda_{12}\right)'\left(\int_0^1 B_2^{n*}(r)B_2^{n*}(r)'dr\right)^{-1}\left(\int_0^1 B_2^{n*}(r)B_1^{n*}(r)dr\right) \\
&= \int_0^1 B_1^{n*}(r)dB_1^n(r) - \left(\int_0^1 B_2^{n*}(r)dB_1^n(r)\right)'\left(\int_0^1 B_2^{n*}(r)B_2^{n*}(r)'dr\right)^{-1}\left(\int_0^1 B_2^{n*}(r)B_1^{n*}(r)dr\right) \\
&\qquad + \lambda_{11} - \lambda_{12}'\left(\int_0^1 B_2^{n*}(r)B_2^{n*}(r)'dr\right)^{-1}\left(\int_0^1 B_2^{n*}(r)B_1^{n*}(r)dr\right) \\
&\qquad = \int_0^1 J(r)dB_1^n(r) \\
&\qquad \lambda_{11} - \lambda_{12}'\left(\int_0^1 B_2^{n*}(r)B_2^{n*}(r)'dr\right)^{-1}\left(\int_0^1 B_2^{n*}(r)B_1^{n*}(r)dr\right),
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=2}^{T} e_{2t}y_{t-1} - \left(\tfrac{1}{\sqrt{T}}\sum_{t=2}^{T} e_{2t} \quad \tfrac{1}{T}\sum_{t=2}^{T} e_{2t}X_{t-1}'\right)\begin{pmatrix}\frac{1}{\sqrt{T}}\hat{\mu}_T \\ \hat{\beta}_T\end{pmatrix} \\
&\xrightarrow{d} \int_0^1 B_1^{n*}(r)dB_2^n(r)' - \left(\int_0^1 B_2^{n*}(r)dB_2^n(r)\right)'\left(\int_0^1 B_2^{n*}(r)B_2^{n*}(r)'dr\right)^{-1}\left(\int_0^1 B_2^{n*}(r)B_1^{n*}(r)dr\right) \\
&\qquad + \lambda_{21}' - \lambda_{22}'\left(\int_0^1 B_2^{n*}(r)B_2^{n*}(r)'dr\right)^{-1}\left(\int_0^1 B_2^{n*}(r)B_1^{n*}(r)dr\right) \\
&\qquad = \int_0^1 J(r)dB_2^n(r)' \\
&\qquad + \lambda_{21}' - \lambda_{22}'\left(\int_0^1 B_2^{n*}(r)B_2^{n*}(r)'dr\right)^{-1}\left(\int_0^1 B_2^{n*}(r)B_1^{n*}(r)dr\right).
\end{aligned}
$$

Therefore,

$$
\frac{1}{T}\Delta_T \xrightarrow{d} \int_0^1 J(r)dJ(r) + \eta'\lambda\eta,
$$

where we define

$$
\eta = \begin{pmatrix} 1 \\ -\left(\int_0^1 B_2^{n*}(r)B_2^{n*}(r)'dr\right)^{-1}\left(\int_0^1 B_2^{n*}(r)B_1^{n*}(r)dr\right)\end{pmatrix}.
$$

The convergence results above all hold jointly, so

$$
T(\hat{\rho}_T - 1) = T\left(\frac{N_T}{D_T} - 1\right) = T\cdot\frac{\Delta_T}{D_T} = \frac{\frac{1}{T}\Delta_T}{\frac{1}{T^2}D_T}
$$

$$\xrightarrow{d} \frac{\int_0^1 J(r)dJ(r)}{\int_0^1 J(r)^2 dr} + \frac{1}{\int_0^1 J(r)^2 dr}\eta'\lambda\eta.$$

Examining the nuisance parameter term $\eta'\lambda\eta$, we find that

$$\begin{aligned}
\eta'\lambda\eta &= \eta'\left(\sum_{j=1}^{\infty}\Gamma(j)'\right)\eta \\
&= \eta'\left(\Sigma_u - \Gamma(0)\right)\eta - \eta'\left(\sum_{j=1}^{\infty}\Gamma(j)\right)\eta \\
&= \eta'\left(\Sigma_u - \Gamma(0)\right)\eta - \eta'\lambda\eta.
\end{aligned}$$

Therefore,

$$\eta'\lambda\eta = \frac{1}{2}\eta'\left(\Sigma_u - \Gamma(0)\right)\eta.$$

Step 3: Simplifying the Asymptotic Distributions

Define $L_{22} = \Sigma_{u,22}^{\frac{1}{2}}$ as the Cholesky factor of $\Sigma_{u,22}$, and

$$\begin{aligned}
l_{11} &= \sqrt{\Sigma_{u,1\cdot 2}} := \sqrt{\Sigma_{u,11} - \Sigma_{u,12}\Sigma_{u,22}^{-1}\Sigma_{u,21}} > 0 \\
L_{12} &= \Sigma_{u,12}\left(\Sigma_{u,22}^{\frac{1}{2}\prime}\right)^{-1}
\end{aligned}$$

Then, letting

$$L = \begin{pmatrix} l_{11} & L_{12} \\ O & L_{22} \end{pmatrix} \in \mathbb{R}^{n\times n},$$

we have

$$LL' = \begin{pmatrix} l_{11}^2 + L_{12}L_{12}' & L_{12}L_{22}' \\ L_{22}L_{12}' & L_{22}L_{22}' \end{pmatrix} = \Sigma_u.$$

Therefore, LW^n and B^n are both n-dimensional Brownian motions with variance Σ_u; it follows that $LW^n \sim B^n$.

This implies that

$$\begin{pmatrix} B_1^n \\ B_2^n \end{pmatrix} = \begin{pmatrix} l_{11} & L_{12} \\ O & L_{22} \end{pmatrix}\begin{pmatrix} W_1^n \\ W_2^n \end{pmatrix} = \begin{pmatrix} l_{11}W_1^n + L_{12}W_2^n \\ L_{22}W_2^n \end{pmatrix},$$

so that

$$\left(\int_0^1 B_2^{n*}(r)B_2^{n*}(r)'dr\right)^{-1}\left(\int_0^1 B_2^{n*}(r)B_1^{n*}(r)dr\right)$$

$$\sim L_{22}'^{-1}\left(\int_0^1 W_2^{n*}(r)W_2^{n*}(r)'\right)^{-1}\left(l_{11}\cdot\int_0^1 W_2^{n*}(r)W_1^{n*}(r)dr+\int_0^1 W_2^{n*}(r)W_2^{n*}(r)drL_{12}'\right)$$

$$=l_{11}L_{22}'^{-1}\left(\int_0^1 W_2^{n*}(r)W_2^{n*}(r)'\right)^{-1}\left(\int_0^1 W_2^{n*}(r)W_1^{n*}(r)dr\right)+L_{22}'^{-1}L_{12}'.$$

It follows that

$$J(r)=B_1^{n*}(r)-\left(\int_0^1 B_1^{n*}(s)B_2^{n*}(s)'ds\right)\left(\int_0^1 B_2^{n*}(s)B_2^{n*}(s)'dr\right)^{-1}B_2^{n*}(r)$$

$$\sim l_{11}\left[W_1^{n*}(r)-\left(\int_0^1 W_2^{n*}(r)'W_1^{n*}(r)dr\right)\left(\int_0^1 W_2^{n*}(r)W_2^{n*}(r)'\right)^{-1}W_2^{n*}(r)\right]=l_{11}Q(r)$$

for any $r\in[0,1]$. The asymptotic distribution of the test statistic is now given as

$$T(\hat{\rho}_T-1)\xrightarrow{d}\frac{\int_0^1 Q(r)dQ(r)}{\int_0^1 Q(r)^2dr}+\frac{1}{\int_0^1 Q(r)^2dr}\frac{1}{2\Sigma_{u,1\cdot 2}}\eta'\left(\Sigma_u-\Gamma(0)\right)\eta.$$

Q.E.D.

As expected, the test statistic $\hat{\rho}_T$ is superconsistent for 1 under the null of no cointegration; if there is no cointegration, no linear combination of the variables yields a stationary process. However, a nuisance term is present on the right hand side, which prevents us from actually implementing the test in practice. For this reason, Phillips and Ouliaris propose a bias correction in the same vein as the PP test.

The bias correction term is based on the observation that

$$\begin{pmatrix}1\\-\hat{\beta}_T\end{pmatrix}$$

is a consistent estimator of the random vector η. Thus, provided that there exists a consistent estimator $\hat{\Sigma}_u$ of Σ_u and a consistent estimator $\hat{\sigma}^2$ of $\Gamma(0)$, it follows that

$$\begin{pmatrix}1 & -\hat{\beta}_T'\end{pmatrix}\left(\hat{\Sigma}_u-\hat{\sigma}^2\right)\begin{pmatrix}1\\-\hat{\beta}_T\end{pmatrix}\xrightarrow{p}\eta'\left(\Sigma_u-\Gamma(0)\right)\eta.$$

Furthermore, we already saw that

$$\frac{1}{T^2}\sum_{t=2}^T\hat{u}_{t-1}^2\xrightarrow{d}\int_0^1 J(r)^2dr\sim\Sigma_{u,1\cdot 2}\cdot\int_0^1 Q(r)^2dr.$$

Therefore, we can construct the bias-corrected Phillips-Ouliaris test statistic

$$\hat{Z}_T = T(\hat{\rho}_T - 1) - \frac{1}{2}\begin{pmatrix} 1 & -\hat{\beta}_T' \end{pmatrix}\left(\hat{\Sigma}_u - \hat{\sigma}^2\right)\begin{pmatrix} 1 \\ -\hat{\beta}_T \end{pmatrix} \cdot \left(\frac{1}{T^2}\sum_{t=2}^{T}\hat{u}_{t-1}^2\right)^{-1}.$$

By the result in the preceding theorem,

$$\hat{Z}_T \xrightarrow{d} \frac{\int_0^1 Q(r)dQ(r)}{\int_0^1 Q(r)^2 dr},$$

where the right hand side is a pivotal distribution.

5.2 Estimating Cointegrating Relationships

Here we study methods to estimate for cointegrating relationships in the case that the cointegration rank is 1. We start with the Static OLS (SOLS) approach to estimating the cointegrating vector, that is, via ordinary least squares from the first equation in the triangular representation. Due to the presence of nuisance parameters in the asymptotic distribution of the SOLS estimator, we introduce the Fully Modified OLS (FM-OLS) estimator, which provides a non-parametric correction that yields a pivotal asymptotic distribution.

First we study Phillips' triangular representation, which provides the basis for the estimation methods studied later.

5.2.1 Phillips' Triangular Representation

Recall that the augmented cointegrating space is a linear subspace. This indicates that any linear combination of cointegrating vectors is again a cointegrating vector, and we can exploit this feature to derive a convenient representation for cointegrated systems.

Transforming the Cointegrating Basis

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a non-stationary and cointegrated n-dimensional I(1) process. Letting $\{\Delta Y_t\}_{t\in\mathbb{Z}}$ be a weakly stationary causal linear process with one summable coefficients $\{\Psi_j\}_{j\in\mathbb{N}}$ and mean $\delta \in \mathbb{R}^n$, we showed above that the augmented cointegrating space $\tilde{V}$ is the intersection of the null spaces of $\delta' \in \mathbb{R}^{1\times n}$ and $\Psi(1)' \in \mathbb{R}^{n\times n}$:

$$\tilde{V} = N_{\delta'} \cap N_{\Psi(1)'}.$$

Let the cointegration rank be $0 < q < n$ and choose a cointegrating basis $\{\beta_1, \cdots, \beta_q\} \subset \tilde{V}$.

We proceed by induction. Since $\beta_1 \neq \mathbf{0}$, it has a non-zero element; assume that the elements $Y_{1t}, \cdots, Y_{nt}$ are ordered so that $\beta_{11} \neq 0$. Then, we can define

$$\{\beta_1^{(1)}, \cdots, \beta_q^{(1)}\} \subset \mathbb{R}^n$$

as

$$\beta_1^{(1)} = \frac{1}{\beta_{11}}\beta_1 \quad \text{and} \quad \beta_i^{(1)} = \beta_i - \frac{\beta_{1i}}{\beta_{11}}\beta_1 \quad \text{for any } 2 \leq i \leq q.$$

Then, since each $\beta_i^{(1)}$ is a linear combination of $\beta_1, \cdots, \beta_q$, they are contained in $\tilde{V}$. To see that $\{\beta_1^{(1)}, \cdots, \beta_q^{(1)}\}$ is linearly independent, let

$$\sum_{i=1}^{q} r_i \cdot \beta_i^{(1)} = \mathbf{0}$$

for some $r_1, \cdots, r_n \in \mathbb{R}$; then,

$$\sum_{i=1}^{q} r_i \cdot \beta_i^{(1)} = \frac{1}{\beta_{11}} \left(r_1 - \sum_{i=2}^{q} r_i \right) \beta_1 + \sum_{i=2}^{q} r_i \cdot \beta_i = \mathbf{0},$$

and since the cointegrating basis consists of linearly independent vectors,

$$r_2 = \cdots = r_q = \frac{1}{\beta_{11}} \left(r_1 - \sum_{i=2}^{q} r_i \right) = 0,$$

which implies that $r_1 = 0$ as well. By definition, $\{\beta_1^{(1)}, \cdots, \beta_q^{(1)}\}$ is a collection of q linearly independent subset of $\tilde{V}$; it is thus a cointegrating basis. In particular, the first column of the $q \times n$ matrix

$$B^{(1)} = \begin{pmatrix} \beta_1^{(1)\prime} \\ \vdots \\ \beta_q^{(1)\prime} \end{pmatrix}$$

is the first standard basis vector in $\mathbb{R}^q$.

Now suppose, for some $1 \leq k < q$, that we have constructed a cointegrating basis $\{\beta_1^{(k)}, \cdots, \beta_q^{(k)}\} \subset \tilde{V}$ such that the first k columns of the $q \times n$ matrix

$$B^{(k)} = \begin{pmatrix} \beta_1^{(k)\prime} \\ \vdots \\ \beta_q^{(k)\prime} \end{pmatrix}$$

are the first k standard basis vectors in $\mathbb{R}^q$. Since $\beta_{k+1}^{(k)}$ is a non-zero vector, it contains a non-zero value; since the first k elements of $\beta_{k+1}^{(k)}$ are equal to 0, the non-zero element must be found among the last $n-k$ elements of this vector. Without loss of generality, assume that $Y_{k+1,t}, \cdots, Y_{nt}$ have been arranged so that $\beta_{k+1,k+1}^{(k)} \neq 0$. We now proceed identically as above; define

$$\{\beta_1^{(k+1)}, \cdots, \beta_q^{(k+1)}\} \subset \mathbb{R}^n$$

as

$$\beta_{k+1}^{(k+1)} = \frac{1}{\beta_{k+1,k+1}^{(k)}} \beta_{k+1}^{(k)} \quad \text{and} \quad \beta_i^{(k+1)} = \beta_i^{(k)} - \frac{\beta_{k+1,i}^{(k)}}{\beta_{k+1,k+1}^{(k)}} \beta_{k+1}^{(k)} \quad \text{for any } i \neq k+1.$$

Since the first k elements of $\beta_{k+1}^{(k)}$ are equal to 0, the first k elements of $\beta_1^{(k+1)}, \cdots, \beta_q^{(k+1)}$ are identical to those of $\beta_1^{(k)}, \cdots, \beta_q^{(k)}$. The same line of reasoning as above leads us to conclude that $\{\beta_1^{(k+1)}, \cdots, \beta_q^{(k+1)}\}$ is a cointegrating basis, this time with the property that the first $k+1$

columns of

$$B^{(k+1)} = \begin{pmatrix} \beta_1^{(k+1)\prime} \\ \vdots \\ \beta_q^{(k+1)\prime} \end{pmatrix}$$

are equal to the first $k+1$ standard basis vectors of $\mathbb{R}^q$.

Proceeding in this manner, by induction there exists a cointegrating basis $\{\beta_1^{(q)}, \cdots, \beta_q^{(q)}\} \subset \tilde{V}$ such that

$$B^{(q)} = \begin{pmatrix} \beta_1^{(q)\prime} \\ \vdots \\ \beta_q^{(q)\prime} \end{pmatrix} = \begin{pmatrix} I_q & -\Gamma \end{pmatrix}$$

for some $\Gamma \in \mathbb{R}^{q \times n-q}$.

Deriving the Triangular Representation

Define

$$Y_t^{(1)} = \begin{pmatrix} Y_{1t} \\ \vdots \\ Y_{qt} \end{pmatrix} \quad \text{and} \quad Y_t^{(2)} = \begin{pmatrix} Y_{q+1,t} \\ \vdots \\ Y_{nt} \end{pmatrix}$$

for any $t \in \mathbb{Z}$. Letting $\{\eta_t\}_{t\in\mathbb{Z}}$ be the cycle component of Y_t, we can see that

$$B^{(q)} Y_t = Y_{1t} - \Gamma Y_{2t} = B^{(q)} \eta_t + B^{(q)}(Y_0 - \eta_0) := \mu + u_t$$

where u_t is a q-dimensional weakly stationary and causal linear process with mean zero, and we assume $\mu = B^{(q)}(Y_0 - \eta_0)$ is a degenerate q-dimensional random vector.

Letting v_t collect the last $n-q$ entries in $\Psi(L)\varepsilon_t$ for any $t \in \mathbb{Z}$ and $\delta^{(2)}$ the last $n-q$ entries in δ, we finally have the triangular representation

$$Y_t^{(1)} = \mu + \Gamma Y_t^{(2)} + u_t \quad \text{and} \quad \Delta Y_t^{(2)} = \delta^{(2)} + v_t$$

for any $t \in \mathbb{N}$. In other words, we have partitioned the data into two components: the first component is $Y_t^{(2)}$, which collects the $n-q$ common trends that drive the n series comprising Y_t. Each variable in $Y_t^{(2)}$ represents a distinct common trend because all the cointegration relationships are collected in the equation relating $Y_t^{(1)}$ to $Y_t^{(2)}$. Speaking of that equation, it shows that $Y_t^{(1)}$ contains the common trends $Y_t^{(2)}$ and a stationary noise comopnent $\mu + u_t$.

A More General Triangular Representation

Suppose $\eta_t = \alpha(L)\varepsilon_t$ for any $t \in \mathbb{Z}$, where $\{\alpha_j\}_{j\in\mathbb{N}}$ is an absolutely summable sequence of coefficients. Defining the absolutely summable sequences $\{\Theta_j\}_{j\in\mathbb{N}}$ and $\{\Lambda_j\}_{j\in\mathbb{N}}$ as

$$\Theta_j = \begin{pmatrix} I_q & -\Gamma \end{pmatrix} \alpha_j \quad \text{and} \quad \Lambda_j = J\Psi_j$$

for any $j \in \mathbb{N}$, where $J = \begin{pmatrix} O & I_{n-q} \end{pmatrix} \in \mathbb{R}^{(n-q)\times n}$, we can see that

$$\begin{aligned} u_t &= \begin{pmatrix} I_q & -\Gamma \end{pmatrix} \eta_t = \Theta(L)\varepsilon_t \\ v_t &= J\Psi(L)\varepsilon_t = \Lambda(L)\varepsilon_t \end{aligned}$$

for any $t \in \mathbb{Z}$, and as such that

$$e_t = \begin{pmatrix} u_t \\ v_t \end{pmatrix} = \sum_{j=0}^{\infty} \begin{pmatrix} \Theta_j \\ \Lambda_j \end{pmatrix} \varepsilon_{t-j}.$$

It follows that $\{e_t\}_{t\in\mathbb{Z}}$ is a weakly stationary causal mean zero linear process with absolutely summable coefficients.

Therefore, we can see that any cointegrated non-stationary system $\{Y_t\}_{t\in\mathbb{Z}}$ can be represented in terms of two equations

$$Y_t^{(1)} = \mu + \Gamma Y_t^{(2)} + e_t^{(1)} \quad \text{and} \quad \Delta Y_t^{(2)} = \delta^{(2)} + e_t^{(2)},$$

where $e_t = (e_t^{(1)\prime}, e_t^{(2)\prime})'$ is a weakly stationary causal mean zero linear process with absolutely summable coefficients and we assume that $\begin{pmatrix} I_q & -\Gamma \end{pmatrix}(Y_0 - \eta_0)$ is degenerate.

5.2.2 SOLS Estimation

The exposition here is based on Phillips and Hansen (1990). Suppose $\{Y_t\}_{t\in\mathbb{Z}}$ is an I(1) process that is non-stationarity and cointegrated, with a single cointegrating relationship. As stated in the previous section, this indicates that Y_t can be partitioned as

$$Y_t = \begin{pmatrix} y_t \\ X_t \end{pmatrix}$$

and that there exists a weakly stationary causal linear process $\{e_t = (u_t, v_t')'\}_{t\in\mathbb{Z}}$ with absolutely summable coefficients such that

$$y_t = \mu + \beta' X_t + u_t \quad \text{and} \quad \Delta X_t = \delta + v_t$$

for some $\beta, \delta \in \mathbb{R}^{n-1}$ and $\mu \in \mathbb{R}$. The intuitive approach is to estimate β via least squares, by regressing y_t on X_t. However, the fact that Y_t is I(1) implies that this least squares estimate has a non-standard asymptotic distribution; this is demonstrated below. We consider a case with no deterministic time trend for simplicity. An extension to models with deterministic time trends can be found in Hansen (1992).

Theorem (Asymptotic Distribution of the SOLS Estimator)
Let $\{Y_t = (y_t, X_t')'\}_{t\in\mathbb{Z}}$ be an n-dimensional I(1) process that is non-stationary and cointegrated, with cointegrating rank 1. Let the triangular representation of the system be

$$\begin{aligned} y_t &= \mu + \beta' X_t + u_t \\ \Delta X_t &= v_t, \end{aligned}$$

for some $\mu \in \mathbb{R}$ and $\delta \in \mathbb{R}^{n-1}$, where $\{e_t = (u_t, v_t')'\}_{t\in\mathbb{Z}}$ is a weakly stationary causal linear process with absolutely summable coefficients $\{\Phi_j\}_{j\in\mathbb{N}}$ and i.i.d. innovation process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ with positive definite variance $\Sigma \in \mathbb{R}^{n\times n}$ and finite fourth moments. Let $\Gamma : \mathbb{Z} \to \mathbb{R}^{n\times n}$ be the autocovariance function of $\{e_t\}_{t\in\mathbb{Z}}$, and assume in addition that the coefficients are one-summable and that $\Phi(1)$ has full rank n, so that the long run variance $\Sigma_u = \Phi(1)\Sigma\Phi(1)'$ is positive definite. Lastly, define

$$\lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} = \sum_{j=0}^{\infty} \Gamma(j)' = \sum_{j=0}^{\infty} \begin{pmatrix} \mathbb{E}[u_0 u_j] & \mathbb{E}\left[u_0 v_j'\right] \\ \mathbb{E}[v_0 u_j] & \mathbb{E}\left[v_0 v_j'\right] \end{pmatrix}.$$

Denoting by $(\hat{\mu}_T, \hat{\beta}_T')$ the OLS estimator of $(\mu, \beta')'$ obtained using the sample up to time T, we can see that

$$\begin{pmatrix} \sqrt{T}(\hat{\mu}_T - \mu) \\ T \cdot \left(\hat{\beta}_T - \beta\right) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 1 & \int_0^1 B_2^n(r)' dr \\ \int_0^1 B_2^n(r) dr & \int_0^1 B_2^n(r) B_2^n(r)' dr \end{pmatrix}^{-1} \begin{pmatrix} B_1^n(1) \\ \int_0^1 B_2^n(r) dB_1^n(r) + \lambda_{21} \end{pmatrix},$$

where $\{B^n(r) = (B_1^n(r), B_2^n(r)')'\}_{r\in[0,1]}$ is an n-dimensional Brownian motion with variance Σ_u.

Proof) The OLS estimator of $(\mu, \beta')'$ is given by

$$\begin{aligned} \begin{pmatrix} \hat{\mu}_T \\ \hat{\beta}_T \end{pmatrix} &= \begin{pmatrix} T & \sum_{t=1}^T X_t' \\ \sum_{t=1}^T X_t & \sum_{t=1}^T X_t X_t' \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^T y_t \\ \sum_{t=1}^T X_t y_t \end{pmatrix} \\ &= \begin{pmatrix} \mu \\ \beta \end{pmatrix} + \begin{pmatrix} T & \sum_{t=1}^T X_t' \\ \sum_{t=1}^T X_t & \sum_{t=1}^T X_t X_t' \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^T u_t \\ \sum_{t=1}^T X_t u_t \end{pmatrix}. \end{aligned}$$

We can scale the estimators above as

$$\begin{pmatrix} \sqrt{T}(\hat{\mu}_T - \mu) \\ T \cdot \left(\hat{\beta}_T - \beta\right) \end{pmatrix} = \begin{pmatrix} \sqrt{T} & O \\ O & T \cdot I_{n-1} \end{pmatrix} \begin{pmatrix} \hat{\mu}_T - \mu \\ \hat{\beta}_T - \beta \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \frac{1}{T^{3/2}} \sum_{t=1}^{T} X_t' \\ \frac{1}{T^{3/2}} \sum_{t=1}^{T} X_t & \frac{1}{T^2} \sum_{t=1}^{T} X_t X_t' \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_t \\ \frac{1}{T} \sum_{t=1}^{T} X_t u_t \end{pmatrix}.$$

We investigate below the asymptotic distribution of each of these terms.

Define the process $\{Z_t = (Z_{1t}, Z_{2t}')'\}_{t\in\mathbb{N}}$ as $Z_0 = 0$ and

$$Z_t = \sum_{s=1}^{t} e_s + Z_0 = \begin{pmatrix} \sum_{s=1}^{t} u_s \\ \sum_{s=1}^{t} v_s \end{pmatrix}$$

for any $t \in \mathbb{N}$. By the asymptotic results derived above we know that

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_t &\overset{d}{\to} B^n(1) \\ \frac{1}{T} \sum_{t=1}^{T} e_t e_t' &\overset{p}{\to} \Gamma(0) \\ \frac{1}{T} \sum_{t=1}^{T} Z_{t-1} e_t' &\overset{d}{\to} \int_0^1 B^n(r) dB^n(r)' + \Sigma_u - \sum_{j=0}^{\infty} \Gamma(j) \\ \frac{1}{T^{3/2}} \sum_{t=1}^{T} Z_{t-1} &\overset{d}{\to} \int_0^1 B^n(r) dr \\ \frac{1}{T^2} \sum_{t=1}^{T} Z_{t-1} Z_{t-1}' &\overset{d}{\to} \int_0^1 B^n(r) B^n(r)' dr. \end{aligned}$$

Since Σ_u is nonsingular by assumption, $\int_0^1 B^n(r)B^n(r)'dr$ is positive definite-valued. Since $Z_t = Z_{t-1} + e_t$ for any $t \in \mathbb{N}$, we have

$$\begin{aligned} \frac{1}{T^{3/2}} \sum_{t=1}^{T} Z_t &= \frac{1}{T^{3/2}} \sum_{t=1}^{T} Z_{t-1} + \frac{1}{T^{3/2}} \sum_{t=1}^{T} e_t \overset{d}{\to} \int_0^1 B^n(r) dr \\ \frac{1}{T} \sum_{t=1}^{T} Z_t e_t' &= \frac{1}{T} \sum_{t=1}^{T} Z_{t-1} e_t' + \frac{1}{T} \sum_{t=1}^{T} e_t e_t' \overset{d}{\to} \int_0^1 B^n(r) dB^n(r)' + \underbrace{\Sigma_u - \sum_{j=1}^{\infty} \Gamma(j)}_{\lambda} \end{aligned}$$

and

$$\begin{aligned} \frac{1}{T^2} \sum_{t=1}^{T} Z_t Z_t' = \frac{1}{T^2} \sum_{t=1}^{T} Z_{t-1} Z_{t-1}' + \frac{1}{T^2} \sum_{t=1}^{T} Z_{t-1} e_t' + \frac{1}{T^2} \sum_{t=1}^{T} e_t Z_{t-1}' + \frac{1}{T^2} \sum_{t=1}^{T} e_t e_t' \\ \overset{d}{\to} \int_0^1 B^n(r) B^n(r)' dr. \end{aligned}$$

Note that $Z_{2t} = X_t - X_0$, so that the asymptotic results for Z_{2t} apply for X_t as well. Specifically,

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} X_t \overset{d}{\to} \int_0^1 B_2^n(r) dr$$

$$\frac{1}{T}\sum_{t=1}^{T} X_t \cdot u_t \xrightarrow{d} \int_0^1 B_2^n(r) dB_1^n(r) + \lambda_{21}$$
$$\frac{1}{T^2}\sum_{t=1}^{T} X_t X_t' \xrightarrow{d} \int_0^1 B_2^n(r) B_2^n(r)' dr.$$

It now follows easily that

$$\begin{pmatrix} \sqrt{T}\left(\hat{\mu}_T - \mu\right) \\ T \cdot \left(\hat{\beta}_T - \beta\right) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{T^{3/2}}\sum_{t=1}^{T} X_t' \\ \frac{1}{T^{3/2}}\sum_{t=1}^{T} X_t & \frac{1}{T^2}\sum_{t=1}^{T} X_t X_t' \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t \\ \frac{1}{T}\sum_{t=1}^{T} X_t u_t \end{pmatrix}$$
$$\xrightarrow{d} \begin{pmatrix} 1 & \int_0^1 B_2^n(r)' dr \\ \int_0^1 B_2^n(r) dr & \int_0^1 B_2^n(r) B_2^n(r)' dr \end{pmatrix}^{-1} \begin{pmatrix} B_1^n(1) \\ \int_0^1 B_2^n(r) dB_1^n(r) + \lambda_{21} \end{pmatrix}.$$

Q.E.D.

Remarkably, the OLS estimator $\hat{\beta}_T$ of the cointegrating vector is superconsistent for the true value β. However, the formulation above is intractable for a number of reasons For one, due to the dependence between the Brownian motions $B_2^n(r)$ and $B_1^n(r)$, the long run variance cannot be easily disentangled from the asymptotic distribution. Moreover, the asymptotic distribution contains a nuisance parameter λ_{21}, which represents the sum of the covariances between v_0 and u_j for any $j \in \mathbb{N}$.

Thus, we require a modification of the estimator that has a pivotal asymptotic distribution. There are many ways to do this; among the most famous is the Dynamic OLS (DOLS) estimator introduced in Saikkonen (1991), which furnishes a parametric correction for the nuisance parameter by adding lags and leads of ΔX_t into the regression. The technical details, however, are complicated, so we instead focus on the FM-OLS estimator introduced in Phillips and Hansen (1990), which provides a non-parametric correction. The relationship between the DOLS and FM-OLS estimators is akin to that of the ADF and PP tests for unit roots.

5.2.3 FM-OLS Estimation

We retain the notations introduced in the previous theorem. The FM-OLS estimator is based on the modification of the error u_t to form the new error process

$$u_t^+ = u_t - \Sigma_{u,12}\Sigma_{u,22}^{-1} v_t = \begin{pmatrix} 1 & -\Sigma_{u,12}\Sigma_{u,22}^{-1} \end{pmatrix} e_t$$

for any $t \in \mathbb{Z}$, where $\{u_t^+\}_{t\in\mathbb{Z}}$ is a mean zero absolutely summable linear process. Given a consistent estimator $\hat{\Sigma}_u$ for the long run variance Σ_u and a consistent estimator $\hat{\lambda}$ for $\sum_{j=0}^{\infty}\Gamma(j)'$, define

$$y_t^+ = y_t - \hat{\Sigma}_{u,12}\hat{\Sigma}_{u,22}^{-1}v_t$$

and

$$\hat{\lambda}^+ = \hat{\lambda}_{21} - \hat{\Sigma}_{u,12}\hat{\Sigma}_{u,22}^{-1}\hat{\lambda}_{22},$$

where $v_t = \Delta X_t$. Then, the FM-OLS estimator is the OLS estimate from regressing y_t^+ on X_t and an intercept term with an additional correction term $\hat{\lambda}^+$; formally, we define the FM-OLS estimator $(\mu_T^+, \beta_T^{+\prime})'$ of $(\mu,\beta')'$ as

$$\begin{aligned}\begin{pmatrix}\mu_T^+\\ \beta_T^+\end{pmatrix} &= \left[\sum_{t=1}^{T}\begin{pmatrix}1\\ X_t\end{pmatrix}\begin{pmatrix}1 & X_t'\end{pmatrix}\right]^{-1}\left[\sum_{t=1}^{T}\begin{pmatrix}y_t^+\\ X_t y_t^+\end{pmatrix} - \begin{pmatrix}0\\ T\cdot\hat{\lambda}^+\end{pmatrix}\right]\\ &= \begin{pmatrix}T & \sum_{t=1}^{T}X_t'\\ \sum_{t=1}^{T}X_t & \sum_{t=1}^{T}X_tX_t'\end{pmatrix}^{-1}\begin{pmatrix}\sum_{t=1}^{T}y_t^+\\ \sum_{t=1}^{T}X_ty_t^+ - T\cdot\hat{\lambda}^+\end{pmatrix}.\end{aligned}$$

We show below that the FM-OLS estimator converges to a pivotal distribution.

Theorem (Asymptotic Distribution of the FM-OLS Estimator)
Let $\{Y_t = (y_t, X_t')'\}_{t\in\mathbb{Z}}$ be an n-dimensional I(1) process that is non-stationary and cointegrated, with cointegrating rank 1. Let the triangular representation of the system be

$$\begin{aligned} y_t &= \mu + \beta' X_t + u_t\\ \Delta X_t &= v_t,\end{aligned}$$

where we retain the assumptions and notations of the previous theorem. Then,

$$\begin{aligned}&\begin{pmatrix}\sqrt{T}\left(\mu_T^+ - \mu\right)\\ T\cdot\left(\beta_T^+ - \beta\right)\end{pmatrix}\\ &\xrightarrow{d} \sigma_1\cdot\begin{pmatrix}1 & \left(\int_0^1 W_2^n(r)'dr\right)L_{22}'\\ L_{22}\left(\int_0^1 W_2^n(r)dr\right) & L_{22}\left(\int_0^1 W_2^n(r)W_2^n(r)'dr\right)L_{22}'\end{pmatrix}^{-1}\begin{pmatrix}W_1^n(1)\\ L_{22}\left(\int_0^1 W_2^n(r)dW_1^n(r)\right)\end{pmatrix},\end{aligned}$$

where $\{W^n(r) = (W_1^n(r), W_2^n(r)')'\}_{r\in[0,1]}$ is an n-dimensional standard Wiener process on $[0,1]$,

$$L_{22} = \Sigma_{u,22}^{\frac{1}{2}} \quad \text{and} \quad \sigma_1^2 = \Sigma_{u,1\cdot 2} = \Sigma_{u,11} - \Sigma_{u,12}\Sigma_{u,22}^{-1}\Sigma_{u,21}.$$

Proof) By definition,

$$\begin{aligned}
y_t^+ &= y_t - \hat{\Sigma}_{u,12}\hat{\Sigma}_{u,22}^{-1} v_t \\
&= \mu + \beta' X_t + u_t - \hat{\Sigma}_{u,12}\hat{\Sigma}_{u,22}^{-1} v_t \\
&= \mu + \beta' X_t + u_t^+ + \left(\Sigma_{u,12}\Sigma_{u,22}^{-1} - \hat{\Sigma}_{u,12}\hat{\Sigma}_{u,22}^{-1}\right) v_t
\end{aligned}$$

for any $t \in \mathbb{N}$, so that

$$\begin{pmatrix} \mu_T^+ \\ \beta_T^+ \end{pmatrix} = \begin{pmatrix} T & \sum_{t=1}^T X_t' \\ \sum_{t=1}^T X_t & \sum_{t=1}^T X_t X_t' \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^T y_t^+ \\ \sum_{t=1}^T X_t y_t^+ - T \cdot \hat{\lambda}^+ \end{pmatrix}$$

$$\begin{aligned}
= \begin{pmatrix} \mu \\ \beta \end{pmatrix} &+ \begin{pmatrix} T & \sum_{t=1}^T X_t' \\ \sum_{t=1}^T X_t & \sum_{t=1}^T X_t X_t' \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^T u_t^+ \\ \sum_{t=1}^T X_t u_t^+ \end{pmatrix} \\
&- \begin{pmatrix} 1 & \sum_{t=1}^T X_t' \\ \sum_{t=1}^T X_t & \sum_{t=1}^T X_t X_t' \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ T \cdot \hat{\lambda}^+ \end{pmatrix} \\
&+ \begin{pmatrix} T & \sum_{t=1}^T X_t' \\ \sum_{t=1}^T X_t & \sum_{t=1}^T X_t X_t' \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^T v_t' \\ \sum_{t=1}^T X_t v_t' \end{pmatrix} \left(\Sigma_{u,12}\Sigma_{u,22}^{-1} - \hat{\Sigma}_{u,12}\hat{\Sigma}_{u,22}^{-1}\right)'.
\end{aligned}$$

This reveals that

$$\begin{aligned}
\begin{pmatrix} \sqrt{T}\left(\mu_T^+ - \mu\right) \\ T\left(\beta_T^+ - \beta\right) \end{pmatrix} &= \begin{pmatrix} 1 & \frac{1}{T^{3/2}}\sum_{t=1}^T X_t' \\ \frac{1}{T^{3/2}}\sum_{t=1}^T X_t & \frac{1}{T^2}\sum_{t=1}^T X_t X_t' \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^T u_t^+ \\ \frac{1}{T}\sum_{t=1}^T X_t \cdot u_t^+ \end{pmatrix} \\
&- \begin{pmatrix} 1 & \frac{1}{T^{3/2}}\sum_{t=1}^T X_t' \\ \frac{1}{T^{3/2}}\sum_{t=1}^T X_t & \frac{1}{T^2}\sum_{t=1}^T X_t X_t' \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \hat{\lambda}^+ \end{pmatrix} \\
&+ \begin{pmatrix} 1 & \frac{1}{T^{3/2}}\sum_{t=1}^T X_t' \\ \frac{1}{T^{3/2}}\sum_{t=1}^T X_t & \frac{1}{T^2}\sum_{t=1}^T X_t X_t' \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^T v_t' \\ \frac{1}{T}\sum_{t=1}^T X_t v_t' \end{pmatrix} \left(\Sigma_{u,12}\Sigma_{u,22}^{-1} - \hat{\Sigma}_{u,12}\hat{\Sigma}_{u,22}^{-1}\right)'.
\end{aligned}$$

Letting $\{B^n(r) = (B_1^n(r), B_2^n(r)')'\}_{r\in[0,1]}$ be an n-dimensional Brownian motion with variance Σ_u, by the asymptotic results in the previous thereom we have

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^T e_t \xrightarrow{d} B^n(1) \\
&\frac{1}{T^{3/2}}\sum_{t=1}^T Z_t \xrightarrow{d} \int_0^1 B^n(r)dr \\
&\frac{1}{T}\sum_{t=1}^T Z_t e_t' \xrightarrow{d} \int_0^1 B^n(r)dB^n(r)' + \underbrace{\Sigma_u - \sum_{j=1}^{\infty}\Gamma(j)}_{\lambda}
\end{aligned}$$

$$\frac{1}{T^2}\sum_{t=1}^{T} Z_t Z_t' \xrightarrow{d} \int_0^1 B^n(r)B^n(r)'dr$$

for the stochastic process $\{Z_t = (Z_{1t}, Z_{2t}')'\}_{t\in\mathbb{N}}$ with $Z_0 = 0$ and defined as

$$Z_t = \sum_{s=1}^{t} e_s + Z_0 = \begin{pmatrix} \sum_{s=1}^{t} u_s \\ \sum_{s=1}^{t} v_s \end{pmatrix}$$

for any $t \in \mathbb{N}$. We noted earlier that, since $Z_{2t} = X_t - X_0$,

$$\begin{aligned}
\frac{1}{T^{3/2}}\sum_{t=1}^{T} X_t &\xrightarrow{d} \int_0^1 B_2^n(r)dr \\
\frac{1}{T}\sum_{t=1}^{T} X_t e_t' &\xrightarrow{d} \int_0^1 B_2^n(r)dB^n(r)' + \begin{pmatrix}\lambda_{21} & \lambda_{22}\end{pmatrix} \\
\frac{1}{T^2}\sum_{t=1}^{T} X_t X_t' &\xrightarrow{d} \int_0^1 B_2^n(r)B_2^n(r)'dr.
\end{aligned}$$

In particular,

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t^+ &= \begin{pmatrix}1 & -\Sigma_{u,12}\Sigma_{u,22}^{-1}\end{pmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_t \\
&\xrightarrow{d} B^n(1)' \begin{pmatrix} 1 \\ -\Sigma_{u,22}^{-1}\Sigma_{u,21} \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} X_t \cdot u_t^+ &= \left(\frac{1}{T}\sum_{t=1}^{T} X_t \cdot e_t'\right)\cdot \begin{pmatrix} 1 \\ -\Sigma_{u,22}^{-1}\Sigma_{u,21} \end{pmatrix} \\
&\xrightarrow{d} \left(\int_0^1 B_2^n(r)dB^n(r)'\right)\begin{pmatrix} 1 \\ -\Sigma_{u,22}^{-1}\Sigma_{u,21} \end{pmatrix} \\
&\quad + \left(\lambda_{21} - \lambda_{22}'\Sigma_{u,22}^{-1}\Sigma_{u,21}\right).
\end{aligned}$$

These asymptotic results tell us that

$$\begin{aligned}
&\begin{pmatrix} 1 & \frac{1}{T^{3/2}}\sum_{t=1}^{T} X_t' \\ \frac{1}{T^{3/2}}\sum_{t=1}^{T} X_t & \frac{1}{T^2}\sum_{t=1}^{T} X_t X_t' \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t^+ \\ \frac{1}{T}\sum_{t=1}^{T} X_t \cdot u_t^+ \end{pmatrix} \\
&\xrightarrow{d} \begin{pmatrix} 1 & \int_0^1 B_2^n(r)'dr \\ \int_0^1 B_2^n(r)dr & \int_0^1 B_2^n(r)B_2^n(r)'dr \end{pmatrix}^{-1} \begin{pmatrix} B^n(1)' \\ \int_0^1 B_2^n(r)dB^n(r)' \end{pmatrix} \begin{pmatrix} 1 \\ -\Sigma_{u,22}^{-1}\Sigma_{u,21} \end{pmatrix} \\
&\qquad + \begin{pmatrix} 1 & \int_0^1 B_2^n(r)'dr \\ \int_0^1 B_2^n(r)dr & \int_0^1 B_2^n(r)B_2^n(r)'dr \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \lambda_{21} - \lambda_{22}'\Sigma_{u,22}^{-1}\Sigma_{u,21} \end{pmatrix}
\end{aligned}$$

and

$$\begin{pmatrix} 1 & \frac{1}{T^{3/2}}\sum_{t=1}^T X_t' \\ \frac{1}{T^{3/2}}\sum_{t=1}^T X_t & \frac{1}{T^2}\sum_{t=1}^T X_t X_t' \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \hat{\lambda}^+ \end{pmatrix}$$

$$\xrightarrow{d} \begin{pmatrix} 1 & \int_0^1 B_2^n(r)'dr \\ \int_0^1 B_2^n(r)dr & \int_0^1 B_2^n(r)B_2^n(r)'dr \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \lambda_{21} - \lambda_{22}'\Sigma_{u,22}^{-1}\Sigma_{u,21} \end{pmatrix}.$$

As for

$$\begin{pmatrix} 1 & \frac{1}{T^{3/2}}\sum_{t=1}^T X_t' \\ \frac{1}{T^{3/2}}\sum_{t=1}^T X_t & \frac{1}{T^2}\sum_{t=1}^T X_t X_t' \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^T v_t' \\ \frac{1}{T}\sum_{t=1}^T X_t v_t' \end{pmatrix} \left(\Sigma_{u,12}\Sigma_{u,22}^{-1} - \hat{\Sigma}_{u,12}\hat{\Sigma}_{u,22}^{-1}\right)',$$

since the first two terms are $O_p(1)$ and the last one is $o_p(1)$, the entire term converges to 0 in probability. Thus, we have

$$\begin{pmatrix} \sqrt{T}\left(\mu_T^+ - \mu\right) \\ T\left(\beta_T^+ - \beta\right) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 1 & \int_0^1 B_2^n(r)'dr \\ \int_0^1 B_2^n(r)dr & \int_0^1 B_2^n(r)B_2^n(r)'dr \end{pmatrix}^{-1} \begin{pmatrix} B^n(1)' \\ \int_0^1 B_2^n(r)dB^n(r)' \end{pmatrix} \begin{pmatrix} 1 \\ -\Sigma_{u,22}^{-1}\Sigma_{u,21} \end{pmatrix}.$$

Define the univariate stochastic process $\{J(r)\}_{r\in[0,1]}$ as

$$J(r) = \begin{pmatrix} 1 & -\Sigma_{u,12}\Sigma_{u,22}^{-1} \end{pmatrix} B^n(r) = B_1^n(r) - \Sigma_{u,12}\Sigma_{u,22}^{-1}B_2^n(r)$$

for any $r \in [0,1]$. We can see that $\{J(r)\}_{r\in[0,1]}$ is a Brownian motion with variance equal to

$$\begin{pmatrix} 1 & -\Sigma_{u,12}\Sigma_{u,22}^{-1} \end{pmatrix} \Sigma_u \begin{pmatrix} 1 \\ -\Sigma_{u,22}^{-1}\Sigma_{u,21} \end{pmatrix} = \Sigma_{u,1\cdot 2} := \Sigma_{u,11} - \Sigma_{u,12}\Sigma_{u,22}^{-1}\Sigma_{u,21}.$$

In addition,

$$\begin{aligned} \mathbb{E}\left[B_2^n(r)J(r)\right] &= \mathbb{E}\left[B_2^n(r)B^n(r)'\right] \begin{pmatrix} 1 \\ -\Sigma_{u,22}^{-1}\Sigma_{u,21} \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_{u,21} & \Sigma_{u,22} \end{pmatrix} \begin{pmatrix} 1 \\ -\Sigma_{u,22}^{-1}\Sigma_{u,21} \end{pmatrix} = \mathbf{0}, \end{aligned}$$

so that $\{J(r)\}_{r\in[0,1]}$ and $\{B_2^n(r)\}_{r\in[0,1]}$ are independent Brownian motions with variances equal to $\Sigma_{u,1\cdot 2}$ and $\Sigma_{u,22}$, respectively. Therefore, we can write $\Sigma_{u,1\cdot 2}^{\frac{1}{2}} \cdot W_1^n$ and $\Sigma_{u,22}^{\frac{1}{2}} \cdot W_2^n$ in place of J and B_2^n, which yields

$$\begin{pmatrix} \sqrt{T}\left(\mu_T^+ - \mu\right) \\ T\left(\beta_T^+ - \beta\right) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 1 & \int_0^1 B_2^n(r)'dr \\ \int_0^1 B_2^n(r)dr & \int_0^1 B_2^n(r)B_2^n(r)'dr \end{pmatrix}^{-1} \begin{pmatrix} J(1) \\ \int_0^1 B_2^n(r)dJ(r) \end{pmatrix}$$

$$\sim \sigma_1 \cdot \begin{pmatrix} 1 & \left(\int_0^1 W_2^n(r)' dr\right) L_{22}' \\ L_{22}\left(\int_0^1 W_2^n(r)dr\right) & L_{22}\left(\int_0^1 W_2^n(r)W_2^n(r)'dr\right) L_{22}' \end{pmatrix}^{-1} \begin{pmatrix} W_1^n(1) \\ L_{22}\left(\int_0^1 W_2^n(r)dW_1^n(r)\right) \end{pmatrix},$$

where

$$L_{22} = \Sigma_{u,22}^{\frac{1}{2}} \quad \text{and} \quad \sigma_1^2 = \Sigma_{u,1\cdot 2}$$

Q.E.D.

It is clear that the resulting asymptotic distribution is free from any nuisance parameters and, since we have estimates of σ_1^2 and $\Sigma_{u,22}$, pivotal. Using this distribution, we can obtain the limiting distribution of the usual Wald-type test that imposes linear restrictions on the cointegrating vector under the null.

First note that

$$T(\beta_T^+ - \beta) \xrightarrow{d} \sigma_1 \left[L_{22}\left(\int_0^1 W_2^{n*}(r)W_2^{n*}(r)dr\right) L_{22}'\right]^{-1} \left(L_{22} \cdot \int_0^1 W_2^{n*}(r)dW_1^n(r)\right),$$

where $\{W_2^{n*}(r)\}_{r\in[0,1]}$ is the Brownian bridge defined as

$$W_2^{n*}(r) = W_2^n(r) - \int_0^1 W_2^n(s)ds$$

for any $r \in [0,1]$. We denote

$$H = L_{22}\left(\int_0^1 W_2^{n*}(r)W_2^{n*}(r)dr\right) L_{22}'$$
$$J = L_{22} \cdot \int_0^1 W_2^{n*}(r)dW_1^n(r).$$

From the proof above, it is clear that

$$\frac{1}{T^2}\sum_{t=1}^{T} X_t^d X_t^{d\prime} \xrightarrow{d} H,$$

where the superscript d indicates that the variable has been demeaned.

Now consider testing the null hypothesis $H_0 : R\beta = q$ against the alternative $H_0 : R\beta \neq q$, where $R \in \mathbb{R}^{m\times(n-1)}$ is a matrix of full rank m. Then, under the null,

$$T(R\beta_T^+ - q) = R \cdot T(\beta_T^+ - \beta),$$

so we can consider the Wald statistic

$$\mathcal{W}_T = \frac{1}{\hat{\sigma}_1^2}(R\beta_T^+ - q)' \left[R\left(\sum_{t=1}^{T} X_t^d X_t^{d\prime}\right)^{-1} R'\right]^{-1} (R\beta_T^+ - q),$$

where $\hat{\sigma}_1^2$ is a consistent estimator of σ_1^2. Then,

$$\mathcal{W}_T = \frac{1}{\hat{\sigma}_1^2}\left[T(\beta_T^+ - \beta)\right]' R' \left[R\left(\frac{1}{T^2}\sum_{t=1}^{T} X_t^d X_t^{d\prime}\right)R'\right]^{-1} R\left[T(\beta_T^+ - \beta)\right]$$

$$\xrightarrow{d} J'R'\left[RHR'\right]^{-1} RJ.$$

Given W_2^n, due to the independence of W_2^n and W_1^n we can see that the distribution of J is normal with mean 0 and variance H:

$$J \mid W_2^n \sim N\left(\mathbf{0}, H\right).$$

Therefore, given W_2^n, $J'R'\left[RHR'\right]^{-1} RJ$ follows a chi-squared distribution with m degrees of freedom. Since this distribution does not depend on W_2^n, we can conclude that the unconditional distribution of $J'R'\left[RHR'\right]^{-1} RJ$ is χ_m^2, and as such that

$$\mathcal{W}_T \xrightarrow{d} \chi_m^2.$$

This demonstrates that we the Wald test statistic constructed using the FM-OLS estimator has the same asymptotic distribution as the usual Wald statistic.

5.3 Vector Error Correction Models

So far, we have only considered the estimation of cointegrating relationships under the triangular representation, and even then only under the assumption that there is at most one cointegrating relationship. Here, we study a model that imposes a semi-parametric structure on the time series of interest in exchange for allowing for the consistent estimation of the cointegrating space itself under a fixed cointegrating rank (which is allowed to be greater than 1). The structure is identical to that of the VAR case, except that now we use an equivalent error correction representation, to be defined below, as a means of accounting for the non-stationarity of the process.

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional time series that follows the (reduced form) VAR(p) process

$$Y_t = \delta + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t,$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a white noise process with positive definite covariance $\Sigma \in \mathbb{R}^{n\times n}$, $\delta \in \mathbb{R}^n$, and $\Phi_1, \cdots, \Phi_p \in \mathbb{R}^{n\times n}$. Denote the AR polynomial corresponding to this process by

$$A(L) = I_n - \Phi_1 \cdot L - \cdots - \Phi_p \cdot L^p.$$

We saw in the section on vector autoregressions that $\{Y_t\}_{t\in\mathbb{Z}}$ is stationary with a causal linear process representation if the eigenvalues of the polynomial

$$|A(z)| = \det(I_n - \Phi_1 \cdot z - \cdots - \Phi_p \cdot z^p)$$

lie outside the unit circle. We are now interested in what happens if $|A(z)|$ possess a unit root, that is, when $|A(1)| = 0$. It can be shown that there is a very close connection between the number of unit roots of $|A(z)|$ and the non-stationarity of $\{Y_t\}_{t\in\mathbb{Z}}$.

In what follows, we deal with the case where the roots of $|A(z)|$ are either on or outside the unit circle. When $|A(z)|$ has roots within the unit circle, the companion matrix of $\{Y_t\}_{t\in\mathbb{Z}}$ has eigenvalues greater than 1 in magnitude and thus it becomes an explosive process; we are precluding this case. In this context, we first show that a necessary condition for $\{Y_t\}_{t\in\mathbb{Z}}$ to be I(1) is for $|A(z)|$ to have at most n unit roots and at least one unit root.

Lemma Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional time series that follows the VAR(p) process

$$Y_t = \delta + \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t,$$

for an n-dimensional i.i.d. process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a white noise process with positive definite covariance $\Sigma \in \mathbb{R}^{n\times n}$. Letting $A(z)$ be the AR polynomial corresponding to the above VAR process, suppose $|A(z)|$ has roots on or outside the unit circle.

In this case, if $\{Y_t\}_{t\in\mathbb{Z}}$ is I(1) with innovation process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, then $|A(z)|$ has at least one unit root and at most n unit roots.

Proof) Suppose initially that $\{Y_t\}_{t\in\mathbb{Z}}$ is I(1) with innovation process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. Then, $\{\Delta Y_t\}_{t\in\mathbb{Z}}$ is I(0) with innovation process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, that is, there exists a one-summable sequence $\{\Psi_j\}_{j\in\mathbb{N}}$ and $\mu \in \mathbb{R}^n$ such that

$$\Delta Y_t = (1-L)Y_t = \mu + \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j} = \mu + \Psi(L)\varepsilon_t$$

for any $t \in \mathbb{Z}$, where $\Psi(1) = O$. Since $A(L)Y_t = \delta + \varepsilon_t$ and $(1-L)\delta = \mathbf{0}$, we have

$$(1-L)\varepsilon_t = (1-L)A(L)Y_t = A(L)\Delta Y_t = A(1)\mu + A(L)\Psi(L)\varepsilon_t,$$

which implies that $A(1)\mu = \mathbf{0}$ and

$$(1-z)\cdot I_n = A(z)\Psi(z)$$

for any $z \in \mathbb{C}$. Therefore,

$$(1-z)^n = |A(z)\Psi(z)| = |A(z)|\cdot|\Psi(z)|.$$

for any $z \in \mathbb{C}$, and $|A(z)|$ can have at most n unit roots.

On the other hand, if $|A(z)|$ has no unit roots, then

$$O = A(1)\Psi(1)$$

and $|A(1)| \neq 0$. By implication, $A(1)$ is nonsingular and $\Psi(1) = O$, which contradicts the no MA unit root condition of I(0) processes. Thus, $|A(z)|$ must have at least one unit root.

Q.E.D.

Since we are interested in I(1) processes, the above lemma shows that we need only consider the case where $|A(z)|$ has at most n unit roots. The case where $|A(z)|$ has no unit roots (it has roots outside the unit circle) was already considered in the section on stationary vector autoregressions.

5.3.1 VAR Processes with Finite Starting Times

In this section we briefly touch on VAR processes that start from a finite time, say, $t_0 \in \mathbb{Z}$. We show that, in this case, there exist appropriate initial values such that the eigenvalue condition is sufficient to ensure the stationarity of the VAR process; the L^2-boundedness condition, which usually cannot be assumed a priori when it comes to non-stationary processes, can be omitted.

Suppose $\{Y_t\}_{t\geq t_0}$ is an n-dimensional time series that follows a mean zero VAR(p) process

$$Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t$$

for any $t \geq t_0 + p$, where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a white noise process with positive definite covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$. $t_0 \in \mathbb{Z}$ serves as the finite starting time for this VAR process. The companion matrix is, as usual, given by

$$F = \begin{pmatrix} \Phi_1 & \cdots & \Phi_{p-1} & \Phi_p \\ I_n & \cdots & O & O \\ \vdots & \ddots & \vdots & \vdots \\ O & \cdots & I_n & O \end{pmatrix} \in \mathbb{R}^{np\times np},$$

and the companion form of the VAR process is

$$Z_t = \begin{pmatrix} Y_t \\ \vdots \\ Y_{t-p+1} \end{pmatrix} = F Z_{t-1} + \underbrace{\begin{pmatrix} \varepsilon_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}}_{u_t}$$

for any $t \geq t_0 + p$. We can now show the following result:

Theorem (Stationarity of VAR Processes Started at Finite Time)
Let $\{Y_t\}_{t\geq t_0}$ be the VAR(p) process defined above. If the eigenvalues of F are all contained within the unit circle and the initial values $Y_{t_0+p-1}, \cdots, Y_{t_0}$ are determined as

$$\begin{pmatrix} Y_{t_0+p-1} \\ \vdots \\ Y_{t_0} \end{pmatrix} = Z_{t_0+p-1} = \sum_{j=0}^{\infty} F^j \cdot u_{t_0+p-1-j}.$$

Then, $\{Y_t\}_{t\geq t_0}$ is a stationary VAR(p) process with causal linear process representation

$$Y_t = \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j}$$

for any $t \geq t_0$, where $\{\Psi_j\}_{j\in\mathbb{N}}$ is a one-summable sequence of $n \times n$ matrices.

Proof) We showed when proving the sufficiency of the eigenvalue condition for stationarity that, if the eigenvalues of F are all contained within the unit circle, then

$$\sum_{j=0}^{\infty} j \cdot \left\| F^j \right\| < +\infty.$$

and $I_n - F$ is nonsingular. It follows that the vector

$$\sum_{j=0}^{\infty} F^j \cdot u_{t-j}$$

is well defined for any $t \in \mathbb{Z}$ as the almost sure and L^2 limit of the sequence

$$\left\{ \sum_{j=0}^{m} F^j \cdot u_{t-j} \right\}_{m \in N_+}.$$

As such, the statement

$$Z_{t_0+p-1} = \sum_{j=0}^{\infty} F^j \cdot u_{t_0+p-1-j}$$

is well-defined. For any $t \geq t_0 + p$, it now follows that

$$\begin{aligned} Z_t &= F Z_{t-1} + u_t \\ &= \sum_{j=0}^{t-t_0-p} F^j u_{t-j} + F^{t-t_0+1-p} Z_{t_0+p-1} \\ &= \sum_{j=0}^{t-t_0-p} F^j u_{t-j} + \sum_{j=1}^{\infty} F^{(t-t_0-p)+j} \cdot u_{t-(t-t_0-p)-j} \\ &= \sum_{j=0}^{\infty} F^j \cdot u_{t-j}. \end{aligned}$$

Thus, $\{Z_t\}_{t \geq t_0+p-1}$ is a weakly stationary causal linear process with one-summable coefficients.

Letting Ψ_j be the $n \times n$ matrix in the (1,1) position of F^j for any $j \in \mathbb{N}$, $\{\Psi_j\}_{j \in \mathbb{N}}$ is a one-summable sequence of coefficients such that

$$Y_t = \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j}.$$

for any $t \geq t_0 + p - 1$. This shows us that $\{Y_t\}_{t \geq t_0+p-1}$ has a causal linear process representation with one-summable coefficients.

It remains to show that the initial values chosen above have the same linear process representations. For any $j \in \mathbb{N}$, by definition Ψ_j is the $n \times n$ matrix in the (1,1) position of F^j.

Now suppose that, for some $1 \leq k < p$, that the $n \times n$ matrix in the $(k,1)$ position of F^{j+k-1} equals Ψ_j. Then, since

$$F^{j+k} = \begin{pmatrix} \Phi_1 & \cdots & \Phi_{p-1} & \Phi_p \\ I_n & \cdots & O & O \\ \vdots & \ddots & \vdots & \vdots \\ O & \cdots & I_n & O \end{pmatrix} \cdot F^{j+k-1},$$

and the $n \times n$ matrix in the $(k,1)$ position of F^{j+k-1} is Ψ_j by the inductive hypothesis, the $n \times n$ matrix in the $((k+1),1)$ position of F^{j+k} is Ψ_j as well.

By induction, for any $j \in \mathbb{N}$ the first n columns of F^j are given as

$$\begin{pmatrix} \Psi_j \\ \Psi_{j-1} \\ \vdots \\ \Psi_{j-p+1} \end{pmatrix},$$

where we define $\Psi_i = O$ for any $i < 0$. Therefore, our initial values can be written as

$$\begin{pmatrix} Y_{t_0+p-1} \\ \vdots \\ Y_{t_0} \end{pmatrix} = \sum_{j=0}^{\infty} F^j \cdot u_{t_0+p-1-j} = \sum_{j=0}^{\infty} \begin{pmatrix} \Psi_j \varepsilon_{t_0+p-1-j} \vdots \\ \Psi_{j-p+1} \cdot \varepsilon_{t_0+p-1-j} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t_0+p-1-j} \\ \sum_{j=1}^{\infty} \Psi_{j-1} \cdot \varepsilon_{t_0+p-1-j} \\ \vdots \\ \sum_{j=p-1}^{\infty} \Psi_{j-(p-1)} \cdot \varepsilon_{t_0+p-1-j} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t_0+p-1-j} \\ \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t_0+p-2-j} \\ \vdots \\ \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t_0-j} \end{pmatrix}.$$

This shows that

$$Y_t = \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j}$$

for any $t \geq t_0$, so that $\{Y_t\}_{t \geq t_0}$ follows a stationary VAR(p) process that has a causal linear process representation with one-summable coefficients.

Q.E.D.

The preceding theorem shows that the L^2-boundedness condition need not be assumed a priori when proving the stationarity of a VAR(p) process if the VAR process is started at some finite time with appropriate initial values. In this case, the process has the same causal linear process representation as in the case where it is assumed to be L^2-bounded and does not start at some finite time.

While we defined I(0) processes as doubly infinite processes, that is, processes with time index $\mathbb{Z}$, we now allow processes started at some finite time to also be I(0) processes. Specifically, we say that an n-dimensional time series $\{Y_t\}_{t \geq t_0}$ started at some finite time $t_0 \in \mathbb{Z}$ is I(0) if

$$Y_t = \mu + \sum_{j=0}^{\infty} \Psi_j \cdot \varepsilon_{t-j}$$

for any $t \geq t_0$, where $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is an m-dimensional i.i.d. process, $\mu \in \mathbb{R}^n$, and $\{\Psi_j\}_{j \in \mathbb{N}}$ is a one-summable sequence of $n \times m$ matrices. In light of this extension, the preceding theorem tells us that the initial values of any intercept-less VAR(p) process started at some finite time, whose companion matrix has eigenvalues within the unit circle and whose innovation process is i.i.d., can be chosen so that the process becomes I(0).

A final comment is that the eigenvalue condition can, as usual, be replaced with the condition that the roots of

$$|A(z)| = \det(I_n - \Phi_1 \cdot z - \cdots - \Phi_p \cdot z^p)$$

lie outside the unit circle.

5.3.2 Error Correction Form and the Granger Representation Theorem

For now, suppose that there is no intercept in the VAR process above, that is, put $\delta = \mathbf{0}$. One advantage of doing so is that we can apply the result in the previous section and work with VAR(p) processes started at some finite time.

First, we require a different, more convenient representation of the VAR process, called the error correction representation. It is derived as follows:

$$\begin{aligned} Y_t &= \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t \\ &= (\Phi_1 + \cdots + \Phi_p) Y_{t-1} + \Phi_2 (Y_{t-2} - Y_{t-1}) + \cdots + \Phi_p (Y_{t-p} - Y_{t-1}) + \varepsilon_t \\ &= (I_n - A(1)) Y_{t-1} + \sum_{j=2}^{p} \Phi_j \left(-\sum_{i=2}^{j} \Delta Y_{t-i+1} \right) + \varepsilon_t \\ &= (I_n - A(1)) Y_{t-1} - (\Phi_2 + \cdots + \Phi_p) \Delta Y_{t-1} - \cdots - \Phi_p \cdot \Delta Y_{t-p+1} + \varepsilon_t \\ &= (I_n - A(1)) Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t, \end{aligned}$$

where

$$\Gamma_j = -\sum_{i=j+1}^{p} \Phi_i$$

for $1 \leq j \leq p-1$. Rearranging terms reveals that

$$\begin{aligned} \Delta Y_{t-1} &= -A(1) Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t \\ &= \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t, \end{aligned}$$

where we define $\Pi = -A(1)$. This is called the vector error correction model (VECM) representation of the original VAR process.

The VECM AR polynomial is defined as

$$\Gamma(L) = I_n - \Gamma_1 L - \cdots - \Gamma_{p-1} L^{p-1}$$

and satisfies

$$\begin{aligned} \Gamma(L) \Delta Y_t &= (1-L) \Gamma(L) \cdot Y_t \\ &= \delta - A(1) Y_{t-1} + \varepsilon_t = \delta + \varepsilon_t - A(1) L \cdot Y_t = (A(L) - A(1) L)\, Y_t. \end{aligned}$$

This implies that

$$(1-z)\Gamma(z) = A(z) - A(1) z$$

for any $z \in \mathbb{C}$, and, for any $z \neq 1$,

$$\Gamma(z) = \frac{A(z) - A(1)z}{1-z}.$$

$\Gamma(\cdot)$, being a polynomial function, is continuous on $\mathbb{C}$, so

$$\Gamma(1) = \lim_{z\to 1}\Gamma(z) = \lim_{z\to 1}\frac{A(z)-A(1)z}{1-z} = -\lim_{z\to 1}\left(A'(z) - A(1)\right) = -A'(1) + A(1),$$

where $A'(\cdot)$ is the first derivative of $A(\cdot)$.

Suppose $|A(1)| = 0$, or equivalently, $|A(z)|$ has a unit root. The $n \times n$ matrix $A(1)$, and by extension Π, has rank $0 \leq r < n$. We now present a convenient decomposition of Π.

Lemma (Decomposition of Reduced Rank Matrices)
Let $\Pi \in \mathbb{R}^{n\times n}$ be a matrix of rank $1 \leq r \leq n$. Then, there exist matrices $\alpha, \beta \in \mathbb{R}^{n\times r}$ of rank r such that

$$\Pi = \alpha\beta'.$$

Proof) Let $\{v_1, \cdots, v_r\} \subset \mathbb{R}^n$ be the basis of the range $R(\Pi)$ of Π, and define

$$\alpha = \begin{pmatrix} v_1 & \cdots & v_r \end{pmatrix};$$

note that α is an $n \times r$ matrix of full rank. Letting $\{e_1, \cdots, e_n\} \subset \mathbb{R}^n$ be the standard basis of $\mathbb{R}^n$ and $\Pi_1, \cdots, \Pi_n$ the columns of Π, for any $1 \leq i \leq n$

$$\Pi_i = \Pi \cdot e_i \in R(\Pi),$$

so that there exists an r-dimensional vector $b_i \in \mathbb{R}^r$ such that

$$\Pi_i = \alpha \cdot b_i.$$

Defining

$$\beta = \begin{pmatrix} b_1' \\ \vdots \\ b_n' \end{pmatrix},$$

β is an $n \times r$ matrix such that

$$\Pi = \begin{pmatrix} \Pi_1 & \cdots & \Pi_n \end{pmatrix} = \alpha\beta'.$$

Since α has full rank, $\alpha'\alpha$ is nonsingular, implying that

$$\beta' = (\alpha'\alpha)^{-1}\alpha'\Pi.$$

The range of β' is the same as that of $\alpha'\Pi$. Note that, if $\Pi v = \mathbf{0}$ for some $v \in \mathbb{R}^n$, then $\beta' v = \alpha'\Pi v = \mathbf{0}$; conversely, if $\beta' v = \mathbf{0}$ for some $v \in \mathbb{R}^n$, then

$$\Pi v = \alpha\beta' v = \mathbf{0}.$$

This indicates that the β' and Π share the same null space and thus the same nullity; because they are both linear transformations on $\mathbb{R}^n$, by the dimension theorem they have the same rank r.

Q.E.D.

In the case when $r = 0$, we let $\alpha = \beta = \mathbf{0}$. Thus, the error correction representation can be written as

$$\Delta Y_t = \alpha\beta' \cdot Y_{t-1} + \Theta_1 \Delta Y_{t-1} + \cdots + \Theta_{p-1}\Delta Y_{t-p+1} + \varepsilon_t.$$

Orthogonal Complements of Full Rank Matrices

We reuiqre addition concepts to understand the representation theorem below.

For any $A \in \mathbb{R}^{n\times r}$ with full rank $0 < r < n$, we let the orthogonal complement $A_\perp \in \mathbb{R}^{n\times(n-r)}$ be the matrix satisfying $A' A_\perp = O$.

If $r = 0$, so that A is the $n \times 1$ zero vector, we define $A_\perp$ as any nonsingular $n\times n$ matrix, and if $r = n$, we define $A_\perp = \mathbf{0}$.

In the case $0 < r < n$, we can construct $A_\perp$, and thus see that it exists, as follows. Since A has rank r, the columns $\{A_1, \cdots, A_r\} \subset \mathbb{R}^n$ of A are linearly independent. Letting V be the vector subspace of $\mathbb{R}^n$ spanned by the columns of A, we can think of the orthogonal complement $V^\perp$ of V; since $\mathbb{R}^n$ is a finite dimensional vector space, $V \bigoplus V^\perp = \mathbb{R}^n$. This indicates that the sum of the dimensions of V and $V^\perp$ must equal the dimension of $\mathbb{R}^n$, which is n. In other words, $V^\perp$ is an $n-r$ dimensional vector subspace of $\mathbb{R}^n$, and thus has a basis $\{B_1, \cdots, B_{n-r}\} \subset \mathbb{R}^n$. Defining

$$A_\perp = \begin{pmatrix} B_1 & \cdots & B_{n-r} \end{pmatrix},$$

$A_\perp$ is a matrix of rank $n-r$ because its columns are linearly independent, and we can immediately see that

$$A'_\perp A = \begin{pmatrix} B'_1 \\ \vdots \\ B'_{n-r} \end{pmatrix} \begin{pmatrix} A_1 & \cdots & A_r \end{pmatrix} = \begin{pmatrix} B'_1 A_1 & \cdots & B'_1 A_r \\ \vdots & \ddots & \vdots \\ B'_{n-r} A_1 & \cdots & B'_{n-r} A_r \end{pmatrix} = O$$

because the columns of $A_\perp$ are all contained in the space orthogonal to V. Since the basis of $V^\perp$ is not unique, neither is $A_\perp$; in most cases, we use a convenient normalization of $A_\perp$ that puts as many entries equal to 0 or 1.

Granger's Representation Theorem

The following is our main result, called Granger's representation theorem. The proof is adapted from Johansen (1991).

Theorem (Granger's Representation Theorem)
Let $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional time series that follows the VAR(p) process

$$Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t$$

with error correction representation

$$\Delta Y_t = \Pi \cdot Y_{t-1} + \Gamma_1 \cdot \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \cdot \Delta Y_{t-p+1} + \varepsilon_t$$

for an n-dimensional i.i.d. process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ with positive definite covariance $\Sigma \in \mathbb{R}^{n\times n}$. Let $A(z)$ be the AR polynomial corresponding to the above VAR process, and $\Gamma(z)$ the VECM AR polynomial.

Assume the following:

i) $|A(z)|$ has roots on or outside the unit circle.

ii) Π has rank $0 \leq r < n$ with decomposition $\Pi = \alpha\beta'$, where α, β are full rank $n \times r$ matrices.

iii) Letting $\Psi = \Gamma(1)$, the matrix $\alpha'_\perp \Psi \beta_\perp$ is nonsingular.

Then, there exist appropriate initial values $Y_0, \cdots, Y_{1-p}$ such that:

i) Defining

$$C = \beta_\perp \left(\alpha'_\perp \Psi \beta_\perp\right)^{-1} \alpha'_\perp,$$

there exists an absolutely summable seqeunce $\{C_j\}_{j\in\mathbb{N}} \subset \mathbb{R}^{n\times n}$ such that

$$Y_t = C \cdot \left(\sum_{s=1}^{t} \varepsilon_s\right) + C(L)\varepsilon_t - C(L)\varepsilon_0 + Y_0$$

almost surely for any $t \in \mathbb{N}$.

ii) $\{\Delta Y_t\}_{t\in\mathbb{N}}$ and $\{\beta' Y_t\}_{t\in\mathbb{N}}$ are zero-mean I(0) processes.

iii) $|A(z)|$ has exactly $0 < n - r \leq n$ unit roots, so that there are $n - r$ common trends.

iv) The cointegrating rank is r; in other words, there are exactly r cointegrating relationships.

Proof) **The Case $r = 0$**

If $r = 0$, then $\alpha = \beta = \mathbf{0}$, so that $\{\Delta Y_t\}_{t\in\mathbb{Z}}$ follows a VAR(p-1) process with AR polynomial $\Gamma(L)$. $\alpha_\perp, \beta_\perp$ are nonsingular, so by assumption, $\Gamma(1)$ is nonsingular and thus $|\Gamma(z)|$ has no roots outside the unit circle. Furthermore, in this case $A(1) = O$, and

$$\Gamma(z) = \frac{A(z) - zA(1)}{1-z}$$

so that

$$|\Gamma(z)| = |A(z)| \cdot (1-z)^{-n}$$

for any $z \neq 1$. Premultiplying both sides by $(1-z)^n$ yields

$$(1-z)^n |\Gamma(z)| = |A(z)|,$$

and since this holds for $z = 1$ as well, this holds for any $z \in \mathbb{C}$. Since $|\Gamma(z)|$ has no unit roots, $|A(z)|$ has exactly n unit roots, and because $|A(z)|$ has roots outside or on the unit circle, $|\Gamma(z)|$ must have all roots outside the unit circle.

Therefore, we can choose the initial values $Y_{p-1}, \cdots, Y_0$ so that $\{\Delta Y_t\}_{t\in\mathbb{N}}$ follows a stationary VAR(p-1) process with causal linear process representation

$$\Delta Y_t = \Gamma(L)^{-1}\varepsilon_t$$

for any $t \in \mathbb{N}$. Here, the MA(∞) coefficients are one-summable, so that $\{\Delta Y_t\}_{t\in\mathbb{N}}$ is an I(0) process. We denote $\Psi(L) = \Gamma(L)^{-1}$. That $\{\beta' Y_t\}_{t\in\mathbb{N}}$ is I(0) is trivial because $\beta = \mathbf{0}$.

By the Beveridge-Nelson decomposition, there exists a sequence of absolutely summable coefficients $\{C_j\}_{j\in\mathbb{N}}$ such that

$$Y_t = \Psi(1)\left(\sum_{s=1}^{t} \varepsilon_s\right) + C(L)\varepsilon_t - C(L)\varepsilon_0 + Y_0$$

with probability 1 for any $t \in \mathbb{N}$. Here,

$$\Psi(1) = \Gamma(1)^{-1} = \beta_\perp \left(\alpha_\perp' \Gamma(1)\beta_\perp\right)^{-1} \alpha_\perp'$$

since $\alpha_\perp, \beta_\perp$ are nonsingular $n \times n$ matrices, which proves the representation part of the theorem. Finally, since the cointegrating space is the null space of $\Psi(1)$, which is a nonsingular matrix, the cointegrating space consists only of the zero vector. This shows us that the cointegrating rank is 0.

The Case $0 < r < n$

Now suppose that $0 < r < n$. Recall that the relationship

$$A(z) = (1-z)\Gamma(z) + z \cdot A(1) = (1-z)\Gamma(z) - z \cdot \alpha\beta'$$

holds for any $z \in \mathbb{C}$ between the AR and VECM AR polynomials, in light of the equation

$$A(L)Y_t = \Gamma(L) \cdot \Delta Y_t - \alpha\beta' \cdot Y_{t-1} = \varepsilon_t.$$

Pre-multiplying both sides by α' and then $\alpha'_\perp$ yields the equations

$$\begin{aligned} -\alpha'\alpha\beta' \cdot Y_{t-1} + \alpha'\Gamma(L) \cdot \Delta Y_t &= \alpha'\varepsilon_t \\ \alpha'_\perp\Gamma(L) \cdot \Delta Y_t &= \alpha'_\perp\varepsilon_t, \end{aligned}$$

since $\alpha'_\perp\alpha = O$.

To facilitate the proof, we define

$$\bar{a} = a(a'a)^{-1}$$

for any $a \in \mathbb{R}^{n\times k}$ of full rank, where $1 \le k \le n$. Note that $a\bar{a}' = a(a'a)^{-1}a'$ is a matrix that orthogonally projects any vector $v \in \mathbb{R}^n$ onto the vector subspace of $\mathbb{R}^n$ spanned by the columns of a. Letting $a_\perp$ be the orthogonal complement of a, since $\mathbb{R}^n$ is the direct sum of the vector space spanned by the columns of $a_\perp$ and that spanned by the columns of a, which in turn are orthogonal complements, we can see that

$$a_\perp\bar{a_\perp}' + a\bar{a}' = a_\perp(a'_\perp a_\perp)^{-1}a'_\perp + a(a'a)^{-1}a' = I_n.$$

Now define

$$Z_t = \bar{\beta}'Y_t \quad \text{and} \quad X_t = \bar{\beta}'_\perp\Delta Y_t$$

for any $t \in \mathbb{Z}$. It follows that

$$\Delta Y_t = (\beta\bar{\beta}' + \beta_\perp\bar{\beta}'_\perp)\Delta Y_t = \beta \cdot \Delta Z_t + \beta_\perp \cdot X_t,$$

and substituting this into the two equations above yields

$$\begin{aligned} \left(-\alpha'\alpha\beta' L + \alpha'\Gamma(L)(1-L)\right)\beta Z_t + \alpha'\Gamma(L)\beta_\perp \cdot X_t &= \alpha'\varepsilon_t \\ \alpha'_\perp\Gamma(L)(1-L) \cdot \beta Z_t + \alpha'_\perp\Gamma(L)(1-L)\beta_\perp \cdot X_t &= \alpha'_\perp\varepsilon_t, \end{aligned}$$

where we used the fact that

$$\begin{aligned} -\alpha'\alpha\beta' \cdot Y_{t-1} &= -\alpha'\alpha\beta'\beta(\beta'\beta)^{-1}\beta' \cdot Y_{t-1} \\ &= -\alpha'\alpha\beta'\beta \cdot \bar{\beta}'Y_{t-1} = -\alpha'\alpha\beta'\beta L \cdot Z_t. \end{aligned}$$

In matrix form, the equations become

$$\tilde{A}(L)\begin{pmatrix} Z_t \\ X_t \end{pmatrix} = \begin{pmatrix} \alpha' \\ \alpha'_\perp \end{pmatrix}\varepsilon_t,$$

where we define the polynomial

$$\tilde{A}(z) = \begin{pmatrix} -\alpha'\alpha\beta'\beta\cdot z + \alpha'\Gamma(z)(1-z)\beta & \alpha'\Gamma(z)\beta_\perp \\ \alpha'_\perp\Gamma(z)(1-z)\beta & \alpha'_\perp\Gamma(z)\beta_\perp \end{pmatrix}.$$

We can now proceed in steps to show each of the claims above:

Claim 1: $|A(z)|$ has exactly $n-r$ unit roots

Note that

$$\tilde{A}(1) = \begin{pmatrix} -\alpha'\alpha\beta'\beta & \alpha'\Gamma(1)\beta_\perp \\ O & \alpha'_\perp\Gamma(1)\beta_\perp \end{pmatrix},$$

so that

$$\left|\tilde{A}(1)\right| = |\alpha'\alpha||\beta'\beta||\alpha'_\perp\Gamma(1)\beta_\perp|.$$

All three matrices on the right hand side are nonsingular, so $\left|\tilde{A}(1)\right| \neq 0$, meaning that $\tilde{A}(z)$ does not have a unit root.

On the other hand, if $z \neq 1$, then

$$\begin{aligned}\tilde{A}(z) &= \begin{pmatrix} \alpha' \\ \alpha'_\perp \end{pmatrix}\left[-\alpha\beta'\cdot z + \Gamma(z)(1-z)\right]\begin{pmatrix} \beta & \beta_\perp(1-z)^{-1} \end{pmatrix} \\ &= \begin{pmatrix} \alpha' \\ \alpha'_\perp \end{pmatrix} A(z)\begin{pmatrix} \beta & \beta_\perp(1-z)^{-1} \end{pmatrix},\end{aligned}$$

so that

$$\left|\tilde{A}(z)\right| = \left|\begin{pmatrix} \alpha' \\ \alpha'_\perp \end{pmatrix}\right|\cdot|A(z)|\cdot\left|\begin{pmatrix} \beta & \beta_\perp \end{pmatrix}\right|(1-z)^{-(n-r)}.$$

By the nonsingularity of $\begin{pmatrix} \alpha' \\ \alpha'_\perp \end{pmatrix}$ and $\begin{pmatrix} \beta & \beta_\perp \end{pmatrix}$, this quantity equals 0 if and only if $|A(z)| = 0$, and since all the roots of $|A(z)|$ are outside or on the unit circle, if $z \neq 1$ and $\left|\tilde{A}(z)\right| = 0$ then z must lie outside the unit circle. We have shown that all the roots of $\tilde{A}(z)$ are outside the unit circle.

The preceding analysis also shows that

$$(1-z)^{n-r}\left|\tilde{A}(z)\right| = \left|\begin{pmatrix}\alpha' \\ \alpha'_{\perp}\end{pmatrix}\right| \cdot |A(z)| \cdot \left|\begin{pmatrix}\beta & \beta_{\perp}\end{pmatrix}\right|$$

for any $z \neq 1$, and since the equation holds trivially for $z = 1$, we can say that it holds for any $z \in \mathbb{C}$. Since the roots of $\left|\tilde{A}(z)\right|$ all lie oustide the unit circle, it follows that $|A(z)|$ has exactly $n-r$ roots that equal 1.

Claim 2: $\{\Delta Y_t\}_{t\in\mathbb{N}}$ and $\{\beta' Y_t\}_{t\in\mathbb{N}}$ are mean zero I(0)

Defining $Q = \begin{pmatrix}\alpha' \\ \alpha'_{\perp}\end{pmatrix}$ and the polynomial $B(z)$ as

$$B(z) = Q^{-1}\tilde{A}(z)$$

for any $z \in \mathbb{C}$, we can see that $B(L)$ is a finite order AR lag polynomial:

$$\begin{aligned}\tilde{A}(z) &= \begin{pmatrix}\alpha' \\ \alpha'_{\perp}\end{pmatrix} A(z)\begin{pmatrix}\beta & \beta_{\perp}\end{pmatrix} - \begin{pmatrix}\alpha' \\ \alpha'_{\perp}\end{pmatrix} z\Gamma(z)\begin{pmatrix}\beta & O\end{pmatrix} \\ &= \begin{pmatrix}\alpha' \\ \alpha'_{\perp}\end{pmatrix}\Big[(I_n - \Phi_1 z - \cdots - \Phi_p z^p)\begin{pmatrix}\beta & \beta_{\perp}\end{pmatrix} - \left(z\cdot I_n - \Gamma_1 z^2 - \cdots - \Gamma_{p-1}z^p\right)\begin{pmatrix}\beta & O\end{pmatrix}\Big],\end{aligned}$$

so that

$$B(z) = Q^{-1}\tilde{A}(z) = I_n - \sum_{j=1}^{p}\Big[\Phi_j\begin{pmatrix}\beta & \beta_{\perp}\end{pmatrix} - \Gamma_{j-1}\begin{pmatrix}\beta & O\end{pmatrix}\Big] z^j,$$

where we define $\Gamma_0 = I_n$.

By design,

$$B(L)\begin{pmatrix}Z_t \\ X_t\end{pmatrix} = \varepsilon_t.$$

Since all the roots of $B(z)$ lie outside the unit circle due to the fact that all the roots of $\tilde{A}(z)$ also lie outside the unit circle, the process $\{(Z'_t, X'_t)'\}_{t\in\mathbb{N}}$ is weakly stationary for appropriate initial values, and has the causal linear process representation

$$\begin{pmatrix}Z_t \\ X_t\end{pmatrix} = B(L)^{-1}\varepsilon_t = \underbrace{\sum_{j=0}^{\infty}\Theta_j \cdot \varepsilon_{t-j}}_{\Theta(L)\varepsilon_t}$$

for any $t \in \mathbb{N}$, where $\{\Theta_j\}_{j\in\mathbb{N}}$ is a one-summable sequence of $n \times n$ matrices. Since $\beta' Y_t = (\beta'\beta) Z_t$ for any $t \in \mathbb{Z}$, the above result tells us that $\{\beta' Y_t\}_{t\in\mathbb{N}}$ is a mean zero I(0) process under our initial values, with one-summable coefficient process $\{(\beta'\beta)\Theta_{1,j}\}_{j\in\mathbb{N}}$ of $r \times n$ matrices, where $\Theta_{1,j}$ collects the first r rows of Θ_j for each $j \in \mathbb{N}$.

As for the first difference process,

$$\begin{aligned}\Delta Y_t = \begin{pmatrix}\beta(1-L) & \beta_\perp\end{pmatrix}\begin{pmatrix}Z_t \\ X_t\end{pmatrix} &= \sum_{j=0}^{\infty}\begin{pmatrix}\beta(1-L) & \beta_\perp\end{pmatrix}\Theta_j \cdot \varepsilon_{t-j} \\ &= \sum_{j=0}^{\infty}\Psi_j \cdot \varepsilon_{t-j}\end{aligned}$$

for any $t \in \mathbb{N}$. Since $\{\Psi_j\}_{j\in\mathbb{N}}$ is a one-summable sequence due to the one-summability of $\{\Theta_j\}_{j\in\mathbb{N}}$, by definition $\{\Delta Y_t\}_{t\in\mathbb{N}}$ is a mean zero I(0) process.

Claim 3: Deriving the Granger Representation

The Beveridge-Nelson decomposition now tells us that

$$Y_t = \Psi(1)\cdot\left(\sum_{s=1}^{t}\varepsilon_s\right) + C(L)\varepsilon_t - C(L)\varepsilon_0 + Y_0$$

with probability 1 for any $t \in \mathbb{N}$, where the absolutely summable sequence $\{C_j\}_{j\in\mathbb{N}}$ is defined as

$$C_j = -\sum_{i=j+1}^{\infty}\Psi_i$$

for any $j \in \mathbb{N}$ and

$$\begin{aligned}\Psi(1) = \sum_{j=0}^{\infty}\Psi_j &= \begin{pmatrix}O & \beta_\perp\end{pmatrix}\cdot\sum_{j=0}^{\infty}\Theta_j \\ &= \begin{pmatrix}O & \beta_\perp\end{pmatrix}B(1)^{-1} \\ &= \begin{pmatrix}O & \beta_\perp\end{pmatrix}\tilde{A}(1)^{-1}\begin{pmatrix}\alpha' \\ \alpha'_\perp\end{pmatrix} \\ &= \begin{pmatrix}O & \beta_\perp\end{pmatrix}\begin{pmatrix}-\alpha'\alpha\beta'\beta & \alpha'\Gamma(1)\beta_\perp \\ O & \alpha'_\perp\Gamma(1)\beta_\perp\end{pmatrix}^{-1}\begin{pmatrix}\alpha' \\ \alpha'_\perp\end{pmatrix} \\ &= \begin{pmatrix}O & \beta_\perp\end{pmatrix}\begin{pmatrix}-(\beta'\beta)^{-1}(\alpha'\alpha)^{-1} & (\beta'\beta)^{-1}(\alpha'\alpha)^{-1}\alpha'\Gamma(1)\beta_\perp\left(\alpha'_\perp\Gamma(1)\beta_\perp\right)^{-1} \\ O & \left(\alpha'_\perp\Gamma(1)\beta_\perp\right)^{-1}\end{pmatrix}\begin{pmatrix}\alpha' \\ \alpha'_\perp\end{pmatrix} \\ &= \beta_\perp\left(\alpha'_\perp\Gamma(1)\beta_\perp\right)^{-1}\alpha_\perp.\end{aligned}$$

Therefore,

$$Y_t = C\left(\sum_{s=1}^{t}\varepsilon_s\right) + C(L)\varepsilon_t - C(L)\varepsilon_0 + Y_0$$

with probability 1 for any $t \in \mathbb{N}$, where

$$C = \beta_\perp \left(\alpha'_\perp \Gamma(1)\beta_\perp\right)^{-1} \alpha_\perp.$$

Claim 4: Cointegration Properties of $\{Y_t\}_{t\in\mathbb{Z}}$

The above representation tells us that the null space $N_{C'}$ of the linear operator C' is the (augmented) cointegration space. Since $C' = \alpha_\perp \left(\beta'_\perp \Gamma(1)'\alpha_\perp\right)^{-1} \beta'_\perp$ has rank $n-r$, by the dimension theorem the nullity of C' is r, so that the cointegrating rank is exactly r. Furthermore, since

$$C'\beta = O$$

by the definition of the orthogonal complement of β, the columns of β are r linearly independent vectors in $\mathbb{R}^n$ contained in the null space $N_{C'}$. Since $N_{C'}$ has dimension r, it follows that the columns of β form a basis of $N_{C'}$, and as such they form a cointegrating basis for Y_t.

Q.E.D.

Granger's representation theorem tells us that the cointegration properties of an interceptless VAR(p) process $\{Y_t\}_{t\in\mathbb{Z}}$ is determined by the rank r of $A(1) = -\Pi$, where $A(z)$ is the AR polynomial. We can consider the following three cases:

i) $\mathbf{r = 0}$

In this case, $\Pi = O$ and the first-difference process $\{\Delta Y_t\}_{t\in\mathbb{N}}$ follows a stationary VAR(p-1) process. Thus, we can simply apply the results of the section on stationary vector autoregressions to study the properties of the first-difference process. This is made possible because there are no cointegrating relationships among the variables, as seen by the fact that the cointegrating rank is 0.

ii) $\mathbf{0 < r < n}$

In this case, there are exactly r cointegrating relationships and $\{Y_t\}_{t\in\mathbb{N}}$ is an I(1) process with trend-cycle decomposition

$$Y_t = C\left(\sum_{s=1}^{t} \varepsilon_s\right) + C(L)\varepsilon_t + Y_0^*,$$

where Y_0^* contains initial values. C is a matrix of rank $n-r$, which tells us that there are $n-r$ common trends driving the dynamics of Y_t. Since $\Pi \neq O$, we cannot just estimate a stationary VAR with first differences.

iii) $\mathbf{r = n}$

In this case, $\Pi = -A(1)$ is nonsingular, meaning that $|A(z)|$ only has roots outside the unit circle. By implication, $\{Y_t\}_{t\in\mathbb{N}}$ is an I(0) process under the appropriate initial values. Analysis proceeds by estimating the levels VAR using the methods studied in the section on stationary vector autoregressions.

5.4 Estimation of VECMs

Suppose $\{Y_t\}_{t\in\mathbb{Z}}$ is an n-dimensional time series generated by the VAR(p) process

$$Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t$$

with VECM representation

$$\Delta Y_t = \Pi \cdot Y_{t-1} + \Gamma_1 \cdot \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \cdot \Delta Y_{t-p+1} + \varepsilon_t.$$

We assume that the rank $0 \le r < n$ of Π is known, so that it may be decomposed as $\Pi = \alpha\beta'$ for $n \times r$ matrices α, β of full rank. Note that we exclude the case $r = n$ because in that case $\{Y_t\}_{t\in\mathbb{N}}$ is weakly stationary and thus different asymptotic rules apply. On the other hand, if $0 \le r < n$, then Granger's representation theorem shows us that $\{Y_t\}_{t\in\mathbb{N}}$ is I(1), and the asymptotic results follow that of chapter 3.

Suppose the sample size is T, so that we have the sample observations $Y_1, \cdots, Y_T$. Then, we define

$$Y_{-1} = \begin{pmatrix} Y_p' \\ \vdots \\ Y_{T-1}' \end{pmatrix}, \quad \Delta Y = \begin{pmatrix} \Delta Y_{p+1}' \\ \vdots \\ \Delta Y_T' \end{pmatrix}$$

$$X_t = \begin{pmatrix} \Delta Y_{t-1} \\ \vdots \\ \Delta Y_{t-p+1} \end{pmatrix}, \quad \text{for any } p+1 \le t \le T$$

$$X = \begin{pmatrix} X_{p+1}' \\ \vdots \\ X_T' \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_{p+1}' \\ \vdots \\ \varepsilon_T' \end{pmatrix}$$

$$\Gamma = \begin{pmatrix} \Gamma_1' \\ \vdots \\ \Gamma_{p-1}' \end{pmatrix}.$$

Then, for any $p+1 \le t \le T$

$$\Delta Y_t = \Pi \cdot Y_{t-1} + \Gamma' \cdot X_t + \varepsilon_t,$$

and stacking these observations yields

$$\Delta Y = Y_{-1} \cdot \Pi' + X \cdot \Gamma + \varepsilon.$$

Finally, we denote by $A(z)$ and $\Gamma(z)$ the AR polynomials corresponding to the VAR in levels and the VECM, respectively.

5.4.1 Assumptions and Preliminary Asymptotic Results

We make the following assumptions:

A1. Cointegration Properties

We assume that $\{Y_t\}_{t\in\mathbb{Z}}$ be an n-dimensional time series that follows the VAR(p) process

$$Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t$$

with error correction representation

$$\Delta Y_t = \Pi \cdot Y_{t-1} + \Gamma_1 \cdot \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \cdot \Delta Y_{t-p+1} + \varepsilon_t$$

for an n-dimensional i.i.d. process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ with positive definite covariance $\Sigma \in \mathbb{R}^{n\times n}$. We assume the following hold:

i) $|A(z)|$ has roots on or outside the unit circle.

ii) Π has rank $0 \leq r < n$ with decomposition $\Pi = \alpha\beta'$, where α, β are full rank $n \times r$ matrices.

iii) The matrix $\alpha_\perp' \Gamma(1) \beta_\perp$ is nonsingular.

For notational convenience, we put $\beta_\perp = I_n$ when $r = 0$. The form of $\alpha_\perp$ will be specified later.

Suppose the initial values have been chosen so that the conclusions of the Granger representation theorem hold, namely that $\{\Delta Y_t\}_{t\in\mathbb{N}}$ and $\{\beta' Y_t\}_{t\in\mathbb{N}}$ are mean zero I(0) processes and that

$$Y_t = C\left(\sum_{s=1}^{t} \varepsilon_s\right) + C(L)\varepsilon_t + Y_0^*,$$

for any $t \in \mathbb{N}$, where Y_0^* collects initial values, $\{C_j\}_{j\in\mathbb{N}}$ is one-summable and

$$C = \beta_\perp \left(\alpha_\perp' \Gamma(1) \beta_\perp\right)^{-1} \alpha_\perp'.$$

A2. Nonsingular Population and Sample Moments

By Granger's representation theorem, $\{\Delta Y_t\}_{t\in\mathbb{N}}$ and $\{\beta' Y_t\}_{t\in\mathbb{N}}$ are mean zero I(0) processes given appropriate initial values. Letting $G:\mathbb{Z}\to\mathbb{R}^{n\times n}$ be the autocovariance function of the first-difference process, we assume that

$$\begin{pmatrix} G(0) & \cdots & G(p-1) \\ \vdots & \ddots & \vdots \\ G(p-1)' & \cdots & G(0) \end{pmatrix}$$

is a positive definite $np\times np$ matrix. This implies that the submatrices

$$\begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0) \end{pmatrix}$$

and $G(0)$ of the above matrix are also positive definite.

We also assume that $\Delta Y, Y_{-1}$ and X have linearly independent columns almost everywhere. This ensures that the matrix

$$\begin{pmatrix} \Delta Y' \\ Y_{-1}' \\ X' \end{pmatrix} \begin{pmatrix} \Delta Y & Y_{-1} & X \end{pmatrix}$$

is nonsingular for large enough T.

For later use, we let $G_\beta:\mathbb{Z}\to\mathbb{R}^{n\times n}$ as the autocovariance function of $\{\beta' Y_t\}_{t\in\mathbb{N}}$. We assume that the variance $G_\beta(0)$ of $\beta' Y_t$ is positive definite if $r>0$; it equals 0 if $r=0$ and $\beta=\mathbf{0}$.

A3. I.I.D. Innovations

We assume that the innovation process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ has finite fourth moments. Since $\{\Delta Y_t\}_{t\in\mathbb{N}}$ is a one-summable causal linear process with innovation process ε_t, it follows from one-summability that $\{\Delta Y_t\}_{t\in\mathbb{N}}$ also has finite fourth moments.

The main asymptotic results concerning these quantities are given below:

Theorem (Preliminary Asymptotic Results)
Maintain assumptions A1 to A3. Define the long run variance $\Sigma_u = C\Sigma C'$ and $\Sigma_u^{\frac{1}{2}} = C\Sigma^{\frac{1}{2}}$, and let $\{W^n(r)\}_{r\in[0,1]}$ be the standard n-dimensional Brownian motion. Then, the following convergence results hold jointly:

$$\frac{1}{T}\begin{pmatrix}\Delta Y' \\ X'\end{pmatrix}\begin{pmatrix}\Delta Y & X\end{pmatrix} \xrightarrow{p} \begin{pmatrix} G(0) & \cdots & G(p-1) \\ \vdots & \ddots & \vdots \\ G(p-1)' & \cdots & G(0)\end{pmatrix}$$

$$\frac{1}{T}\beta' Y_{-1}'\varepsilon \xrightarrow{p} O$$

$$\frac{1}{T}\beta' Y_{-1}' Y_{-1}\beta \xrightarrow{p} G_\beta(0)$$

$$\frac{1}{T}\beta' Y_{-1}'\begin{pmatrix}\Delta Y & X\end{pmatrix} \xrightarrow{p} \begin{pmatrix}\beta'\Lambda_0 & \cdots & \beta'\Lambda_{p-1}\end{pmatrix}$$

$$\frac{1}{T}Y_{-1}'(\Delta Y) \xrightarrow{d} \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Lambda' + \Lambda_0$$

$$\frac{1}{T}Y_{-1}'X \xrightarrow{d} \iota_{p-1}' \bigotimes \left[\Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Lambda'\right] + \begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}$$

$$\frac{1}{T}Y_{-1}'\varepsilon \xrightarrow{d} \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}$$

$$\frac{1}{T^2}Y_{-1}'Y_{-1} \xrightarrow{d} \Lambda\left(\int_0^1 W^n(r)W^n(r)'dr\right)\Lambda'$$

$$\frac{1}{T}Y_{-1}'Y_{-1}\beta \xrightarrow{d} \left[\Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}(\Gamma(1)C - I_n)' + \Lambda_0 - \begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\Gamma\right]\bar{\alpha}$$

where

$$\Lambda_h = \Sigma_u - \sum_{j=h}^{\infty} G(j)$$

for any $0 \le h \le p-1$ and

$$\bar{\alpha} = \begin{cases}\alpha(\alpha'\alpha)^{-1} & \text{if } r > 0 \\ O & \text{if } r = 0\end{cases}.$$

In addition, we can obtain the following rate of convergence, which need not hold jointly with the above:

$$X'\varepsilon = O_p(T^{1/2}).$$

The following matrix equalities also hold:

$$\Sigma = G(0) - \Lambda_0' \cdot \beta\alpha' - \begin{pmatrix} G(1) & \cdots & G(p-1) \end{pmatrix} \Gamma$$

$$\Gamma = \begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0) \end{pmatrix}^{-1} \left[\begin{pmatrix} G(1)' \\ \vdots \\ G(p-1)' \end{pmatrix} - \begin{pmatrix} \Lambda_1' \\ \vdots \\ \Lambda_{p-1}' \end{pmatrix} \beta\alpha' \right]$$

$$\begin{aligned} \Lambda_0 \alpha_\perp &= \begin{pmatrix} \Lambda_1 & \cdots & \Lambda_{p-1} \end{pmatrix} \Gamma \alpha_\perp \\ G_\beta(0)\alpha' &= \beta'\Lambda_0 - \beta' \begin{pmatrix} \Lambda_1 & \cdots & \Lambda_{p-1} \end{pmatrix} \Gamma. \end{aligned}$$

Proof) Note that, under our assumptions, there exist one-summable coefficients $\{\Psi_j\}_{j\in\mathbb{N}}$, such that

$$\Delta Y_t = \Psi(L)\varepsilon_t$$

for any $t \in \mathbb{N}$. Define the doubly infinite processes $\{u_t\}_{t\in\mathbb{Z}}$ as

$$u_t = \Psi(L)\varepsilon_t$$

for any $t \in \mathbb{Z}$; we have $u_t = \Delta Y_t$ for any $t \in \mathbb{N}$. Furthermore, let us define $\{S_t\}_{t\in\mathbb{N}}$ as

$$S_t = Y_t - Y_0 = \sum_{s=1}^{t} \Delta Y_s = \sum_{s=1}^{t} u_s$$

for any $t \in \mathbb{N}$. Since $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ has finite fourth moments and is i.i.d., we can see that the main asymptotic results derived in the previous chapter hold for $\{S_t\}_{t\in\mathbb{N}}$ as well. For the sake of completeness, they are enumerated below:

$$\begin{aligned} &\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t \xrightarrow{d} \Sigma^{\frac{1}{2}} W^n(1) \\ &\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_t \xrightarrow{d} \Lambda \cdot W^n(1) \\ &\frac{1}{T} \sum_{t=1}^{T} u_t u_{t-h}' \xrightarrow{p} G(h) && \text{for any } h \geq 0 \\ &\frac{1}{T} \sum_{t=h+1}^{T} S_{t-1}\varepsilon_{t-h}' \xrightarrow{d} \Lambda \left(\int_0^1 W^n(r) dW^n(r)' \right) \Sigma^{\frac{1}{2}\prime} + \Sigma && \text{for any } 0 \leq h \leq p-1 \\ &\frac{1}{T} \sum_{t=h+1}^{T} S_{t-1} u_{t-h}' \xrightarrow{d} \Lambda \left(\int_0^1 W^n(r) dW^n(r)' \right) \Lambda' + \Lambda_h && \text{for any } 0 \leq h \leq p-1 \\ &\frac{1}{T^{3/2}} \sum_{t=1}^{T} S_{t-1} \xrightarrow{d} \Lambda \cdot \int_0^1 W^n(r) dr \end{aligned}$$

$$\frac{1}{T^2}\sum_{t=1}^{T} S_{t-1}S_{t-1}' \xrightarrow{d} \Lambda\left(\int_0^1 W^n(r)W^n(r)'dr\right)\Lambda'$$

where $\Sigma_u = \Psi(1)\Sigma\Psi(1)'$, $\Sigma^{\frac{1}{2}}$ is the Cholesky factor of Σ, $\Lambda = \Psi(1)\Sigma^{\frac{1}{2}}$, W^n is the standard n-dimensional Wiener function, and $\{W^n(r)\}_{r\in[0,1]}$ the corresponding Brownian motion.

Therefore,

$$\begin{aligned}
\frac{1}{T}\begin{pmatrix}\Delta Y' \\ X'\end{pmatrix}\begin{pmatrix}\Delta Y & X\end{pmatrix} &= \frac{1}{T}\sum_{t=p+1}^{T}\begin{pmatrix}u_t \\ \vdots \\ u_{t-p+1}\end{pmatrix}\begin{pmatrix}u_t \\ \vdots \\ u_{t-p+1}\end{pmatrix}' \\
&= \begin{pmatrix}\frac{1}{T}\sum_{t=p+1}^{T} u_t u_t' & \cdots & \frac{1}{T}\sum_{t=p+1}^{T} u_t u_{t-p+1}' \\ \vdots & \ddots & \vdots \\ \frac{1}{T}\sum_{t=p+1}^{T} u_{t-p+1} u_t' & \cdots & \frac{1}{T}\sum_{t=p+1}^{T} u_{t-p+1} u_{t-p+1}'\end{pmatrix} \\
&\xrightarrow{p} \begin{pmatrix}G(0) & \cdots & G(p-1) \\ \vdots & \ddots & \vdots \\ G(p-1)' & \cdots & G(0)\end{pmatrix}
\end{aligned}$$

$$\frac{1}{T}Y_{-1}'\varepsilon = \frac{1}{T}\sum_{t=p+1}^{T} Y_{t-1}\varepsilon_t' \xrightarrow{d} \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}$$

$$\begin{aligned}
\frac{1}{T}Y_{-1}'(\Delta Y) &= \frac{1}{T}\sum_{t=p+1}^{T} Y_{t-1}(\Delta Y_t)' = \frac{1}{T}\sum_{t=p+1}^{T} Y_{t-1}u_t' \\
&\xrightarrow{d} \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Lambda' + \Lambda_0
\end{aligned}$$

$$\begin{aligned}
\frac{1}{T}Y_{-1}'X &= \begin{pmatrix}\frac{1}{T}\sum_{t=p+1}^{T} Y_{t-1}u_{t-1}' & \cdots & \frac{1}{T}\sum_{t=p+1}^{T} Y_{t-1}u_{t-p+1}'\end{pmatrix} \\
&\xrightarrow{d} \iota_{p-1}' \bigotimes \left[\Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Lambda'\right] + \begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}
\end{aligned}$$

$$\frac{1}{T^{3/2}}Y_{-1}'\iota_{T-p} = \frac{1}{T^{3/2}}\sum_{t=p+1}^{T} Y_{t-1} \xrightarrow{d} \Lambda\cdot\int_0^1 W^n(r)dr$$

$$\begin{aligned}
\frac{1}{T^2}Y_{-1}'Y_{-1} &= \frac{1}{T^2}\sum_{t=p+1}^{T} Y_{t-1}Y_{t-1}' \\
&= \frac{1}{T^2}\sum_{t=p+1}^{T} S_{t-1}S_{t-1}' + Y_0\left(\frac{1}{T^2}\sum_{t=p+1}^{T} S_{t-1}\right)' \\
&\quad + \left(\frac{1}{T^2}\sum_{t=p+1}^{T} S_{t-1}\right)Y_0' + \frac{T-p}{T^2}Y_0Y_0' \\
&\xrightarrow{d} \Lambda\left(\int_0^1 W^n(r)W^n(r)'dr\right)\Lambda'.
\end{aligned}$$

Moreover, since $\{u_t\}_{t\in\mathbb{Z}}$ is a causal linear process with absolutely summable coefficients and iid innovation process with finite fourth moments,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{vec}\left(X_t\varepsilon_t'\right)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{vec}\left(\begin{pmatrix}u_{t-1}\\ \vdots\\ u_{t-p+1}\end{pmatrix}\varepsilon_t'\right)\xrightarrow{d}N\left[\mathbf{0},\Sigma\bigotimes Q\right].$$

It follows that

$$\frac{1}{\sqrt{T}}X'\varepsilon=\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}X_t\varepsilon_t'=O_p(1).$$

By implication,

$$\frac{1}{T}X'\varepsilon\xrightarrow{p}O.$$

It remains to show the convergence results for terms invovling β. If $r=0$, they hold trivially because $\beta=\mathbf{0}$. Below, we assume that $r>0$. Since

$$C=\beta_\perp(\alpha_\perp'\Gamma(1)\beta_\perp)^{-1}\alpha_\perp',$$

$\beta'\Lambda=(\beta'C)\Sigma^{\frac{1}{2}}=O$ and therefore

$$\frac{1}{T}\beta'Y_{-1}'\varepsilon\xrightarrow{d}O$$

$$\frac{1}{T}\beta'Y_{-1}'(\Delta Y)\xrightarrow{d}\beta'\Lambda_0$$

$$\frac{1}{T}\beta'Y_{-1}'X\xrightarrow{d}\beta'\begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}.$$

The limits on the right hand side are all non-random matrices, so it follows that

$$\frac{1}{T}\beta'Y_{-1}'\varepsilon\xrightarrow{p}O$$

$$\frac{1}{T}\beta'Y_{-1}'(\Delta Y)\xrightarrow{p}\beta'\Lambda_0$$

$$\frac{1}{T}\beta'Y_{-1}'X\xrightarrow{p}\beta'\begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}$$

as well.

Since

$$\Delta Y = Y_{-1} \cdot \beta\alpha' + X \cdot \Gamma + \varepsilon,$$

premultiplying both sides by Y'_{-1} yields

$$Y'_{-1}(\Delta Y) = (Y'_{-1}Y_{-1}\beta)\alpha' + (Y'_{-1}X)\Gamma + Y'_{-1}\varepsilon$$

This tells us that

$$\left(\frac{1}{T}Y'_{-1}Y_{-1}\beta\right)\alpha' = \frac{1}{T}Y'_{-1}(\Delta Y) - \left(\frac{1}{T}Y'_{-1}X\right)\Gamma - \frac{1}{T}Y'_{-1}\varepsilon.$$

Since $\alpha'\alpha$ is nonsingular (α has ful rank), we can therefore express $\frac{1}{T}Y'_{-1}Y_{-1}\beta$ as a continuous function of the random matrices studied above:

$$\frac{1}{T}Y'_{-1}Y_{-1}\beta = \left[\frac{1}{T}Y'_{-1}(\Delta Y) - \left(\frac{1}{T}Y'_{-1}X\right)\Gamma - \frac{1}{T}Y'_{-1}\varepsilon\right]\bar{\alpha}.$$

This means that $\frac{1}{T}Y'_{-1}Y_{-1}\beta$ converges jointly with these quantities, with asymptotic distribution given via the continuous mapping theorem as

$$\begin{aligned}\frac{1}{T}Y'_{-1}Y_{-1}\beta \xrightarrow{d} \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Lambda'\bar{\alpha} + \Lambda_0\bar{\alpha} - \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}\bar{\alpha} \\ - \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Lambda'\left(\sum_{i=1}^{p-1}\Gamma_i\right)'\bar{\alpha} - \begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\Gamma\bar{\alpha} \\ = \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}(\Gamma(1)C - I_n)'\bar{\alpha} + \Lambda_0\bar{\alpha} - \begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\Gamma\bar{\alpha}.\end{aligned}$$

It remains to show the matrix equalities. We first show the equalities for $r > 0$. Premultiplying

$$\left(\frac{1}{T}Y'_{-1}Y_{-1}\beta\right)\alpha' \xrightarrow{p} \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}(\Gamma(1)C - I_n)' + \Lambda_0\bar{\alpha} - \begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\Gamma$$

by β' tells us that

$$\begin{aligned}G_\beta(0)\alpha' &= \beta'\Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}(\Gamma(1)C - I_n)'\bar{\alpha} + \beta'\Lambda_0 - \beta'\begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\Gamma \\ &= \beta'\Lambda_0 - \beta'\begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\Gamma.\end{aligned}$$

Premultiplying both sides of

$$\Delta Y = Y_{-1} \cdot \beta\alpha' + X \cdot \Gamma + \varepsilon,$$

by X' and $(\Delta Y)'$ yield

$$\begin{aligned} X'(\Delta Y) &= X'Y_{-1} \cdot \beta\alpha' + (X'X)\Gamma + X'\varepsilon \\ (\Delta Y)'(\Delta Y) &= (\Delta Y)'Y_{-1} \cdot \beta\alpha' + (\Delta Y)'X \cdot \Gamma + (\Delta Y)'\varepsilon. \end{aligned}$$

Inspecting the last equation, we have

$$\frac{1}{T}(\Delta Y)'\varepsilon = \frac{1}{T}(\Delta Y)'(\Delta Y) - \left(\frac{1}{T}(\Delta Y)'Y_{-1}\right) \cdot \beta\alpha' - \left(\frac{1}{T}(\Delta Y)'X\right)\Gamma.$$

Taking $T \to \infty$ on both sides,

$$\begin{aligned} \frac{1}{T}(\Delta Y)'\varepsilon &\xrightarrow{p} G(0) - \left[\Lambda \left(\int_0^1 W^n(r) dW^n(r)'\right)' \Lambda' + \Lambda_0'\right] \beta\alpha' - \begin{pmatrix} G(1) & \cdots & G(p-1) \end{pmatrix} \Gamma \\ &= G(0) - \Lambda_0' \cdot \beta\alpha' - \begin{pmatrix} G(1) & \cdots & G(p-1) \end{pmatrix} \Gamma. \end{aligned}$$

Meanwhile,

$$\begin{aligned} \frac{1}{T}(\Delta Y)'\varepsilon &= \alpha\beta' \left(\frac{1}{T}Y_{-1}'\varepsilon\right) + \Gamma' \cdot \left(\frac{1}{T}X'\varepsilon\right) + \frac{1}{T}\varepsilon'\varepsilon \\ &\xrightarrow{p} \alpha\beta' \cdot \Lambda \left(\int_0^1 W^n(r) dW^n(r)'\right) \Sigma^{\frac{1}{2}\prime} + \Sigma = \Sigma, \end{aligned}$$

so the uniqueness of probability limits tells us that

$$\Sigma = G(0) - \Lambda_0' \cdot \beta\alpha' - \begin{pmatrix} G(1) & \cdots & G(p-1) \end{pmatrix} \Gamma.$$

Note that we also have

$$\Gamma = (X'X)^{-1}(X'(\Delta Y) - X'Y_{-1} \cdot \beta\alpha' - X'\varepsilon).$$

Taking $T \to \infty$ in the relation

$$\Gamma = \left(\frac{1}{T}X'X\right)^{-1} \left(\frac{1}{T}X'(\Delta Y) - \left(\frac{1}{T}X'Y_{-1}\right) \cdot \beta\alpha' - \frac{1}{T}X'\varepsilon\right)$$

tells us that

$$\Gamma = \begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0) \end{pmatrix}^{-1} \left[\begin{pmatrix} G(1)' \\ \vdots \\ G(p-1)' \end{pmatrix} - \begin{pmatrix} \Lambda_1' \\ \vdots \\ \Lambda_{p-1}' \end{pmatrix} \beta\alpha'\right].$$

Meanwhile, premultipyling

$$\left(\frac{1}{T}Y_{-1}'Y_{-1}\beta\right)\alpha' = \frac{1}{T}Y_{-1}'(\Delta Y) - \left(\frac{1}{T}Y_{-1}'X\right)\Gamma - \frac{1}{T}Y_{-1}'\varepsilon.$$

by $\alpha_\perp$ yields

$$O = \left[\frac{1}{T}Y_{-1}'(\Delta Y) - \left(\frac{1}{T}Y_{-1}'X\right)\Gamma - \frac{1}{T}Y_{-1}'\varepsilon\right]\alpha_\perp,$$

and taking $T \to \infty$ once again yields

$$O = \left[\Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}(\Gamma(1)C - I_n)' + \Lambda_0 - \begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\Gamma\right]\alpha_\perp.$$

Since

$$\alpha_\perp'\Gamma(1)C = \alpha_\perp',$$

we can see that

$$(\Gamma(1)C - I_n)'\alpha_\perp = \alpha_\perp - \alpha_\perp = O,$$

so that

$$\Lambda_0\alpha_\perp = \begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\Gamma\alpha_\perp.$$

When $r = 0$, the equality

$$\Sigma = G(0) - \begin{pmatrix}G(1) & \cdots & G(p-1)\end{pmatrix}\Gamma$$

follows by premultiplying

$$\varepsilon_t = \Delta Y_t - \sum_{i=1}^{p-1}\Gamma_i \cdot \Delta Y_{t-p+1}$$

by ε_t' and taking expectations, while the equality

$$\Gamma = \begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0) \end{pmatrix}^{-1} \begin{pmatrix} G(1)' \\ \vdots \\ G(p-1)' \end{pmatrix}$$

simply represents the Yule-Walker equations. Meanwhile,

$$G_\beta(0)\alpha' = \beta'\Lambda_0 - \beta'\begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\Gamma$$

holds trivially because α, β are zero vectors. Finally, the equality

$$\Lambda_0 = \begin{pmatrix} \Lambda_1 & \cdots & \Lambda_{p-1} \end{pmatrix} \Gamma$$

follows by taking $T \to \infty$ on both sides of

$$\frac{1}{T} Y'_{-1}(\Delta Y) - \left(\frac{1}{T} Y'_{-1} X \right) \Gamma - \frac{1}{T} Y'_{-1} \varepsilon = O$$

and using the fact that $C = \Gamma(1)^{-1}$.

Q.E.D.

5.4.2 The Concentrated Log-Likelihood

As in the section on stationary vector autoregressions, we study the properties of the Gaussian Quasi-MLEs of the model parameters. Assuming $0 < r < n$ for now, the parameters of interest are

$$\alpha, \beta, \Gamma, \Sigma,$$

where α, β are full rank $n \times r$ matrices. To define a parameter space on which the quasi log likelihood is differentiable, we must first show that the set of full rank $n \times r$ matrices is an open subset of $\mathbb{R}^{n\times r}$. We can define the set of full rank $n \times r$ matrices as

$$FR^{n\times r} = \{A \in \mathbb{R}^{n\times r} \mid A'A \in PS^{r\times r}\},$$

where we used the fact that $A'A$ is positive semidefinite and thus has full rank if and only if it is positive definite. Defining the function $f : \mathbb{R}^{n\times r} \to \mathbb{R}^{r\times r}$ as

$$f(A) = A'A$$

for any $A \in \mathbb{R}^{n\times r}$, we can see that

$$FR^{n\times r} = f^{-1}(PS^{r\times r}),$$

that is, $FR^{n\times r}$ is the inverse image of the set of all positive definite $r \times r$ matrices with respect to f. Thus, we can show that $FR^{n\times r}$ is an open subset of $\mathbb{R}^{n\times r}$ if f is continuous with respect to the metric induced by the trace norm on $\mathbb{R}^{n\times r}$ and $\mathbb{R}^{r\times r}$, since we already showed that $PS^{r\times r}$ is an open subset of $\mathbb{R}^{r\times r}$.

The continuity of f follows from the fact that, for any $A \in \mathbb{R}^{n\times r}$ and $\varepsilon > 0$, if $B \in \mathbb{R}^{n\times r}$ satisfies

$$\|A - B\| < \min\left(1, \frac{\varepsilon}{2\|A\| + 1}\right),$$

we have

$$\begin{aligned} \|f(A) - f(B)\| &= \|A'A - B'B\| = \|(B-A)'(B-A) + A'(B-A) + (B-A)'A\| \\ &\le \|B - A\|^2 + 2\|A\| \cdot \|B - A\| \\ &\le (2\|A\| + 1)\|B - A\| < \varepsilon. \end{aligned}$$

Therefore, $FR^{n\times r}$ is an open subset of $\mathbb{R}^{n\times r}$ with respect to the metric induced by the trace norm on $\mathbb{R}^{n\times r}$. It now follows easily that the set

$$\mathcal{R} = \mathrm{vec}\left(FR^{n\times r}\right)$$

is an open subset of $\mathbb{R}^{nr}$ with respect to the euclidean metric (this follows from the same line of reasoning used to show that $\mathcal{A}$ is an open subset of $\mathbb{R}^{n(n+1)/2}$).

The full parameter space is then given as

$$\Theta = \mathcal{R}^2 \times \mathbb{R}^{n^2(p-1)} \times \mathcal{A},$$

where $\mathrm{vec}(\alpha), \mathrm{vec}(\beta) \in \mathcal{R}$, $\mathrm{vec}(\Gamma) \in \mathbb{R}^{n^2(p-1)}$ and $\mathrm{vech}(\Sigma) \in \mathcal{A}$. Clearly, Θ is an open subset of $\mathbb{R}^{2nr+n^2(p-1)+n(n+1)/2}$. We deonte the vectorized parameters by

$$\begin{aligned} a &= \mathrm{vec}(\alpha), \\ b &= \mathrm{vec}(\beta), \\ \gamma &= \mathrm{vec}(\Gamma), \\ \sigma &= \mathrm{vech}(\Sigma). \end{aligned}$$

The Gaussian (quasi) conditional log-likelihood is, in turn, given as

$$\begin{aligned} l(a,b,\gamma,\sigma) = &-\frac{n(T-p)}{2}\log(2\pi) - \frac{T-p}{2}\log|\Sigma| \\ &- \frac{1}{2}\sum_{t=p+1}^{T}(\Delta Y_t - \alpha\beta'\cdot Y_{t-1} - \Gamma' X_t)'\Sigma^{-1}(\Delta Y_t - \alpha\beta'\cdot Y_{t-1} - \Gamma' X_t) \\ = &-\frac{n(T-p)}{2}\log(2\pi) - \frac{T-p}{2}\log|\Sigma| - \frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}(\Delta Y - Y_{-1}\cdot\beta\alpha' - X\cdot\Gamma)'(\Delta Y - Y_{-1}\cdot\beta\alpha' - X\cdot\Gamma)\right]. \end{aligned}$$

In this section, we concentrate out the parameters one by one until the likelihood is a function only of β. In what follows, we assume that the MLEs (given other parameters) always exists, so that we need only inspect the first order conditions.

Concentrating out Γ

The log likelihood can be written as

$$\begin{aligned} l(a,b,\gamma,\sigma) = &-\frac{n(T-p)}{2}\log(2\pi) - \frac{T-p}{2}\log|\Sigma| \\ &- \frac{1}{2}\sum_{t=p+1}^{T}\left[\Delta Y_t - \alpha\beta'\cdot Y_{t-1} - (I_n \bigotimes X_t')\gamma\right]'\Sigma^{-1}\left[\Delta Y_t - \alpha\beta'\cdot Y_{t-1} - (I_n \bigotimes X_t')\gamma\right]. \end{aligned}$$

Given a, b, σ, the MLE of γ, $\hat{\gamma}_T(a,b,\sigma)$, must satisfy the first order condition

$$-\sum_{t=p+1}^{T}(I_n \bigotimes X_t)\Sigma^{-1}\left[\Delta Y_t - \alpha\beta'\cdot Y_{t-1} - (I_n \bigotimes X_t')\hat{\gamma}_T(a,b,\sigma)\right] = \mathbf{0},$$

rearranging which we obtain

$$\begin{aligned}\hat{\gamma}_T(a,b,\sigma) &= \left[I_n \bigotimes \left(\sum_{t=p+1}^{T} X_t X_t'\right)\right]^{-1} \mathrm{vec}\left(\sum_{t=p+1}^{T} X_t(\Delta Y_t - \alpha\beta' \cdot Y_{t-1})'\right) \\ &= (I_n \bigotimes (X'X)^{-1})\mathrm{vec}\left(X'(\Delta Y) - X'Y_{-1} \cdot \beta\alpha'\right) \\ &= \mathrm{vec}\left((X'X)^{-1}X'(\Delta Y - Y_{-1} \cdot \beta\alpha')\right),\end{aligned}$$

analogously to the stationary vector autoregression case. Therefore, the MLE of Γ, $\hat{\Gamma}_T(a,b,\sigma)$, becomes

$$\hat{\Gamma}_T(a,b,\sigma) = (X'X)^{-1}X'(\Delta Y - Y_{-1} \cdot \beta\alpha'),$$

and the concentrated log-likelihood is

$$\begin{aligned}l_{-\gamma}(a,b,\sigma) &= l(a,b,\hat{\gamma}_T(a,b,\sigma),\sigma) \\ &= -\frac{n(T-p)}{2}\log(2\pi) - \frac{T-p}{2}\log|\Sigma| \\ &\quad -\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}(\Delta Y - Y_{-1} \cdot \beta\alpha' - X \cdot \hat{\Gamma}_T(a,b,\sigma))'(\Delta Y - Y_{-1} \cdot \beta\alpha' - X \cdot \hat{\Gamma}_T(a,b,\sigma))\right] \\ &= -\frac{n(T-p)}{2}\log(2\pi) - \frac{T-p}{2}\log|\Sigma| \\ &\quad -\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}(\Delta Y - Y_{-1} \cdot \beta\alpha')'M_X(\Delta Y - Y_{-1} \cdot \beta\alpha')\right],\end{aligned}$$

where $M_X = I_{T-p} - X(X'X)^{-1}X'$ is the residual maker of X. Define the residuals from regressing ΔY and Y_{-1} on X as

$$R_\Delta = \begin{pmatrix} R'_{\Delta,p+1} \\ \vdots \\ R'_{\Delta,T} \end{pmatrix} = M_X(\Delta Y) \quad \text{and} \quad R_{-1} = \begin{pmatrix} R'_{-1,p+1} \\ \vdots \\ R'_{-1,T} \end{pmatrix} = M_X \cdot Y_{-1}.$$

Then, since M_X is symmetric and idempotent, we can write

$$\begin{aligned}l_{-\gamma}(a,b,\sigma) &= -\frac{n(T-p)}{2}\log(2\pi) - \frac{T-p}{2}\log|\Sigma| - \frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}(R_\Delta - R_{-1} \cdot \beta\alpha')'(R_\Delta - R_{-1} \cdot \beta\alpha')\right] \\ &= -\frac{n(T-p)}{2}\log(2\pi) - \frac{T-p}{2}\log|\Sigma| \\ &\quad -\frac{1}{2}\sum_{t=p+1}^{T}(R_{\Delta,t} - \alpha\beta' \cdot R_{-1,t})'\Sigma^{-1}(R_{\Delta,t} - \alpha\beta' \cdot R_{-1,t}).\end{aligned}$$

Concentrating out α

The concentrated log likelihood can be written in terms of the vector a as

$$\begin{aligned} l_{-\gamma}(a,b,\sigma) = &-\frac{n(T-p)}{2}\log(2\pi) - \frac{T-p}{2}\log|\Sigma| \\ &-\frac{1}{2}\sum_{t=p+1}^{T}(R_{\Delta,t} - (R'_{-1,t}\beta \bigotimes I_n)a)'\Sigma^{-1}(R_{\Delta,t} - (R'_{-1,t}\beta \bigotimes I_n)a). \end{aligned}$$

Given b,σ, the MLE of a, $\hat{a}_T(b,\sigma)$, must satisfy the first order condition

$$-\sum_{t=p+1}^{T}(\beta' R_{-1,t} \bigotimes I_n)\Sigma^{-1}(R_{\Delta,t} - (R'_{-1,t}\beta \bigotimes I_n)\hat{a}_T(b,\sigma)) = \mathbf{0},$$

rearranging which we obtain

$$\begin{aligned} \hat{a}_T(b,\sigma) &= \left[\left(\sum_{t=p+1}^{T}\beta' R_{-1,t}R'_{-1,t}\beta\right)\bigotimes I_n\right]^{-1}\text{vec}\left(\sum_{t=p+1}^{T}R_{\Delta,t}R'_{-1,t}\beta\right) \\ &= \left((\beta' R'_{-1}R_{-1}\beta)^{-1}\bigotimes I_n\right)\text{vec}\left(R'_{\Delta}R_{-1}\beta\right) \\ &= \text{vec}\left(R'_{\Delta}R_{-1}\beta(\beta' R'_{-1}R_{-1}\beta)^{-1}\right). \end{aligned}$$

Therefore, the MLE of α, $\hat{\alpha}_T(b,\sigma)$, becomes

$$\hat{\alpha}_T(b,\sigma) = R'_{\Delta}R_{-1}\beta(\beta' R'_{-1}R_{-1}\beta)^{-1},$$

and the concentrated log-likelihood is

$$\begin{aligned} l_{-\gamma,a}(b,\sigma) &= l_{-\gamma}(\hat{a}_T(b,\sigma),b,\sigma) \\ &= -\frac{n(T-p)}{2}\log(2\pi) - \frac{T-p}{2}\log|\Sigma| \\ &\quad -\frac{1}{2}\text{tr}\left[\Sigma^{-1}(R_{\Delta} - R_{-1}\cdot\beta\hat{\alpha}_T(b,\sigma)')'(R_{\Delta} - R_{-1}\cdot\beta\hat{\alpha}_T(b,\sigma)')\right] \\ &= -\frac{n(T-p)}{2}\log(2\pi) - \frac{T-p}{2}\log|\Sigma| - \frac{1}{2}\text{tr}\left(\Sigma^{-1}R'_{\Delta}M_{R_{-1}\beta}R_{\Delta}\right), \end{aligned}$$

where $M_{R_{-1}\beta} = I_{T-p} - R_{-1}\beta(\beta' R'_{-1}R_{-1}\beta)^{-1}\beta' R'_{-1}$ is the residual maker associated with regressions on $R_{-1}\beta$.

Concentrating out Σ

As we derived in the chapter on stationary vector autoregressions,

$$\frac{\partial l_{-\gamma,a}(b,\sigma)}{\partial \Sigma} = -\frac{T-p}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}R_\Delta' M_{R_{-1}\beta} R_\Delta \Sigma^{-1}$$

for any $\Sigma \in PS^{n\times n}$. It follows that, given b, the MLE $\hat{\Sigma}_T(b)$ of Σ is

$$\hat{\Sigma}_T(b) = \frac{1}{T-p} R_\Delta' M_{R_{-1}\beta} R_\Delta,$$

making the concentrated log-likelihood

$$\begin{aligned}
l_{-\gamma,a,\sigma}(b) &= l_{-\gamma,a}(b, \hat{\sigma}_T(b)) \\
&= -\frac{n(T-p)}{2}(\log(2\pi)+1) - \frac{T-p}{2}\log\left|\hat{\Sigma}_T(b)\right| \\
&= -\frac{n(T-p)}{2}(\log(2\pi)+1) - \frac{T-p}{2}\log\left|\frac{1}{T-p}R_\Delta' M_{R_{-1}\beta} R_\Delta\right|.
\end{aligned}$$

5.4.3 Maximum Likelihood Estimates when $r > 0$

Using the concentrated likelihood derived above, we can derive the maximized log-likelihood and the MLE of β. Afterward, we can use the MLE of β to recover the estimates of α, Γ and Σ using the formulas derived in the previous section.

For notational convenience, we define the quantities

$$\begin{aligned}
S_\Delta &= \frac{1}{T-p} R'_\Delta R_\Delta \\
S_{\Delta,-1} &= \frac{1}{T-p} R_{\Delta'} R_{-1} \\
S_{-1} &= \frac{1}{T-p} R'_{-1} R_{-1}.
\end{aligned}$$

Note that

$$\begin{pmatrix} S_\Delta & S_{\Delta,-1} \\ S'_{\Delta,-1} & S_{-1} \end{pmatrix} = \frac{1}{T-p} \begin{pmatrix} R'_\Delta \\ R'_{-1} \end{pmatrix} \begin{pmatrix} R_\Delta & R_{-1} \end{pmatrix} = \frac{1}{T-p} \begin{pmatrix} \Delta Y' \\ Y'_{-1} \end{pmatrix} M_X \begin{pmatrix} \Delta Y & Y_{-1}. \end{pmatrix}$$

Since M_X has rank $T-p-n(p-1)$, for large enough T this is larger than $2n$ and therefore the $2n \times 2n$ random matrix above is almost surely positive definite. It follows that, for large enough T, S_Δ and S_{-1} are positive definite, and that the Schur complement

$$S_{-1} - S'_{\Delta,-1} S_\Delta^{-1} S_{\Delta,-1}$$

is also positive definite.

Returning to the concentrated log-likelihood, we have

$$l_{-\gamma,a,\sigma}(b) = -\frac{n(T-p)}{2}(\log(2\pi)+1) - \frac{T-p}{2} \log \left| S_\Delta - S_{\Delta,-1}\beta(\beta' S_{-1}\beta)^{-1}\beta' S'_{\Delta,-1} \right|.$$

Since $S_\Delta - S_{\Delta,-1}\beta(\beta' S_{-1}\beta)^{-1}\beta' S'_{\Delta,-1}$ is the Schur complement of the block matrix

$$\begin{pmatrix} S_\Delta & S_{\Delta,-1}\beta \\ \beta' S'_{\Delta,-1} & \beta' S_{-1}\beta \end{pmatrix},$$

it follows that

$$\begin{aligned}
\left| \begin{pmatrix} S_\Delta & S_{\Delta,-1}\beta \\ \beta' S'_{\Delta,-1} & \beta' S_{-1}\beta \end{pmatrix} \right| &= |\beta' S_{-1}\beta| \cdot \left| S_\Delta - S_{\Delta,-1}\beta(\beta' S_{-1}\beta)^{-1}\beta' S'_{\Delta,-1} \right| \\
&= |S_\Delta| \cdot \left| \beta' S_{-1}\beta - \beta' S'_{\Delta,-1} S_\Delta^{-1} S_{\Delta,-1}\beta \right|.
\end{aligned}$$

Since $\beta' S_{-1}\beta$ is positive definite, we now have

$$\begin{aligned} l_{-\gamma,a,\sigma}(b) &= -\frac{n(T-p)}{2}(\log(2\pi)+1) - \frac{T-p}{2}\log\frac{|S_\Delta|\cdot\left|\beta' S_{-1}\beta - \beta' S'_{\Delta,-1}S_\Delta^{-1}S_{\Delta,-1}\beta\right|}{|\beta' S_{-1}\beta|} \\ &= -\frac{n(T-p)}{2}(\log(2\pi)+1) - \frac{T-p}{2}\log\frac{\left|\beta'(S_{-1} - S'_{\Delta,-1}S_\Delta^{-1}S_{\Delta,-1})\beta\right|}{|\beta' S_{-1}\beta|} - \frac{T-p}{2}\log|S_\Delta|. \end{aligned}$$

Since the only term with β is the second one, the QMLE of β is the solution to the maximization problem

$$\max_{\beta \in FR^{n\times r}} \quad V(\beta) = \log\left|\beta'(S_{-1} - S'_{\Delta,-1}S_\Delta^{-1}S_{\Delta,-1})\beta\right| - \log|\beta' S_{-1}\beta|.$$

Necessary Conditions for Maximization

For any symmetric $n\times n$ matrix M, define the function $g_M : \mathcal{R} \to PS^{r\times r}$ as

$$g_M(b) = \log|\beta' M\beta|$$

for any $b \in \mathcal{R}$, where $b = \text{vec}(\beta)$ for $\beta \in FR^{n\times r}$. Note that

$$\frac{\partial \beta' M\beta}{\partial x} = \left(\frac{\partial \beta}{\partial x}\right)' M\beta + \beta' M\left(\frac{\partial \beta}{\partial x}\right),$$

which implies that

$$\frac{\partial \text{vec}(\beta' M\beta)}{\partial x} = \left[(\beta' M \bigotimes I_r)K_{nr} + (I_r \bigotimes \beta' M)\right]\frac{\partial \text{vec}(\beta)}{\partial x}$$

and thus

$$\frac{\partial \text{vec}(\beta' M\beta)}{\partial \text{vec}(\beta)'} = (\beta' M \bigotimes I_r)K_{nr} + (I_r \bigotimes \beta' M).$$

In addition,

$$\frac{\partial \log|A|}{\partial A} = A^{-1}$$

for any $A \in PS^{r\times r}$, so

$$\frac{\partial \log|A|}{\partial \text{vec}(A)} = \text{vec}\left(A^{-1}\right).$$

By the chain rule, it now follows that

$$
\begin{aligned}
\frac{\partial g_M(b)}{\partial b'} &= \frac{\partial \log|\beta' M\beta|}{\partial \text{vec}\,(\beta' M\beta)'} \cdot \frac{\partial \text{vec}\,(\beta' M\beta)}{\partial \text{vec}\,(\beta)'} \\
&= \text{vec}\left((\beta' M\beta)^{-1}\right)' \left[(\beta' M \bigotimes I_r)K_{nr} + (I_r \bigotimes \beta' M)\right] \\
&= \text{vec}\left((\beta' M\beta)^{-1}\beta' M\right)' K_{nr} + \text{vec}\left(M\beta(\beta' M\beta)^{-1}\right)' \\
&= \text{vec}\left(M\beta(\beta' M\beta)^{-1}\right)'.
\end{aligned}
$$

Therefore,

$$
\frac{\partial V(\beta)}{\partial b} = \text{vec}\left(\mathcal{S}\beta(\beta'\mathcal{S}\beta)^{-1} - S_{-1}\beta(\beta' S_{-1}\beta)^{-1}\right),
$$

where we define

$$
\mathcal{S} = S_{-1} - S'_{\Delta,-1} S_{\Delta}^{-1} S_{\Delta,-1}.
$$

Therefore, the QMLE $\hat{\beta}_T$ of β must satisfy the first order condition

$$
\mathcal{S}\hat{\beta}_T(\hat{\beta}'_T \mathcal{S}\hat{\beta}_T)^{-1} = S_{-1}\hat{\beta}_T(\hat{\beta}'_T S_{-1}\hat{\beta}_T)^{-1}.
$$

Maximizing the Likelihood

Define

$$
\hat{C}_T = S_{-1}^{\frac{1}{2}\prime}\hat{\beta}_T,
$$

where $S_{-1}^{\frac{1}{2}}$ is the Cholesky factor of S_{-1}. Then,

$$
\begin{aligned}
S_{-1}\hat{\beta}_T(\hat{\beta}'_T S_{-1}\hat{\beta}_T)^{-1} &= S_{-1}^{\frac{1}{2}} \cdot \hat{C}_T(\hat{C}'_T\hat{C}_T)^{-1} \\
&= S\hat{\beta}_T(\hat{\beta}'_T S\hat{\beta}_T)^{-1} = \mathcal{S} \cdot S_{-1}^{-\frac{1}{2}\prime}\hat{C}_T\left[\hat{C}'_T(S_{-1}^{-\frac{1}{2}} \cdot \mathcal{S} \cdot S_{-1}^{-\frac{1}{2}\prime})\hat{C}_T\right]^{-1}.
\end{aligned}
$$

Defining

$$
\begin{aligned}
M &= S_{-1}^{-\frac{1}{2}} \cdot \mathcal{S} \cdot S_{-1}^{-\frac{1}{2}\prime} \\
&= I_n - S_{-1}^{-\frac{1}{2}} S'_{\Delta,-1} S_{\Delta}^{-1} S_{\Delta,-1} S_{-1}^{-\frac{1}{2}\prime},
\end{aligned}
$$

which is positive definite since $\mathcal{S}$ is, we have

$$
\hat{C}_T(\hat{C}'_T\hat{C}_T)^{-1} = M\hat{C}_T(\hat{C}'_T M\hat{C}_T)^{-1}.
$$

This means that the maximized log-likelihood becomes

$$\hat{l}_T = -\frac{n(T-p)}{2}\left(\log(2\pi)+1+\frac{1}{n}\log|S_\Delta|\right) - \frac{T-p}{2}\log\frac{\left|\hat{\beta}_T'(S_{-1}-S_{\Delta,-1}'S_\Delta^{-1}S_{\Delta,-1})\hat{\beta}_T\right|}{\left|\hat{\beta}_T'S_{-1}\hat{\beta}_T\right|}$$

$$= -\frac{n(T-p)}{2}\left(\log(2\pi)+1+\frac{1}{n}\log|S_\Delta|\right) - \frac{T-p}{2}\log\frac{\left|\hat{C}_T'M\hat{C}_T\right|}{\left|\hat{C}_T'\hat{C}_T\right|}.$$

We now derive an expression for the rightmost term, and, through that expression, obtain an estimator for the cointegrating space. To proceed, we require the following algebraic results:

Lemma (Properties of Self-Adjoint Linear Operators)
Let $(V, \langle\cdot,\cdot\rangle)$ be a finite-dimensional inner product space over the real field, and $T \in L(V)$ a self-adjoint linear operator on V, that is, a linear operator on V such that

$$\langle Tv, u\rangle = \langle v, Tu\rangle$$

for any $v, u \in V$. Then, the following hold true:

i) The eigenvalues of T are real.

ii) There exists a basis of V consisting of orthonormal eigenvectors of T.

Proof) i) Let $\lambda \in \mathbb{C}$ be an eigenvalue of T with corresponding eigenvector $v \in V$, which must be non-zero by definition. It then holds that

$$\lambda\langle v,v\rangle = \langle \lambda\cdot v, v\rangle = \langle Tv, v\rangle = \langle v, Tv\rangle = \langle v, \lambda\cdot v\rangle = \overline{\lambda}\langle v,v\rangle$$

by the bilinearity of the inner product, as well as the definition of self-adjointness and eigenvectors. Since $\langle v,v\rangle \neq 0$ due to the fact that v is non-zero, we have $\lambda = \overline{\lambda}$, and thus $\lambda \in \mathbb{R}$.

ii) We proceed by induction on the dimension n of V. Denote the norm induced by $\langle\cdot,\cdot\rangle$ as $\|\cdot\|$. When $n=1$, letting $\{v\}$ be a basis of V, there must exist some $\lambda \in \mathbb{R}$ such that $Tv = \lambda v$, since Tv which belongs to V and is thus a scalar multiple of v. Therefore, $\{\frac{v}{\|v\|}\}$ is a basis of V that consists of orthonormal eigenvectors of T.

Now suppose that the claim holds for any real inner product space of dimension $n \geq 1$ and self-adjoint operator on that space. Suppose that $(V, \langle\cdot,\cdot\rangle)$ is a real inner product space of dimension $n+1$ and that T is a self-adjoint operator on V. Choose any eigenvector $v_{n+1} \in V$ of T normalized to have norm 1; there exists a

$\lambda \in \mathbb{R}$ such that $Tv_{n+1} = \lambda \cdot v_{n+1}$, in light of the preceding result. Defining the subspace $W = \text{span}(\{v_{n+1}\})$ of V, since W and V are both finite-dimensional spaces, we have $V = W \bigoplus W^{\perp}$, where $W^{\perp}$ is the orthogonal complement of V. Since W has dimension 1 (v_{n+1} is non-zero), and V is an $n+1$-dimensional space, the fact that V is the direct sum of W and $W^{\perp}$ tells us that $W^{\perp}$ must have dimension n.

Let $\tilde{T} \in L(W^{\perp}, V)$ be the restriction of T on $W^{\perp}$; since T is self-adjoint on V, $\tilde{T}$ must be self-adjoint on $W^{\perp}$. By the inductive hypothesis, there exists a basis $\{v_1, \cdots, v_n\} \subset V$ of $W^{\perp}$ that is comprised of orthonormal eigenvectors of $\tilde{T}$. $V = W \bigoplus W^{\perp}$ tells us once again that $\{v_1, \cdots, v_n, v_{n+1}\}$ is a basis of V. Since

$$\langle v_{n+1}, v_i \rangle = 0$$

for any $1 \leq i \leq n$ since v_{n+1} belongs to the orthogonal complement of $W^{\perp}$, $\|v_{n+1}\| = 1$ and for any $1 \leq i \leq n$ there exists a $\lambda_i \in \mathbb{R}$ such that

$$Tv_i = \tilde{T}v_i = \lambda_i \cdot v_i$$

since v_i is an eigenvector of $\tilde{T}$ of norm 1, it follows that $\{v_1, \cdots, v_n, v_{n+1}\}$ is a basis of V comprised of orthonormal eigenvectors of T.

Q.E.D.

Lemma Let $\{y_1, \cdots, y_r\} \subset \mathbb{R}^n$ be a linearly independent set of n-dimensional vectors that are collected into a matrix

$$y = \begin{pmatrix} y_1 & \cdots & y_r \end{pmatrix} \in \mathbb{R}^{n \times r}$$

of full rank r. Let $A \in \mathbb{R}^{n \times n}$ be a positive definite matrix, and suppose that

$$y(y'y)^{-1} = Ay(y'Ay)^{-1}$$

holds. Then, the vector space $V = \text{span}(\{y_1, \cdots, y_r\})$ is invariant under A, that is, $Av \in V$ for any $v \in V$. In addition, there exists an $n \times r$ matrix E whose columns are orthogonal eigenvectors of A, and a nonsingular matrix $P \in \mathbb{R}^{r \times r}$, such that $y = EP$.

Proof) We first show that V is invariant under A. The orthogonal projection (as a linear operator) onto V is defined as

$$\text{proj}_V = y(y'y)^{-1}y';$$

we can easily see that, for any $v \in \mathbb{R}^n$,

$$(v - y(y'y)^{-1}y'v)'y = 0.$$

Choose any $v \in V$. Since there exists an $\mathbf{a} \in \mathbb{R}^r$ such that $v = y \cdot \mathbf{a}$, it follows that

$$\text{proj}_V(v) = y(y'y)^{-1}y'v = v.$$

Now note that

$$\text{proj}_V(Av) = y(y'y)^{-1}y'Av = y(y'y)^{-1}(y'Ay)\mathbf{a}.$$

Since

$$y(y'y)^{-1}y'Ay = Ay$$

by assumption, we have

$$\text{proj}_V(Av) = y(y'y)^{-1}(y'Ay)\mathbf{a} = Ay \cdot \mathbf{a} = Av.$$

Therefore, $Av \in V$ and the subspace V is invariant under A.

Let T be the left multiplication transformation corresponding to A, and $\tilde{T}$ its restriction to V. Since A, and by extension T, is invariant on V, it follows that $\tilde{T}$ is a linear operator on V. Furthermore, it is self-adjoint because A is a symmetric matrix; by the previous lemma, there exists a basis $\{e_1, \cdots, e_r\} \subset V$ of V that is comprised of orthonormal eigenvectors of T and thus of A. Defining

$$E = \begin{pmatrix} e_1 & \cdots & e_r, \end{pmatrix}$$

since $\{y_1, \cdots, y_r\}$ is a basis of V and thus contained in V, there exists a vector $P_i \in \mathbb{R}^r$ such that

$$y_i = E \cdot P_i$$

for each $1 \leq i \leq r$. Defining

$$P = \begin{pmatrix} P_1 & \cdots & P_r \end{pmatrix},$$

it follows that $y = EP$.

To see that P is non-singular, note that, by the same reasoning as above, there exists a $\Lambda \in \mathbb{R}^{r \times r}$ such that

$$E = y\Lambda.$$

Thus,

$$I_r = E'E = E'y\Lambda = (E'E)P\Lambda = P\Lambda,$$

where the first equality follows because the columns of E are orthonormal. This shows us that $\Lambda = P^{-1}$.

Q.E.D.

Since

$$\hat{C}_T(\hat{C}_T'\hat{C}_T)^{-1} = M\hat{C}_T(\hat{C}_T'M\hat{C}_T)^{-1},$$

everywhere on Ω, the preceding lemma tells us that, for any outcome $\omega \in \Omega$, there exist orthonormal eigenvectors $e_1, \cdots, e_r$ of M and a nonsingular $r \times r$ matrix P such that

$$\hat{C}_T = \underbrace{\begin{pmatrix} e_1 & \cdots & e_r \end{pmatrix}}_{E} P.$$

Letting $\mu_i \in \mathbb{R}$ be the eigenvalue of M corresponding to e_i (it is real because the eigenvalues of symmetric matrices are real), it follows that

$$M \cdot E = E \cdot \begin{pmatrix} \mu_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mu_r \end{pmatrix}$$

and thus

$$\begin{aligned} \frac{\left|\hat{C}_T'M\hat{C}_T\right|}{\left|\hat{C}_T'\hat{C}_T\right|} &= \frac{|P'E'MEP|}{|P'E'EP|} \\ &= |E'ME| = \left|\begin{pmatrix} \mu_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mu_r \end{pmatrix}\right| = \prod_{i=1}^{r} \mu_i. \end{aligned}$$

Finally, note that the eigenvalues $\mu_1, \cdots, \mu_r$ are solutions to the equation

$$\begin{aligned} 0 = |M - \mu \cdot I_n| &= \left|(1-\mu) \cdot I_n - S_{-1}^{-\frac{1}{2}} S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1} S_{-1}^{-\frac{1}{2}\prime}\right| \\ &= \left|(1-\mu) \cdot S_{-1} - S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1}\right| \left|S_{-1}^{-1}\right|. \end{aligned}$$

Thus, the maximized log-likelihood must be equal to

$$\hat{l}_T = -\frac{n(T-p)}{2}\left(\log(2\pi)+1+\frac{1}{n}\log|S_\Delta|\right)-\frac{T-p}{2}\log\frac{\left|\hat{C}_T' M \hat{C}_T\right|}{\left|\hat{C}_T'\hat{C}_T\right|}$$

$$= -\frac{n(T-p)}{2}\left(\log(2\pi)+1+\frac{1}{n}\log|S_\Delta|\right)-\frac{T-p}{2}\sum_{i=1}^{r}\log\left(1-\hat{\lambda}_i\right)$$

for the r largest solutions $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_r > 0$ that solve the equation

$$\left|\lambda \cdot S_{-1} - S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\right| = 0.$$

Note that $1 \geq \hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_r > 0$ are the r largest sample canonical correlations between R_Δ and R_{-1} (for more details, consult section 2.1 in the factor model text). Heuristically, in the case that there are no lagged differences (no X), this means that the cointegrating vectors in $\hat{\beta}_T$ must be determined so that the sample correlation between ΔY and Y_{-1} is maximized. This is in accordance to our usual conception of cointegration that Y_{-1} must drive long run changes in ΔY if the variables in Y_t are cointegrated.

It remains to obtain a tractable expression for the cointegrating basis $\hat{\beta}_T$. We saw above that the columns of $\hat{C}_T = S_{-1}^{\frac{1}{2}\prime}\hat{\beta}_T$ at the maximum can be chosen to be any linearly independent eigenvectors corresponding to the eigenvalues $1-\hat{\lambda}_1, \cdots, 1-\hat{\lambda}_r$ of M. To simplify things, note first that, for any eigenvector v of M with eigenvalue μ, the quantity v is an eigenvector of the positive definite matrix $S_{-1}^{-\frac{1}{2}} S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1} S_{-1}^{-\frac{1}{2}\prime}$ with eigenvalue $1-\mu$; this follows by noting that

$$Mv = v - S_{-1}^{-\frac{1}{2}} S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1} S_{-1}^{-\frac{1}{2}\prime} \cdot v = \mu \cdot v,$$

so that we have

$$S_{-1}^{-\frac{1}{2}} S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1} S_{-1}^{-\frac{1}{2}\prime} \cdot v = (1-\mu)v.$$

Thus, it follows that the columns of $\hat{C}_T = S_{-1}^{\frac{1}{2}\prime}\hat{\beta}_T$ can be chosen to be orthonormal eigenvectors of $S_{-1}^{-\frac{1}{2}} S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1} S_{-1}^{-\frac{1}{2}\prime}$ corresponding to its r largest eigenvalues $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_r > 0$. Note that we have imposed the normalization

$$\hat{C}_T'\hat{C}_T = \hat{\beta}_T' S_{-1}\hat{\beta}_T = I_r.$$

Summary of QMLE Values

Given the quantities

$$
\begin{aligned}
R_\Delta &= M_X(\Delta Y) = \Delta Y - X(X'X)^{-1}X'(\Delta Y)\\
R_{-1} &= M_X \cdot Y_{-1} = Y_{-1} - X(X'X)^{-1}X'Y_{-1}\\
S_\Delta &= \frac{1}{T-p}R'_\Delta R_\Delta\\
S_{-1} &= \frac{1}{T-p}R'_{-1}R_{-1}\\
S_{\Delta,-1} &= \frac{1}{T-p}R'_\Delta R_{-1},
\end{aligned}
$$

the Gaussian QMLEs of the model parameters are given as follows:

$$\hat{\beta}_T = S_{-1}^{-\frac{1}{2}\prime}\begin{pmatrix}\hat{C}_{1,T} & \cdots & \hat{C}_{r,T}\end{pmatrix},$$

where $\hat{C}_{1,T},\cdots,\hat{C}_{r,T}$ are orthonormal eigenvectors of $S_{-1}^{-\frac{1}{2}}S'_{\Delta,-1}S_\Delta^{-1}S_{\Delta,-1}S_{-1}^{-\frac{1}{2}\prime}$ corresponding to its r largest eigenvalues $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_r > 0$

$$\hat{\alpha}_T = R'_\Delta R_{-1}\hat{\beta}_T(\hat{\beta}'_T R'_{-1}R_{-1}\hat{\beta}_T)^{-1}$$

$$\hat{\Sigma}_T = S_\Delta - S_{\Delta,-1}\hat{\beta}_T\left(\hat{\beta}'_T S_{-1}\hat{\beta}_T\right)^{-1}\hat{\beta}'_T S'_{\Delta,-1}$$

$$\hat{\Gamma}_T = (X'X)^{-1}X'\left(\Delta Y - Y_{-1}\cdot\hat{\beta}_T\hat{\alpha}'_T\right).$$

The maximized log-likelihood is

$$\hat{l}_T = -\frac{n(T-p)}{2}\left(\log(2\pi) + 1 + \frac{1}{n}\log|S_\Delta|\right) - \frac{T-p}{2}\sum_{i=1}^{r}\log\left(1-\hat{\lambda}_i\right).$$

Note that, if the eigenvalues of $S_{-1}^{-1}S'_{\Delta,-1}S_\Delta^{-1}S_{\Delta,-1}$ are distinct, then $\hat{\beta}_T$ is unique up to sign changes in its columns.

5.4.4 Maximum Likelihood Estimates when $r = 0$

So far, we have studied the (quasi) maximum likelihood estimates of the model when $0 < r < n$. The case when $r = 0$, that is, when there is no cointegrating relationships, is much simpler to analyze. $r = 0$ is equivalent to the claim that $\Pi = \alpha\beta' = O$, so that the VECM becomes a VAR in first-differences:

$$\Delta Y_t = \Gamma_1 \cdot \Delta Y_{t-1} + \cdots + \Gamma_{p-1} \cdot \Delta Y_{t-p+1} + \varepsilon_t.$$

Therefore, the maximized log-likelihood is the same as in the stationary VAR case, given by

$$\hat{l}_T = -\frac{n(T-p)}{2}\left(\log(2\pi) + 1 + \frac{1}{n}\log|S_\Delta|\right),$$

and the QMLEs of Γ and Σ are

$$\begin{aligned}
\hat{\Gamma}_T &= (X'X)^{-1}X'\Delta Y \\
\hat{\Sigma}_T &= S_\Delta = \frac{1}{T-p}\left(\Delta Y - X \cdot \hat{\Gamma}_T\right)'\left(\Delta Y - X \cdot \hat{\Gamma}_T\right).
\end{aligned}$$

Since the assumptions on the first-difference process $\{\Delta Y_t\}_{t\in\mathbb{Z}}$ and the innovation process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are identical to those for stationary vector autoregressions, the asymptotic results proved there continue to hold; specifically,

$$\begin{gathered}
\hat{\Sigma}_T \xrightarrow{p} \Sigma \\
\sqrt{T}\left(\mathrm{vec}\left(\hat{\Gamma}_T\right) - \mathrm{vec}(\Gamma)\right) \xrightarrow{d} N\left[\mathbf{0}, \Sigma \bigotimes Q\right],
\end{gathered}$$

where

$$Q = \begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0) \end{pmatrix}.$$

5.4.5 Specifying $\alpha_\perp$

When studying the asymptotic properties of the QMLEs derived above, it is conveninent to choose the following specific form for the orthogonal complement $\alpha_\perp$. Assuming that $r > 0$, define

$$P_\alpha(M) = \alpha(\alpha' M^{-1}\alpha)^{-1}\alpha' M^{-1}$$

for any positive definite matrix M. Since the matrices $I_n - P_\alpha(\Sigma)$ and $P_\alpha(\Sigma)$ are idempotent, their rank equals their traces, so that

$$rank(I_n - P_\alpha(\Sigma)) = \operatorname{tr}(I_n - P_\alpha(\Sigma)) = n - r.$$

Consider the matrix

$$\Sigma^{-1}(I_n - P_\alpha(\Sigma)).$$

This matrix clearly has rank $n-r$, and

$$\Sigma^{-1}(I_n - P_\alpha(\Sigma)) = \Sigma^{-1} - \Sigma^{-1}\alpha(\alpha'\Sigma^{-1}\alpha)^{-1}\alpha'\Sigma^{-1}$$

shows us that $\Sigma^{-1}(I_n - P_\alpha(\Sigma))$ is positive semidefinite. Therefore, it has the eigendecompsition

$$\Sigma^{-1}(I_n - P_\alpha(\Sigma)) = PDP'$$

for orthogonal matrix $P \in \mathbb{R}^{n\times n}$ and diagonal matrix D with diagonal entries equal to the eigenvalues of $\Sigma^{-1}(I_n - P_\alpha(\Sigma))$, which are non-negative; the last r entries in D are 0. Denoting

$$D = \begin{pmatrix} \tilde{D} & O \\ O & O \end{pmatrix},$$

where $\tilde{D} \in \mathbb{R}^{(n-r)\times(n-r)}$ collects the non-zero elements of D, define

$$\alpha_\perp = P\begin{pmatrix} \tilde{D}^{\frac{1}{2}} \\ O \end{pmatrix} \in \mathbb{R}^{n\times n-r}.$$

We can then see that

$$\begin{aligned} \alpha_\perp\alpha_\perp' &= \Sigma^{-1}(I_n - P_\alpha(\Sigma)) \\ \alpha_\perp'\alpha_\perp &= \tilde{D}. \end{aligned}$$

By implication,

$$\alpha_\perp\alpha_\perp'\alpha = \Sigma^{-1}(I_n - P_\alpha(\Sigma))\alpha = O$$

and thus

$$\alpha_\perp' \alpha = \tilde{D}^{-1} O = O,$$

so that $\alpha_\perp$ truly does act like the orthogonal complement of α. Another important property that this $\alpha_\perp$ possesses is that, because

$$\Sigma \alpha_\perp \alpha_\perp' = I_n - P_\alpha(\Sigma),$$

and the matrix on the right hand side is idempotent,

$$\Sigma \alpha_\perp (\alpha_\perp' \Sigma \alpha_\perp) \alpha_\perp' = \Sigma \alpha_\perp \alpha_\perp',$$

which implies that

$$\alpha_\perp' \Sigma \alpha_\perp = I_{n-r}.$$

When $r = 0$, we define

$$\alpha_\perp = \Sigma^{-\frac{1}{2}\prime},$$

the inverse of the Cholesky factor of Σ.

Define the $n-r$-dimensional Brownian motion $\{B(s)\}_{s\in[0,1]}$ as

$$B(s) = \alpha_\perp' \Sigma^{\frac{1}{2}} \cdot W^n(s)$$

for any $s \in [0,1]$, where $\{W^n(s)\}_{s\in[0,1]}$ is the standard n-dimensional Wiener process. Since $\alpha_\perp' \Sigma \alpha_\perp = I_{n-r}$, it follows that $\{B(s)\}_{s\in[0,1]}$ has variance I_{n-r} and therefore is identically distributed to the standard $n-r$-dimensional Wiener process. This establishes that

$$\alpha_\perp' \Sigma^{\frac{1}{2}} \cdot W^n(s) \sim W^{n-r}(s)$$

under our choice of $\alpha_\perp$.

5.4.6 Consistency of QMLEs

Now that the QMLEs have been derived, we can study their asymptotic properties. We start with the r largest sample canonical correlations $\hat{\lambda}_1, \cdots, \hat{\lambda}_r$ of R_Δ and R_{-1}, and then make our way back up to $\hat{\Sigma}_T$ and $\hat{\Gamma}_T$.

Note first that the canonical correlations, being real ordered eigenvalues of the positive definite matrix $S_{-1}^{-\frac{1}{2}} S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1} S_{-1}^{-\frac{1}{2}\prime}$, are continuous functions of $S_{-1}^{-\frac{1}{2}} S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1} S_{-1}^{-\frac{1}{2}\prime}$ (for a formal proof of this result, consult section 1.2.3 in the factor model text). This ensures that $\hat{\lambda}_1, \cdots, \hat{\lambda}_r$ are well-defined random variables, and that we can use the continuous mapping theorem to derive the asymptotic distribution of the canonical correlations. Formally, we will denote

$$(\hat{\lambda}_1, \cdots, \hat{\lambda}_r) = eig_n^r \left(S_{-1}^{-\frac{1}{2}} S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1} S_{-1}^{-\frac{1}{2}\prime} \right),$$

so that $eig_n^r(\cdot)$ is a function that extracts the r largest eigenvalues from an $n \times n$ matrix with real ordered eigenvalues.

The quantities $S_\Delta, S_{-1}, S_{\Delta,-1}$ defined above have the following asymptotic properties:

Lemma (Asymptotic Results for QMLE)
Under assumptions A1 to A3, the following convergence results hold jointly:

$$S_\Delta \xrightarrow{p} \mu_\Delta$$

$$\frac{1}{T-p} S_{-1} \xrightarrow{d} \Lambda \left(\int_0^1 W^n(r) W^n(r)' dr \right) \Lambda'$$

$$\frac{1}{T-p} \beta_\perp' S_{-1} \beta_\perp \xrightarrow{d} (\beta_\perp' \beta_\perp)(\alpha_\perp' \Gamma(1) \beta_\perp)^{-1} \left(\int_0^1 W^{n-r}(s) W^{n-r}(s)' ds \right) (\beta_\perp' \Gamma(1)' \alpha_\perp)^{-1} (\beta_\perp' \beta_\perp)$$

$$\left(\frac{1}{T-p} \beta_\perp' S_{-1} \beta_\perp \right)^{-1} = O_p(1)$$

$$S_{-1} \beta = O_p(1)$$

$$\beta' S_{-1} \beta \xrightarrow{p} \mu_{-1}$$

$$S_{\Delta,-1} \xrightarrow{d} \mu_{\Delta,-1} + \Lambda \left(\int_0^1 W^n(r) dW^n(r)' \right)' \Lambda' - \begin{pmatrix} G(1) & \cdots & G(p-1) \end{pmatrix} \begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0) \end{pmatrix}^{-1} \begin{pmatrix} \Lambda \left(\int_0^1 W^n(r) dW^n(r)' \right)' \Lambda' \\ \vdots \\ \Lambda \left(\int_0^1 W^n(r) dW^n(r)' \right)' \Lambda' \end{pmatrix}$$

$$S_{\Delta,-1} \beta \xrightarrow{p} \mu_{\Delta,-1} \beta$$

$$\beta' S_{\Delta,-1} = O_p(1)$$

$$S'_{\Delta,-1}\alpha_\perp \xrightarrow{d} \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}\alpha_\perp,$$

where

$$\mu_\Delta = G(0) - \begin{pmatrix} G(1) & \cdots & G(p-1) \end{pmatrix} \begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0) \end{pmatrix}^{-1} \begin{pmatrix} G(1)' \\ \vdots \\ G(p-1)' \end{pmatrix}$$

$$\mu_{-1} = G_\beta(0) - \beta'\begin{pmatrix} \Lambda_1 & \cdots & \Lambda_{p-1} \end{pmatrix} \begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0) \end{pmatrix}^{-1} \begin{pmatrix} \Lambda_1' \\ \vdots \\ \Lambda_{p-1}' \end{pmatrix}\beta$$

$$\mu_{\Delta,-1} = \Lambda_0' - \begin{pmatrix} G(1) & \cdots & G(p-1) \end{pmatrix} \begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0) \end{pmatrix}^{-1} \begin{pmatrix} \Lambda_1' \\ \vdots \\ \Lambda_{p-1}' \end{pmatrix}.$$

μ_Δ, μ_{-1} and $\mu_{\Delta,-1}$ are related to one another as

$$\mu_\Delta = \Sigma + \mu_{\Delta,-1}\beta\alpha' \quad \text{and} \quad \mu_{-1}\alpha' = \beta'\mu'_{\Delta,-1}.$$

Proof) We can recover the probability limit of S_Δ as

$$\begin{aligned} S_\Delta &= \frac{1}{T-p}R'_\Delta R_\Delta = \frac{1}{T-p}(\Delta Y)'M_X(\Delta Y) \\ &= \frac{1}{T-p}(\Delta Y)'(\Delta Y) - \left(\frac{1}{T-p}(\Delta Y)'X\right)\left(\frac{1}{T-p}X'X\right)^{-1}\left(\frac{1}{T-p}(\Delta Y)'X\right)' \\ &\xrightarrow{p} G(0) - \begin{pmatrix} G(1) & \cdots & G(p-1) \end{pmatrix} \begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0) \end{pmatrix}^{-1} \begin{pmatrix} G(1)' \\ \vdots \\ G(p-1)' \end{pmatrix} = \mu_\Delta. \end{aligned}$$

Note that μ_Δ is the Schur complement of the positive definite matrix

$$\begin{pmatrix} G(0) & \cdots & G(p-1) \\ \vdots & \ddots & \vdots \\ G(p-1)' & \cdots & G(0) \end{pmatrix},$$

so that it is also positive definite.

Likewise, we can conclude that

$$\beta'S_{-1}\beta = \frac{1}{T-p}\beta'R'_{-1}R_{-1}\beta = \frac{1}{T-p}\beta'Y'_{-1}M_XY_{-1}\beta$$

$$= \frac{1}{T-p}\beta' Y_{-1}' Y_{-1}\beta - \left(\frac{1}{T-p}\beta' Y_{-1}' X\right)\left(\frac{1}{T-p}X'X\right)^{-1}\left(\frac{1}{T-p}\beta' Y_{-1}' X\right)'$$

$$\xrightarrow{p} G_\beta(0) - \beta'\begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}\Lambda_1' \\ \vdots \\ \Lambda_{p-1}'\end{pmatrix}\beta = \mu_{-1}.$$

We also have

$$\begin{aligned} S_{\Delta,-1} &= \frac{1}{T-p}(\Delta Y)' M_X Y_{-1} \\ &= \frac{1}{T-p}(\Delta Y)' Y_{-1} - \left(\frac{1}{T-p}(\Delta Y)' X\right)\left(\frac{1}{T-p}X'X\right)^{-1}\left(\frac{1}{T-p}Y_{-1}' X\right)' \\ &\xrightarrow{d} \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)'\Lambda' + \Lambda_0' \\ &\quad - \begin{pmatrix} G(1) & \cdots & G(p-1)\end{pmatrix}\begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}\Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)'\Lambda' \\ \vdots \\ \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)'\Lambda'\end{pmatrix} \\ &\quad - \begin{pmatrix} G(1) & \cdots & G(p-1)\end{pmatrix}\begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}\Lambda_1' \\ \vdots \\ \Lambda_{p-1}'\end{pmatrix}. \end{aligned}$$

When $r > 0$, using the fact that

$$\Gamma = \begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix} G(1)' \\ \vdots \\ G(p-1)'\end{pmatrix} - \begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}\Lambda_1' \\ \vdots \\ \Lambda_{p-1}'\end{pmatrix}\beta\alpha',$$

the above limit can also be reformulated as

$$\begin{aligned} S_{\Delta,-1} &\xrightarrow{d} \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)'\Lambda' + \Lambda_0' - \Gamma'\begin{pmatrix}\Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)'\Lambda' \\ \vdots \\ \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)'\Lambda'\end{pmatrix} \\ &\quad - \alpha\beta'\begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}\Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)'\Lambda' \\ \vdots \\ \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)'\Lambda'\end{pmatrix} \\ &\quad - \begin{pmatrix} G(1) & \cdots & G(p-1)\end{pmatrix}\begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}\Lambda_1' \\ \vdots \\ \Lambda_{p-1}'\end{pmatrix} \end{aligned}$$

$$= \Gamma(1)'\Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)'\Lambda' + \Lambda_0'$$

$$- \alpha\beta'\begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\begin{pmatrix}G(0) & \cdots & G(p-2)\\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}\Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)'\Lambda' \\ \vdots \\ \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)'\Lambda'\end{pmatrix}$$

$$- \begin{pmatrix}G(1) & \cdots & G(p-1)\end{pmatrix}\begin{pmatrix}G(0) & \cdots & G(p-2)\\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}\Lambda_1' \\ \vdots \\ \Lambda_{p-1}'\end{pmatrix}.$$

It now follows easily that

$$S_{\Delta,-1}\beta \xrightarrow{p} \left[\Lambda_0' - \begin{pmatrix}G(1) & \cdots & G(p-1)\end{pmatrix}\begin{pmatrix}G(0) & \cdots & G(p-2)\\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}\Lambda_1' \\ \vdots \\ \Lambda_{p-1}'\end{pmatrix}\right]\beta$$

$$= \mu_{\Delta,-1}\beta$$

since $\beta'\Lambda = O$, where the convergence in distribution changes into convergence in probability because the limit is non-random. Similarly,

$$\beta' S_{\Delta,-1} \xrightarrow{d} \beta'\Lambda_0' - \beta'\begin{pmatrix}G(1) & \cdots & G(p-1)\end{pmatrix}\begin{pmatrix}G(0) & \cdots & G(p-2)\\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}\Lambda_1' \\ \vdots \\ \Lambda_{p-1}'\end{pmatrix}$$

$$- \beta'\begin{pmatrix}G(1) & \cdots & G(p-1)\end{pmatrix}\begin{pmatrix}G(0) & \cdots & G(p-2)\\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}\Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)'\Lambda' \\ \vdots \\ \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)'\Lambda'\end{pmatrix}.$$

where the convergence this time is in distribution.

If $r > 0$, then

$$\begin{aligned}\Lambda'\Gamma(1)'\alpha_\perp &= \Sigma^{\frac{1}{2}\prime} C'\Gamma(1)\alpha_\perp \\ &= \Sigma^{\frac{1}{2}\prime}\alpha_\perp(\beta_\perp'\Gamma(1)'\alpha_\perp)^{-1}\beta_\perp'\Gamma(1)'\alpha_\perp \\ &= \Sigma^{\frac{1}{2}\prime}\alpha_\perp,\end{aligned}$$

while if $r = 0$, because $\Lambda = C\Sigma^{\frac{1}{2}} = \Gamma(1)^{-1}\Sigma^{\frac{1}{2}}$ and $\alpha_\perp = I_n$, we still have $\Lambda'\Gamma(1)'\alpha_\perp = \Sigma^{\frac{1}{2}\prime}\alpha_\perp$. Therefore,

$$S_{\Delta,-1}'\alpha_\perp \xrightarrow{d} \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}\alpha_\perp + \Lambda_0\alpha_\perp$$

$$
-\begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\begin{pmatrix}G(0) & \cdots & G(p-2)\\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}G(1)'\\ \vdots \\ G(p-1)'\end{pmatrix}\alpha_\perp
$$

$$
= \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}\alpha_\perp + \Lambda_0\alpha_\perp - \begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\Gamma\alpha_\perp
$$

$$
= \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}\alpha_\perp,
$$

where the last equality follows because

$$
\Lambda_0\alpha_\perp = \begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\Gamma\alpha_\perp.
$$

Meanwhile,

$$
\begin{aligned}
\frac{1}{T-p}S_{-1} &= \frac{1}{(T-p)^2}Y_{-1}'M_XY_{-1}\\
&= \frac{1}{(T-p)^2}Y_{-1}'Y_{-1} - \left(\frac{1}{(T-p)^{3/2}}Y_{-1}'X\right)\left(\frac{1}{T-p}X'X\right)^{-1}\left(\frac{1}{(T-p)^{3/2}}Y_{-1}'X\right)'\\
&\xrightarrow{d} \Lambda\left(\int_0^1 W^n(r)W^n(r)'dr\right)\Lambda',
\end{aligned}
$$

since $\frac{1}{(T-p)^{3/2}}Y_{-1}'X = o_p(1)$. Pre- and post-multiplying both sides by $\beta_\perp$ implies

$$
\frac{1}{T-p}\beta_\perp' S_{-1}\beta_\perp \xrightarrow{d} \beta_\perp'\Lambda\left(\int_0^1 W^n(r)W^n(r)'dr\right)\Lambda'\beta_\perp.
$$

Here,

$$
\beta_\perp\Lambda = (\beta_\perp'\beta_\perp)(\alpha_\perp'\Gamma(1)\beta_\perp)^{-1}\alpha_\perp'\Sigma^{\frac{1}{2}},
$$

so that

$$
\frac{1}{T-p}\beta_\perp' S_{-1}\beta_\perp \xrightarrow{d} (\beta_\perp'\beta_\perp)(\alpha_\perp'\Gamma(1)\beta_\perp)^{-1}\left(\int_0^1 B(s)B(s)'ds\right)(\beta_\perp'\Gamma(1)'\alpha_\perp)^{-1}(\beta_\perp'\beta_\perp),
$$

where we define $B(s) = \alpha_\perp'\Sigma^{\frac{1}{2}}\cdot W^n(s)$ for any $s\in[0,1]$. Since $B(s)\sim W^{n-r}(s)$, we can see that

$$
\frac{1}{T-p}\beta_\perp' S_{-1}\beta_\perp \xrightarrow{d} (\beta_\perp'\beta_\perp)(\alpha_\perp'\Gamma(1)\beta_\perp)^{-1}\left(\int_0^1 W^{n-r}(s)W^{n-r}(s)'ds\right)(\beta_\perp'\Gamma(1)'\alpha_\perp)^{-1}(\beta_\perp'\beta_\perp).
$$

Clearly, the limit has full rank $n-r$, so that

$$
\left(\frac{1}{T-p}\beta_\perp' S_{-1}\beta_\perp\right)^{-1} = O_p(1).
$$

We can also see that

$$
\begin{aligned}
S_{-1}\beta &= \frac{1}{T-p} Y_{-1}' M_X Y_{-1}\beta \\
&= \frac{1}{T-p} Y_{-1}' Y_{-1}\beta - \left(\frac{1}{T-p} Y_{-1}' X\right)\left(\frac{1}{T-p} X'X\right)^{-1}\left(\frac{1}{T-p}\beta' Y_{-1}' X\right)' \\
&\xrightarrow{d} \left[\Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)\Sigma^{\frac{1}{2}\prime}(\Gamma(1)C - I_n)' + \Lambda_0 - \begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\Gamma\right]\bar{\alpha} \\
&\quad - \left[\beta'\begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}\Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)'\Lambda' \\ \vdots \\ \Lambda\left(\int_0^1 W^n(r)dW^n(r)'\right)'\Lambda'\end{pmatrix}\right]' \\
&\quad - \left[\beta'\begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}\Lambda_1' \\ \vdots \\ \Lambda_{p-1}'\end{pmatrix}\right]'.
\end{aligned}
$$

Therefore, we can conclude that $S_{-1}\beta = O_p(T)$.

The relationship between μ_Δ and $\mu_{\Delta,-1}$ can be seen by noting that

$$
\begin{aligned}
\Sigma &= G(0) - \Lambda_0'\beta\alpha' - \begin{pmatrix}G(1) & \cdots & G(p-1)\end{pmatrix}\Gamma \\
&= G(0) - \Lambda_0'\cdot\beta\alpha' - \begin{pmatrix}G(1) & \cdots & G(p-1)\end{pmatrix}\begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}G(1)' \\ \vdots \\ G(p-1)'\end{pmatrix} \\
&\quad - \begin{pmatrix}G(1) & \cdots & G(p-1)\end{pmatrix}\begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}\Lambda_1' \\ \vdots \\ \Lambda_{p-1}'\end{pmatrix}\beta\alpha' \\
&= \mu_\Delta - \mu_{\Delta,-1}\beta\alpha'.
\end{aligned}
$$

The relationship between μ_{-1} and $\mu_{\Delta,-1}$ follows from

$$
\mu_{-1}\alpha' = G_\beta(0)\alpha' - \beta'\begin{pmatrix}\Lambda_1 & \cdots & \Lambda_{p-1}\end{pmatrix}\left[\begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0)\end{pmatrix}^{-1}\begin{pmatrix}G(1)' \\ \vdots \\ G(p-1)'\end{pmatrix} - \Gamma\right]
$$

$$= \beta' \left[\Lambda_0 - \begin{pmatrix} \Lambda_1 & \cdots & \Lambda_{p-1} \end{pmatrix} \begin{pmatrix} G(0) & \cdots & G(p-2) \\ \vdots & \ddots & \vdots \\ G(p-2)' & \cdots & G(0) \end{pmatrix}^{-1} \begin{pmatrix} G(1)' \\ \vdots \\ G(p-1)' \end{pmatrix} \right]$$

$$= \beta' \mu'_{\Delta,-1}.$$

Q.E.D.

Suppose $r > 0$. Note that

$$\begin{pmatrix} \mu_{-1} & \mu_{\Delta,-1}\beta \\ \beta'\mu'_{\Delta,-1} & \mu_\Delta \end{pmatrix}$$

is the probability limit of

$$\begin{pmatrix} \beta' S_{-1}\beta & S_{\Delta,-1}\beta \\ \beta' S'_{\Delta,-1} & S_\Delta \end{pmatrix} = \begin{pmatrix} \beta' Y'_{-1} \\ (\Delta Y)' \end{pmatrix} \left(\frac{1}{T-p} M_X \right) \begin{pmatrix} Y_{-1}\beta & \Delta Y \end{pmatrix}.$$

If we assume that the smallest eigenvalue of this matrix is bounded below at a level greater than 0, it follows that

$$\begin{pmatrix} \mu_{-1} & \mu_{\Delta,-1}\beta \\ \beta'\mu'_{\Delta,-1} & \mu_\Delta \end{pmatrix}$$

must also be positive definite. By impliction, μ_{-1}, μ_Δ and the Schur complement $\mu_{-1} - \beta'\mu'_{\Delta,-1}\mu_\Delta^{-1}\mu_{\Delta,-1}\beta$ must also be positive definite.

If $r = 0$, then we can still assume that μ_Δ is positive definite.

We can also show that $P_\alpha(\mu_\Delta)$ and $P_\alpha(\Sigma)$ are related in the following manner:

Lemma (Relationship between $P_\alpha(\mu_\Delta)$ and $P_\alpha(\Sigma)$)
Under assumptions A1 to A3, if $r > 0$, then the following hold true:

$$P_\alpha(\Sigma) = P_\alpha(\mu_\Delta) \quad \text{and} \quad \Sigma^{-1}(I_n - P_\alpha(\Sigma)) = \mu_\Delta^{-1}(I_n - P_\alpha(\mu_\Delta)).$$

Proof) The above results make use of the fact that

$$\mu_\Delta = \Sigma + \mu_{\Delta,-1}\beta\alpha' = \Sigma + \alpha\mu_{-1}\alpha'.$$

First we show that $P_\alpha(\Sigma) = P_\alpha(\mu_\Delta)$. To this end, note that

$$\mu_\Delta^{-1} = \Sigma^{-1} - \Sigma^{-1}\alpha\mu_{-1}\alpha'\mu_\Delta^{-1};$$

this can be seen by direct verification through the identity $\mu_\Delta = \Sigma + \alpha\mu_{-1}\alpha'$. It follows that

$$\alpha'\mu_\Delta^{-1}\alpha = \alpha'\Sigma^{-1}\alpha - \alpha'\Sigma^{-1}\alpha\mu_{-1}\alpha'\mu_\Delta^{-1}\alpha$$
$$\alpha'\mu_\Delta^{-1} = \alpha'\Sigma^{-1} - \alpha'\Sigma^{-1}\alpha\mu_{-1}\alpha'\mu_\Delta^{-1};$$

direct verification via the first equation leads us to the conclusion that

$$\left(\alpha'\mu_\Delta^{-1}\alpha\right)^{-1} = \left(\alpha'\Sigma^{-1}\alpha\right)^{-1} + \mu_{-1}.$$

Therefore,

$$\begin{aligned}
\alpha\left(\alpha'\mu_\Delta^{-1}\alpha\right)^{-1}\alpha'\mu_\Delta^{-1} &= \alpha\left[\left(\alpha'\Sigma^{-1}\alpha\right)^{-1} + \mu_{-1}\right]\left[\alpha'\Sigma^{-1} - \alpha'\Sigma^{-1}\alpha\mu_{-1}\alpha'\mu_\Delta^{-1}\right] \\
&= \alpha\left[\left(\alpha'\Sigma^{-1}\alpha\right)^{-1}\alpha'\Sigma^{-1} + \mu_{-1}\alpha'\Sigma^{-1} - \mu_{-1}\alpha'\mu_\Delta^{-1} - \mu_{-1}\alpha'\Sigma^{-1}\alpha\mu_{-1}\alpha'\mu_\Delta^{-1}\right] \\
&= P_\alpha(\Sigma) + \left(\alpha\mu_{-1}\alpha'\right)\Sigma^{-1} - (\alpha\mu_{-1}\alpha')^{-1}\mu_\Delta^{-1} - (\alpha\mu_{-1}\alpha')\Sigma^{-1}(\alpha\mu_{-1}\alpha')\mu_\Delta^{-1} \\
&= P_\alpha(\Sigma) + \left[(\alpha\mu_{-1}\alpha')^{-1}\Sigma^{-1} + I_n\right]\left[I_n - (\alpha\mu_{-1}\alpha')^{-1}\mu_\Delta^{-1}\right] - I_n.
\end{aligned}$$

Since $\alpha\mu_{-1}\alpha' = \mu_\Delta - \Sigma$, we can see that

$$(\alpha\mu_{-1}\alpha')^{-1}\Sigma^{-1} = \mu_\Delta\Sigma^{-1} - I_n$$
$$(\alpha\mu_{-1}\alpha')^{-1}\mu_\Delta^{-1} = I_n - \Sigma\mu_\Delta^{-1},$$

so that

$$P_\alpha(\mu_\Delta) = \alpha\left(\alpha'\mu_\Delta^{-1}\alpha\right)^{-1}\alpha'\mu_\Delta^{-1} = P_\alpha(\Sigma).$$

It remains to show that

$$\Sigma^{-1}\left(I_n - P_\alpha(\Sigma)\right) = \mu_\Delta^{-1}\left(I_n - P_\alpha(\mu_\Delta)\right).$$

Since $P_\alpha(\mu_\Delta) = P_\alpha(\Sigma)$, we can see that

$$\begin{aligned}
\mu_\Delta^{-1}\left(I_n - P_\alpha(\mu_\Delta)\right) &= \mu_\Delta^{-1} - \mu_\Delta^{-1}P_\alpha(\Sigma) \\
&= \mu_\Delta^{-1} - \underbrace{\mu_\Delta^{-1}\alpha\left(\alpha'\mu_\Delta^{-1}\alpha\right)^{-1}\alpha'}_{P_\alpha(\mu_\Delta)'}\mu_\Delta^{-1} \\
&= \Sigma^{-1} - \Sigma^{-1}(\alpha\mu_{-1}\alpha')\mu_\Delta^{-1} - P_\alpha(\Sigma)'\mu_\Delta^{-1} \\
&= \Sigma^{-1} - \Sigma^{-1}(\alpha\mu_{-1}\alpha')\mu_\Delta^{-1} - \Sigma^{-1}\alpha\left(\alpha'\Sigma^{-1}\alpha\right)^{-1}\alpha'\mu_\Delta^{-1} \\
&= \Sigma^{-1} - \Sigma^{-1}\alpha\left[\mu_{-1} + \left(\alpha'\Sigma^{-1}\alpha\right)^{-1}\right]\alpha'\mu_\Delta^{-1}
\end{aligned}$$

$$
\begin{aligned}
&= \Sigma^{-1} - \Sigma^{-1} \underbrace{\alpha\left(\alpha' \mu_\Delta^{-1}\alpha\right)^{-1} \alpha' \mu_\Delta^{-1}}_{P_\alpha(\mu_\Delta)} \\
&= \Sigma^{-1} - \Sigma^{-1} P_\alpha(\Sigma) = \Sigma^{-1}\left(I_n - P_\alpha(\Sigma)\right).
\end{aligned}
$$

Q.E.D.

The preceding results allow us to prove a consistency result for the sample canonical correlations $\hat{\lambda}_1, \cdots, \hat{\lambda}_r$ obtained as the r largest solutions to the eigenvalue equation

$$
\left|\lambda S_{-1} - S'_{\Delta,-1} S_\Delta^{-1} S_{\Delta,-1}\right| = 0.
$$

Theorem (Consistency of Eigenvalues)
Maintain assumptions A1 to A3. Let $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_n > 0$ be the real ordered solutions to the eigenvalue equation

$$
\left|\lambda S_{-1} - S'_{\Delta,-1} S_\Delta^{-1} S_{\Delta,-1}\right| = 0.
$$

If $r > 0$, then

$$
(\hat{\lambda}_1, \cdots, \hat{\lambda}_n) \xrightarrow{p} (\lambda_1, \cdots, \lambda_r, 0, \cdots, 0),
$$

where $\lambda_1 \geq \cdots \geq \lambda_r \geq 0$ are the ordered eigenvalues that solve the problem

$$
\left|\lambda \mu_{-1} - \beta' \mu'_{\Delta,-1} \mu_\Delta^{-1} \mu_{\Delta,-1} \beta\right| = 0.
$$

On the other hand, if $r = 0$, then

$$
(\hat{\lambda}_1, \cdots, \hat{\lambda}_n) \xrightarrow{p} \left(\underbrace{0, \cdots, 0}_{n}\right).
$$

Proof) **The Case $r = 0$**

We first prove the result for the case $r = 0$. In this case, because $\beta_\perp = I_n$, we have

$$
\left(\frac{1}{T-p} S_{-1}\right)^{-1} = O_p(1),
$$

and by implication,

$$S_{-1}^{-1} = \frac{1}{T-p}\left(\frac{1}{T-p}S_{-1}\right)^{-1} = o_p(1).$$

Since $\hat{\lambda}_1, \cdots, \hat{\lambda}_n$ are the solutions to the equation

$$\left|\lambda \cdot I_n - S_{-1}^{-\frac{1}{2}} S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1} S_{-1}^{-\frac{1}{2}\prime}\right| = 0,$$

and

$$S_{-1}^{-\frac{1}{2}} S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1} S_{-1}^{-\frac{1}{2}\prime} = o_p(1),$$

by the continuity of ordered eigenvalues $\hat{\lambda}_1, \cdots, \hat{\lambda}_n$ converge in probability to the solutions of the equation

$$|\lambda \cdot I_n| = 0,$$

that is, to the n-dimensional zero vector.

The Case $r > 0$

Recall that, if $r > 0$, we assumed that

$$\begin{pmatrix} \mu_{-1} & \mu_{\Delta,-1}\beta \\ \beta'\mu_{\Delta,-1}' & \mu_{\Delta} \end{pmatrix}$$

is positive definite. Thus, $\lambda_1, \cdots, \lambda_r$ are the eigenvalues of the positive definite matrix

$$\mu_{-1}^{-\frac{1}{2}} \beta' \mu_{\Delta,-1}' \mu_{\Delta}^{-1} \mu_{\Delta,-1} \beta \mu_{-1}^{-\frac{1}{2}\prime},$$

so that $\lambda_1, \cdots, \lambda_r$ are real, ordered and non-zero.

For any $\lambda \in \mathbb{R}$, defining $M_T = (\beta_\perp' S_{-1} \beta_\perp)^{-\frac{1}{2}}$, the inverse of the Cholesky factor of $\beta_\perp' S_{-1} \beta_\perp$, we have

$$\begin{aligned}
&|(\beta,\beta_\perp)'| \cdot \left|\lambda S_{-1} - S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1}\right| \cdot |(\beta,\beta_\perp)| \\
&\qquad = \left|\begin{pmatrix} \beta' \\ \beta_\perp \end{pmatrix} \left(\lambda S_{-1} - S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1}\right) \begin{pmatrix} \beta & \beta_\perp \end{pmatrix}\right| \\
&= \left|\lambda \begin{pmatrix} \beta' S_{-1}\beta & \beta' S_{-1}\beta_\perp \\ \beta_\perp' S_{-1}\beta & \beta_\perp' S_{-1}\beta_\perp \end{pmatrix} - \begin{pmatrix} \beta' S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1}\beta & \beta' S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1}\beta_\perp \\ \beta_\perp' S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1}\beta & \beta_\perp' S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1}\beta_\perp \end{pmatrix}\right|
\end{aligned}$$

$$= \left|\lambda\begin{pmatrix}\beta' S_{-1}\beta & \beta' S_{-1}\beta_\perp M_T' \\ M_T\beta_\perp' S_{-1}\beta & M_T\beta_\perp' S_{-1}\beta_\perp M_T'\end{pmatrix} - \begin{pmatrix}\beta' S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\beta & \beta' S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\beta_\perp M_T' \\ M_T\beta_\perp' S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\beta & M_T\beta_\perp' S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\beta_\perp M_T'\end{pmatrix}\right| \times \left|\begin{pmatrix} I_r & O \\ O & M_T^{-1}\end{pmatrix}\right|.$$

Since $(\beta, \beta_\perp)'$ and M_T are nonsingular matrices, it follows that the solutions to the eigenvalue equation $\left|\lambda S_{-1} - S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\right| = 0$ are solutions to the equation

$$\left|\lambda\begin{pmatrix}\beta' S_{-1}\beta & \beta' S_{-1}\beta_\perp M_T' \\ M_T\beta_\perp' S_{-1}\beta & M_T\beta_\perp' S_{-1}\beta_\perp M_T'\end{pmatrix} - \begin{pmatrix}\beta' S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\beta & \beta' S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\beta_\perp M_T' \\ M_T\beta_\perp' S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\beta & M_T\beta_\perp' S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\beta_\perp M_T'\end{pmatrix}\right| = 0.$$

The preceding result tells us that

$$\left(\frac{1}{T-p}\beta_\perp' S_{-1}\beta_\perp\right)^{-1} = O_p(1),$$

so that

$$(\beta_\perp' S_{-1}\beta_\perp)^{-1} = \frac{1}{T-p}\left(\frac{1}{T-p}\beta_\perp' S_{-1}\beta_\perp\right)^{-1} = o_p(1),$$

implying that

$$M_T = (\beta_\perp' S_{-1}\beta_\perp)^{-\frac{1}{2}} = o_p(1)$$

as well by the continuity of the Cholesky operation. Therefore,

$$\begin{pmatrix}\beta' S_{-1}\beta & \beta' S_{-1}\beta_\perp M_T' \\ M_T\beta_\perp' S_{-1}\beta & M_T\beta_\perp' S_{-1}\beta_\perp M_T'\end{pmatrix} \xrightarrow{p} \begin{pmatrix}\mu_{-1} & O \\ O & I_{n-r}\end{pmatrix}$$

and

$$\begin{pmatrix}\beta' S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\beta & \beta' S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\beta_\perp M_T' \\ M_T\beta_\perp' S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\beta & M_T\beta_\perp' S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\beta_\perp M_T'\end{pmatrix} \xrightarrow{p} \begin{pmatrix}\beta'\mu_{\Delta,-1}'\mu_\Delta^{-1}\mu_{\Delta,-1}\beta & O \\ O & O\end{pmatrix}.$$

By the continuity of real ordered eigenvalues and the continuous mapping theorem, $(\hat\lambda_1, \cdots, \hat\lambda_n)$ should converge in probability to the solutions of the eigenvalue equation

$$\left|\lambda\begin{pmatrix}\mu_{-1} & O \\ O & I_{n-r}\end{pmatrix} - \begin{pmatrix}\beta'\mu_{\Delta,-1}'\mu_\Delta^{-1}\mu_{\Delta,-1}\beta & O \\ O & O\end{pmatrix}\right| = \left|\lambda\mu_{-1} - \beta'\mu_{\Delta,-1}'\mu_\Delta^{-1}\mu_{\Delta,-1}\beta\right| \cdot |\lambda \cdot I_{n-r}|.$$

This equation has exactly r non-zero solutions, denoted $\lambda_1 \geq \cdots \geq \lambda_r$, and $n-r$ solutions equal to 0.

Q.E.D.

While the uniqueness of the ordered sample canonical correlates allows us to obtain probability limits for them, the potential non-uniqueness of the corresponding eigenvectors of $S_{-1}^{-\frac{1}{2}}S'_{\Delta,-1}S_{\Delta}^{-1}S_{\Delta,-1}S_{-1}^{-\frac{1}{2}\prime}$ makes it more difficult to obtain limiting results for $\hat{\beta}_T$. Fortunately, we can establish that the cointegrating space, at the very least, is consistently estimated, in a sense to be discussed below.

First we decompose $\hat{\beta}_T$ as the sum of the projections of $\hat{\beta}_T$ onto the space spanned by the columns of β and its orthogonal complement; specifically, we let

$$\hat{\beta}_T = \beta \underbrace{(\beta'\beta)^{-1}\beta'\hat{\beta}_T}_{\hat{x}_T} + \beta_\perp \underbrace{(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'\hat{\beta}_T}_{\hat{y}_T}.$$

Since the column space of β is the (augmented) cointegrating space, we can say that the cointegrating space is consistently estimated if $\hat{y}_T = o_p(1)$, that is, if the projection of $\hat{\beta}_T$ onto the orthogonal complement of the cointegrating space vanishes as $T \to \infty$. In other words, the cointegrating space is consistently estimated if $\hat{\beta}_T$ belongs to the cointegrating space (the column space of β) with probability 1 as $T \to \infty$.

To show this consistency result, we require more preliminary asymptotic results. First, define

$$S(\lambda) = \lambda S_{-1} - S'_{\Delta,-1}S_{\Delta}^{-1}S_{\Delta,-1}$$

for any $\lambda \in \mathbb{R}$. Note that, since $\hat{\lambda}_1, \cdots, \hat{\lambda}_n$ are the solutions to the equation $|S(\lambda)| = 0$, $S(\hat{\lambda}_i)$ is singular for any $1 \le i \le n$.

Lemma (Rates of Convergence of Eigenvectors)
Maintain assumptions A1 to A3, and suppose that $r > 0$. Then, the following results hold:

i) $\hat{\beta}_T = O_p(1)$, $\hat{x}_T = O_p(1)$ and $\hat{x}_T^{-1} = O_p(1)$.

ii) $\hat{y}_T = O_p(T^{-1})$ and $\beta' S(\hat{\lambda}_i)\beta \cdot \hat{x}_{i,T} = O_p(T^{-1})$ for any $1 \le i \le r$.

iii) For any $1 \le i \le r$,

$$\beta_\perp' S(\hat{\lambda}_i)\beta \cdot \hat{x}_{i,T} = -\beta_\perp' \left(\frac{1}{T-p}Y'_{-1}\varepsilon\right)\mu_\Delta^{-1}\mu_{\Delta,-1}\beta\hat{x}_{i,T} + o_p(1).$$

Proof) We proceed in steps.

Step 1: $\hat{\beta}_T = O_p(1)$

We first show that $\hat{\beta}_T = O_p(1)$. This follows from the normalization condition

$$\hat{\beta}_T' S_{-1}\hat{\beta}_T = I_r.$$

We transform S_{-1} in the same manner as in the preceding theorem; defining $M_T = (\beta_\perp' S_{-1}\beta_\perp)^{-1}$ as above, so that $M_T = o_p(1)$, note that

$$\zeta = \begin{pmatrix} \beta' S_{-1}\beta & \beta' S_{-1}\beta_\perp M_T' \\ M_T\beta_\perp' S_{-1}\beta & M_T\beta_\perp' S_{-1}\beta_\perp M_T' \end{pmatrix} = \begin{pmatrix} I_r & O \\ O & M_T \end{pmatrix} \begin{pmatrix} \beta' \\ \beta_\perp' \end{pmatrix} S_{-1} \begin{pmatrix} \beta & \beta_\perp \end{pmatrix} \begin{pmatrix} I_r & O \\ O & M_T \end{pmatrix}.$$

Letting $LL' = \zeta$ be the Cholesky decomposition of ζ, we can now see that

$$\hat{\beta}_T' S_{-1}\hat{\beta}_T = \hat{\beta}_T' \begin{pmatrix} \beta' \\ \beta_\perp' \end{pmatrix}^{-1} \begin{pmatrix} I_r & O \\ O & M_T^{-1} \end{pmatrix} LL' \begin{pmatrix} I_r & O \\ O & M_T^{-1} \end{pmatrix} \begin{pmatrix} \beta & \beta_\perp \end{pmatrix}^{-1} \hat{\beta}_T = I_r,$$

or, defining

$$\hat{\delta}_T = L' \begin{pmatrix} I_r & O \\ O & M_T^{-1} \end{pmatrix} \begin{pmatrix} \beta & \beta_\perp \end{pmatrix}^{-1} \hat{\beta}_T,$$

we have

$$\hat{\delta}_T' \hat{\delta}_T = I_r.$$

This means that the columns of $\hat{\delta}_T$ are orthonormal and have norm 1 for any $T \in N_+$, so that $\left\| \hat{\delta}_T \right\| \leq \sqrt{r}$ and thus $\hat{\delta}_T = O_p(1)$. Furthermore,

$$\hat{\beta}_T = \begin{pmatrix} \beta & \beta_\perp \end{pmatrix} \begin{pmatrix} I_r & O \\ O & M_T \end{pmatrix} (L')^{-1} \cdot \hat{\delta}_T,$$

where

$$\begin{pmatrix} I_r & O \\ O & M_T \end{pmatrix} = O_p(1)$$

because $M_T = o_p(1)$. Furthermore, since

$$\zeta \xrightarrow{p} \begin{pmatrix} \mu_{-1} & O \\ O & I_{n-r} \end{pmatrix},$$

a nonsingular matrix, $\zeta^{-1} = O_p(1)$ and $(L')^{-1}$, being the Cholesky factor of ζ^{-1}, is also $O_p(1)$. Therefore, $\hat{\beta}_T$ is the product of a nonrandom matrix $\begin{pmatrix} \beta & \beta_\perp \end{pmatrix}$ and three $O_p(1)$ matrices, so that it is also $O_p(1)$.

Since $\hat{x}_T = (\beta'\beta)^{-1}\beta'\hat{\beta}_T$, by implication $\hat{x}_T = O_p(1)$ as well.

Step 2: $\hat{y}_{i,T} = O_p(T^{-1})$

Recall that the columns of

$$\hat{C}_T = S_{-1}^{\frac{1}{2}\prime} \hat{\beta}_T = S_{-1}^{\frac{1}{2}\prime} \begin{pmatrix} \hat{\beta}_{1,T} & \cdots & \hat{\beta}_{r,T} \end{pmatrix}$$

are eigenvectors of $S_{-1}^{-\frac{1}{2}} S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1} S_{-1}^{-\frac{1}{2}\prime}$ corresponding to the eigenvalues $\hat{\lambda}_1, \cdots, \hat{\lambda}_r$. Therefore, for any $1 \leq i \leq r$,

$$S_{-1}^{-\frac{1}{2}} S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1} \cdot \hat{\beta}_{i,T} = \hat{\lambda}_i \cdot S_{-1}^{\frac{1}{2}\prime} \hat{\beta}_{i,T},$$

or equivalently,

$$S(\hat{\lambda}_i) \cdot \hat{\beta}_{i,T} = \left[\hat{\lambda}_i S_{-1} - S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1} \right] \hat{\beta}_{i,T} = \mathbf{0}.$$

For any $1 \leq i \leq r$,

$$\hat{\beta}_{i,T} = \beta \hat{x}_{i,T} + \beta_\perp \hat{y}_{i,T}$$

and the equations

$$\beta' S(\hat{\lambda}_i) \beta \cdot \hat{x}_{i,T} + \beta' S(\hat{\lambda}_i) \beta_\perp \cdot \hat{y}_{i,T} = \mathbf{0}$$

$$\beta_\perp' S(\hat{\lambda}_i) \beta \cdot \hat{x}_{i,T} + \beta_\perp' S(\hat{\lambda}_i) \beta_\perp \cdot \hat{y}_{i,T} = \mathbf{0}$$

hold. From the asymptotic results for QMLE and the consistency results for the eigenvalues shown above,

$$\beta' S(\hat{\lambda}_i) \beta = \hat{\lambda}_i \cdot \beta' S_{-1} \beta - \beta' S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1} \beta = O_p(1),$$

$$\beta' S(\hat{\lambda}_i) \beta_\perp = \hat{\lambda}_i \cdot \beta' S_{-1} \beta_\perp - \beta' S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1} \beta_\perp = O_p(1)$$

$$\beta_\perp' S(\hat{\lambda}_i) \beta = \hat{\lambda}_i \cdot \beta_\perp' S_{-1} \beta - \beta_\perp' S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1} \beta = O_p(1)$$

$$\begin{aligned} \frac{1}{T} \beta_\perp' S(\hat{\lambda}_i) \beta_\perp &= \hat{\lambda}_i \cdot \frac{\beta_\perp' S_{-1} \beta_\perp}{T} - \frac{1}{T} \beta_\perp' S_{\Delta,-1}' S_{\Delta}^{-1} S_{\Delta,-1} \beta_\perp \\ &\xrightarrow{d} \lambda_i \cdot \beta_\perp' \Lambda \left(\int_0^1 W^n(r) W^n(r)' dr \right) \Lambda' \beta_\perp, \end{aligned}$$

where $\lambda_i > 0$ is the ith largest solution to the equation $\left| \lambda \mu_{-1} - \beta' \mu_{\Delta,-1}' \mu_{\Delta}^{-1} \mu_{\Delta,-1} \beta \right| = 0$. By implication,

$$\left(\frac{1}{T} \beta_\perp' S(\hat{\lambda}_i) \beta_\perp \right)^{-1} = O_p(1)$$

as well.

We showed above that $\hat{x}_T = O_p(1)$. Therefore,

$$T\hat{y}_{i,T} = -\left(\frac{1}{T}\beta_\perp' S(\hat{\lambda}_i)\beta_\perp\right)^{-1} \beta_\perp' S(\hat{\lambda}_i)\beta \cdot \hat{x}_{i,T} = O_p(1),$$

which tells us that

$$\hat{y}_{i,T} = O_p(T^{-1}).$$

By implication,

$$\beta' S(\hat{\lambda}_i)\beta \cdot \hat{x}_{i,T} = -\beta' S(\hat{\lambda}_i)\beta_\perp \cdot \hat{y}_{i,T} = O_p(T^{-1})$$

as well.

Step 3: $\hat{x}_T^{-1} = O_p(1)$

Using the normalization condition, we can see that

$$\begin{aligned} I_r = \hat{\beta}_T' S_{-1}\hat{\beta}_T &= (\beta\hat{x}_T + \beta_\perp\hat{y}_T)' S_{-1}(\beta\hat{x}_T + \beta_\perp\hat{y}_T) \\ &= \hat{x}_T' \cdot \beta' S_{-1}\beta \cdot \hat{x}_T + \hat{x}_T' \cdot \beta' S_{-1}\beta_\perp \cdot \hat{y}_T + \hat{y}_T' \cdot \beta_\perp' S_{-1}\beta \cdot \hat{x}_T + \hat{y}_T' \cdot \beta_\perp' S_{-1}\beta_\perp \cdot \hat{y}_T. \end{aligned}$$

Since $\hat{y}_T = O_p(T^{-1})$ and $\hat{x}_T = O_p(1)$,

$$\hat{x}_T' \cdot \beta' S_{-1}\beta_\perp \cdot \hat{y}_T = \hat{x}_T' \cdot \left(\frac{1}{T}\beta' S_{-1}\beta_\perp\right) \cdot T\hat{y}_T = o_p(1)$$

since $\beta' S_{-1} = O_p(1)$. Similarly,

$$\hat{y}_T' \cdot \beta_\perp' S_{-1}\beta_\perp \cdot \hat{y}_T = T\hat{y}_T' \cdot \left(\frac{1}{T^2}\beta_\perp' S_{-1}\beta_\perp\right) \cdot T\hat{y}_T = o_p(1)$$

since $\beta_\perp' S_{-1}\beta_\perp = O_p(T)$. It follows that

$$I_r = \hat{\beta}_T' S_{-1}\hat{\beta}_T = \hat{x}_T' \cdot \beta' S_{-1}\beta \cdot \hat{x}_T + o_p(1).$$

It follows that

$$\hat{x}_T' \cdot \beta' S_{-1}\beta \cdot \hat{x}_T \xrightarrow{p} I_r.$$

Because the determinant is a continuous function, by the continuous mapping theorem we have

$$|\hat{x}_T|^2 \cdot |\beta' S_{-1}\beta| \xrightarrow{p} |I_r| = 1.$$

$|\beta' S_{-1}\beta| \xrightarrow{p} |\mu_{-1}| > 0$, so

$$|\hat{x}_T| \xrightarrow{p} \left(\frac{1}{|\mu_{-1}|}\right)^{\frac{1}{2}} > 0.$$

The elements of the adjugate of a matrix A are polynomials of the elements of A. Therefore, the adjugate of a random matrix that is bounded in probability is also bounded in probability. We just showed that $\frac{1}{|\hat{x}_T|} = O_p(1)$, so

$$\hat{x}_T^{-1} = \frac{1}{|\hat{x}_T|}\mathrm{adj}(\hat{x}_T) = O_p(1).$$

Step 4: An Expression for $\beta'_\perp S(\hat{\lambda}_i)\beta \cdot \hat{x}_{i,T}$

The final result follows by noting that

$$\begin{aligned}
\beta'_\perp S'_{\Delta,-1} - \beta'_\perp S_{-1}\beta\alpha' &= \beta'_\perp \left[\frac{1}{T-p} Y'_{-1} M_X (\Delta Y - Y_{-1} \cdot \beta\alpha')\right] \\
&= \beta'_\perp \left[\frac{1}{T-p} Y'_{-1} M_X (X \cdot \Gamma + \varepsilon)\right] \\
&= \beta'_\perp \left(\frac{1}{T-p} Y'_{-1}\varepsilon\right) - \beta'_\perp \left(\frac{1}{T-p} Y'_{-1}X\right)\left(\frac{1}{T-p}X'X\right)^{-1}\left(\frac{1}{T-p}X'\varepsilon\right).
\end{aligned}$$

Since $\frac{1}{T-p}X'\varepsilon = o_p(1)$,

$$\beta'_\perp S'_{\Delta,-1} - (\beta'_\perp S_{-1}\beta)\,\alpha' = \beta'_\perp \left(\frac{1}{T-p} Y'_{-1}\varepsilon\right) + o_p(1).$$

The asymptotic results for QMLE tell us that

$$\alpha' = \mu_{-1}^{-1}(\mu_{\Delta,-1}\beta)',$$

and $\beta' S_{-1}\beta$ and $S_{\Delta,-1}\beta$ is consistent for μ_{-1} and $\mu_{\Delta,-1}\beta$, so

$$\beta'_\perp S'_{\Delta,-1} - \beta'_\perp S_{-1}\beta(\beta' S_{-1}\beta)^{-1}\beta' S'_{\Delta,-1} = \beta'_\perp \left(\frac{1}{T-p} Y'_{-1}\varepsilon\right) + o_p(1),$$

where we used the fact that $S_{-1}\beta$ is $O_p(1)$. Therefore, for any $1 \le i \le r$,

$$\begin{aligned}
\beta'_\perp S(\hat{\lambda}_i)\beta \cdot \hat{x}_{i,T} &= \hat{\lambda}_i \cdot \beta'_\perp S_{-1}\beta \cdot \hat{x}_{i,T} - \beta'_\perp S'_{\Delta,-1} S_\Delta^{-1} S_{\Delta,-1}\beta \cdot \hat{x}_{i,T} \\
&= \hat{\lambda}_i \cdot \beta'_\perp S_{-1}\beta \cdot \hat{x}_{i,T} - \beta'_\perp S_{-1}\beta(\beta' S_{-1}\beta)^{-1}\beta' S'_{\Delta,-1} S_\Delta^{-1} S_{\Delta,-1}\beta \cdot \hat{x}_{i,T} \\
&\quad - \beta'_\perp \left(\frac{1}{T-p} Y'_{-1}\varepsilon\right) S_\Delta^{-1} S_{\Delta,-1}\beta \cdot \hat{x}_{i,T} + o_p(1)
\end{aligned}$$

$$
\begin{aligned}
&= \beta_\perp' S_{-1}\beta\,(\beta' S_{-1}\beta)^{-1}\left[\hat{\lambda}_i \cdot \beta' S_{-1}\beta - \beta' S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\beta\right] \cdot \hat{x}_{i,T} \\
&\quad - \beta_\perp' \left(\frac{1}{T-p} Y_{-1}' \varepsilon\right) S_\Delta^{-1} S_{\Delta,-1}\beta \cdot \hat{x}_{i,T} + o_p(1) \\
&= \beta_\perp' S_{-1}\beta\,(\beta' S_{-1}\beta)^{-1} \beta' S(\hat{\lambda}_i)\beta \cdot \hat{x}_{i,T} \\
&\quad - \beta_\perp' \left(\frac{1}{T-p} Y_{-1}' \varepsilon\right) S_\Delta^{-1} S_{\Delta,-1}\beta \cdot \hat{x}_{i,T} + o_p(1).
\end{aligned}
$$

We saw above that $\beta' S(\hat{\lambda}_i)\beta \cdot \hat{x}_{i,T} = O_p(T^{-1})$ and thus $o_p(1)$, so that

$$
\beta_\perp' S(\hat{\lambda}_i)\beta \cdot \hat{x}_{i,T} = -\beta_\perp' \left(\frac{1}{T-p} Y_{-1}' \varepsilon\right) S_\Delta^{-1} S_{\Delta,-1}\beta \cdot \hat{x}_{i,T} + o_p(1).
$$

Finally, since $S_\Delta^{-1} \xrightarrow{p} \mu_\Delta^{-1}$ and $S_{\Delta,-1}\beta \xrightarrow{p} \mu_{\Delta,-1}\beta$, we have

$$
\beta_\perp' S(\hat{\lambda}_i)\beta \cdot \hat{x}_{i,T} = -\beta_\perp' \left(\frac{1}{T-p} Y_{-1}' \varepsilon\right) \mu_\Delta^{-1} \mu_{\Delta,-1}\beta \cdot \hat{x}_{i,T} + o_p(1).
$$

Q.E.D.

The preceding lemma can be directly used to establish several consistency results:

Theorem (Consistency of Parameter Estimates)
Maintain assumptions A1 to A3, and suppose that $r > 0$. Then, the following hold true:

i) The cointegration space is consistently estimated:

$$
\beta_\perp \hat{y}_T = O_p(T^{-1}).
$$

ii) $\hat{\Pi}_T = \hat{\alpha}_T \cdot \hat{\beta}_T'$ is a consistent estimator for Π:

$$
\hat{\Pi}_T \xrightarrow{p} \Pi = \alpha\beta'.
$$

iii) $\hat{\Sigma}_T$ is a consistent estimator for Σ:

$$
\hat{\Sigma}_T \xrightarrow{p} \Sigma.
$$

iv) $\hat{\Gamma}_T$ is a consistent estimator for Γ:

$$
\hat{\Gamma}_T \xrightarrow{p} \Gamma.
$$

Proof) i) This follows immediately from the preceding lemma. Heuristically, it tells us that the projection of $\hat{\beta}_T$ on the orthogonal complement of β vanishes as $T \to \infty$, so that, for large T, $\hat{\beta}_T$ lies in the cointegrating space with probability close to 1.

ii) Note that

$$\hat{\beta}_T = \beta \hat{x}_T + \beta_\perp \hat{y}_T,$$

so that

$$\hat{\beta}_T \hat{x}_T^{-1} - \beta = \beta_\perp \hat{y}_T \hat{x}_T^{-1} = O_p(T^{-1}),$$

since $\hat{y}_T = O_p(T^{-1})$ and $\hat{x}_T = O_p(1)$. Therefore,

$$\begin{aligned}
\hat{\Pi}_T = \hat{\alpha}_T \hat{\beta}_T' &= \left(\frac{1}{T-p} R_\Delta' R_{-1}\right) \hat{\beta}_T \left[\hat{\beta}_T' \left(\frac{1}{T-p} R_{-1}' R_{-1}\right) \hat{\beta}_T\right]^{-1} \hat{\beta}_T' \\
&= S_{\Delta,-1} \hat{\beta}_T \left(\hat{\beta}_T' S_{-1} \hat{\beta}_T\right)^{-1} \hat{\beta}_T' \\
&= S_{\Delta,-1} \hat{\beta}_T \hat{x}_T^{-1} \left(\hat{x}_T^{-1\prime} \hat{\beta}_T' S_{-1} \hat{\beta}_T \hat{x}_T^{-1}\right)^{-1} \hat{x}_T^{-1\prime} \hat{\beta}_T' \\
&\xrightarrow{p} \mu_{\Delta,-1} \beta \mu_{-1}^{-1} \beta' = \alpha\beta' = \Pi.
\end{aligned}$$

iii) Similarly to the preceding result,

$$\begin{aligned}
\hat{\Sigma}_T &= S_\Delta - S_{\Delta,-1} \hat{\beta}_T \left(\hat{\beta}_T' S_{-1} \hat{\beta}_T\right)^{-1} \hat{\beta}_T' S_{\Delta,-1}' \\
&= S_\Delta - S_{\Delta,-1} \hat{\beta}_T \hat{x}_T^{-1} \left(\hat{x}_T^{-1\prime} \hat{\beta}_T' S_{-1} \hat{\beta}_T \hat{x}_T^{-1}\right)^{-1} \hat{x}_T^{-1\prime} \hat{\beta}_T' S_{\Delta,-1}' \\
&\xrightarrow{p} \mu_\Delta - \mu_{\Delta,-1} \beta \mu_{-1}^{-1} \beta' \mu_{\Delta,-1} \\
&= \Sigma + \mu_{\Delta,-1} \beta \alpha' - \mu_{\Delta,-1} \beta \alpha' = \Sigma.
\end{aligned}$$

iv) Finally, using the consistency of $\hat{\Pi}_T$, we can see that

$$\begin{aligned}
\hat{\Gamma}_T &= \Gamma + (X'X)^{-1} X' Y_{-1} \left(\Pi' - \hat{\Pi}_T'\right) + (X'X)^{-1} X'\varepsilon \\
&= \Gamma + \left(\frac{1}{T} X'X\right)^{-1} \left(\frac{1}{T} X' Y_{-1}\right) \left(\Pi' - \hat{\Pi}_T'\right) + \left(\frac{1}{T} X'X\right)^{-1} \left(\frac{1}{T} X'\varepsilon\right) \\
&\xrightarrow{p} \Gamma,
\end{aligned}$$

since $\left(\frac{1}{T} X'X\right)^{-1} = O_p(1)$, $\frac{1}{T} X' Y_{-1} = O_p(1)$, $\Pi' - \hat{\Pi}_T' = o_p(1)$ and $\frac{1}{T} X'\varepsilon = o_p(1)$.

Q.E.D.

That $\hat{\beta}_T \hat{x}_T^{-1} - \beta = O_p(T^{-1})$ can also be expressed as

$$\hat{\beta}_T - \beta \hat{x}_T = O_p(T^{-1}).$$

This is reminiscent of the result in factor models that the factor estimates are consistent only for a rotation of the true factors. As in factor models, because the cointegrating relationships are non-unique, our estimates of the cointegrating relationships collected in $\hat{\beta}_T$ is consistent only for a rotation of the true cointegrating basis β. In this sense, too, is the cointegrating space consistently estimated.

5.4.7 Testing for the Cointegrating Rank

Here we state and derive the limiting distribution of two test statistics designed to test for the cointegrating rank r. To this end, we first derive the limiting distribution of the last $n-r$ sample canonical correlations, $\hat{\lambda}_{r+1},\cdots,\hat{\lambda}_n$, under the assumption that the true cointegration rank is $0<r<n$.

It turns out that $\hat{\lambda}_{r+1},\cdots,\hat{\lambda}_n$ converge at a rate of T to their limiting distributions. The formal statement and proof are given below:

Theorem (Asymptotic Distribution of Eigenvalues)
Maintain assumptions A1 to A3. Then,

$$(T\hat{\lambda}_{r+1},\cdots,T\hat{\lambda}_n)\xrightarrow{d}(\eta_1,\cdots,\eta_{n-r}),$$

where $\eta_1\geq\cdots\geq\eta_{n-r}$ are the ordered eigenvalues that solve the equation

$$\left|\lambda\cdot\int_0^1 W^{n-r}(s)W^{n-r}(s)'ds-\left(\int_0^1 W^{n-r}(s)dW^{n-r}(s)'\right)\left(\int_0^1 W^{n-r}(s)dW^{n-r}(s)'\right)'\right|=0,$$

or equivalently, the ordered eigenvalues of the positive definite valued random matrix

$$\left(\int_0^1 W^{n-r}(s)dW^{n-r}(s)'\right)'\left(\int_0^1 W^{n-r}(s)W^{n-r}(s)'ds\right)^{-1}\left(\int_0^1 W^{n-r}(s)dW^{n-r}(s)'\right).$$

In addition,

$$T\left(\sum_{i=r+1}^{n}\hat{\lambda}_i\right)\xrightarrow{d}\text{tr}\left[\left(\int_0^1 W^{n-r}(s)dW^{n-r}(s)'\right)'\left(\int_0^1 W^{n-r}(s)W^{n-r}(s)'ds\right)^{-1}\left(\int_0^1 W^{n-r}(s)dW^{n-r}(s)'\right)\right].$$

Proof) **The Case $r=0$**

Again, we deal with the case where there is no cointegrating relationships, so that $r=0$. In this case, $T\hat{\lambda}_1\geq\cdots\geq T\hat{\lambda}_n>0$ are the ordered solutions to the eigenvalue equation

$$\left|\lambda\left(\frac{1}{T}S_{-1}\right)-S_{\Delta,-1}'S_{\Delta}^{-1}S_{\Delta,-1}\right|=0.$$

Letting F be the weak limit of $S_{\Delta,-1}$, the continuous mapping theorem tells us that $T\hat{\lambda}_1\geq\cdots\geq T\hat{\lambda}_n>0$ converge weakly to the solutions of the equation

$$\left|\lambda\cdot\Gamma(1)^{-1}\Sigma^{\frac{1}{2}}\left(\int_0^1 W^n(s)W^n(s)'ds\right)\Sigma^{\frac{1}{2}\prime}\Gamma(1)^{-1\prime}-F'\mu_{\Delta}^{-1}F\right|=0.$$

Since $\mu_\Delta=\Sigma$ and $\alpha_\perp=\Sigma^{-\frac{1}{2}\prime}$ when $r=0$, the above equation can be written as

$$\left|\lambda\cdot\Gamma(1)^{-1}\Sigma^{\frac{1}{2}\prime}\left(\int_0^1 W^n(s)W^n(s)'ds\right)\Sigma^{\frac{1}{2}}\Gamma(1)^{-1\prime}-F'\alpha_\perp\alpha_\perp'F\right|=0.$$

Since $F'\alpha_\perp$ and

$$\Gamma(1)^{-1}\Sigma^{\frac{1}{2}}\left(\int_0^1 W^n(r)dW^n(r)'\right)$$

are both limits of $S'_{\Delta,-1}\alpha_\perp$ when $r=0$, the uniqueness of weak limits tells us that they are identically distributed and thus that $T\hat{\lambda}_1,\cdots,T\hat{\lambda}_n$ converge weakly to the solutions of the equation

$$\left|\lambda\cdot\Gamma(1)^{-1}\Sigma^{\frac{1}{2}}\left(\int_0^1 W^n(s)W^n(s)'ds\right)\Sigma^{\frac{1}{2}\prime}\Gamma(1)^{-1\prime}-\Gamma(1)^{-1}\Sigma^{\frac{1}{2}}\left(\int_0^1 W^n(r)dW^n(r)'\right)\left(\int_0^1 W^n(r)dW^n(r)'\right)'\Sigma^{\frac{1}{2}\prime}\Gamma(1)^{-1\prime}\right|=0,$$

or equivalently, the equation

$$\left|\lambda\cdot\int_0^1 W^n(s)W^n(s)'ds-\left(\int_0^1 W^n(r)dW^n(r)'\right)\left(\int_0^1 W^n(r)dW^n(r)'\right)'\right|=0.$$

The Case $r>0$

Note that $(T\hat{\lambda}_n)^{-1}\geq\cdots\geq(T\hat{\lambda}_1)^{-1}>0$ are the ordered solutions to the equation

$$\left|\frac{1}{T}S_{-1}-\eta\cdot S'_{\Delta,-1}S_\Delta^{-1}S_{\Delta,-1}\right|=0.$$

As above, we can use the fact that $\begin{pmatrix}\beta & \beta_\perp\end{pmatrix}$ is nonsingular to conclude that $(T\hat{\lambda}_n)^{-1}\geq\cdots\geq(T\hat{\lambda}_1)^{-1}>0$ are also the ordered solutions to the eigenvalue equation

$$\left|\begin{pmatrix}\frac{\beta' S_{-1}\beta}{T} & \frac{\beta' S_{-1}\beta_\perp}{T}\\ \frac{\beta'_\perp S_{-1}\beta}{T} & \frac{\beta'_\perp S_{-1}\beta_\perp}{T}\end{pmatrix}-\mu\cdot\begin{pmatrix}\beta' S'_{\Delta,-1}S_\Delta^{-1}S_{\Delta,-1}\beta & \beta' S'_{\Delta,-1}S_\Delta^{-1}S_{\Delta,-1}\beta_\perp\\ \beta'_\perp S'_{\Delta,-1}S_\Delta^{-1}S_{\Delta,-1}\beta & \beta'_\perp S'_{\Delta,-1}S_\Delta^{-1}S_{\Delta,-1}\beta_\perp\end{pmatrix}\right|=0.$$

Note that

$$\begin{pmatrix}\beta' S'_{\Delta,-1}S_\Delta^{-1}S_{\Delta,-1}\beta & \beta' S'_{\Delta,-1}S_\Delta^{-1}S_{\Delta,-1}\beta_\perp\\ \beta'_\perp S'_{\Delta,-1}S_\Delta^{-1}S_{\Delta,-1}\beta & \beta'_\perp S'_{\Delta,-1}S_\Delta^{-1}S_{\Delta,-1}\beta_\perp\end{pmatrix}\xrightarrow{d}\begin{pmatrix}\beta'\mu'_{\Delta,-1}\mu_\Delta^{-1}\mu_{\Delta,-1}\beta & \beta'\mu'_{\Delta,-1}\mu_\Delta^{-1}F\\ F'\mu_\Delta^{-1}\mu_{\Delta,-1}\beta & F'\mu_\Delta^{-1}F\end{pmatrix},$$

a positive definite matrix, where F is the limiting distribution of $S_{\Delta,-1}\beta_\perp$, and likewise,

$$\begin{pmatrix}\frac{\beta' S_{-1}\beta}{T} & \frac{\beta' S_{-1}\beta_\perp}{T}\\ \frac{\beta'_\perp S_{-1}\beta}{T} & \frac{\beta'_\perp S_{-1}\beta_\perp}{T}\end{pmatrix}\xrightarrow{p}\begin{pmatrix}O & O\\ O & \beta'_\perp\Lambda\left(\int_0^1 W^n(r)W^n(r)'dr\right)\Lambda'\beta_\perp\end{pmatrix}.$$

By the continuity of ordered eigenvalues, $\left((T\hat{\lambda}_n)^{-1},\cdots,(T\hat{\lambda}_1)^{-1}\right)$ converge in distribution to the ordered solutions of the equation

$$\left|\begin{pmatrix} O & O \\ O & \beta_\perp' \Lambda \left(\int_0^1 W^n(r) W^n(r)' dr\right) \Lambda' \beta_\perp \end{pmatrix} - \eta \cdot \begin{pmatrix} \beta' \mu_{\Delta,-1}' \mu_\Delta^{-1} \mu_{\Delta,-1} \beta & \beta' \mu_{\Delta,-1}' \mu_\Delta^{-1} F \\ F' \mu_\Delta^{-1} \mu_{\Delta,-1} \beta & F' \mu_\Delta^{-1} F \end{pmatrix}\right| = 0.$$

The matrix on the right is positive definite, so the solutions to this equation are non-negative and real valued (they are the eigenvalues of a positive semidefinite matrix). If $\eta > 0$ is a non-zero solution to the above equation, it is also a solution to the equation

$$\begin{aligned} 0 &= \left|\eta \cdot \beta' \mu_{\Delta,-1}' \mu_\Delta^{-1} \mu_{\Delta,-1} \beta\right| \\ &\times \left|\beta_\perp' \Lambda \left(\int_0^1 W^n(r) W^n(r)' dr\right) \Lambda' \beta_\perp - \eta \cdot F' \left[\mu_\Delta^{-1} - \mu_\Delta^{-1} \mu_{\Delta,-1} \beta \left(\beta' \mu_{\Delta,-1}' \mu_\Delta^{-1} \mu_{\Delta,-1} \beta\right)^{-1} \beta' \mu_{\Delta,-1}' \mu_\Delta^{-1}\right] F\right|. \end{aligned}$$

Since the first term is always non-zero, it follows that any non-zero solutions to the eigenvalue equation of interest must also solve the equation

$$\left|\beta_\perp' \Lambda \left(\int_0^1 W^n(r) W^n(r)' dr\right) \Lambda' \beta_\perp - \eta \cdot F' \left[\mu_\Delta^{-1} - \mu_\Delta^{-1} \mu_{\Delta,-1} \beta \left(\beta' \mu_{\Delta,-1}' \mu_\Delta^{-1} \mu_{\Delta,-1} \beta\right)^{-1} \beta' \mu_{\Delta,-1}' \mu_\Delta^{-1}\right] F\right| = 0.$$

This equation has $n-r$ positive roots, so the equation

$$\left|\begin{pmatrix} O & O \\ O & \beta_\perp' \Lambda \left(\int_0^1 W^n(r) W^n(r)' dr\right) \Lambda' \beta_\perp \end{pmatrix} - \eta \cdot \begin{pmatrix} \beta' \mu_{\Delta,-1}' \mu_\Delta^{-1} \mu_{\Delta,-1} \beta & \beta' \mu_{\Delta,-1}' \mu_\Delta^{-1} F \\ F' \mu_\Delta^{-1} \mu_{\Delta,-1} \beta & F' \mu_\Delta^{-1} F \end{pmatrix}\right| = 0$$

must have $n-r$ positive solutions and r solutions equal to 0. It follows that

$$\left((T\hat{\lambda}_r)^{-1}, \cdots, (T\hat{\lambda}_1)^{-1}\right) \xrightarrow{p} (0, \cdots, 0),$$

where the convergence is in probability because the limit is non-random.

It remains to find the non-zero roots to the above equation. Note that

$$\begin{aligned} &\mu_\Delta^{-1} - \mu_\Delta^{-1} \mu_{\Delta,-1} \beta \left(\beta' \mu_{\Delta,-1}' \mu_\Delta^{-1} \mu_{\Delta,-1} \beta\right)^{-1} \beta' \mu_{\Delta,-1}' \mu_\Delta^{-1} \\ &\qquad = \mu_\Delta^{-1} - \mu_\Delta^{-1} \alpha \mu_{-1} \left(\mu_{-1} \alpha' \mu_\Delta^{-1} \alpha \mu_{-1}\right)^{-1} \mu_{-1} \alpha' \mu_\Delta^{-1} \\ &\qquad = \mu_\Delta^{-1} \left(I_n - \alpha \left(\alpha' \mu_\Delta^{-1} \alpha\right)^{-1} \alpha' \mu_\Delta^{-1}\right) \\ &\qquad = \mu_\Delta^{-1} \left(I_n - P_\alpha(\mu_\Delta)\right)^{-1} = \Sigma^{-1} \left(I_n - P_\alpha(\Sigma)\right)^{-1} = \alpha_\perp \alpha_\perp' \end{aligned}$$

where the last two equality follows from results shown earlier. Therefore, the non-zero solutions to the eigenvalue equation are also solutions to the equation

$$\left|\beta_\perp' \Lambda \left(\int_0^1 W^n(r) W^n(r)' dr\right) \Lambda' \beta_\perp - \eta \cdot F' \alpha_\perp \alpha \perp' F\right| = 0.$$

Since $\beta_\perp' F' \alpha_\perp$ and

$$\beta_\perp' \Lambda \left(\int_0^1 W^n(r) dW^n(r)'\right) \Sigma^{\frac{1}{2}\prime} \alpha_\perp$$

are both limits of $S_{\Delta,-1}' \alpha_\perp$, the uniqueness of weak limits tells us that they are identically distributed and thus that $(T\hat{\lambda}_n)^{-1}, \cdots, (T\hat{\lambda}_{r+1})^{-1}$ converge weakly to the solutions of the equation

$$\left| \beta_\perp' \Lambda \left(\int_0^1 W^n(r) W^n(r)' dr \right) \Lambda' \beta_\perp - \eta \cdot \beta_\perp' \Lambda \left(\int_0^1 W^n(s) dW^n(s)' \right) \Sigma^{\frac{1}{2}\prime} \alpha_\perp \alpha_\perp' \left(\int_0^1 W^n(s) dW^n(s)' \right)' \Lambda' \beta_\perp \right| = 0.$$

Since

$$\beta_\perp \Lambda \left(\int_0^1 W^n(s) dW^n(s)' \right) \Sigma^{\frac{1}{2}\prime} \alpha_\perp \sim (\beta_\perp' \beta_\perp)(\alpha_\perp' \Gamma(1) \beta_\perp)^{-1} \left(\int_0^1 W^{n-r}(s) dW^{n-r}(s)' \right)$$

$$\beta_\perp' \Lambda \left(\int_0^1 W^n(r) W^n(r)' dr \right) \Lambda' \beta_\perp \sim (\beta_\perp' \beta_\perp)(\alpha_\perp' \Gamma(1) \beta_\perp)^{-1} \left(\int_0^1 W^{n-r}(s) W^{n-r}(s)' ds \right) (\beta_\perp' \Gamma(1)' \alpha_\perp)^{-1} (\beta_\perp' \beta_\perp),$$

we can say that $(T\hat{\lambda}_n)^{-1}, \cdots, (T\hat{\lambda}_{r+1})^{-1}$ converge weakly to the solutions $\tilde{\eta}_1 \geq \cdots \geq \tilde{\eta}_{n-r} > 0$ of the equation

$$\left| \int_0^1 W^{n-r}(s) W^{n-r}(s)' ds - \eta \cdot \left(\int_0^1 W^{n-r}(s) dW^{n-r}(s)' \right) \left(\int_0^1 W^{n-r}(s) dW^{n-r}(s)' \right)' \right| = 0.$$

Letting $\eta_1 \geq \cdots \geq \eta_{n-r} > 0$ be the ordered solutions to the equation

$$\left| \lambda \cdot \int_0^1 W^{n-r}(s) W^{n-r}(s)' ds - \left(\int_0^1 W^{n-r}(s) dW^{n-r}(s)' \right) \left(\int_0^1 W^{n-r}(s) dW^{n-r}(s)' \right)' \right| = 0,$$

since $\eta_i = \frac{1}{\tilde{\eta}_i}$ for any $1 \leq i \leq n-r$, it follows from the continuous mapping theorem that the first claim of the theorem holds true.

To see the second claim, note that the trace of a positive definite matrix is given as the sum of its eigenvalues. Therefore, since $\eta_1 \geq \cdots \geq \eta_{n-r} > 0$ are the ordered eigenvalues of the positive-definite valued random matrix

$$\left(\int_0^1 W^{n-r}(s) dW^{n-r}(s)' \right)' \left(\int_0^1 W^{n-r}(s) W^{n-r}(s)' ds \right)^{-1} \left(\int_0^1 W^{n-r}(s) dW^{n-r}(s)' \right),$$

we can see that

$$T \left(\sum_{i=r+1}^{n} \hat{\lambda}_i \right) \xrightarrow{d} \sum_{i=1}^{n-r} \eta_i$$

$$= \operatorname{tr} \left[\left(\int_0^1 W^{n-r}(s) dW^{n-r}(s)' \right)' \left(\int_0^1 W^{n-r}(s) W^{n-r}(s)' ds \right)^{-1} \left(\int_0^1 W^{n-r}(s) dW^{n-r}(s)' \right) \right].$$

Q.E.D.

We now consider two tests for the cointegrating rank.

The Trace Test

Initially, consider testing the null hypothesis

$$H_0 : rank(\Pi) \leq r < n$$

against the alternative hypothesis

$$H_1 : rank(\Pi) > r$$

for some $0 \leq r < n$. The likelihood ratio test statistic is given as

$$\begin{aligned}
\hat{LR}_T &= -2\left[\sup_{H_0} \quad l(\alpha,\beta,\Pi,\Sigma) - \sup_{H_1} \quad l(\alpha,\beta,\Pi,\Sigma)\right] \\
&= n(T-p)\left(\log(2\pi)+1+\frac{1}{n}\log|S_\Delta|\right) + (T-p)\sum_{i=1}^{r}\log\left(1-\hat{\lambda}_i\right) \\
&\quad - \left[n(T-p)\left(\log(2\pi)+1+\frac{1}{n}\log|S_\Delta|\right) + (T-p)\sum_{i=1}^{n}\log\left(1-\hat{\lambda}_i\right)\right] \\
&= -(T-p)\sum_{i=r+1}^{n}\log\left(1-\hat{\lambda}_i\right).
\end{aligned}$$

Suppose the null hypothesis is true. The stochastic version of a second order Taylor approximation around 0 tells us that

$$\log\left(1-\hat{\lambda}_i\right) = -\hat{\lambda}_i - \frac{1}{2}\tilde{\lambda}_i^2,$$

where $\tilde{\lambda}_i$ is a convex combination of 0 and $\hat{\lambda}_i$. Since $T\hat{\lambda}_i = O_p(1)$, it follows that

$$(T-p)\hat{\lambda}_i^2 = \frac{T-p}{T^2}(T\hat{\lambda}_i)^2 = o_p(1)$$

and, by implication, $(T-p)\tilde{\lambda}_i^2 = o_p(1)$ as well. Therefore,

$$\begin{aligned}
\hat{LR}_T &= (T-p)\sum_{i=r+1}^{n}\hat{\lambda}_i + \frac{1}{2}\sum_{i=r+1}^{n}(T-p)\tilde{\lambda}_i^2 \\
&= (T-p)\sum_{i=r+1}^{n}\hat{\lambda}_i + o_p(1).
\end{aligned}$$

Slutsky's theorem now tells us that

$$\hat{LR}_T \xrightarrow{d} \text{tr}\left[\left(\int_0^1 W^{n-r}(s)dW^{n-r}(s)'\right)'\left(\int_0^1 W^{n-r}(s)W^{n-r}(s)'ds\right)^{-1}\left(\int_0^1 W^{n-r}(s)dW^{n-r}(s)'\right)\right].$$

This is called the trace test due to the form of the limiting distribution.

The Maximum Eigenvalue Test

An alternative to the trace test considers testing the null hypothesis

$$H_0 : rank(\Pi) = r$$

against the alternative hypothesis

$$H_1 : rank(\Pi) = r+1$$

sequentially for $r = 0, \cdots, n-1$, stopping only when the null can no longer be rejected. In this case, the likelihood ratio test statistic is given as

$$\begin{aligned}
\hat{LR}_T &= -2\left[l(\alpha,\beta,\Pi,\Sigma \mid H_0) - l(\alpha,\beta,\Pi,\Sigma \mid H_1)\right] \\
&= n(T-p)\left(\log(2\pi) + 1 + \frac{1}{n}\log|S_\Delta|\right) + (T-p)\sum_{i=1}^{r}\log\left(1-\hat{\lambda}_i\right) \\
&\quad - \left[n(T-p)\left(\log(2\pi) + 1 + \frac{1}{n}\log|S_\Delta|\right) + (T-p)\sum_{i=1}^{r+1}\log\left(1-\hat{\lambda}_i\right)\right] \\
&= -(T-p)\log\left(1-\hat{\lambda}_{r+1}\right).
\end{aligned}$$

Suppose the null is true. Relying on the same second degree Taylor approximation as before tells us that

$$\hat{LR}_T = (T-p)\hat{\lambda}_{r+1} + o_p(1) \xrightarrow{d} \eta_1,$$

where η_1 is the largest eigenvalue of

$$\left(\int_0^1 W^{n-r}(s)dW^{n-r}(s)'\right)'\left(\int_0^1 W^{n-r}(s)W^{n-r}(s)'ds\right)^{-1}\left(\int_0^1 W^{n-r}(s)dW^{n-r}(s)'\right).$$

This is called the maximum eigenvalue test, again due to the form of the limiting distribution.

5.4.8 Asymptotic Distribution of QMLEs

Here we derive the asymptotic distribution of the QMLEs of β, Π and Γ. We first use a previous result to establish the limit of $\hat{\beta}_T$, or at least a rotation of it.

Theorem (Asymptotic Distribution of Cointegrating Relationships)
Maintain assumptions A1 to A3, and suppose that $r > 0$. Then,

$$T\left(\hat{\beta}_T \hat{x}_T^{-1} - \beta\right) \xrightarrow{d} \beta_\perp \left(\int_0^1 B^{n-r}(s) B^{n-r}(s)' ds\right)^{-1} \int_0^1 B^{n-r}(s) dV^n(s)',$$

where $\{B^{n-r}(s)\}_{s\in[0,1]}$ and $\{V^n(s)\}_{s\in[0,1]}$ are $n-r$ and n-dimensional Brownian motions defined as

$$B^{n-r}(s) = (\beta_\perp' \beta_\perp)(\alpha_\perp' \Gamma(1) \beta_\perp)^{-1} \cdot W^{n-r}(s)$$

$$V^n(s) = \mu_{-1} \left(\beta' \mu_{\Delta,-1}' \mu_\Delta^{-1} \mu_{\Delta,-1} \beta\right)^{-1} \beta' \mu_{\Delta,-1}' \mu_\Delta^{-1} \Sigma^{\frac{1}{2}} \cdot W^n(s)$$

for any $s \in [0,1]$.

Proof) For any $1 \le i \le r$, recall that

$$\hat{\beta}_i = \beta \cdot \hat{x}_{i,T} + \beta_\perp \hat{y}_{i,T}$$

and

$$\beta_\perp' S(\hat{\lambda}_i) \beta \cdot \hat{x}_{i,T} + \beta_\perp' S(\hat{\lambda}_i) \beta_\perp \cdot \hat{y}_{i,T} = \mathbf{0}$$

Therefore,

$$\begin{aligned} T\left(\hat{\beta}_i - \beta \hat{x}_{i,T}\right) &= T \cdot \beta_\perp \hat{y}_{i,T} \\ &= -\beta_\perp \left(\frac{1}{T} \beta_\perp' S(\hat{\lambda}_i) \beta_\perp\right)^{-1} \beta_\perp' S(\hat{\lambda}_i) \beta \cdot \hat{x}_{i,T}. \end{aligned}$$

We saw above that

$$\frac{1}{T} \beta_\perp' S(\hat{\lambda}_i) \beta_\perp \xrightarrow{d} \lambda_i \cdot \beta_\perp' \Lambda \left(\int_0^1 W^n(r) W^n(r)' dr\right) \Lambda' \beta_\perp,$$

where $\lambda_i > 0$ is the ith largest solution to

$$\left|\lambda \mu_{-1} - \beta' \mu_{\Delta,-1}' \mu_\Delta^{-1} \mu_{\Delta,-1} \beta\right| = 0;$$

by implication,

$$\left(\frac{1}{T} \beta_\perp' S(\hat{\lambda}_i) \beta_\perp\right)^{-1} \xrightarrow{d} \lambda_i^{-1} (\beta_\perp' \beta_\perp)^{-1} (\beta_\perp' \Gamma(1)' \alpha_\perp) \left(\int_0^1 W^{n-r}(s) W^{n-r}(s)' ds\right)^{-1} (\alpha_\perp' \Gamma(1) \beta_\perp)(\beta_\perp' \beta_\perp)^{-1}.$$

It was also shown that

$$\beta_\perp' S(\hat{\lambda}_i)\beta \cdot \hat{x}_{i,T} = -\beta_\perp' \left(\frac{1}{T-p} Y_{-1}'\varepsilon\right) \mu_\Delta^{-1}\mu_{\Delta,-1}\beta\hat{x}_{i,T} + o_p(1).$$

Therefore,

$$\begin{aligned}
T\left(\hat{\beta}_i - \beta\hat{x}_{i,T}\right) &= \beta_\perp \left(\frac{1}{T}\beta_\perp' S(\hat{\lambda}_i)\beta_\perp\right)^{-1} \beta_\perp' \left(\frac{1}{T-p} Y_{-1}'\varepsilon\right) \mu_\Delta^{-1}\mu_{\Delta,-1}\beta \cdot \hat{x}_{i,T} + o_p(1) \\
&= \beta_\perp(\beta_\perp'\beta_\perp)^{-1}(\beta_\perp'\Gamma(1)'\alpha_\perp)\left(\int_0^1 W^{n-r}(s)W^{n-r}(s)'ds\right)^{-1}(\alpha_\perp'\Gamma(1)\beta_\perp)(\beta_\perp'\beta_\perp)^{-1}\beta_\perp' \\
&\quad \times \left(\frac{1}{T-p} Y_{-1}'\varepsilon\right)\mu_\Delta^{-1}\mu_{\Delta,-1}\beta \cdot \hat{x}_{i,T}\lambda_i^{-1} + o_p(1).
\end{aligned}$$

Defining

$$D_r = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_r, \end{pmatrix}$$

we can now see that

$$\begin{aligned}
T\left(\hat{\beta}_T - \beta\hat{x}_T\right) &= \beta_\perp(\beta_\perp'\beta_\perp)^{-1}(\beta_\perp'\Gamma(1)'\alpha_\perp)\left(\int_0^1 W^{n-r}(s)W^{n-r}(s)'ds\right)^{-1}(\alpha_\perp'\Gamma(1)\beta_\perp)(\beta_\perp'\beta_\perp)^{-1}\beta_\perp' \\
&\quad \times \left(\frac{1}{T-p} Y_{-1}'\varepsilon\right)\mu_\Delta^{-1}\mu_{\Delta,-1}\beta \cdot \hat{x}_T D_r^{-1} + o_p(1).
\end{aligned}$$

Recall that, for any $1 \le i \le r$,

$$\beta' S(\hat{\lambda}_i)\beta \cdot \hat{x}_{i,T} = \beta' S_{-1}\beta \cdot \hat{x}_{i,T}\hat{\lambda}_i - \beta' S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\beta \cdot \hat{x}_{i,T} = o_p(1).$$

Since

$$\begin{aligned}
\beta' S_{-1}\beta \cdot \hat{\lambda}_i - \mu_{-1} \cdot \lambda_i &= o_p(1) \\
\beta' S_{\Delta,-1}' S_\Delta^{-1} S_{\Delta,-1}\beta - \beta'\mu_{\Delta,-1}'\mu_\Delta^{-1}\mu_{\Delta,-1}\beta &= o_p(1)
\end{aligned}$$

and $\hat{x}_{i,T} = O_p(1)$, we have

$$\mu_{-1} \cdot \hat{x}_{i,T}\lambda_i - \beta'\mu_{\Delta,-1}'\mu_\Delta^{-1}\mu_{\Delta,-1}\beta \cdot \hat{x}_{i,T} = o_p(1)$$

as well, so that

$$\mu_{-1} \cdot \hat{x}_T D_r - \beta'\mu_{\Delta,-1}'\mu_\Delta^{-1}\mu_{\Delta,-1}\beta \cdot \hat{x}_T = o_p(1).$$

$\hat{x}_T = O_p(1)$ and $\hat{x}_T^{-1} = O_p(1)$, so that

$$\hat{x}_T D_r^{-1} \hat{x}_T^{-1} \xrightarrow{p} \left(\beta' \mu_{\Delta,-1}' \mu_\Delta^{-1} \mu_{\Delta,-1} \beta\right)^{-1} \mu_{-1}.$$

This reveals that

$$\begin{aligned} T\left(\hat{\beta}_T \hat{x}_T^{-1} - \beta\right) &= \beta_\perp (\beta_\perp' \beta_\perp)^{-1} (\beta_\perp' \Gamma(1)' \alpha_\perp) \left(\int_0^1 W^{n-r}(s) W^{n-r}(s)' ds\right)^{-1} (\alpha_\perp' \Gamma(1) \beta_\perp)(\beta_\perp' \beta_\perp)^{-1} \beta_\perp' \\ &\quad \times \left(\frac{1}{T-p} Y_{-1}' \varepsilon\right) \mu_\Delta^{-1} \mu_{\Delta,-1} \beta \cdot \hat{x}_T D_r^{-1} \hat{x}_T^{-1} + o_p(1) \\ &= \beta_\perp (\beta_\perp' \beta_\perp)^{-1} (\beta_\perp' \Gamma(1)' \alpha_\perp) \left(\int_0^1 W^{n-r}(s) W^{n-r}(s)' ds\right)^{-1} (\alpha_\perp' \Gamma(1) \beta_\perp)(\beta_\perp' \beta_\perp)^{-1} \beta_\perp' \\ &\quad \times \left(\frac{1}{T-p} Y_{-1}' \varepsilon\right) \mu_\Delta^{-1} \mu_{\Delta,-1} \beta \left(\beta' \mu_{\Delta,-1}' \mu_\Delta^{-1} \mu_{\Delta,-1} \beta\right)^{-1} \mu_{-1} + o_p(1). \end{aligned}$$

Finally, we know that

$$\frac{1}{T-p} Y_{-1}' \varepsilon \xrightarrow{d} \Lambda \left(\int_0^1 W^n(r) dW^n(r)'\right) \Sigma^{\frac{1}{2}\prime},$$

so

$$\begin{aligned} &\beta_\perp' \left(\frac{1}{T-p} Y_{-1}' \varepsilon\right) \mu_\Delta^{-1} \mu_{\Delta,-1} \beta \left(\beta' \mu_{\Delta,-1}' \mu_\Delta^{-1} \mu_{\Delta,-1} \beta\right)^{-1} \mu_{-1} \\ &\xrightarrow{d} (\beta_\perp' \beta_\perp)(\alpha_\perp' \Gamma(1) \beta_\perp)^{-1} \left(\int_0^1 W^{n-r}(s) dW^n(s)'\right) \Sigma^{\frac{1}{2}\prime} \mu_\Delta^{-1} \mu_{\Delta,-1} \beta \left(\beta' \mu_{\Delta,-1}' \mu_\Delta^{-1} \mu_{\Delta,-1} \beta\right)^{-1} \mu_{-1}. \end{aligned}$$

Therefore,

$$T\left(\hat{\beta}_T \hat{x}_T^{-1} - \beta\right) \xrightarrow{d} \beta_\perp \left(\int_0^1 B^{n-r}(s) B^{n-r}(s)' ds\right)^{-1} \int_0^1 B^{n-r}(s) dV^n(s)',$$

where the processes $\{B^{n-r}(s)\}_{s\in[0,1]}$ and $\{V^n(s)\}_{s\in[0,1]}$ are the $n-r$ and n-dimensional Brownian motions defined above.

Q.E.D.

Estimating Structural Break Points

Bai and Perron (1998)

In this paper, the authors estimate multiple break points at once under the assumption of structural changes in regression coefficients, and also present a way to test for the number of break points. In this chapter we deal with the former problem, namely the estimation of multiple break points through simultaneous and sequential means.

Assuming that there are m break points $T_1, \cdots, T_m$ in the sample, so that there are $m+1$ regimes, the model in question is given as

$$y_t = \underbrace{x_t'}_{1\times p} \cdot \underbrace{\beta}_{p\times 1} + \underbrace{z_t'}_{1\times k} \cdot \underbrace{\delta_j}_{k\times 1} + u_t$$

for any $T_{j-1}+1 \le t \le T_j$ and $1 \le j \le m+1$, where $T_0 = 0$ and $T_{m+1} = T$. Due to the presence of regime-independent coefficients β, this is a model of partial structural change; if $p = 0$, then every slope coefficient becomes regime-dependent.
Our objective is to estimate the structural break dates $T_1, \cdots, T_m$. The true dates and parametes are denoted with the superscript 0.

There exists a convenient way to organize the data. Define $Y = (y_1, \cdots, y_T)'$, $X = (x_1, \cdots, x_T)'$, $U = (u_1, \cdots, u_T)'$ and

$$\bar{Z} = \text{diag}(Z_1, \cdots, Z_{m+1}) \quad \text{where} \quad Z_i = \begin{pmatrix} z'_{T_{j-1}+1} \\ \vdots \\ z'_{T_j} \end{pmatrix} \text{ for any } 1 \le j \le m+1,$$

with $\delta = (\delta_1, \cdots, \delta_{m+1})'$.
Then, the model can be expressed in matrix form as

$$\underbrace{Y}_{T\times 1} = \underbrace{X}_{T\times p} \underbrace{\beta}_{p\times 1} + \underbrace{\bar{Z}}_{T\times k(m+1)} \underbrace{\delta}_{k(m+1)\times 1} + \underbrace{U}_{T\times 1}.$$

6.1 Estimation of the Break Points

To estimate the break points $T_1,\cdots,T_m$, the authors use a least squares method. Specifically, for some $q>0$, denote by $B_{q,T}$ the set of all potential break dates $\{T_j\}=(T_1,\cdots,T_m)$ in $\{1,\cdots,T\}$ such that $|T_j-T_{j-1}|\geq q$ for $1\leq j\leq m+1$. For any $\{T_j\}\in B_{q,T}$, define the sum of squared deviations given break points $\{T_j\}$ as

$$\begin{aligned} S_T(\{T_j\},\beta,\delta) &= \sum_{j=1}^{m+1}\sum_{t=T_{j-1}+1}^{T_j}(y_t-x_t'\beta-z_t'\delta_j)^2 \\ &= (Y-X\beta-\bar{Z}\delta)'(Y-X\beta-\bar{Z}\delta). \end{aligned}$$

The minimizers of the above function with respect to β,δ are given by the least squares estimates

$$\begin{pmatrix}\hat{\beta}(\{T_j\})\\ \hat{\delta}(\{T_j\})\end{pmatrix}=\begin{pmatrix}X'X & X'\bar{Z}\\ \bar{Z}'X & \bar{Z}'\bar{Z}\end{pmatrix}^{-1}\begin{pmatrix}X'\\ \bar{Z}'\end{pmatrix}Y,$$

which implies that

$$\begin{aligned} \hat{\beta}(\{T_j\}) &= (X'M_{\bar{Z}}X)^{-1}X'M_{\bar{Z}}Y \\ \hat{\delta}(\{T_j\}) &= \left(\bar{Z}'M_X\bar{Z}\right)^{-1}\bar{Z}'M_XY, \end{aligned}$$

where $M_{\bar{Z}}$ and M_X are the residual makers associated with $\bar{Z}$ and X.
Defining $V(\{T_j\})=(X,\bar{Z})$, the concentrated sum of squared deviations is then given by

$$\tilde{S}_T(\{T_j\})=S_T(\{T_j\},\hat{\beta}(\{T_j\}),\hat{\delta}(\{T_j\})=Y'M_{V(\{T_j\})}Y.$$

We define our break point estimators $\{\hat{T}_j\}=(\hat{T}_1,\cdots,\hat{T}_m)$ as the solutions to the minimzation problem

$$\begin{aligned} \min_{\{T_j\}\in B_{q,T}} \quad & \tilde{S}_T(\{T_j\})=Y'M_{V(\{T_j\})}Y \\ \text{subject to} \quad & W=(X,\bar{Z}). \end{aligned}$$

Our estimators of β and δ are then given by

$$\begin{aligned} \hat{\beta} &= \hat{\beta}(\{\hat{T}_j\}) \\ \hat{\delta} &= \hat{\delta}(\{\hat{T}_j\}). \end{aligned}$$

6.2 Assumptions

The following are the assumptions made to ensure that the estimators of the break points are consistent and well-behaved estimators.

Formally, we assume the following:

(1) **Existence of Break Fractions**
We assume that there exist fractions $0 < \lambda_1^0 < \cdots < \lambda_m^0 < 1$ such that

$$T_j^0 = \lfloor T\lambda_j^0 \rfloor$$

for any $1 \leq j \leq m$ and $T \in N_+$.
It follows that

$$\frac{T_j^0}{T} \leq \lambda_j^0 < \frac{T_j^0}{T} + \frac{1}{T}$$

for any $T \in N_+$, impyling that $\frac{T_j^0}{T} \xrightarrow{p} \lambda_j^0$ as $T \to \infty$ for each $1 \leq j \leq m+1$.

(2) **Asymptotic Properties of Sample Covariances**
Denote $w_t = (x_t', z_t')'$ for any $t \in N_+$, and define $W = (w_1, \cdots, w_T)'$. Let $\bar{W}^0$ be the diagonal partition of W at the true break points $T_1^0, \cdots, T_m^0$, that is,

$$\bar{W}^0 = \text{diag}(W_1^0, \cdots, W_{m+1}^0) \quad \text{where} \quad W_i^0 = \begin{pmatrix} w'_{T_{j-1}^0+1} \\ \vdots \\ w'_{T_j^0} \end{pmatrix} \text{ for any } 1 \leq i \leq m+1.$$

Then, we assume that, for any $1 \leq j \leq m+1$,

$$\frac{1}{T_j^0 - T_{j-1}^0} W_j^{0\prime} W_j^0 = \frac{1}{T_j^0 - T_{j-1}^0} \sum_{t=T_{j-1}^0+1}^{T_j^0} w_t w_t' \xrightarrow{p} Q_j^0 = \begin{pmatrix} Q_{j,x}^0 & Q_{j,xz}^0 \\ Q_{j,zx}^0 & Q_{j,z}^0 \end{pmatrix}$$

for some positive definite $(p+k) \times (p+k)$ matrix Q_j^0.
By implication,

$$\frac{1}{T} \sum_{t=1}^{T} w_t w_t' = \sum_{j=1}^{m+1} \frac{T_j^0 - T_{j-1}^0}{T} \cdot \left(\frac{1}{T_j^0 - T_{j-1}^0} \sum_{t=T_{j-1}^0+1}^{T_j^0} w_t w_t' \right)$$

for any $T \in N_+$; as such, we can conclude that

$$\frac{1}{T}\sum_{t=1}^{T} w_t w_t' \xrightarrow{p} \sum_{j=1}^{m+1}\left(\lambda_j^0 - \lambda_{j-1}^0\right) Q_j^0 := Q,$$

where Q is positive definite because each Q_j^0 is.

(3) **Identification Condition for Break Points**

We assume that there exists some $l_0 > 0$ and $\rho_{\min} > 0$ such that, for any $l > l_0$ and $1 \leq j \leq m+1$, the minimum eigenvalues of

$$A_{j-1,l} = \frac{1}{l}\sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+l} w_t w_t' \quad \text{and} \quad A_{jl}^* = \frac{1}{l}\sum_{t=T_j^0-l}^{T_j^0} w_t w_t'$$

are greater than or equal to $\rho_{\min}$. In other words, the matrices $A_{j-1,l}$ and A_{jl}^* are positive definite matrices that are bounded away from 0.

(4) **Nonsingularity of Regressors**

For any $i < l$ such that $l - i \geq k$,

$$B_{il} = \sum_{t=i}^{l} z_t z_t'$$

is nonsingular.

(5) **Uncorrelated Errors**

Since we are interested in investigating structural breaks when the regressors include lagged versions of the dependent variable, we assume that the error process $\{u_t\}_{t\in\mathbb{Z}}$ is a Martingale Difference Sequence (MDS) with respect to the filtration

$$\mathcal{F} = \{\mathcal{F}_t \mid t \in \mathbb{Z}\}$$

on $\mathbb{Z}$ where

$$\mathcal{F}_t = \sigma\left(\{w_s\}_{s\in\mathbb{Z}} \cup \{u_s\}_{s\leq t}\right)$$

for each $t \in \mathbb{Z}$, such that

$$\sup_{t\in\mathbb{Z}} \mathbb{E}|u_t|^{4+c} < +\infty$$

for some $c > 0$.
Furthermore, by the definition of an MDS $\mathbb{E}[u_t] = 0$, and we assume that $\mathbb{E}\left[u_t^2 \mid \mathcal{F}_{t-1}\right] = \sigma^2$ for any $t \in \mathbb{Z}$.

(6) **An FCLT for Martingale Difference Sequences**
Let the stochastic process $\{v_t\}_{t\in\mathbb{Z}}$ be defined as

$$v_t = w_t u_t$$

for any $t \in \mathbb{Z}$. We assumed above that $\{u_t\}_{t\in\mathbb{Z}}$ is an MDS with respect to the filtration $\mathcal{F}$; since $v_t = w_t u_t$ is $\mathcal{F}_t$-measurable for any $t \in \mathbb{Z}$ by the definition of $\mathcal{F}_t$, and

$$\begin{aligned}\mathbb{E}[v_t \mid \mathcal{F}_{t-1}] &= w_t \cdot \mathbb{E}[u_t \mid \mathcal{F}_{t-1}] && (w_t \text{ is } \mathcal{F}_{t-1}\text{-measurable})\\ &= w_t \cdot 0 = \mathbf{0}, && (\text{MDS property of } u_t)\end{aligned}$$

it follows that $\{v_t\}_{t\in\mathbb{Z}}$ is also an MDS with respect to $\mathcal{F}$.
For any $1 \le i,j \le p+k$, $\{V_{ij,t} = (u_t^2 - \sigma^2)w_{it}w_{jt}\}_{t\in\mathbb{Z}}$ is a mutually uncorrelated sequence of random variables with finite mean; to see uncorrelatedness, note that, for any $t,s \in \mathbb{Z}$ such that $s < t$,

$$\begin{aligned}\mathbb{E}[V_{ij,t}V_{ij,s}] &= \mathbb{E}[\mathbb{E}[V_{ij,t} \mid \mathcal{F}_{t-1}] \cdot V_{ij,s}]\\ &= \mathbb{E}\left[\left(\mathbb{E}\left[u_t^2 \mid \mathcal{F}_{t-1}\right] - \sigma^2\right) w_{it}w_{jt}V_{ij,s}\right] = 0,\end{aligned}$$

and for finite mean, note that

$$\mathbb{E}[V_{ij,t}] = \mathbb{E}\left[\left(\mathbb{E}\left[u_t^2 \mid \mathcal{F}_{t-1}\right] - \sigma^2\right) w_{it}w_{jt}\right] = 0.$$

It follows from the WLLN for uncorrelated sequences that

$$\frac{1}{T}\sum_{t=1}^{T} V_{ij,t} \xrightarrow{p} 0,$$

and because this holds for any $1 \le i,j \le p+k$, we have

$$\frac{1}{T}\sum_{t=1}^{T}(u_t^2 - \sigma^2)w_t w_t' \xrightarrow{p} O.$$

Finally,

$$\sigma^2 \cdot \frac{1}{T}\sum_{t=1}^{T} w_t w_t' \xrightarrow{p} \sigma^2 Q$$

by assumption (2), so we have

$$\frac{1}{T}\sum_{t=1}^{T} v_t v_t' = \frac{1}{T}\sum_{t=1}^{T} u_t^2 w_t w_t' \xrightarrow{p} \sigma^2 Q.$$

Therefore, it makes sense to assume that $\{v_t\}_{t\in\mathbb{Z}}$ follows some sort of FCLT result. Specifically, define the stochastic processes $\{S_T(r)\}_{r\in[0,1]}$ and with continuous paths as

$$S_T(r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor} v_t + \frac{1}{\sqrt{T}}(Tr - \lfloor Tr\rfloor)v_{\lfloor Tr\rfloor+1}$$

for any $r \in [0,1]$. Letting S_T be the random function in $\mathcal{C}^{p+k}[0,1]$ corresponding to $\{S_T(r)\}_{r\in[0,1]}$, we assume that

$$S_T \xrightarrow{d} \sigma Q^{\frac{1}{2}} \cdot W^{p+k},$$

where $Q^{\frac{1}{2}}$ is the Cholesky factor of Q.
Defining $\{V_T(r)\}_{r\in[0,1]}$ as the process collecting the lower k rows of $\{S_T(r)\}_{r\in[0,1]}$ and V_T the random function corresponding to $\{V_T(r)\}_{r\in[0,1]}$, it follows that

$$V_T(r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor} z_t u_t + \frac{1}{\sqrt{T}}(Tr - \lfloor Tr\rfloor)z_{\lfloor Tr\rfloor+1}u_{\lfloor Tr\rfloor+1}$$

for any $r \in [0,1]$ and

$$V_T \xrightarrow{d} B^k$$

for some k-dimensional random function B^k corresponding to a Brownian motion process.

(7) **Uniform Convergence of Sample Covariances**

Assuming that the regressor z_t contains lagged values of the dependent variable, so that $\{z_t\}_{t\in\mathbb{Z}}$ is not strictly exogenous, our break point estimators are found by solving the minimzation problem

$$\begin{aligned} \min_{\{T_j\}\in B_{\varepsilon T,T}} \quad & \tilde{S}_T(\{T_j\}) = Y' M_{V(\{T_j\})} Y \\ \text{subject to} \quad & W = (X, \bar{Z}) \end{aligned}$$

for some $\varepsilon > 0$.

We assume that

$$\frac{1}{T} \sum_{t=\lfloor Ts \rfloor+1}^{\lfloor Tr \rfloor} z_t z_t' \xrightarrow{p} \Omega(r) - \Omega(s)$$

uniformly on the set of all $(r,s) \in [0,1]^2$, where $\Omega(0) = O$, $\Omega(r) - \Omega(s)$ is positive definite for any $0 \le s < r \le 1$, and

$$\sup_{(v,u)\in[0,1]^2,\ v-u\ge\varepsilon} \left\| \left(\Omega(v) - \Omega(u)\right)^{-1} \right\| < +\infty.$$

Note that, for any $T \in N_+$ such that $T \ge \frac{k-1}{\varepsilon}$, $\frac{1}{T}\sum_{t=\lfloor Ts \rfloor+1}^{\lfloor Tr \rfloor} z_t z_t'$ is positive definite for any $(r,s) \in [0,1]^2$ such that $r - s \ge \varepsilon$ by assumption (4), since

$$\lfloor Tr \rfloor - \lfloor Ts \rfloor \ge Tr + 1 - Ts = T(r-s) + 1 \ge \varepsilon T + 1 \ge k$$

in this case.

6.3 Preliminary Results

The following are results that Bai and Perron establish prior to the proof:

i) **Rate of Convergence of $X_j^{0\prime} M_{Z_j} X_j^0$**
For any $1 \leq j \leq m+1$, define $X_1^0, \cdots, X_{m+1}^0$ and $Z_1^0, \cdots, Z_{m+1}^0$ as

$$X_j^0 = \begin{pmatrix} x'_{T_{j-1}^0+1} \\ \vdots \\ x'_{T_j^0} \end{pmatrix} \quad \text{and} \quad Z_j^0 = \begin{pmatrix} z'_{T_{j-1}^0+1} \\ \vdots \\ z'_{T_j^0} \end{pmatrix}$$

for any $1 \leq j \leq m+1$, so that $W_j^0 = (X_j^0, Z_j^0)$.
Choose any $1 \leq j \leq m+1$, and note that

$$\begin{aligned} \frac{X_j^{0\prime} M_{Z_j^0} X_j^0}{T} &= \frac{1}{T} X_j^{0\prime} X_j^0 - \frac{1}{T} X_j^{0\prime} Z_j^0 \left(\frac{1}{T} Z_j^{0\prime} Z_j^0 \right)^{-1} \frac{1}{T} Z_j^{0\prime} X_j^0 \\ &= \frac{T_j^0 - T_{j-1}^0}{T} \cdot \frac{1}{T_j^0 - T_{j-1}^0} \sum_{t=T_{j-1}^0+1}^{T_j^0} x_t x_t' \\ -\frac{T_j^0 - T_{j-1}^0}{T} & \left(\frac{1}{T_j^0 - T_{j-1}^0} \sum_{t=T_{j-1}^0+1}^{T_j^0} x_t z_t' \right) \left(\frac{1}{T_j^0 - T_{j-1}^0} \sum_{t=T_{j-1}^0+1}^{T_j^0} z_t z_t' \right)^{-1} \left(\frac{1}{T_j^0 - T_{j-1}^0} \sum_{t=T_{j-1}^0+1}^{T_j^0} z_t x_t' \right). \end{aligned}$$

By assumption (1),

$$\frac{T_j^0}{T} \xrightarrow{p} \lambda_j^0$$

as $T \to \infty$ for any $1 \leq j \leq m+1$. This, together with assumption (2), implies that

$$\frac{X_j^{0\prime} M_{Z_j^0} X_j^0}{T} \xrightarrow{p} (\lambda_j^0 - \lambda_{j-1}^0) \left(Q_{j,x}^0 - Q_{j,xz}^0 \cdot Q_{j,z}^{0,-1} \cdot Q_{j,zx} \right),$$

where the right hand side is nonsingular due to the nonsingularity of Q_j. Therefore,

$$\frac{X_j^{0\prime} M_{Z_j^0} X_j^0}{T} \quad \text{and} \quad \left(\frac{X_j^{0\prime} M_{Z_j^0} X_j^0}{T} \right)^{-1}$$

are $O_p(1)$.

ii) **The Sizes of Submatrices**

Let there be full rank matrices S_1, V_1 with m and n columns, respectively, and r rows. Now let S, V be full rank matrices with the same number of columns as S_1, V_1, but now with $r+s$ rows. We will show that $S'M_V S - S_1' M_{V_1} S_1$ is positive semidefinite.

To this end, let the matrices S_2, V_2, with the same number of columns as S_1, V_1, collect the lower s rows of S and V. Choose any $\alpha \in \mathbb{R}^m$, and define $y = S\alpha \in \mathbb{R}^{r+s}$, $y_1 = S_1\alpha \in \mathbb{R}^r$ and $y_2 = S_2\alpha \in \mathbb{R}^s$, so that $y = (y_1', y_2')'$.

We approach this problem in terms of projections. Defining

$$\hat{b} = (V'V)^{-1}V'y \quad \text{and} \quad \hat{b}_1 = (V_1'V_1)^{-1}V_1'y_1,$$

it is easy to see that $V\hat{b}$ is the projection of y on the column space of V and $V_1\hat{b}_1$ the projection of y_1 on the column space of V_1. By definition,

$$\left|y_1 - V_1\hat{b}_1\right|^2 = \inf_{b\in\mathbb{R}^n} |y_1 - V_1 b|^2.$$

Note that

$$\left|y - V\hat{b}\right|^2 = \left|\begin{pmatrix} y_1 - V_1\hat{b} \\ y_2 - V_2\hat{b} \end{pmatrix}\right|^2 = \left|y_1 - V_1\hat{b}\right|^2 + \left|y_2 - V_2\hat{b}\right|^2 \geq \left|y_1 - V_1\hat{b}\right|^2.$$

Suppose that $\left|y - V\hat{b}\right|^2 < \left|y_1 - V_1\hat{b}_1\right|^2$. Then,

$$\left|y_1 - V_1\hat{b}\right|^2 \leq \left|y_1 - V_1\hat{b}\right|^2 + \left|y_2 - V_2\hat{b}\right|^2 = \left|y - V\hat{b}\right|^2 < \left|y_1 - V_1\hat{b}_1\right|^2,$$

which contradicts the assumption above that $\left|y_1 - V_1\hat{b}_1\right|^2$ is the infimum of $|y_1 - V_1 b|^2$ over the set of all n-dimensional vectors b. Therefore, it must be the case that

$$\left|y_1 - V_1\hat{b}_1\right|^2 \leq \left|y - V\hat{b}\right|^2.$$

Since

$$\left|y_1 - V_1\hat{b}_1\right|^2 = (y_1 - V_1\hat{b}_1)'(y_1 - V_1\hat{b}_1) = y_1' M_{V_1} y_1$$

and $\left|y - V\hat{b}\right|^2 = y'M_V y$, we finally have the inequality

$$\begin{aligned} \alpha'(S'M_V S - S_1' M_{V_1} S_1)\alpha &= y'M_V y - y_1' M_{V_1} y_1 \\ &= \left|y - V\hat{b}\right|^2 - \left|y_1 - V_1\hat{b}_1\right|^2 \geq 0. \end{aligned}$$

This holds for any $\alpha \in \mathbb{R}^m$, so $S'M_V S - S_1' M_{V_1} S_1$ is positive semidefinite.

iii) **Rate of Convergence of $X'M_{\bar{Z}}X$ over $B_{\varepsilon,T,T}$**

We will now show that

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left\| \left(\frac{X'M_{\bar{Z}}X}{T} \right)^{-1} \right\| = O_p(1).$$

under the stated assumptions.

Choose any $\{T_j\} \in B_{\varepsilon T,T}$, so that $0 < T_1 < \cdots < T_m < 1$ and $T_j - T_{j-1} \geq \varepsilon T$ for any $1 \leq j \leq m+1$, and construct $\bar{Z}$ as the diagonal partition of $Z = (z_1, \cdots, z_T)'$ according to $\{T_j\}$. It follows that

$$P_{\bar{Z}} = \bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}',$$

where $\bar{Z}'\bar{Z}$ is invertible for T large enough so that $\varepsilon T > k$ due to assumption (4). Using block matrix operations, we now have

$$\begin{aligned}
P_{\bar{Z}} &= \begin{pmatrix} Z_1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & Z_{m+1} \end{pmatrix} \begin{pmatrix} Z_1'Z_1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & Z_{m+1}'Z_{m+1} \end{pmatrix}^{-1} \begin{pmatrix} Z_1' & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & Z_{m+1}' \end{pmatrix} \\
&= \begin{pmatrix} Z_1(Z_1'Z_1)^{-1}Z_1' & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & Z_{m+1}(Z_{m+1}'Z_{m+1})^{-1}Z_{m+1}' \end{pmatrix} = \mathrm{diag}\left(P_{Z_1}, \cdots, P_{Z_{m+1}}\right).
\end{aligned}$$

and as such,

$$M_{\bar{Z}} = I_T - P_{\bar{Z}} = \mathrm{diag}\left(M_{Z_1}, \cdots, M_{Z_{m+1}}\right).$$

Define

$$X_j = \begin{pmatrix} x'_{T_{j-1}+1} \\ \vdots \\ x'_{T_j} \end{pmatrix}$$

for any $1 \leq j \leq m+1$, so that

$$X'M_{\bar{Z}}X = \begin{pmatrix} X_1' & \cdots & X_{m+1}' \end{pmatrix} \begin{pmatrix} M_{Z_1} & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & M_{Z_{m+1}} \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_{m+1} \end{pmatrix} = \sum_{j=1}^{m+1} X_j'M_{Z_j}X_j.$$

Let $\mathcal{R}_1,\cdots,\mathcal{R}_{m+1}$ be the regimes under $\{T_j\}$, that is,

$$\mathcal{R}_j=\{T_{j-1}+1,\cdots,T_j\}$$

for any $1\le j\le m+1$. Let $\mathcal{R}^0_1,\cdots,\mathcal{R}^0_{m+1}$ be the true regimes.
Suppose that none of the true regimes are contained in a single regime under $\{T_j\}$, that is, suppose that, for any $1\le j\le m+1$, there does not exist an $1\le i\le m+1$ such that

$$\mathcal{R}^0_j\subset\mathcal{R}_i.$$

In this case, because $T^0_0=T_0=1$, $T^0_1\ge T_1+1$, since otherwise $\mathcal{R}^0_1$ will be contained in $\mathcal{R}_1$.
Now suppose, for some $1\le j<m+1$, that $T^0_j\ge T_j+1$. Then, it must be the case that $T^0_{j+1}\ge T_{j+1}+1$, since otherwise, the regime $\mathcal{R}^0_{j+1}=\{T^0_j+1,\cdots,T^0_{j+1}\}$ would be contained in $\mathcal{R}_{j+1}=\{T_{j+1},\cdots,T_{j+1}\}$.
By induction, $T^0_{m+1}=T\ge T_{m+1}+1=T+1$, a contradiction. Therefore, it must be the case that there exists at least one $1\le j\le m+1$ and $1\le i\le m+1$ such that $\mathcal{R}^0_j\subset\mathcal{R}_i$.
This implies that $T_{i-1}+1\le T^0_{j-1}+1<T^0_j\le T_i$, so that (X^0_j,Z^0_j) is contained within (X_i,Z_i). By the above result on submatrices, it follows that

$$X_i'M_{Z_i}X_i-X^{0\prime}_jM_{Z^0_j}X^0_j$$

is positive semidefinite, and because each $X_l'M_{Z_l}X_l$ is positive semidefinite, we can see that

$$X'M_{\bar{Z}}X-X^{0\prime}_jM_{Z^0_j}X^0_j=\sum_{l=1}^{m+1}X_l'M_{Z_l}X_l-X^{0\prime}_jM_{Z^0_j}X^0_j$$

is also positive semidefinite.
Since $X^0_jM_{Z^0_j}X^0_j$ is a positive definite matrix, this implies that $X'M_{\bar{Z}}X$ must have full rank (otherwise, there exists a vector α such that $\alpha'X'M_{\bar{Z}}\alpha=0<\alpha'X^0_jM_{Z^0_j}X^0_j\alpha$, a contradiction). Therefore, $X'M_{\bar{Z}}X$ is nonsingular and by extension positive definite.
Furthermore, the above implies that

$$\left(\frac{X^{0\prime}_jM_{Z^0_j}X^0_j}{T}\right)^{-1}-\left(\frac{X'M_{\bar{Z}}X}{T}\right)^{-1}$$

is positive semidefinite, and as such the maximum eigenvalue of $\left(X^{0\prime}_jM_{Z^0_j}X^0_j\right)^{-1}$ is larger than or equal to that of $(X'M_{\bar{Z}}X)^{-1}$. Since the operator norm of positive semidefinite matrices is equal to their maximum eigenvalue, it follows that

$$\left\|\left(\frac{X'M_{\bar{Z}}X}{T}\right)^{-1}\right\|\le\left\|\left(\frac{X^{0\prime}_jM_{Z^0_j}X^0_j}{T}\right)^{-1}\right\|.$$

The above holds for any $\{T_j\} \in B_{\varepsilon T,T}$ for some $1 \leq j \leq m+1$, so we have

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left\| \left(\frac{X' M_{\bar{Z}} X}{T} \right)^{-1} \right\| \leq \max_{1\leq j\leq m+1} \left\| \left(\frac{X_j^0 M_{Z_j^0} X_j^0}{T} \right)^{-1} \right\|.$$

We saw above that, for each $1 \leq j \leq m+1$,

$$\left(\frac{X_j^0 M_{Z_j^0} X_j^0}{T} \right)^{-1} = O_p(1),$$

so it follows that

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left\| \left(\frac{X' M_{\bar{Z}} X}{T} \right)^{-1} \right\| = O_p(1).$$

iv) **Rate of Convergence of $X'M_{\bar{Z}}\bar{Z}^0$ over $B_{\varepsilon T,T}$**

We will show here that

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left\| \frac{X'M_{\bar{Z}}\bar{Z}^0}{T} \right\| = O_p(1).$$

Choose any $\{T_j\} \in B_{\varepsilon T,T}$, and let $\bar{Z}$ be the diagonal partition of Z according to $\{T_j\}$. Then, the residual maker $M_{\bar{Z}} = I_T - \bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}'$ is symmetric and idempotent, so that it is positive semidefinite (its eigenvalues are either 1 or 0 due to idempotence) with $\text{rank}(M_{\bar{Z}}) = \text{tr}(M_{\bar{Z}}) = T-(p+k) > 0$ for large enough T. Therefore, the operator norm of $M_{\bar{Z}}$ is equal to its largest eigenvalue, which is 1, and it follows that

$$\left\| \frac{X'M_{\bar{Z}}\bar{Z}^0}{T} \right\| \le \frac{1}{T}\|X\| \cdot \left\|\bar{Z}^0\right\| \le \text{tr}\left(\frac{1}{T}X'X\right)^{\frac{1}{2}} \text{tr}\left(\frac{1}{T}\bar{Z}^{0\prime}\bar{Z}^0\right)^{\frac{1}{2}}.$$

This holds for any $\{T_j\} \in B_{\varepsilon T,T}$, so

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left\| \frac{X'M_{\bar{Z}}\bar{Z}^0}{T} \right\| \le \text{tr}\left(\frac{1}{T}X'X\right)^{\frac{1}{2}} \text{tr}\left(\frac{1}{T}\bar{Z}^{0\prime}\bar{Z}^0\right)^{\frac{1}{2}}.$$

Since

$$\frac{1}{T}X'X = \sum_{j=1}^{m+1} \frac{T_{j-1}^0 - T_j^0}{T} \left(\frac{1}{T_{j-1}^0 - T_j^0} \sum_{t=T_{j-1}^0+1}^{T_j^0} x_t x_t' \right) \xrightarrow{p} \sum_{j=1}^{m+1} \left(\lambda_j^0 - \lambda_{j-1}^0\right) Q_{j,x}^0$$

and similarly,

$$\frac{1}{T}Z_j^{0\prime}Z^0 = \frac{T_{j-1}^0 - T_j^0}{T} \left(\frac{1}{T_{j-1}^0 - T_j^0} \sum_{t=T_{j-1}^0+1}^{T_j^0} z_t z_t' \right) \xrightarrow{p} \left(\lambda_j^0 - \lambda_{j-1}^0\right) Q_{j,z}^0$$

for any $1 \le j \le m+1$ by assumption (2), together with the condition

$$\left\| \frac{1}{T}\bar{Z}^{0\prime}\bar{Z} \right\| \le \sum_{j=1}^{m+1} \left\| \frac{1}{T}Z_j^{0\prime}Z_j^0 \right\|$$

we can see that $\frac{1}{T}X'X$ and $\frac{1}{T}\bar{Z}^{0\prime}\bar{Z}^0$ are $O_p(1)$. It follows that

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left\| \frac{X'M_{\bar{Z}}\bar{Z}^0}{T} \right\| = O_p(1).$$

v) **Rate of Convergence of $P_{\bar{Z}}U$ over $B_{\varepsilon T,T}$**

We will show here that

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}} |P_{\bar{Z}}U|^2 = O_p(1).$$

First note that, for any $\{T_j\} \in B_{\varepsilon T,T}$, we can write

$$\begin{aligned}|P_{\bar{Z}}U|^2 &= U'P_{\bar{Z}}U = U'\bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}'U \\ &= \left(\frac{1}{\sqrt{T}}\bar{Z}'U\right)'\left(\frac{1}{T}\bar{Z}'\bar{Z}\right)^{-1}\left(\frac{1}{\sqrt{T}}\bar{Z}'U\right),\end{aligned}$$

so that

$$\begin{aligned}U'P_{\bar{Z}}U &= |U'P_{\bar{Z}}U| \\ &\le \left|\frac{1}{\sqrt{T}}\bar{Z}'U\right|^2 \cdot \left\|\left(\frac{1}{T}\bar{Z}'\bar{Z}\right)^{-1}\right\|.\end{aligned}$$

We now examine the terms on the right hand side one by one.

Step 1: The $\left|\frac{1}{\sqrt{T}}\bar{Z}'U\right|^2$ Ordinate

Since

$$\bar{Z}'U = \begin{pmatrix} Z_1' & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & Z_{m+1}' \end{pmatrix}\begin{pmatrix} u_1 \\ \vdots \\ u_T \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^{T_1} z_t u_t \\ \vdots \\ \sum_{t=T_m+1}^{T} z_t u_t \end{pmatrix},$$

we have

$$\begin{aligned}\left|\frac{1}{\sqrt{T}}\bar{Z}'U\right|^2 &= \sum_{j=1}^{m+1}\left|\frac{1}{\sqrt{T}}\sum_{t=T_{j-1}+1}^{T_j} z_t u_t\right|^2 \\ &= \sum_{j=1}^{m+1}\left|V_T\left(\frac{T_j}{T}\right) - V_T\left(\frac{T_{j-1}}{T}\right)\right|^2,\end{aligned}$$

where the last equality follows because $T\frac{T_j}{T} = \lfloor T\frac{T_j}{T}\rfloor$ for any $0 \le j \le m+1$.
Define the set

$$J_\varepsilon = \{(v,u) \in [0,1]^2 \mid v - u \ge \varepsilon\}.$$

By the way in which we defined $B_{\varepsilon T,T}$, $T_j - T_{j-1} > \varepsilon T$, which implies $\frac{T_j}{T} - \frac{T_{j-1}}{T} > \varepsilon$; as such,

$$\left|\frac{1}{\sqrt{T}}\bar{Z}'U\right|^2 \le (m+1)\cdot \sup_{(v,u)\in J_\varepsilon} |V_T(v)-V_T(u)|^2.$$

Because this holds for any $\{T_j\} \in B_{\varepsilon T,T}$, we in fact have

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left|\frac{1}{\sqrt{T}}\bar{Z}'U\right|^2 \le (m+1)\cdot \sup_{(v,u)\in J_\varepsilon} |V_T(v)-V_T(u)|^2.$$

We now show that the function of V_T defined above is continuous on $\mathcal{C}^n[0,1]$, and therefore that the term itself is $O_p(1)$ by the FCLT assumption and the continuous mapping theorem.

Let the function $g:\mathcal{C}^n[0,1]\to\mathbb{R}_+$ be defined as

$$g(f) = \sup_{(v,u)\in J_\varepsilon} |f(v)-f(u)|$$

for any $f\in\mathcal{C}^n[0,1]$.
Now choose any $f,h\in\mathcal{C}^n[0,1]$, and assume without loss of generality that

$$g(f) = \sup_{(v,u)\in J_\varepsilon} |f(v)-f(u)| > \sup_{(v,u)\in J_\varepsilon} |h(v)-h(u)| = g(h).$$

Then, define $\bar{f}: J_\varepsilon \to \mathbb{R}_+$ as

$$\bar{f}(v,u) = |f(v)-f(u)|$$

for any $(v,u)\in J_\varepsilon$. J_ε is closed and bounded, so it is compact, and $\bar{f}$ is a continuous function on the compact set J_ε; by the extreme value theorem, there exists a $(v^*,u^*)\in J_\varepsilon$ such that

$$|f(v^*)-f(u^*)| = \bar{f}(v^*,u^*) = \sup_{(v,u)\in J_\varepsilon} \bar{f}(v,u) = \sup_{(v,u)\in J_\varepsilon} |f(v)-f(u)|.$$

Therefore,

$$\begin{aligned}
|g(f)-g(h)| = g(f)-g(h) &= \sup_{(v,u)\in J_\varepsilon} |f(v)-f(u)| - \sup_{(v,u)\in J_\varepsilon} |h(v)-h(u)| \\
&= |f(v^*)-f(u^*)| - \sup_{(v,u)\in J_\varepsilon} |h(v)-h(u)| \\
&\le |f(v^*)-f(u^*)| - |h(v^*)-h(u^*)| \\
&\le |f(v^*)-f(u^*)-(h(v^*)-h(u^*))| \\
&\le |f(v^*)-h(v^*)| + |f(u^*)-h(u^*)| \le 2\cdot\|f-h\|_{\mathcal{C}}.
\end{aligned}$$

Therefore, g is Lipschitz continuous on $\mathcal{C}^n[0,1]$, and by the continuous mapping theorem,

$$\sup_{(v,u)\in J_\varepsilon} |V_T(v) - V_T(u)|^2 = g(V_T) \xrightarrow{d} g(W^k) = \sup_{(v,u)\in J_\varepsilon} \left|W^k(v) - W^k(u)\right|^2.$$

It follows that

$$\sup_{(v,u)\in J_\varepsilon} |V_T(v) - V_T(u)|^2 = O_p(1)$$

and as such that

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left|\frac{1}{\sqrt{T}}\bar{Z}'U\right|^2 = O_p(1).$$

Step 2: The $\left\|\left(\frac{1}{T}\bar{Z}'\bar{Z}\right)^{-1}\right\|$ Ordinate
Because

$$\frac{1}{T}\bar{Z}'\bar{Z} = \begin{pmatrix} \frac{1}{T}Z_1'Z_1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & \frac{1}{T}Z_{m+1}'Z_{m+1} \end{pmatrix},$$

we have

$$\left(\frac{1}{T}\bar{Z}'\bar{Z}\right)^{-1} = \begin{pmatrix} \left(\frac{1}{T}Z_1'Z_1\right)^{-1} & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & \left(\frac{1}{T}Z_{m+1}'Z_{m+1}\right)^{-1} \end{pmatrix}$$

and thus

$$\begin{aligned} \left\|\left(\frac{1}{T}\bar{Z}'\bar{Z}\right)^{-1}\right\| &\leq \sum_{j=1}^{m+1} \left\|\left(\frac{1}{T}\sum_{t=T_{j-1}+1}^{T_j} z_t z_t'\right)^{-1}\right\| \\ &\leq (m+1)\cdot \sup_{(v,u)\in J_\varepsilon} \left\|\left(\frac{1}{T}\sum_{t=\lfloor Tu\rfloor+1}^{\lfloor Tv\rfloor} z_t z_t'\right)^{-1}\right\|, \end{aligned}$$

where the inverse functions exist for large enough T by assumption (4) because we restrict our attention to $(v,u)\in[0,1]^2$ such that $v-u\geq\varepsilon$. This holds for any $\{T_j\}\in B_{\varepsilon T,T}$, so

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left\|\left(\frac{1}{T}\bar{Z}'\bar{Z}\right)^{-1}\right\| \leq (m+1)\cdot \sup_{(v,u)\in J_\varepsilon} \left\|\left(\frac{1}{T}\sum_{t=\lfloor Tu\rfloor+1}^{\lfloor Tv\rfloor} z_t z_t'\right)^{-1}\right\|.$$

By assumption (7), which states that

$$\frac{1}{T}\sum_{t=\lfloor Ts\rfloor+1}^{\lfloor Tr\rfloor} z_t z_t' \xrightarrow{p} \Omega(r)-\Omega(s)$$

as $T\to\infty$ uniformly on J_ε,

$$(m+1)\sup_{(v,u)\in J_\varepsilon}\left\|\left(\frac{1}{T}\sum_{t=\lfloor Tu\rfloor+1}^{\lfloor Tv\rfloor} z_t z_t'\right)^{-1}\right\| \xrightarrow{p} (m+1)\sup_{(v,u)\in J_\varepsilon}\left\|\left(\Omega\left(v\right)-\Omega\left(u\right)\right)^{-1}\right\|,$$

where the inverse matrices on the right hand side exist because $v-u\geq varepsilon$, which means $Q\left(v\right)-Q\left(u\right)$ is positive definite by assumption (7).
Since the value on the right hand side above is assumed to be finite, we can conclude that

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left\|\left(\frac{1}{T}\bar{Z}'\bar{Z}\right)^{-1}\right\| = O_p(1).$$

It now follows that

$$\sup_{\{T_j\}B_{\varepsilon T,T}}|P_{\bar{Z}}U|^2 \leq \left(\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{1}{\sqrt{T}}\bar{Z}'U\right|^2\right)\left(\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left\|\left(\frac{q}{T}\bar{Z}'\bar{Z}\right)^{-1}\right\|\right),$$

and as such that

$$\sup_{\{T_j\}B_{\varepsilon T,T}}|P_{\bar{Z}}U|^2 = O_p(1).$$

vi) **Rate of Convergence of $X'P_{\bar{Z}}U$ and $\bar{Z}^{0\prime}P_{\bar{Z}}U$ over $B_{\varepsilon T,T}$**

We will show here that

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{1}{\sqrt{T}}X'P_{\bar{Z}}U\right|, \sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{1}{\sqrt{T}}\bar{Z}^{0\prime}P_{\bar{Z}}U\right| = O_p(1).$$

This follows easily from the preceding result.

Note that

$$\left|\frac{1}{\sqrt{T}}X'P_{\bar{Z}}U\right| \leq |P_{\bar{Z}}U|\cdot\left\|\frac{1}{\sqrt{T}}X\right\|$$

for any $\{T_j\}\in B_{\varepsilon T,T}$. Thus,

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{1}{\sqrt{T}}X'P_{\bar{Z}}U\right| \leq \left(sup_{\{T_j\}\in B_{\varepsilon T,T}}|P_{\bar{Z}}U|\right)\cdot\operatorname{tr}\left(\frac{X'X}{T}\right)^{\frac{1}{2}}$$

We already know that both terms on the right hand side are $O_p(1)$, so that

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{1}{\sqrt{T}}X'P_{\bar{Z}}U\right| = O_p(1).$$

Likewise,

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{1}{\sqrt{T}}\bar{Z}^{0\prime}P_{\bar{Z}}U\right| \leq \left(sup_{\{T_j\}\in B_{\varepsilon T,T}}|P_{\bar{Z}}U|\right)\cdot\operatorname{tr}\left(\frac{\bar{Z}^{0\prime}\bar{Z}^{0}}{T}\right)^{\frac{1}{2}},$$

and by the same reasoning as above,

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{1}{\sqrt{T}}\bar{Z}^{0\prime}P_{\bar{Z}}U\right| = O_p(1).$$

6.4 Consistency of the Break Fraction Estimators

We first prove that the estimators $0 < \hat{\lambda}_1 < \cdots < \hat{\lambda}_m < 1$ of the break fractions, which are defined as

$$\hat{\lambda}_j = \frac{\hat{T}_j}{T}$$

for any $1 \le j \le m$, are consistent for the true break fractions $\lambda_1^0, \cdots, \lambda_m^0$.
This is shown in several steps:

6.4.1 Step 1: Convergence of the Sum of Squared Differences

For any $1 \le t \le T$, define $\hat{u}_t$ and d_t as

$$\begin{aligned} \hat{u}_t &= y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_j \quad \text{and} \\ d_t &= x_t'\left(\hat{\beta} - \beta^0\right) + z_t'\left(\hat{\delta}_j - \delta_l^0\right) \end{aligned}$$

for any $1 \le t \le T$ such that $\hat{T}_{j-1} + 1 \le t \le \hat{T}_j$ and $T_{l-1}^0 + 1 \le t \le T_l^0$ for some $1 \le j \le m+1$, $1 \le l \le m+1$. That is, $\hat{u}_t$ is the residual for the tth observation and d_t the difference between the fitted values and their true counterparts.
By definition, $\hat{\beta}(\{T_j^0\})$ and $\hat{\delta}(\{T_j^0\})$ minimize the sum of squared deviations

$$S_T(\{T_j^0\}, \beta, \delta) = \sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} (y_t - x_t'\beta - z_t'\delta_j)^2$$

with respect to β, δ. Therefore,

$$\tilde{S}_T(\{T_j^0\}) = S_T(\{T_j^0\}, \hat{\beta}(\{T_j^0\}), \hat{\delta}(\{T_j^0\})) \le S_T(\{T_j^0\}, \beta^0, \delta^0) = \sum_{t=1}^{T} u_t^2.$$

In addition, by definition $\{\hat{T}_j\}$ is the minimizer of $\tilde{S}_T(\{T_j\})$ on $B_{\varepsilon T,T}$. Since $\{T_j^0\} \in B_{\varepsilon T,T}$ due to our assumption that

$$\lambda_j^0 - \lambda_{j-1}^0 > \varepsilon$$

for any $1 \le j \le m+1$, it follows that

$$\sum_{t=1}^{T} \hat{u}_t^2 = \tilde{S}_T(\{\hat{T}_j\}) \le \tilde{S}_T(\{T_j^0\}) \le \sum_{t=1}^{T} u_t^2.$$

Note that $u_t - \hat{u}_t = d_t$ for any $1 \leq t \leq T$. By implication,

$$\sum_{t=1}^{T} u_t^2 \geq \sum_{t=1}^{T} \hat{u}_t^2 = \sum_{t=1}^{T} u_t^2 + \sum_{t=1}^{T} d_t^2 - 2 \cdot \sum_{t=1}^{T} u_t d_t,$$

so that

$$0 \leq \frac{1}{T} \sum_{t=1}^{T} d_t^2 \leq 2 \cdot \frac{1}{T} \sum_{t=1}^{T} u_t d_t.$$

An easy consequence is that

$$\left| \frac{1}{T} \sum_{t=1}^{T} u_t d_t \right| = \frac{1}{T} \sum_{t=1}^{T} u_t d_t.$$

We wll now show that this term converges to 0 in probability.

Define $\hat{Z}_1, \cdots, \hat{Z}_{m+1}$ as

$$\hat{Z}_j = \begin{pmatrix} z'_{\hat{T}_{j-1}+1} \\ \vdots \\ z'_{\hat{T}_j} \end{pmatrix}$$

for any $1 \leq j \leq m+1$, and let $\bar{Z}^* = \operatorname{diag}\left(\hat{Z}_1, \cdots, \hat{Z}_{m+1}\right)$. Let $Z_1^0, \cdots, Z_{m+1}^0$ and $\bar{Z}^0$ be the population counterparts.
Note that

$$\begin{aligned}
\sum_{t=1}^{T} u_t d_t &= \left(\sum_{t=1}^{T} u_t x_t' \right) (\hat{\beta} - \beta^0) + \sum_{j=1}^{m+1} \left(\sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} u_t z_t' \right) \hat{\delta}_j - \sum_{j=1}^{m+1} \left(\sum_{t=T_{j-1}^0+1}^{T_j^0} u_t z_t' \right) \delta_j^0 \\
&= U' X (\hat{\beta} - \beta^0) + U' \begin{pmatrix} \hat{Z}_1 \cdot \hat{\delta}_1 \\ \vdots \\ \hat{Z}_{m+1} \cdot \hat{\delta}_{m+1} \end{pmatrix} - U' \begin{pmatrix} Z_1^0 \cdot \delta_1 \\ \vdots \\ Z_{m+1}^0 \cdot \delta_{m+1} \end{pmatrix} \\
&= U' X (\hat{\beta} - \beta^0) + U' \bar{Z}^* \hat{\delta} - U' Z^0 \delta^0,
\end{aligned}$$

so we have

$$\frac{1}{T} \sum_{t=1}^{T} u_t d_t = \frac{1}{T} U' X (\hat{\beta} - \beta^0) + \frac{1}{T} U' \bar{Z}^* \hat{\delta} - \frac{1}{T} U' Z^0 \delta^0.$$

For any fixed $\{T_j\} \in B_{\varepsilon T,T}$, let $\bar{Z}$ be the diagonal partition of Z under $\{T_j\}$. Now we have the following results:

- $\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left|\frac{1}{T}U'X(\hat{\beta}(\{T_j\}) - \beta^0)\right| = o_p(1)$

  By definition,

$$\begin{aligned}\hat{\beta}(\{T_j\}) &= \left(X'M_{\bar{Z}}X\right)^{-1} X'M_{\bar{Z}}Y \\ &= \beta^0 + \left(X'M_{\bar{Z}}X\right)^{-1} X'M_{\bar{Z}}\bar{Z}^0\delta^0 + \left(X'M_{\bar{Z}}X\right)^{-1} X'M_{\bar{Z}}U.\end{aligned}$$

  It follows that

$$\frac{1}{T}U'X\left(\hat{\beta}(\{T_j\}) - \beta^0\right) = \frac{1}{T}U'X\left(X'M_{\bar{Z}}X\right)^{-1} X'M_{\bar{Z}}\bar{Z}^0\delta^0 + \frac{1}{T}U'X\left(X'M_{\bar{Z}}X\right)^{-1} X'M_{\bar{Z}}U.$$

  The first term above is majorized as

$$\left|\frac{1}{T}U'X\left(X'M_{\bar{Z}}X\right)^{-1} X'M_{\bar{Z}}\bar{Z}^0\delta^0\right| \le \left|\frac{1}{T}U'X\right| \cdot \left\|\left(\frac{X'M_{\bar{Z}}X}{T}\right)^{-1}\right\| \cdot \left\|\frac{X'M_{\bar{Z}}\bar{Z}^0}{T}\right\| \cdot \left|\delta^0\right|,$$

  so that

$$\begin{aligned}&\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left|\frac{1}{T}U'X\left(X'M_{\bar{Z}}X\right)^{-1} X'M_{\bar{Z}}\bar{Z}^0\delta^0\right| \\ &\qquad \le \left|\frac{1}{T}X'U\right| \cdot \left(\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left\|\left(\frac{X'M_{\bar{Z}}X}{T}\right)^{-1}\right\|\right) \cdot \left(\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left\|\frac{X'M_{\bar{Z}}\bar{Z}^0}{T}\right\|\right) \cdot \left|\delta^0\right|,\end{aligned}$$

  We already saw that all the terms on the right hand side are $O_p(1)$, except for $\left|\frac{1}{T}X'U\right|$. This term can be majorized as

$$\left|\frac{1}{T}X'U\right| = \left|\frac{1}{T}\sum_{t=1}^{T} x_t u_t\right| \le \left|\frac{1}{\sqrt{T}} S_T(1)\right|,$$

  and because

$$S_T(1) \xrightarrow{d} N\left(\mathbf{0}, \sigma^2 Q\right)$$

  by assumption (6) and the continuous mapping theorem, we can see that

$$\left|\frac{1}{T}X'U\right| = o_p(1).$$

  Therefore,

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left|\frac{1}{T}U'X\left(X'M_{\bar{Z}}X\right)^{-1} X'M_{\bar{Z}}\bar{Z}^0\delta^0\right| = o_p(1).$$

On the other hand, the second term is majorized as

$$\left|\frac{1}{T}U'X\left(X'M_{\bar{Z}}X\right)^{-1}X'M_{\bar{Z}}U\right| \leq \left|\frac{1}{T}U'X\right| \cdot \left\|\left(\frac{X'M_{\bar{Z}}X}{T}\right)^{-1}\right\| \cdot \left|\frac{X'M_{\bar{Z}}U}{T}\right|,$$

so that

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{1}{T}U'X\left(X'M_{\bar{Z}}X\right)^{-1}X'M_{\bar{Z}}U\right|$$
$$\leq \left|\frac{1}{T}X'U\right| \cdot \left(\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left\|\left(\frac{X'M_{\bar{Z}}X}{T}\right)^{-1}\right\|\right) \cdot \left(\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{X'M_{\bar{Z}}U}{T}\right|\right).$$

The first term is $o_p(1)$, and the second is $O_p(1)$. As for the third term,

$$\left|\frac{X'M_{\bar{Z}}U}{T}\right| \leq \left|\frac{1}{T}X'U\right| + \left|\frac{1}{T}X'P_{\bar{Z}}U\right|$$

for any $\{T_j\} \in B_{\varepsilon T,T}$, so that

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{X'M_{\bar{Z}}U}{T}\right| \leq \left|\frac{1}{T}X'U\right| + \frac{1}{\sqrt{T}}\left(\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{1}{\sqrt{T}}X'P_{\bar{Z}}U\right|\right);$$

both terms are $o_p(1)$, so

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{X'M_{\bar{Z}}U}{T}\right| = o_p(1).$$

It follows that

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{1}{T}U'X\left(X'M_{\bar{Z}}X\right)^{-1}X'M_{\bar{Z}}U\right| = o_p(1).$$

Putting these two results together,

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{1}{T}U'X\left(\hat{\beta}(\{T_j\}) - \beta^0\right)\right| = o_p(1).$$

- $\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left| \frac{1}{T} U' \left(\bar{Z}\hat{\delta}(\{T_j\}) - \bar{Z}^0\delta^0 \right) \right| = o_p(1)$

  Using the fact that

$$\begin{aligned}\hat{\delta}(\{T_j\}) &= \left(\bar{Z}'M_X\bar{Z}\right)^{-1}\bar{Z}'M_XY \\ &= \left(\bar{Z}'M_X\bar{Z}\right)^{-1}\bar{Z}'M_X\bar{Z}^0\delta^0 + \left(\bar{Z}'M_X\bar{Z}\right)^{-1}\bar{Z}'M_XU,\end{aligned}$$

  we can see that

$$\begin{aligned}&\frac{1}{T}U'\left(\bar{Z}\hat{\delta}(\{T_j\}) - \bar{Z}^0\delta^0\right) \\ &\qquad = \frac{1}{T}U'\bar{Z}\left(\bar{Z}'M_X\bar{Z}\right)^{-1}\bar{Z}'M_X\bar{Z}^0\delta^0 + \frac{1}{T}U'\bar{Z}\left(\bar{Z}'M_X\bar{Z}\right)^{-1}\bar{Z}'M_XU - \frac{1}{T}U'\bar{Z}^0\delta^0\end{aligned}$$

  Note that

$$\bar{W}'\bar{W} = \begin{pmatrix} X' \\ \bar{Z}' \end{pmatrix}\begin{pmatrix} X & \bar{Z} \end{pmatrix} = \begin{pmatrix} X'X & X'\bar{Z} \\ \bar{Z}'X & \bar{Z}'\bar{Z} \end{pmatrix},$$

  so that the lower (2,2) block matrix in $\bar{W}'\bar{W}$ is given in two different ways:

$$\left(\bar{Z}'M_X\bar{Z}\right)^{-1} = \left(\bar{Z}'\bar{Z}\right)^{-1} + \left(\bar{Z}'\bar{Z}\right)^{-1}\bar{Z}'X\left(X'M_{\bar{Z}}X\right)^{-1}X'\bar{Z}\left(\bar{Z}'\bar{Z}\right)^{-1}.$$

  This implies that

$$\bar{Z}\left(\bar{Z}'M_X\bar{Z}\right)^{-1}\bar{Z}'M_X = \left[I_T + P_{\bar{Z}}X\left(X'M_{\bar{Z}}X\right)^{-1}X'\right]P_{\bar{Z}}M_X,$$

  and as such that

$$\begin{aligned}&\frac{1}{T}U'\bar{Z}\left(\bar{Z}'M_X\bar{Z}\right)^{-1}\bar{Z}'M_X\left(\bar{Z}^0\delta^0 + M_XU\right) \\ &\qquad = \frac{1}{T}U'\left[I_T + P_{\bar{Z}}X\left(X'M_{\bar{Z}}X\right)^{-1}X'\right]P_{\bar{Z}}M_X\left(\bar{Z}^0\delta^0 + M_XU\right) \\ &= \left(\frac{1}{\sqrt{T}}U'P_{\bar{Z}}M_X + \frac{1}{\sqrt{T}}U'P_{\bar{Z}}X\left(X'M_{\bar{Z}}X\right)^{-1}X'P_{\bar{Z}}M_X\right)\left(\frac{1}{\sqrt{T}}\bar{Z}^0\cdot\delta^0 + M_X\cdot\frac{1}{\sqrt{T}}U\right).\end{aligned}$$

  Each term is majorized as follows:

$$\begin{aligned}\left|\frac{1}{\sqrt{T}}U'P_{\bar{Z}}M_X\right| &\le \frac{1}{\sqrt{T}}\left(\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|U'P_{\bar{Z}}\right|\right) \\ \left|\frac{1}{\sqrt{T}}U'P_{\bar{Z}}X\left(X'M_{\bar{Z}}X\right)^{-1}X'P_{\bar{Z}}M_X\right| &\le \frac{1}{\sqrt{T}}\left(\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|U'P_{\bar{Z}}\right|\right)\cdot\left\|\frac{1}{\sqrt{T}}X\right\|^2 \\ &\quad \times\left(\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left\|\left(\frac{X'M_{\bar{Z}}X}{T}\right)^{-1}\right\|\right)\end{aligned}$$

$$\left|\frac{1}{\sqrt{T}}\bar{Z}^0\cdot\delta^0\right| \leq \left\|\frac{1}{\sqrt{T}}\bar{Z}^0\right\|\cdot\left|\delta^0\right|$$
$$\left|M_X\cdot\frac{1}{\sqrt{T}}U\right| \leq \left|\frac{1}{\sqrt{T}}U\right|.$$

The only term whose rates of convergence are unknown is $\frac{1}{\sqrt{T}}U$. But this can be easily seen to be $O_p(1)$; for any $\delta > 0$,

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T}u_t^2-\sigma^2\right|>\delta\right) &\leq \frac{1}{\delta^2}\mathbb{E}\left|\frac{1}{T}\sum_{t=1}^{T}(u_t^2-\sigma^2)\right|^2 \\
&= \frac{1}{T^2\delta^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left[(u_t^2-\sigma^2)(u_s^2-\sigma^2)\right] \\
&= \frac{1}{T^2\delta^2}\sum_{t=1}^{T}\mathbb{E}\left[(u_t^2-\sigma^2)^2\right] \\
&\qquad (\{u_t^2-\sigma^2\}_{t\in\mathbb{Z}} \text{ is an MDS with respect to } \mathcal{F}) \\
&\leq \frac{1}{T^2\delta^2}\sum_{t=1}^{T}\mathbb{E}\left[u_t^4\right] \\
&\leq \frac{1}{T\delta^2}\left(\sup_{t\in\mathbb{Z}}\mathbb{E}\left[u_t^4\right]\right).
\end{aligned}$$

Since $\sup_{t\in\mathbb{Z}}\mathbb{E}\left[u_t^4\right] < +\infty$ by assumption,

$$\lim_{T\to\infty}\mathbb{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T}u_t^2-\sigma^2\right|>\delta\right)=0,$$

and because this holds for any $\delta > 0$,

$$\left|\frac{1}{\sqrt{T}}U\right|^2 = \frac{1}{T}\sum_{t=1}^{T}u_t^2 \xrightarrow{p} \sigma^2.$$

Therefore, $\left|\frac{1}{\sqrt{T}}U\right| = O_p(1)$, and we can conclude that

$$\begin{aligned}
\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{1}{\sqrt{T}}U'P_{\bar{Z}}M_X\right| &= o_p(1) \\
\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{1}{\sqrt{T}}U'P_{\bar{Z}}X\left(X'M_{\bar{Z}}X\right)^{-1}X'P_{\bar{Z}}M_X\right| &= o_p(1) \\
\left|\frac{1}{\sqrt{T}}\bar{Z}^0\cdot\delta^0\right| &= O_p(1) \\
\left|M_X\cdot\frac{1}{\sqrt{T}}U\right| &= O_p(1).
\end{aligned}$$

By implication,

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}}\left|\frac{1}{T}U'\bar{Z}\left(\bar{Z}'M_X\bar{Z}\right)^{-1}\bar{Z}'M_X\left(\bar{Z}^0\delta^0+M_XU\right)\right| = o_p(1).$$

Finally,

$$\left|\frac{1}{T}U'\bar{Z}^0\delta^0\right| \leq \left|\frac{1}{T}\sum_{t=1}^{T} z_t u_t\right| \cdot \left|\delta^0\right|.$$

Here,

$$\left|\frac{1}{T}\sum_{t=1}^{T} z_t u_t\right| \leq \frac{1}{\sqrt{T}}|S_T(1)|,$$

where $S_T(1) = O_p(1)$, so we have

$$\left|\frac{1}{T}U'\bar{Z}^0\delta^0\right| = o_p(1)$$

as well.

Putting all the pieces together,

$$\begin{aligned}
&\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left|\frac{1}{T}U'\left(\bar{Z}\hat{\delta}(\{T_j\}) - \bar{Z}^0\delta^0\right)\right| \\
&\qquad \leq \sup_{\{T_j\}\in B_{\varepsilon T,T}} \left|\frac{1}{T}U'\bar{Z}\left(\bar{Z}'M_X\bar{Z}\right)^{-1}\bar{Z}'M_X\left(\bar{Z}^0\delta^0 + M_X U\right)\right| + \left|\frac{1}{T}U'\bar{Z}^0\delta^0\right| = o_p(1).
\end{aligned}$$

Therefore, we are able to conclude that

$$\sup_{\{T_j\}\in B_{\varepsilon T,T}} \left| \frac{1}{T}U'X(\hat{\beta}(\{T_j\})-\beta^0)+\frac{1}{T}U'\bar{Z}\hat{\delta}(\{T_j\})-\frac{1}{T}U'Z^0\delta^0 \right| = o_p(1),$$

and because $\{\hat{T}_j\} \in B_{\varepsilon T,T}$, it follows that

$$\left|\frac{1}{T}\sum_{t=1}^{T} u_t d_t\right| = \left|\frac{1}{T}U'X(\hat{\beta}-\beta^0)+\frac{1}{T}U'\bar{Z}^*\hat{\delta}-\frac{1}{T}U'Z^0\delta^0\right| = o_p(1).$$

For any $\delta > 0$,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T} d_t^2\right| > \delta\right) &= \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T} d_t^2 > \delta\right)\\ &\leq \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T} u_t d_t > \delta\right) = \mathbb{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T} u_t d_t\right| > \delta\right),\end{aligned}$$

and because the rightmost term goes to 0 as $T \to \infty$ by the definition of an $o_p(1)$ process,

$$\lim_{T\to\infty} \mathbb{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T} d_t^2\right| > \delta\right) = 0.$$

This holds for any $\delta > 0$, so

$$\frac{1}{T}\sum_{t=1}^{T} d_t^2 \xrightarrow{p} 0.$$

6.4.2 Step 2: Deriving a Contradiction from Inconsistency

In this section we will use the convergence result proven above to show that, if

$$\min_{1\leq l\leq m}\left|\hat{\lambda}_l-\lambda_j^0\right|\xrightarrow{p}0$$

does not hold for some $1\leq j\leq m$, then we have a contradiction. To make the notation clearer, we explicitly state the dependence of the break fraction estimators $\hat{\lambda}_j$ on the sample size T, as $\hat{\lambda}_{T,j}$.

We again proceed in steps:

The Implication of Inconsistency

Suppose, for some $1\leq j\leq m$, that

$$\min_{1\leq l\leq m}\left|\hat{\lambda}_{T,l}-\lambda_j^0\right|\overset{p}{\not\to}0$$

By the definition of convergence in probability, there exists an $\eta>0$ such that $\eta<\frac{\varepsilon}{2}$ and

$$\mathbb{P}\left(\min_{1\leq l\leq m}\left|\hat{\lambda}_{T,l}-\lambda_j^0\right|>\eta\right)\not\to0,$$

as $T\to\infty$, and this in turn means that there exists a $\varepsilon_0>0$ such that, for any $\bar{T}\in N_+$, there exists a $T\geq\bar{T}$ such that

$$\mathbb{P}\left(\min_{1\leq l\leq m}\left|\hat{\lambda}_{T,l}-\lambda_j^0\right|>\eta\right)\geq\varepsilon_0.$$

This means that there exists a subsequence N of N_+ such that

$$\mathbb{P}\left(\min_{1\leq l\leq m}\left|\hat{\lambda}_{T,l}-\lambda_j^0\right|>\eta\right)\geq\varepsilon_0.$$

for any $T\in N$.

Characteristics of the Interval $[T(\lambda_j^0-\eta),T(\lambda_j^0+\eta)]$

We now investigate how large T must be for T_j^0 to be the only true break point to fall in $[T(\lambda_j^0-\eta),T(\lambda_j^0+\eta)]$.

Since we assumed $\eta<\frac{\varepsilon}{2}$,

$$\lambda_j^0-\eta>\lambda_j^0-\frac{\varepsilon}{2}>\lambda_{j-1}^0,$$

which in turn implies that

$$T\left(\lambda_j^0-\eta\right)>T\lambda_{j-1}^0\geq\lfloor T\lambda_{j-1}^0\rfloor=T_{j-1}^0.$$

Likewise,

$$\lambda_j^0+\eta<\lambda_j^0+\frac{\varepsilon}{2}<\lambda_{j+1}^0-\frac{\varepsilon}{2},$$

which implies that

$$T\left(\lambda_j^0+\eta\right)<T\cdot\lambda_{j+1}^0-T\frac{\varepsilon}{2};$$

for T such that $T\frac{\varepsilon}{2}>1$, or $T>\frac{2}{\varepsilon}$, we now have

$$T\left(\lambda_j^0+\eta\right)<T\cdot\lambda_{j+1}^0-1\leq T_{j+1}^0.$$

This shows us that

$$[T(\lambda_j^0-\eta),T(\lambda_j^0+\eta)]\subset[T_{j-1}^0+1,T_{j+1}^0]$$

if $T>\frac{2}{\varepsilon}$.
Since

$$T_j^0=\lfloor T\lambda_j^0\rfloor\in[T\lambda_j^0,T\lambda_j^0+1),$$

if $T\eta>1$, then

$$T\lambda_j^0+1<T\left(\lambda_j^0+\eta\right),$$

so that

$$T_j^0\in[T(\lambda_j^0-\eta),T(\lambda_j^0+\eta)].$$

Since $T>\frac{1}{\eta}$ implies $T>\frac{2}{\varepsilon}$, it follows that, for any $T>\frac{1}{\eta}$, T_j^0 is the only true break point to be contained in

$$[T(\lambda_j^0-\eta),T(\lambda_j^0+\eta)].$$

Deriving the Contradiction

For $T \in N$ such that $T > \frac{1}{\eta}$, let ω be an outcome contained in the event

$$\bigcap_{l=1}^{m}\{\left|\hat{\lambda}_{T,l}-\lambda_j^0\right|>\eta\}\in\mathcal{H}.$$

In this case,

$$\hat{\lambda}_{T,l}(\omega)\notin[\lambda_j^0-\eta,\lambda_j^0+\eta],$$

for any $1\le l\le m$, and because

$$\hat{\lambda}_{T,l}(\omega)=\frac{\hat{T}_{T,l}(\omega)}{T}$$

for each $1\le l\le m$, we have

$$\hat{T}_{T,l}(\omega)\notin[T(\lambda_j^0-\eta),T(\lambda_j^0+\eta)].$$

for any $1\le l\le m$. In other words, no estimated break point falls in the interval $[T(\lambda_j^0-\eta),T(\lambda_j^0+\eta)]$ under outcome ω.

Suppose that the above interval is contained in the ith regime under ω, so that

$$\hat{T}_{i-1}(\omega)+1\le T(\lambda_j^0-\eta)<T(\lambda_j^0+\eta)\le\hat{T}_i(\omega).$$

Then, for any $t\in[T(\lambda_j^0-\eta),T(\lambda_j^0+\eta)]$, since $t\le T\lambda_j^0$ implies $t\le T_j^0$ and $t\ge T\lambda_j^0+1$ implies $t>T_j^0$, we have

$$d_t(\omega)=\begin{cases}x_t(\omega)'(\hat{\beta}_T(\omega)-\beta^0)+z_t(\omega)'(\hat{\delta}_{T,i}(\omega)-\delta_j^0) & \text{if } t\in[T(\lambda_j^0-\eta),T\lambda_j^0]\\ x_t(\omega)'(\hat{\beta}_T(\omega)-\beta^0)+z_t(\omega)'(\hat{\delta}_{T,i}(\omega)-\delta_{j+1}^0) & \text{if } t\in[T\lambda_j^0+1,T(\lambda_j^0+\eta)]\end{cases}.$$

Denoting $[T(\lambda_j^0-\eta),T\lambda_j^0]=A_1$ and $[T\lambda_j^0+1,T(\lambda_j^0+\eta)]=A_2$, and suppressing the dependence on the outcome ω for notational brevity, we have

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}d_t^2&\ge\frac{1}{T}\sum_{t\in A_1}d_t^2+\frac{1}{T}\sum_{t\in A_2}d_t^2\\&=\eta\cdot\begin{pmatrix}\hat{\beta}_T-\beta^0\\\hat{\delta}_{T,i}-\delta_j^0\end{pmatrix}'\left[\frac{1}{T\eta}\sum_{t=T_j^0-T\eta}^{T_j^0}w_tw_t'\right]\begin{pmatrix}\hat{\beta}_T-\beta^0\\\hat{\delta}_{T,i}-\delta_j^0\end{pmatrix}\\&\quad+\eta\cdot\begin{pmatrix}\hat{\beta}_T-\beta^0\\\hat{\delta}_{T,i}-\delta_{j+1}^0\end{pmatrix}'\left[\frac{1}{T\eta}\sum_{t=T_j^0+1}^{T_j^0+T\eta}w_tw_t'\right]\begin{pmatrix}\hat{\beta}_T-\beta^0\\\hat{\delta}_{T,i}-\delta_{j+1}^0\end{pmatrix}.\end{aligned}$$

For any d-dimensional vector α and a conformable positive semidefinite matrix A, letting $A = PDP'$ be the eigendecomposition of A,

$$\begin{aligned}\alpha' A\alpha = \alpha' PDP'\alpha = \sum_{i=1}^{d} D_i \cdot (P'\alpha)_i^2 \\ \geq \left(\min_{1\leq i\leq d} D_i\right)\left(\sum_{i=1}^{d}(P'\alpha)_i^2\right) \\ = \left(\min_{1\leq i\leq d} D_i\right)\left|P'\alpha\right|^2 = \left(\min_{1\leq i\leq d} D_i\right)\cdot|\alpha|^2.\end{aligned}$$

Here, $\min_{1\leq i\leq d} D_i$ is the smallest eigenvalue of A.
By the identification condition for the break points (assumption (3)), given that $T\eta > l_0$, the minimum eigenvalues of

$$\frac{1}{T\eta}\sum_{t=T_j^0-T\eta}^{T_j^0} w_t w_t' \quad \text{and} \quad \frac{1}{T\eta}\sum_{t=T_j^0+1}^{T_j^0+T\eta} w_t w_t'$$

are bounded below by $\rho_{\min} > 0$. Therefore,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} d_t^2 \geq \eta\cdot\rho_{\min}\left[\left|\begin{pmatrix}\hat{\beta}_T-\beta^0\\ \hat{\delta}_{T,i}-\delta_j^0\end{pmatrix}\right|^2 + \left|\begin{pmatrix}\hat{\beta}_T-\beta^0\\ \hat{\delta}_{T,i}-\delta_{j+1}^0\end{pmatrix}\right|^2\right] \\ \geq \frac{1}{2}\eta\cdot\rho_{\min}\cdot\left|\begin{pmatrix}\hat{\beta}_T-\beta^0\\ \hat{\delta}_{T,i}-\delta_j^0\end{pmatrix} - \begin{pmatrix}\hat{\beta}_T-\beta^0\\ \hat{\delta}_{T,i}-\delta_{j+1}^0\end{pmatrix}\right|^2 \\ = \frac{1}{2}\eta\cdot\rho_{\min}\left|\delta_{j+1}^0-\delta_j^0\right|^2.\end{aligned}$$

Define $C = \frac{1}{2}\eta\cdot\rho_{\min} > 0$.

Since the above holds for any $T \in N$ such that $T > \max\left(\frac{1}{\eta}, \frac{l_0}{\eta}\right)$ on the set

$$\bigcap_{l=1}^{m}\{\left|\hat{\lambda}_{T,l}-\lambda_j^0\right| > \eta\},$$

it follows that

$$\mathbb{P}\left(\bigcap_{l=1}^{m}\{\left|\hat{\lambda}_{T,l}-\lambda_j^0\right| > \eta\right) \leq \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T} d_t^2 > C\left|\delta_{j+1}^0-\delta_j^0\right|^2\right).$$

Since

$$\bigcap_{l=1}^{m}\{\left|\hat{\lambda}_{T,l}-\lambda_j^0\right| > \eta\} = \{\min_{1\leq l\leq m}\left|\hat{\lambda}_{T,l}-\lambda_j^0\right| > \eta\}$$

and

$$\mathbb{P}\left(\min_{1\le l\le m}\left|\hat{\lambda}_{T,l}-\lambda_j^0\right|>\eta\right)\ge\varepsilon_0$$

for any $T\in N$, it follows that

$$\varepsilon_0\le\mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T}d_t^2>C\left|\delta_{j+1}^0-\delta_j^0\right|^2\right)$$

for any large enough T in the subsequence N of N_+. This contradicts the fact that

$$\frac{1}{T}\sum_{t=1}^{T}d_t^2\xrightarrow{p}0,$$

so it must be the case that

$$\min_{1\le l\le m}\left|\hat{\lambda}_{T,l}-\lambda_j^0\right|\xrightarrow{p}0$$

for any $1\le j\le m$.

6.4.3 Step 3: Matching Estimators with True Break Fractions

Finally, it remains to show that, for any $1 \leq j \leq m$, $\hat{\lambda}_j$ is the break fraction estimator that is consistent for λ_j^0. To make the proof simpler, we prove the result for the case where there are three breaks, that is, $m = 3$.

We showed above that, for any $1 \leq j \leq 3$,

$$\min_{1 \leq l \leq 3} \left| \hat{\lambda}_l - \lambda_j^0 \right| \xrightarrow{p} 0,$$

so that, by definition,

$$\lim_{T \to \infty} \mathbb{P}\left(\min_{1 \leq l \leq 3} \left| \hat{\lambda}_l - \lambda_j^0 \right| > \delta \right) = 0,$$

or equivalently,

$$\lim_{T \to \infty} \mathbb{P}\left(\min_{1 \leq l \leq 3} \left| \hat{\lambda}_l - \lambda_j^0 \right| \leq \delta \right) = 1$$

for any $\delta > 0$ and $1 \leq j \leq 3$. This property allows us to first prove the consistency of $\hat{\lambda}_1$, and then proceed forward until we reach $\hat{\lambda}_3$.

Preliminary Results

Defining

$$\left\{ \left| \hat{\lambda}_l - \lambda_j^0 \right| \leq \frac{\varepsilon}{2} \right\} = A_{T,lj}$$

for $1 \leq l, j \leq 3$, we will show that

$$\mathbb{P}(A_{T,lj}) \to 0$$

as $T \to \infty$ for any $1 \leq l, j \leq m$ such that $l > j$.

- $A_{T,31}$
  If $\left| \hat{\lambda}_3 - \lambda_1^0 \right| \leq \frac{\varepsilon}{2}$, then

$$\frac{\varepsilon}{2} \geq \left| \hat{\lambda}_3 - \lambda_1^0 \right| \geq \hat{\lambda}_3 - \lambda_1^0$$

  implies

$$\hat{\lambda}_3 \leq \lambda_1^0 + \frac{\varepsilon}{2} < \lambda_3^0 - \frac{3}{2}\varepsilon < \lambda_3^0.$$

This means that

$$\hat{\lambda}_1 < \hat{\lambda}_2 < \hat{\lambda}_3 < \lambda_3^0,$$

so that

$$\begin{aligned}\min_{1\le l\le 3}\left|\hat{\lambda}_l - \lambda_3^0\right| = \left|\hat{\lambda}_3 - \lambda_3^0\right| &\ge \left|\lambda_3^0 - \lambda_1^0\right| - \left|\hat{\lambda}_3 - \lambda_1^0\right| \\ &> 2\varepsilon - \frac{\varepsilon}{2} = \frac{3}{2}\varepsilon.\end{aligned}$$

Therefore,

$$A_{T,1} \subset \left\{\min_{1\le l\le 3}\left|\hat{\lambda}_l - \lambda_3^0\right| > \frac{3}{2}\varepsilon\right\},$$

and as such

$$\mathbb{P}\left(A_{T,1}\right) \le \mathbb{P}\left(\min_{1\le l\le 3}\left|\hat{\lambda}_l - \lambda_3^0\right| > \frac{3}{2}\varepsilon\right).$$

The right hand side goes to 0 as $T \to \infty$, so it follows that $\mathbb{P}\left(A_{T,1}\right) \to 0$ as well.

- $A_{T,32}$
  Now suppose that $\left|\hat{\lambda}_3 - \lambda_2^0\right| \le \frac{\varepsilon}{2}$. Then,

$$\frac{\varepsilon}{2} \ge \left|\hat{\lambda}_3 - \lambda_2^0\right| \ge \hat{\lambda}_3 - \lambda_2^0,$$

so that

$$\hat{\lambda}_3 \le \lambda_2^0 + \frac{\varepsilon}{2} < \lambda_3^0 - \frac{\varepsilon}{2} < \lambda_3^0.$$

As above, this implies that

$$\begin{aligned}\min_{1\le l\le 3}\left|\hat{\lambda}_l - \lambda_3^0\right| = \left|\hat{\lambda}_3 - \lambda_3^0\right| &\ge \left|\lambda_3^0 - \lambda_2^0\right| - \left|\hat{\lambda}_3 - \lambda_2^0\right| \\ &> \varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2},\end{aligned}$$

and as such,

$$\mathbb{P}\left(A_{T,2}\right) \le \mathbb{P}\left(\min_{1\le l\le 3}\left|\hat{\lambda}_l - \lambda_3^0\right| > \frac{\varepsilon}{2}\right).$$

Again, this implies $\mathbb{P}\left(A_{T,2}\right) \to 0$ as $T \to \infty$.

- $A_{T,21}$
  Suppose that $\left|\hat{\lambda}_2 - \lambda_1^0\right| < \frac{\varepsilon}{2}$. Then,

$$\frac{\varepsilon}{2} \geq \left|\hat{\lambda}_2 - \lambda_1^0\right| \geq \hat{\lambda}_2 - \lambda_1^0,$$

  so that

$$\hat{\lambda}_2 \leq \lambda_1^0 + \frac{\varepsilon}{2} < \lambda_2^0 - \frac{\varepsilon}{2} < \lambda_2^0.$$

  As above, this means that

$$\hat{\lambda}_1 < \hat{\lambda}_2 < \lambda_2^0,$$

  so that

$$\left|\hat{\lambda}_2 - \lambda_2^0\right| < \left|\hat{\lambda}_1 - \lambda_2^0\right|.$$

  Note that

$$\left|\hat{\lambda}_2 - \lambda_2^0\right| > \left|\lambda_2^0 - \lambda_1^0\right| - \left|\hat{\lambda}_2 - \lambda_1^0\right| > \varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2}.$$

  If, in addition, $\left|\hat{\lambda}_3 - \lambda_2^0\right| > \frac{\varepsilon}{2}$, then

$$\min_{1 \leq j \leq 3} \left|\hat{\lambda}_j - \lambda_2^0\right| = \min\left(\left|\hat{\lambda}_2 - \lambda_2^0\right|, \left|\hat{\lambda}_3 - \lambda_2^0\right|\right) > \frac{\varepsilon}{2}.$$

  Putting these results together, we can see that

$$A_{T,21} \subset \left\{\min_{1 \leq j \leq 3} \left|\hat{\lambda}_j - \lambda_2^0\right| > \frac{\varepsilon}{2}\right\} \cup A_{T,32},$$

  so that

$$\mathbb{P}\left(A_{T,21}\right) \leq \mathbb{P}\left(\min_{1 \leq j \leq 3} \left|\hat{\lambda}_j - \lambda_2^0\right| > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(A_{T,32}\right).$$

  We already saw that the last term on the right converges to 0, and the first term on the right also converges to 0 by the result derived in the previous section, so $\mathbb{P}\left(A_{T,21}\right) \to 0$ as $T \to \infty$.

The Consistency of $\hat{\lambda}_1$

Choose any $0 < \delta \leq \frac{\varepsilon}{2}$. We first state the following decomposition:

$$\mathbb{P}\left(\min_{1\leq j\leq 3}\left|\hat{\lambda}_j - \lambda_1^0\right| \leq \delta\right) = \sum_{j=1}^{3}\mathbb{P}\left(\left\{\left|\hat{\lambda}_j - \lambda_1^0\right| \leq \delta\right\}\cap\left\{\left|\hat{\lambda}_j - \lambda_1^0\right| = \min_{1\leq l\leq 3}\left|\hat{\lambda}_l - \lambda_1^0\right|\right\}\right).$$

For $j = 2,3$,

$$\mathbb{P}\left(\left\{\left|\hat{\lambda}_j - \lambda_1^0\right| \leq \delta\right\}\cap\left\{\left|\hat{\lambda}_j - \lambda_1^0\right| = \min_{1\leq l\leq 3}\left|\hat{\lambda}_l - \lambda_1^0\right|\right\}\right) \leq \mathbb{P}\left(\left|\hat{\lambda}_j - \lambda_1^0\right| \leq \frac{\varepsilon}{2}\right) = A_{T,j1},$$

where the right hand side goes to 0 as $T \to \infty$. Therefore,

$$\mathbb{P}\left(\left\{\left|\hat{\lambda}_j - \lambda_1^0\right| \leq \delta\right\}\cap\left\{\left|\hat{\lambda}_j - \lambda_1^0\right| = \min_{1\leq l\leq 3}\left|\hat{\lambda}_l - \lambda_1^0\right|\right\}\right) \to 0$$

as $T \to \infty$.
Since

$$\mathbb{P}\left(\min_{1\leq j\leq 3}\left|\hat{\lambda}_j - \lambda_1^0\right| \leq \delta\right) \to 1$$

as $T \to \infty$, together with the preceding result it must be the case that

$$\mathbb{P}\left(\left\{\left|\hat{\lambda}_1 - \lambda_1^0\right| \leq \delta\right\}\cap\left\{\left|\hat{\lambda}_1 - \lambda_1^0\right| = \min_{1\leq l\leq 3}\left|\hat{\lambda}_l - \lambda_1^0\right|\right\}\right) \to 1$$

as $T \to \infty$. Finally, because

$$\mathbb{P}\left(\left\{\left|\hat{\lambda}_1 - \lambda_1^0\right| \leq \delta\right\}\cap\left\{\left|\hat{\lambda}_1 - \lambda_1^0\right| = \min_{1\leq l\leq 3}\left|\hat{\lambda}_l - \lambda_1^0\right|\right\}\right) \leq \mathbb{P}\left(\left|\hat{\lambda}_1 - \lambda_1^0\right| \leq \delta\right),$$

and the left hand side goes to 1 as $T \to \infty$, we have

$$\mathbb{P}\left(\left|\hat{\lambda}_1 - \lambda_1^0\right| \leq \delta\right) \to 1$$

as $T \to \infty$.

If $\delta > \frac{\varepsilon}{2}$, then

$$\mathbb{P}\left(\left|\hat{\lambda}_1 - \lambda_1^0\right| \leq \frac{\varepsilon}{2}\right) \leq \mathbb{P}\left(\left|\hat{\lambda}_1 - \lambda_1^0\right| \leq \delta\right),$$

where the left hand side goes to 1 as $T \to \infty$, so

$$\mathbb{P}\left(\left|\hat{\lambda}_1 - \lambda_1^0\right| \leq \delta\right) \to 1$$

as $T \to \infty$.

We have seen that, for any $\delta > 0$,

$$\mathbb{P}\left(\left|\hat{\lambda}_1 - \lambda_1^0\right| \leq \delta\right) \to 1$$

as $T \to \infty$. Thus, by definition,

$$\hat{\lambda}_1 \xrightarrow{p} \lambda_1^0.$$

The Consistency of $\hat{\lambda}_2$

Choose any $0 < \delta \leq \frac{\varepsilon}{2}$. As above, we first state the decomposition

$$\mathbb{P}\left(\min_{1\leq j\leq 3}\left|\hat{\lambda}_j - \lambda_2^0\right| \leq \delta\right) = \sum_{j=1}^{3}\mathbb{P}\left(\left\{\left|\hat{\lambda}_j - \lambda_2^0\right| \leq \delta\right\} \cap \left\{\left|\hat{\lambda}_j - \lambda_2^0\right| = \min_{1\leq l\leq 3}\left|\hat{\lambda}_l - \lambda_2^0\right|\right\}\right).$$

As above,

$$\mathbb{P}\left(\left\{\left|\hat{\lambda}_3 - \lambda_2^0\right| \leq \delta\right\} \cap \left\{\left|\hat{\lambda}_3 - \lambda_2^0\right| = \min_{1\leq l\leq 3}\left|\hat{\lambda}_l - \lambda_2^0\right|\right\}\right) \leq \mathbb{P}\left(\left|\hat{\lambda}_3 - \lambda_2^0\right| \leq \frac{\varepsilon}{2}\right) = A_{T,32},$$

where the right hand side goes to 0 as $T \to \infty$. Therefore,

$$\mathbb{P}\left(\left\{\left|\hat{\lambda}_3 - \lambda_2^0\right| \leq \delta\right\} \cap \left\{\left|\hat{\lambda}_3 - \lambda_2^0\right| = \min_{1\leq l\leq 3}\left|\hat{\lambda}_l - \lambda_2^0\right|\right\}\right) \to 0$$

as $T \to \infty$.

If $\left|\hat{\lambda}_1 - \lambda_2^0\right| \leq \delta$, then

$$\left|\hat{\lambda}_1 - \lambda_1^0\right| \geq \left|\lambda_2^0 - \lambda_1^0\right| - \left|\hat{\lambda}_1 - \lambda_2^0\right| > \varepsilon - \delta \geq \frac{\varepsilon}{2},$$

so that

$$\mathbb{P}\left(\left|\hat{\lambda}_1 - \lambda_2^0\right| \leq \delta\right) \leq \mathbb{P}\left(\left|\hat{\lambda}_1 - \lambda_1^0\right| > \frac{\varepsilon}{2}\right).$$

By the consistency result proved above,

$$\mathbb{P}\left(\left|\hat{\lambda}_1 - \lambda_1^0\right| > \frac{\varepsilon}{2}\right) \to 0$$

as $T \to \infty$, so that

$$\mathbb{P}\left(\left|\hat{\lambda}_1 - \lambda_2^0\right| \leq \delta\right) \to 0$$

as $T \to \infty$ and therefore

$$\mathbb{P}\left(\left\{\left|\hat{\lambda}_1 - \lambda_2^0\right| \leq \delta\right\} \cap \left\{\left|\hat{\lambda}_1 - \lambda_2^0\right| = \min_{1 \leq l \leq 3} \left|\hat{\lambda}_l - \lambda_2^0\right|\right\}\right) \to 0$$

as well.

These results, together with the property that

$$\mathbb{P}\left(\min_{1 \leq j \leq 3} \left|\hat{\lambda}_j - \lambda_2^0\right| \leq \delta\right) \to 1$$

as $T \to \infty$, imply that

$$\lim_{T \to \infty} \mathbb{P}\left(\left\{\left|\hat{\lambda}_2 - \lambda_2^0\right| \leq \delta\right\} \cap \left\{\left|\hat{\lambda}_2 - \lambda_2^0\right| = \min_{1 \leq l \leq 3} \left|\hat{\lambda}_l - \lambda_2^0\right|\right\}\right) = 1.$$

Again, this allows us to deduce that

$$\lim_{T \to \infty} \mathbb{P}\left(\left|\hat{\lambda}_2 - \lambda_2^0\right| \leq \delta\right) = 1.$$

If $\delta > \frac{\varepsilon}{2}$, then

$$\mathbb{P}\left(\left|\hat{\lambda}_2 - \lambda_2^0\right| \leq \frac{\varepsilon}{2}\right) \leq \mathbb{P}\left(\left|\hat{\lambda}_2 - \lambda_2^0\right| \leq \delta\right),$$

where the left hand side goes to 1 as $T \to \infty$, so

$$\mathbb{P}\left(\left|\hat{\lambda}_2 - \lambda_2^0\right| \leq \delta\right) \to 1$$

as $T \to \infty$.

We have seen that, for any $\delta > 0$,

$$\mathbb{P}\left(\left|\hat{\lambda}_2 - \lambda_2^0\right| \leq \delta\right) \to 1$$

as $T \to \infty$. Thus, by definition,

$$\hat{\lambda}_2 \xrightarrow{p} \lambda_2^0.$$

The Consistency of $\hat{\lambda}_3$

Choose any $0 < \delta \leq \frac{\varepsilon}{2}$. As above, we first state the decomposition

$$\mathbb{P}\left(\min_{1\leq j\leq 3}\left|\hat{\lambda}_j - \lambda_3^0\right| \leq \delta\right) = \sum_{j=1}^{3}\mathbb{P}\left(\left\{\left|\hat{\lambda}_j - \lambda_3^0\right| \leq \delta\right\} \cap \left\{\left|\hat{\lambda}_j - \lambda_3^0\right| = \min_{1\leq l\leq 3}\left|\hat{\lambda}_l - \lambda_2^0\right|\right\}\right).$$

If $\left|\hat{\lambda}_2 - \lambda_3^0\right| \leq \delta$, then

$$\left|\hat{\lambda}_2 - \lambda_2^0\right| \geq \left|\lambda_3^0 - \lambda_2^0\right| - \left|\hat{\lambda}_2 - \lambda_3^0\right| > \varepsilon - \delta \geq \frac{\varepsilon}{2},$$

so that

$$\mathbb{P}\left(\left|\hat{\lambda}_2 - \lambda_3^0\right| \leq \delta\right) \leq \mathbb{P}\left(\left|\hat{\lambda}_2 - \lambda_2^0\right| > \frac{\varepsilon}{2}\right).$$

By the consistency result proved above,

$$\mathbb{P}\left(\left|\hat{\lambda}_2 - \lambda_2^0\right| > \frac{\varepsilon}{2}\right) \to 0$$

as $T \to \infty$, so that

$$\mathbb{P}\left(\left|\hat{\lambda}_2 - \lambda_3^0\right| \leq \delta\right) \to 0$$

as $T \to \infty$ and therefore

$$\mathbb{P}\left(\left\{\left|\hat{\lambda}_2 - \lambda_3^0\right| \leq \delta\right\} \cap \left\{\left|\hat{\lambda}_2 - \lambda_3^0\right| = \min_{1\leq l\leq 3}\left|\hat{\lambda}_l - \lambda_3^0\right|\right\}\right) \to 0$$

as well.
Likewise, the consistency of $\hat{\lambda}_1$ for λ_1^0 implies that

$$\mathbb{P}\left(\left\{\left|\hat{\lambda}_1 - \lambda_3^0\right| \leq \delta\right\} \cap \left\{\left|\hat{\lambda}_1 - \lambda_3^0\right| = \min_{1\leq l\leq 3}\left|\hat{\lambda}_l - \lambda_3^0\right|\right\}\right) \to 0.$$

By implication,

$$\mathbb{P}\left(\left\{\left|\hat{\lambda}_3 - \lambda_3^0\right| \leq \delta\right\} \cap \left\{\left|\hat{\lambda}_3 - \lambda_3^0\right| = \min_{1\leq l\leq 3}\left|\hat{\lambda}_l - \lambda_3^0\right|\right\}\right) \to 1,$$

which implies that

$$\lim_{T\to\infty}\mathbb{P}\left(\left|\hat{\lambda}_3 - \lambda_3^0\right| \leq \delta\right) = 1.$$

We can deal with the case $\delta > \frac{\varepsilon}{2}$ as above, so $\hat{\lambda}_3$ is consistent for λ_3^0.

6.5 The Rate of Convergence of the Break Date Estimators

So far, we have shown that the break fraction estimators are consistent for the true break fractions, that is, $\hat{\lambda}_j \xrightarrow{p} \lambda_j$ as $T \to \infty$ for any $1 \leq j \leq m$. We are now interested in the rate of convergence of the break dates; that is, the rate of convergence of

$$\hat{T}_j - T_j^0 = T\left(\hat{\lambda}_j - \lambda_j\right)$$

for each $1 \leq j \leq m$. It will turn out that the above quantity is $O_p(1)$, or equivalently, that $\hat{\lambda}_j - \lambda_j$ converges to 0 at the same rate as $\frac{1}{T}$.

As above, we prove the result for the case where there are three breaks, that is, when $m = 3$, and prove that $\hat{T}_2 - T_2^0$ is bounded in probability. The proofs for $\hat{T}_1 - T_1^0$ and $\hat{T}_3 - T_3^0$ will then be seen to be similar to that for $\hat{T}_2 - T_2^0$, so that we are done.
In addition, we will assume that the model is one of pure structural breaks ($p = 0$) in order to simplify the proof.
We once again proceed in steps.

6.5.1 Step 1: Obtaining a Probability Bound

We want to show that $\hat{T}_2 - T_2^0$ is $O_p(1)$; in other words, that for any $\eta > 0$, there exists a $C > 0$ and $\bar{T} \in N_+$ such that

$$\mathbb{P}\left(\left|\hat{T}_2 - T_2^0\right| > C\right) < \eta$$

for any $T \geq \bar{T}$. To this end, we first derive an upper bound for the probability $\mathbb{P}\left(\left|\hat{T}_2 - T_2^0\right| > C\right)$ for any $C > 0$.

For any $\zeta > 0$, define the set

$$V_\zeta = \left\{(T_1, T_2, T_3) \in B_{\varepsilon T, T} \middle| \left|T_j - T_j^0\right| \leq \zeta T \text{ for any } 1 \leq j \leq 3\right\}.$$

For any $1 \leq j \leq 3$, if $\left|\hat{\lambda}_j - \lambda_j^0\right| < \zeta$, then, because $T_j^0 = \lfloor T\lambda_j^0 \rfloor$,

$$\hat{T}_j - T_j^0 \leq T\hat{\lambda}_j - T\lambda_j^0 \leq \left|T(\hat{\lambda}_j - \lambda_j^0)\right| < T\zeta,$$

and

$$T_j^0 - \hat{T}_j \leq T\lambda_j^0 - T\hat{\lambda}_j + 1 \leq T\left(\left|\hat{\lambda}_j - \lambda_j^0\right| + \frac{1}{T}\right) < T\zeta$$

for large enough T, so that

$$\left|T_j^0 - \hat{T}_j\right| \leq \zeta T$$

for large enough T. Therefore, for large T,

$$\sum_{j=1}^{3}\mathbb{P}\left(\left|\hat{\lambda}_j-\lambda_j^0\right|\leq\zeta\right)-2\leq\mathbb{P}\left(\bigcap_{j=1}^{3}\left\{\left|\hat{\lambda}_j-\lambda_j^0\right|\leq\zeta\right\}\right)$$
$$\leq\mathbb{P}\left((\hat{T}_1,\hat{T}_2,\hat{T}_3)\in V_\zeta\right)$$

which implies that

$$\lim_{T\to\infty}\mathbb{P}\left((\hat{T}_1,\hat{T}_2,\hat{T}_3)\in V_\zeta\right)=1$$

by the consistency of the break fractions $\hat{\lambda}_1,\cdots,\hat{\lambda}_3$.

Now we can start constructing the upper bound. For any $C>0$, note that

$$\mathbb{P}\left(\left|\hat{T}_2-T_2^0\right|>C\right)=\mathbb{P}\left(\hat{T}_2-T_2^0>C\right)+\mathbb{P}\left(\hat{T}_2-T_2^0<-C\right).$$

Because the two terms on the right hand side are symmetric, we focus on the last term. Define

$$V_\varepsilon(C)=\left\{(T_1,T_2,T_3)\in B_{\varepsilon T,T}\middle|\left|T_j-T_j^0\right|\leq\zeta T,\quad T_2-T_2^0<-C\right\}.$$

Clearly, $V_\zeta(C)\subset V_\zeta$, and

$$\mathbb{P}\left(\hat{T}_2-T_2^0<-C\right)\leq\mathbb{P}\left(\{\hat{T}_2-T_2^0<-C\}\cap\{(\hat{T}_1.\hat{T}_2,\hat{T}_3)\in V_\zeta\}\right)+\mathbb{P}\left(\{\hat{T}_2-T_2^0<-C\}\cap\{(\hat{T}_1.\hat{T}_2,\hat{T}_3)\notin V_\zeta\}\right)$$
$$\leq\mathbb{P}\left((\hat{T}_1.\hat{T}_2,\hat{T}_3)\in V_\zeta(C)\right)+\mathbb{P}\left((\hat{T}_1.\hat{T}_2,\hat{T}_3)\notin V_\zeta\right).$$

We focus on bounding the first term. Recall that $\tilde{S}_T(T_1,T_2,T_3)$ is the least squares sum of squared residuals given the break points $(T_1,T_2,T_3)\in B_{\varepsilon T,T}$. By definition, the estimators $(\hat{T}_1,\hat{T}_2,\hat{T}_3)$ minimize this SSR, so

$$\tilde{S}_T(\hat{T}_1,\hat{T}_2,\hat{T}_3)\leq\tilde{S}_T(\hat{T}_1,T_2^0,\hat{T}_3),$$

or equivalently,

$$\tilde{S}_T(\hat{T}_1,\hat{T}_2,\hat{T}_3)-\tilde{S}_T(\hat{T}_1,T_2^0,\hat{T}_3)\leq 0.$$

If $(\hat{T}_1.\hat{T}_2,\hat{T}_3)\in V_\zeta(C)$, then $\hat{T}_2-T_2^0<-C<0$ and

$$\frac{\tilde{S}_T(\hat{T}_1,\hat{T}_2,\hat{T}_3)-\tilde{S}_T(\hat{T}_1,T_2^0,\hat{T}_3)}{T_2^0-\hat{T}_2}\leq 0,$$

which implies that

$$\min_{(T_1,T_2,T_3)\in V_\zeta(C)}\frac{\tilde{S}_T(T_1,T_2,T_3)-\tilde{S}_T(T_1,T_2^0,T_3)}{T_2^0-T_2}\leq 0.$$

Therefore,

$$\mathbb{P}\left((\hat{T}_1.\hat{T}_2,\hat{T}_3)\in V_\zeta(C)\right)\le \mathbb{P}\left(\min_{(T_1,T_2,T_3)\in V_\zeta(C)} \frac{\tilde{S}_T(T_1,T_2,T_3)-\tilde{S}_T(T_1,T_2^0,T_3)}{T_2^0-T_2}\le 0\right),$$

and we can bound $\mathbb{P}\left(\hat{T}_2-T_2^0<-C\right)$ above by

$$\mathbb{P}\left(\hat{T}_2-T_2^0<-C\right)\le \mathbb{P}\left(\min_{(T_1,T_2,T_3)\in V_\zeta(C)} \frac{\tilde{S}_T(T_1,T_2,T_3)-\tilde{S}_T(T_1,T_2^0,T_3)}{T_2^0-T_2}\le 0\right)+\mathbb{P}\left((\hat{T}_1.\hat{T}_2,\hat{T}_3)\notin V_\zeta\right).$$

From the result derived above,

$$\lim_{T\to\infty}\mathbb{P}\left((\hat{T}_1.\hat{T}_2,\hat{T}_3)\notin V_\zeta\right)=0,$$

so there exists a $\bar{T}\in N_+$ such that

$$\mathbb{P}\left((\hat{T}_1.\hat{T}_2,\hat{T}_3)\notin V_\zeta\right)<\frac{\eta}{4}$$

for any $T\ge\bar{T}$.
We need only show that there exists a $C>0$ such that

$$\mathbb{P}\left(\min_{(T_1,T_2,T_3)\in V_\zeta(C)} \frac{\tilde{S}_T(T_1,T_2,T_3)-\tilde{S}_T(T_1,T_2^0,T_3)}{T_2^0-T_2}\le 0\right)<\frac{\eta}{4}$$

for large T to conclude that

$$\mathbb{P}\left(\hat{T}_2-T_2^0<-C\right)<\frac{\eta}{2}$$

for some $C>0$ and large T.

6.5.2 Step 2: Decomposing the Difference of SSRs

For any $C > 0$, choose any $(T_1, T_2, T_3) \in V_\zeta(C)$, and denote $\lambda_j = \frac{T_j}{T}$ for $j = 1, 2, 3$. Define

$$\begin{aligned} SSR_1 &= \tilde{S}_T(T_1, T_2, T_3) \\ SSR_2 &= \tilde{S}_T(T_1, T_2^0, T_3) \\ SSR_3 &= \tilde{S}_T(T_1, T_2, T_2^0, T_3). \end{aligned}$$

Then, we can write

$$\begin{aligned} \tilde{S}_T(T_1, T_2, T_3) - \tilde{S}_T(T_1, T_2^0, T_3) &= SSR_1 - SSR_2 \\ &= (SSR_1 - SSR_3) - (SSR_2 - SSR_3). \end{aligned}$$

Note that each term that comprises the rightmost term is the difference in SSRs of a model with 3 breaks and one with an additional break. Thus, the problem reduces to one of comparing two models, one a restricted version of the other.

Asymptotic Properties of the Model with an Additional Break

Under the break points T_1, T_2, T_2^0, T_3, the model can be written as

$$Y = \bar{Z}^* \delta^{0*} + U,$$

where $\bar{Z}^*$ is the diagonal partition of Z in accordance to T_1, T_2, T_2^0, T_3, that is,

$$\bar{Z}^* = \operatorname{diag}(Z_1^*, Z_2^*, Z_\Delta^*, Z_3^*, Z_4^*)$$

for

$$Z_1^* = \begin{pmatrix} z_1' \\ \vdots \\ z_{T_1}' \end{pmatrix}, \quad Z_2^* = \begin{pmatrix} z_{T_1+1}' \\ \vdots \\ z_{T_2}' \end{pmatrix}, \quad Z_\Delta^* = \begin{pmatrix} z_{T_2+1}' \\ \vdots \\ z_{T_2^0}' \end{pmatrix}, \quad Z_3^* = \begin{pmatrix} z_{T_2^0+1}' \\ \vdots \\ z_{T_3}' \end{pmatrix}, \quad Z_4^* = \begin{pmatrix} z_{T_3+1}' \\ \vdots \\ z_T. \end{pmatrix},$$

and $\delta^{0*} = (\delta_1^{0\prime}, \delta_2^{0\prime}, \delta_2^{0\prime}, \delta_3^{0\prime}, \delta_4^{0\prime})'$, where δ_2^0 is repeated once because T_2 is categorized into the second regime in the true model.
The estimator $\hat{\delta}^* = (\hat{\delta}_1^{*\prime}, \hat{\delta}_2^{*\prime}, \hat{\delta}_\Delta^{*\prime}, \hat{\delta}_3^{*\prime}, \hat{\delta}_4^{*\prime})'$ of δ^* satisfies

$$\hat{\delta}^* = (\bar{Z}^{*\prime} \bar{Z}^*)^{-1} \bar{Z}^{*\prime} Y.$$

Let $\bar{Z}^{0*} = \operatorname{diag}(Z_1^0, Z_2^0, Z_\Delta^0, Z_3^0, Z_4^0)$ be the diagonal partition of Z according to the true break dates with T_2 between the first and second true break dates, specifically the break dates T_1^0, T_2, T_2^0, T_3^0. We choose $\zeta > 0$ small enough so that $T_1^0 < T_2$. In this case, the true model is written as

$$Y = \bar{Z}^{0*}\delta^{0*} + U,$$

and we can expand $\hat{\delta}^*$ as

$$\begin{aligned}
\hat{\delta}^* &= (\bar{Z}^{*\prime}\bar{Z}^*)^{-1}\bar{Z}^{*\prime}Y \\
&= (\bar{Z}^{*\prime}\bar{Z}^*)^{-1}\bar{Z}^{*\prime}\bar{Z}^{0*}\cdot\delta^{0*} + (\bar{Z}^{*\prime}\bar{Z}^*)^{-1}\bar{Z}^{*\prime}U \\
&= \delta^{0*} + \left(\frac{1}{T}\bar{Z}^{*\prime}\bar{Z}^*\right)^{-1}\left(\frac{1}{T}\bar{Z}^{*\prime}U\right) \\
&\quad + \left(\frac{1}{T}\bar{Z}^{*\prime}\bar{Z}^*\right)^{-1}\left[\frac{1}{T}\bar{Z}^{*\prime}\left(\bar{Z}^{0*} - \bar{Z}^*\right)\right].
\end{aligned}$$

$\frac{1}{T}\bar{Z}^{*\prime}U$ is $o_p(1)$ because

$$\frac{1}{T}\sum_{t=\lfloor Ts\rfloor+1}^{\lfloor Tr\rfloor} z_t u_t = \frac{1}{\sqrt{T}}\left(V_T(r) - V_T(s)\right) \xrightarrow{p} \mathbf{0}$$

for any $0 \le s < r \le 1$ by assumption (6), and $\left(\frac{1}{T}\bar{Z}^{*\prime}\bar{Z}^*\right)^{-1}$ is $O_p(1)$ because

$$\left(\frac{1}{T}\sum_{t=\lfloor Ts\rfloor+1}^{\lfloor Tr\rfloor} z_t z_t'\right)^{-1} \xrightarrow{p} \left(\Omega(r) - \Omega(s)\right)^{-1}$$

for any $0 \le s < r \le 1$.

In addition, we can see that

$$\begin{pmatrix}
\min(T_1, T_1^0)\{ & O & O & O & O & O \\
|T_1 - T_1^0|\{ & \mathbf{x} & -\mathbf{x} & O & O & O \\
T_2 - \max(T_1, T_1^0)\Big\{ & O & O & O & O & O \\
T_2^0 - T_2\{ & O & O & O & O & O \\
\min(T_3^0 - T_2^0, T_3 - T_2)\Big\{ & O & O & O & O & O \\
|T_3^0 - T_3|\{ & O & O & O & \mathbf{y} & -\mathbf{y} \\
T - \max(T_3^0, T_3)\{ & O & O & O & O & O
\end{pmatrix} = \bar{Z}^{0*} - \bar{Z}^*,$$

where the values on the left hand side refer to the number of rows associated with each row of blocks and $\mathbf{x}$, $\mathbf{y}$ denote non-zero elements. Therefore,

$$\left\|\frac{1}{\sqrt{T}}\left(\bar{Z}^{0*} - \bar{Z}^*\right)\right\|^2 \le 2\left[\frac{1}{T}\sum_{t=\min(T_1,T_1^0)+1}^{\max(T_1,T_1^0)} z_t' z_t + \frac{1}{T}\sum_{t=\min(T_3,T_3^0)+1}^{\max(T_3,T_3^0)} z_t' z_t\right].$$

Since the first term on the right hand side can be written as

$$\left|\lambda_1^0 - \lambda_1\right| \cdot \left(\frac{1}{\left|T_1^0 - T_1\right|} \sum_{t=\min(T_1,T_1^0)+1}^{\max(T_1,T_1^0)} z_t' z_t \right),$$

where the latter is $O_p(1)$ and $|\lambda_1^0 - \lambda_1| = \frac{1}{T}|T_1^0 - T_1| \leq \zeta$, and the same holds for the second term, we can see that

$$\left\| \frac{1}{\sqrt{T}} \left(\bar{Z}^{0*} - \bar{Z}^* \right) \right\|^2 = \zeta \cdot O_p(1).$$

Therefore, we can see that

$$\begin{aligned}\left|\hat{\delta}^* - \delta^{0*}\right| &\leq \left\| \left(\frac{1}{T} \bar{Z}^{*\prime} \bar{Z}^* \right)^{-1} \right\| \cdot \left| \frac{1}{T} \bar{Z}^{*\prime} U \right| + \left\| \left(\frac{1}{T} \bar{Z}^{*\prime} \bar{Z}^* \right)^{-1} \right\| \cdot \left\| \frac{1}{T} \bar{Z}^{*\prime} \bar{Z}^* \right\|^{\frac{1}{2}} \cdot \left\| \frac{1}{\sqrt{T}} \left(\bar{Z}^{0*} - \bar{Z}^* \right) \right\| \\ &= o_p(1) + \sqrt{\zeta} \cdot O_p(1) = \sqrt{\zeta} \cdot O_p(1).\end{aligned}$$

In other words, for small $\zeta > 0$, large $C > 0$ (the large C is so that $\left(\frac{1}{T} Z_\Delta^{*\prime} Z_\Delta^* \right)^{-1}$ is not too large) and large T, $\hat{\delta}_j^*$ is close to δ_j^0 for $j = 1, 2, \Delta, 3, 4$.

Difference in SSRs via Restricted Linear Regression

The model with break points T_1, T_2, T_3 can be viewed as a special case of the above unrestricted model subject to the restriction $\delta_\Delta = \delta_3$. In other words, $SSR(T_1, T_2, T_2^0, T_3)$ and $SSR(T_1, T_2, T_3)$ are the sum of squared residuals from the unrestricted model above and the model subject to the restriction

$$R\delta^* = \mathbf{0},$$

where R is an $k \times 5k$ full rank matrix defined as

$$R = \begin{pmatrix} O & O & I_k & -I_k & O \end{pmatrix}.$$

The estimator of δ^* subject to the restriction $R\delta^* = \mathbf{0}$ solves the constrained minimization problem

$$\begin{aligned} \min_{\delta^* \in \mathbb{R}^{5k}} \quad & \frac{1}{2}\left(Y - \bar{Z}^*\delta^*\right)'\left(Y - \bar{Z}^*\delta^*\right) \\ \text{subject to} \quad & R\delta^* = \mathbf{0}. \end{aligned}$$

The Lagrangian for this problem is

$$\mathcal{L} = \frac{1}{2}\left(Y - \bar{Z}^*\delta^*\right)'\left(Y - \bar{Z}^*\delta^*\right) - \mu' R\delta^*,$$

where μ is a k-dimensional vector of Lagrangian multipliers. Letting $\tilde{\delta}^*$ be the estimator of δ under the restricted model, we have the f.o.c.

$$\bar{Z}^{*\prime}\left(Y - \bar{Z}^*\tilde{\delta}^*\right) = R'\mu,$$

so that

$$\tilde{\delta}^* = \hat{\delta}^* - \left(\bar{Z}^{*\prime}\bar{Z}^*\right)^{-1} R'\mu.$$

Premultiplying the f.o.c. above by $R\left(\bar{Z}^{*\prime}\bar{Z}^*\right)^{-1}$, we can see that

$$\begin{aligned} R\left(\bar{Z}^{*\prime}\bar{Z}^*\right)^{-1} R'\mu &= R\left(\bar{Z}^{*\prime}\bar{Z}^*\right)^{-1} \bar{Z}^{*\prime}Y - R\tilde{\delta}^* \\ &= R\left(\bar{Z}^{*\prime}\bar{Z}^*\right)^{-1} \bar{Z}^{*\prime}Y, \end{aligned}$$

since $R\tilde{\delta}^* = \mathbf{0}$, and as such that

$$\mu = \left[R\left(\bar{Z}^{*\prime}\bar{Z}^*\right)^{-1} R'\right]^{-1} R\hat{\delta}^*.$$

Therefore,

$$\tilde{\delta}^* = \hat{\delta}^* - \left(\bar{Z}^{*\prime}\bar{Z}^*\right)^{-1} R' \left[R\left(\bar{Z}^{*\prime}\bar{Z}^*\right)^{-1} R'\right]^{-1} R\hat{\delta}^*.$$

Using this formula, we can express $SSR_1 - SSR_3$ as

$$\begin{aligned} SSR_1 - SSR_3 &= \left(Y - \bar{Z}^*\tilde{\delta}^*\right)'\left(Y - \bar{Z}^*\tilde{\delta}^*\right) - \left(Y - \bar{Z}^*\hat{\delta}^*\right)'\left(Y - \bar{Z}^*\hat{\delta}^*\right) \\ &= \tilde{\delta}^{*\prime}\bar{Z}^{*\prime}\bar{Z}^*\tilde{\delta}^* - \hat{\delta}^{*\prime}\bar{Z}^{*\prime}\bar{Z}^*\hat{\delta}^* + 2Y'\bar{Z}^*\left(\hat{\delta}^* - \tilde{\delta}^*\right) \\ &= \hat{\delta}^{*\prime}R'\left[R\left(\bar{Z}^{*\prime}\bar{Z}^*\right)^{-1} R'\right]^{-1} R\hat{\delta}^*. \end{aligned}$$

Under the R defined above,

$$R\hat{\delta}^* = \hat{\delta}^*_\Delta - \hat{\delta}^*_3,$$

and because

$$\left(\bar{Z}^{*\prime}\bar{Z}^*\right)^{-1} = \begin{pmatrix} (Z_1^{*\prime}Z_1^*)^{-1} & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & (Z_4^{*\prime}Z_4^*)^{-1} \end{pmatrix},$$

we can see that

$$\left[R\left(\bar{Z}^{*\prime}\bar{Z}^*\right)^{-1} R'\right]^{-1} = \left[(Z_\Delta^{*\prime}Z_\Delta^*)^{-1} + (Z_3^{*\prime}Z_3^*)^{-1}\right]^{-1}.$$

Therefore, the difference in sum of squares can be written even more compactly as

$$SSR_1 - SSR_3 = \left(\hat{\delta}^*_3 - \hat{\delta}^*_\Delta\right)'\left[(Z_\Delta^{*\prime}Z_\Delta^*)^{-1} + (Z_3^{*\prime}Z_3^*)^{-1}\right]^{-1}\left(\hat{\delta}^*_3 - \hat{\delta}^*_\Delta\right).$$

By a symmetric argument, because SSR_2 is the model with break points T_1, T_2^0, T_3, we have

$$SSR_2 - SSR_3 = \left(\hat{\delta}^*_2 - \hat{\delta}^*_\Delta\right)'\left[(Z_\Delta^{*\prime}Z_\Delta^*)^{-1} + (Z_2^{*\prime}Z_2^*)^{-1}\right]^{-1}\left(\hat{\delta}^*_2 - \hat{\delta}^*_\Delta\right).$$

6.5.3 Step 3: Rate of Convergence of the Difference in SSRs

From the result above, we can see that

$$\begin{aligned}\tilde{S}_T(T_1,T_2,T_3)-\tilde{S}_T(T_1,T_2^0,T_3) &= (SSR_1-SSR_3)-(SSR_2-SSR_3)\\ &= \left(\hat{\delta}_3^*-\hat{\delta}_\Delta^*\right)'\left[(Z_\Delta^{*\prime}Z_\Delta^*)^{-1}+(Z_3^{*\prime}Z_3^*)^{-1}\right]^{-1}\left(\hat{\delta}_3^*-\hat{\delta}_\Delta^*\right)\\ &\quad -\left(\hat{\delta}_2^*-\hat{\delta}_\Delta^*\right)'\left[(Z_\Delta^{*\prime}Z_\Delta^*)^{-1}+(Z_2^{*\prime}Z_2^*)^{-1}\right]^{-1}\left(\hat{\delta}_2^*-\hat{\delta}_\Delta^*\right).\end{aligned}$$

Since

$$\left(\hat{\delta}_2^*-\hat{\delta}_\Delta^*\right)'\left[(Z_\Delta^{*\prime}Z_\Delta^*)^{-1}+(Z_2^{*\prime}Z_2^*)^{-1}\right]^{-1}\left(\hat{\delta}_2^*-\hat{\delta}_\Delta^*\right)\geq\left(\hat{\delta}_2^*-\hat{\delta}_\Delta^*\right)'Z_\Delta^{*\prime}Z_\Delta^*\left(\hat{\delta}_2^*-\hat{\delta}_\Delta^*\right),$$

we can further bound the above quantity below by

$$\begin{aligned}&\tilde{S}_T(T_1,T_2,T_3)-\tilde{S}_T(T_1,T_2^0,T_3)\\ &\quad\geq\left(\hat{\delta}_3^*-\hat{\delta}_\Delta^*\right)'\left[(Z_\Delta^{*\prime}Z_\Delta^*)^{-1}+(Z_3^{*\prime}Z_3^*)^{-1}\right]^{-1}\left(\hat{\delta}_3^*-\hat{\delta}_\Delta^*\right)-\left(\hat{\delta}_2^*-\hat{\delta}_\Delta^*\right)'Z_\Delta^{*\prime}Z_\Delta^*\left(\hat{\delta}_2^*-\hat{\delta}_\Delta^*\right).\end{aligned}$$

Note that

$$\begin{aligned}\frac{1}{T_2^0-T_2}\left[(Z_\Delta^{*\prime}Z_\Delta^*)^{-1}+(Z_3^{*\prime}Z_3^*)^{-1}\right]^{-1} &= \left[\left(\frac{Z_\Delta^{*\prime}Z_\Delta^*}{T_2^0-T_2}\right)^{-1}+\left(\frac{Z_3^{*\prime}Z_3^*}{T_2^0-T_2}\right)^{-1}\right]^{-1}\\ &= \left[\left(\frac{Z_\Delta^{*\prime}Z_\Delta^*}{T_2^0-T_2}\right)^{-1}+\frac{T_2^0-T_2}{T_3-T_2^0}\cdot\left(\frac{Z_3^{*\prime}Z_3^*}{T_3-T_2^0}\right)^{-1}\right]^{-1}.\end{aligned}$$

The eigenvalues of

$$\frac{Z_\Delta^{*\prime}Z_\Delta^*}{T_2^0-T_2}=\frac{1}{T_2^0-T_2}\sum_{t=T_2+1}^{T_2^0}z_tz_t'$$

are bounded below by $\rho_{\min}$, so that the eigenvalues of

$$\left(\frac{Z_\Delta^{*\prime}Z_\Delta^*}{T_2^0-T_2}\right)^{-1}$$

are bounded above by $\rho_{\min}^{-1}$. Likewise, the eigenvalues of

$$\frac{T_2^0-T_2}{T_3-T_2^0}\cdot\left(\frac{Z_3^{*\prime}Z_3^*}{T_3-T_2^0}\right)^{-1}$$

are bounded above by $\rho_{\min}^{-1}\cdot\frac{T_2^0-T_2}{T_3-T_2^0}$, and because two positive definite matrices are simultaneously diagonalizable, this means that the eigenvalues of

$$\left(\frac{Z_\Delta^{*\prime}Z_\Delta^*}{T_2^0-T_2}\right)^{-1}+\frac{T_2^0-T_2}{T_3-T_2^0}\cdot\left(\frac{Z_3^{*\prime}Z_3^*}{T_3-T_2^0}\right)^{-1}$$

are bounded above by

$$\rho_{\min}^{-1}\left(\frac{T_2^0-T_2}{T_3-T_2^0}+1\right)=\rho_{\min}^{-1}\cdot\frac{T_3-T_2}{T_3-T_2^0}.$$

In other words, the eigenvalues of

$$\frac{1}{T_2^0-T_2}\left[\left(Z_\Delta^{*\prime}Z_\Delta^*\right)^{-1}+\left(Z_3^{*\prime}Z_3^*\right)^{-1}\right]^{-1}$$

are bounded below by

$$\rho_{\min}\cdot\frac{T_3-T_2^0}{T_3-T_2}.$$

As such,

$$\begin{aligned}&\frac{\tilde{S}_T(T_1,T_2,T_3)-\tilde{S}_T(T_1,T_2^0,T_3)}{T_2^0-T_2}\\&\qquad\geq\rho_{\min}\frac{T_3-T_2^0}{T_3-T_2}\cdot\left|\hat{\delta}_3^*-\hat{\delta}_\Delta^*\right|^2-\left(\hat{\delta}_2^*-\hat{\delta}_\Delta^*\right)'\frac{Z_\Delta^{*\prime}Z_\Delta^*}{T_2^0-T_2}\left(\hat{\delta}_2^*-\hat{\delta}_\Delta^*\right).\end{aligned}$$

We saw above that

$$\hat{\delta}_j^*-\delta_j^0=\sqrt{\zeta}\cdot O_p(1),$$

for any $j=1,2,\Delta,3,4$, where $\delta_\Delta^0=\delta_2^0$. In addition,

$$\frac{Z_\Delta^{*\prime}Z_\Delta^*}{T_2^0-T_2}=O_p(1)$$

by assumption, so that

$$\left(\hat{\delta}_2^*-\hat{\delta}_\Delta^*\right)'\frac{Z_\Delta^{*\prime}Z_\Delta^*}{T_2^0-T_2}\left(\hat{\delta}_2^*-\hat{\delta}_\Delta^*\right)=\zeta\cdot O_p(1).$$

Finally,

$$T_3-T_2^0=\left|T_3-T_2^0\right|\geq T_3^0-T_2^0-\left|T_3-T_3^0\right|\geq\left(T_3^0-T_2^0\right)-\zeta T$$

and

$$\begin{aligned}T_3-T_2&\leq\left|T_3-T_3^0\right|+\left|T_2-T_2^0\right|+\left(T_3^0-T_2^0\right)\\&\leq\left(T_3^0-T_2^0\right)+2\zeta T,\end{aligned}$$

so that

$$\frac{T_3 - T_2^0}{T_3 - T_2} \geq \frac{(\lambda_3^0 - \lambda_2^0) - \zeta}{(\lambda_3^0 - \lambda_2^0) + 2\zeta}$$

and

$$\rho_{\min} \frac{T_3 - T_2^0}{T_3 - T_2} \cdot \left|\hat{\delta}_3^* - \hat{\delta}_\Delta^*\right|^2 \geq \rho_{\min} \cdot \frac{(\lambda_3^0 - \lambda_2^0) - \zeta}{(\lambda_3^0 - \lambda_2^0) + 2\zeta} \cdot \left|\delta_3^0 - \delta_2^0\right|^2 + \zeta \cdot O_p(1).$$

Putting all the pieces together, we have

$$\frac{\tilde{S}_T(T_1, T_2, T_3) - \tilde{S}_T(T_1, T_2^0, T_3)}{T_2^0 - T_2} \geq \rho_{\min} \cdot \frac{(\lambda_3^0 - \lambda_2^0) - \zeta}{(\lambda_3^0 - \lambda_2^0) + 2\zeta} \cdot \left|\delta_3^0 - \delta_2^0\right|^2 + 2 \cdot \zeta \cdot O_p(1).$$

The first term does not depend on T and very little on ζ, provided that it is small. Thus, for small ζ, large C and large T, there is a large probability that

$$\frac{\tilde{S}_T(T_1, T_2, T_3) - \tilde{S}_T(T_1, T_2^0, T_3)}{T_2^0 - T_2} > 0,$$

and by implication a small probability that

$$\frac{\tilde{S}_T(T_1, T_2, T_3) - \tilde{S}_T(T_1, T_2^0, T_3)}{T_2^0 - T_2} \leq 0,$$

which is precisely what we intended to show.

Testing for Multiple Structural Breaks

Bai and Perron (1998)

Here we consider two ways to test for multiple structural breaks. The first one tests the null of no breaks against the alternative of r breaks, while the second sequentially tests the null of r against the alternative of $r+1$ breaks.

7.1 Testing the Null of No Breaks against the Alternative of m Breaks

We wish to test the null hypothesis

$$H_0: \text{ There are no structural breaks}$$

against the alternative hypothesis

$$H_1: \text{ There are exactly } m \text{ structural breaks.}$$

Under the framework of the structural break model studied in the previous chapter, we can view these hypotheses as representing a restricted model and an unrestricted model, respectively, where there are km restrictions

$$\bar{\delta} := \delta_1 = \delta_j,$$

for $2 \leq j \leq m+1$, so that there are no structural breaks under the null.

To obtain a tractable test statistic for testing H_0 against H_1, define the set

$$\Lambda_\varepsilon = \left\{ (\lambda_1, \cdots, \lambda_m) \in [0,1]^m \mid |\lambda_j - \lambda_{j-1}| > \varepsilon \text{ and } \lambda_j = \frac{n}{d} \text{ for some } n \in N_+ \text{ for any } 1 \leq j \leq m+1 \right\}.$$

for some $d \in N_+$ and small $\varepsilon > 0$, where we put $\lambda_0 = 0$ and $\lambda_{m+1} = 1$ as in the previous chapter. The second requirement, that each λ_j be a rational number with divisor d, is, strictly speaking, not required, but has been included to facilitate the proof and enable an exacty grid search.

Suppose that $(\lambda_1,\cdots,\lambda_m) \in \Lambda_\varepsilon$ represents a collection of true break fractions in the sample. Then, letting the break points be given as $T_1,\cdots,T_m$, where $T_j = \lfloor T\lambda_j \rfloor$ for any $1 \le j \le m$, the model can be written as

$$y_t = x_t'\beta + z_t'\delta_j + u_t$$

for any $T_{j-1}+1 \le t \le T_j$ and $1 \le j \le m+1$, or in matrix form as

$$Y = X\beta + \bar{Z}\delta + U,$$

where $\bar{Z}$ is the diagonal partition of $Z = (z_1,\cdots,z_T)'$ according to the break points $T_1,\cdots,T_m$, and $\delta = (\delta_1',\cdots,\delta_{m+1}')'$, as we defined in the previous chapter.
In this case, the alternative hypothesis H_1 can be viewed as an unrestricted version of the above model, and the null hypothesis H_0 can be viewed as imposing the following linear restrictions on the model:

$$H_0: \quad R\delta = \mathbf{0}$$

for the $km \times k(m+1)$ full rank matrix

$$R = \begin{pmatrix} I_k & -I_k & \cdots & O & O \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & \cdots & I_k & -I_k \end{pmatrix}.$$

Therefore, granted that $(\lambda_1,\cdots,\lambda_m)$ are the true (potential) break fractions, the problem of testing whether there are 0 or m structural breaks reduces to testing for the linear restrictions $R\delta = \mathbf{0}$.
We can naturally think of the following F statistic:

$$F_T(\lambda_1,\cdots,\lambda_m,k) = \frac{(SSR_0 - SSR_m)/km}{SSR_m/(T-(m+1)k-p)},$$

where SSR_0 is the sum of squared residuals under the restricted model of no structural breaks, SSR_m is the sum of squared residuals under the unrestricted model in which the break dates are given by $T_1,\cdots,T_m$, $T-(m+1)k-p$ is the degrees of freedom in the unrestricted model (there are $m+1$ regimes during which the k-dimensional coefficient of z_t changes, along with p coefficients corresponding to the regime-independent exogenous variables x_t), and km is the number of restrictions.

Under the null hypothesis, there are no structural breaks, so to consider all potential break points, our test statistic takes the form

$$\sup F_T(m;k) = \sup_{(\lambda_1,\cdots,\lambda_m)\in\Lambda_\varepsilon} F_T(\lambda_1,\cdots,\lambda_m,k).$$

If the test statistic is large, then it means that there is evidence supporting the structural break model under at least one set of structural break points over the model with no breaks, and therefore we can reject the null of no breaks.

7.1.1 Computation of the Test Statistic

For any $(\lambda_1, \cdots, \lambda_m) \in \Lambda_\varepsilon$, our formulation of $F(\lambda_1, \cdots, \lambda_m, k)$ is given in terms of the sum of squared residuals under the restricted and unrestricted models.
The fact that the null reduces to a set of linear restrictions allows us to use only estimators from the unrestricted model to compute $F(\lambda_1, \cdots, \lambda_m, k)$.

We first consider the unrestricted model. The estimator of β and δ from the unrestricted model are given by

$$\begin{pmatrix} \hat{\beta} \\ \hat{\delta} \end{pmatrix} = \begin{pmatrix} X'X & X'\bar{Z} \\ \bar{Z}'X & \bar{Z}'\bar{Z} \end{pmatrix}^{-1} \begin{pmatrix} X'Y \\ \bar{Z}'Y \end{pmatrix},$$

and by the FWL Theorem, we have

$$\begin{aligned} \hat{\beta} &= (X'M_{\bar{Z}}X)^{-1} X'M_{\bar{Z}}Y = \beta + (X'M_{\bar{Z}}X)^{-1} X'M_{\bar{Z}}U \\ \hat{\delta} &= \left(\bar{Z}'M_X\bar{Z}\right)^{-1} \bar{Z}'M_XY = \delta + \left(\bar{Z}'M_X\bar{Z}\right)^{-1} \bar{Z}'M_XU. \end{aligned}$$

Note that the estimator of δ is found as the OLS estimator in the transformed model

$$M_XY = M_X\bar{Z}\delta + M_XU.$$

In contrast, the estimator of δ in the restricted model solves the restricted minimization problem

$$\begin{aligned} \min_{\delta \in \mathbb{R}^{(m+1)k}} \quad & \frac{1}{2}\left(M_XY - (M_X\bar{Z})\delta\right)'\left(M_XY - (M_X\bar{Z})\delta\right) \\ \text{subject to} \quad & R\delta = \mathbf{0}. \end{aligned}$$

The Lagrangian for this problem is

$$\mathcal{L} = \frac{1}{2}\left(M_XY - (M_X\bar{Z})\delta\right)'\left(M_XY - (M_X\bar{Z})\delta\right) - \mu' \cdot R\delta,$$

where μ is a km-dimensional vector of Lagrangian multipliers. The restricted estimator $\tilde{\delta}$ of δ satisfies the first order condition

$$\bar{Z}'M_X\left(M_XY - (M_X\bar{Z})\tilde{\delta}\right) = R'\mu$$

and the constraint $R\tilde{\delta} = \mathbf{0}$. We can now see that

$$\tilde{\delta} = \hat{\delta} - (\bar{Z}' M_X \bar{Z})^{-1} R' \mu.$$

Premultiplying the f.o.c. by $R\left(\bar{Z}' M_X \bar{Z}\right)^{-1}$ yields

$$R\left(\bar{Z}' M_X \bar{Z}\right)^{-1} \bar{Z}' M_X Y - R\tilde{\delta} = \left(R\left(\bar{Z}' M_X \bar{Z}\right)^{-1} R'\right) \cdot \mu,$$

where $R\left(\bar{Z}' M_X \bar{Z}\right)^{-1}$ is nonsingular because R is of full rank.
Therefore,

$$\begin{aligned}
\tilde{\delta} &= \hat{\delta} - \left(\bar{Z}' M_X \bar{Z}\right)^{-1} R' \left[R\left(\bar{Z}' M_X \bar{Z}\right)^{-1} R'\right]^{-1} R\left(\bar{Z}' M_X \bar{Z}\right)^{-1} \bar{Z}' M_X Y \\
&= \hat{\delta} - \left(\bar{Z}' M_X \bar{Z}\right)^{-1} R' \left[R\left(\bar{Z}' M_X \bar{Z}\right)^{-1} R'\right]^{-1} R\, \hat{\delta}.
\end{aligned}$$

Using the formula above, we can now derive a closed form expression for $SSR_0 - SSR_m$:

$$\begin{aligned}
SSR_0 - SSR_m &= \left(M_X Y - M_X \bar{Z} \cdot \tilde{\delta}\right)' \left(M_X Y - M_X \bar{Z} \cdot \tilde{\delta}\right) - \left(M_X Y - M_X \bar{Z} \cdot \hat{\delta}\right)' \left(M_X Y - M_X \bar{Z} \cdot \hat{\delta}\right) \\
&= \tilde{\delta}' \bar{Z}' M_X \bar{Z} \tilde{\delta} - \hat{\delta}' \bar{Z}' M_X \bar{Z} \hat{\delta}^* - 2Y' M_X \bar{Z} \cdot \left(\tilde{\delta} - \hat{\delta}\right) \\
&= \hat{\delta}' R' \left[R\left(\bar{Z}' M_X \bar{Z}\right)^{-1} R'\right]^{-1} R\hat{\delta}.
\end{aligned}$$

The test statistic can now be expressed entirely in terms of estimators derived from the unrestricted model as:

$$\begin{aligned}
F_T(\lambda_1, \cdots, \lambda_m, k) &= \frac{(SSR_0 - SSR_m)/km}{SSR_m/(T-(m+1)k-p)} \\
&= \left(\frac{T-(m+1)k-p}{km}\right) \cdot \frac{\hat{\delta}' R' \left[R\left(\bar{Z}' M_X \bar{Z}\right)^{-1} R'\right]^{-1} R\hat{\delta}}{SSR_m}.
\end{aligned}$$

7.1.2 Assumptions

To derive the asymptotic distribution of the sup F test statistic above, we need only retain assumptions (5) to (7) above.

Specifically, we assume the following; note that we retain the definiton of $w_t = (x_t', z_t')'$ we made above.

(1) **Uniform Convergence of Sample Covariances**
We assume that there exists a positive definite matrix $Q \in \mathbb{R}^{k\times k}$ such that

$$\frac{1}{T}\sum_{t=\lfloor Ts\rfloor+1}^{\lfloor Tr\rfloor} w_t w_t' \xrightarrow{p} (r-s)Q$$

uniformly on the set of all $(r,s) \in [0,1]^2$ such that $s < r$, and that

$$\sum_{t=i}^{l} w_t w_t'$$

is nonsingular for any $i < l$ such that $i - l \geq k + p$.
Note that this both strengthens assumption (7) above and includes assumption (2) and (4) above as a special case. In addition, this assumption precludes trending regressors.
We allow Q to be decomposed as

$$Q = \begin{pmatrix} Q_x & Q_{xz} \\ Q_{zx} & Q_z \end{pmatrix},$$

where the dimensions of the submatrices are conformable with x_t, z_t.

(2) **Uncorrelated Errors**
We assume that the error process $\{u_t\}_{t\in\mathbb{Z}}$ is a Martingale Difference Sequence (MDS) with respect to the filtration

$$\mathcal{F} = \{\mathcal{F}_t \mid t \in \mathbb{Z}\}$$

on $\mathbb{Z}$ where

$$\mathcal{F}_t = \sigma\left(\{w_s\}_{s\in\mathbb{Z}} \cup \{u_s\}_{s\leq t}\right)$$

for each $t \in \mathbb{Z}$, such that

$$\sup_{t\in\mathbb{Z}} \mathbb{E}|u_t|^{4+c} < +\infty$$

for some $c > 0$.

Furthermore, by the definition of an MDS $\mathbb{E}[u_t] = 0$, and we assume that $\mathbb{E}\left[u_t^2 \mid \mathcal{F}_{t-1}\right] = \sigma^2$ for any $t \in \mathbb{Z}$.

(3) **An FCLT for Martingale Difference Sequences**

Let the stochastic process $\{v_t\}_{t\in\mathbb{Z}}$ be defined as

$$v_t = w_t u_t$$

for any $t \in \mathbb{Z}$. As we showed in the previous chapter, the preceding assumption implies that $\{v_t\}_{t\in\mathbb{Z}}$ is an MDS with respect to the filtration $\mathcal{F}$ such that

$$\frac{1}{T}\sum_{t=1}^{T}(u_t^2 - \sigma^2)w_t w_t' \xrightarrow{p} O.$$

Furthermore, assumption (1) implies that

$$\sigma^2 \cdot \frac{1}{T}\sum_{t=1}^{T} w_t w_t' \xrightarrow{p} \sigma^2 Q,$$

so we have

$$\frac{1}{T}\sum_{t=1}^{T} v_t v_t' = \frac{1}{T}\sum_{t=1}^{T} u_t^2 w_t w_t' \xrightarrow{p} \sigma^2 Q.$$

Therefore, as in the previous chapter, it makes sense to assume that $\{v_t\}_{t\in\mathbb{Z}}$ follows some sort of FCLT result. Specifically, define the stochastic processes $\{S_T(r)\}_{r\in[0,1]}$ and with continuous paths as

$$S_T(r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr \rfloor} v_t + \frac{1}{\sqrt{T}}(Tr - \lfloor Tr \rfloor)v_{\lfloor Tr \rfloor + 1}$$

for any $r \in [0,1]$. Letting S_T be the random function in $\mathcal{C}^{p+k}[0,1]$ corresponding to $\{S_T(r)\}_{r\in[0,1]}$, we assume that

$$S_T \xrightarrow{d} \sigma Q^{\frac{1}{2}} \cdot W^{p+k},$$

where $Q^{\frac{1}{2}}$ is the Cholesky factor of Q.

It follows that, letting $\{A_T(r)\}_{r\in[0,1]}$ and $\{B_T(r)\}_{r\in[0,1]}$ be defined as

$$A_T(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} x_t u_t + \frac{1}{\sqrt{T}} (Tr - \lfloor Tr \rfloor) \left(x_{\lfloor Tr \rfloor + 1} \cdot u_{\lfloor Tr \rfloor + 1} \right)$$
$$B_T(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} z_t u_t + \frac{1}{\sqrt{T}} (Tr - \lfloor Tr \rfloor) \left(z_{\lfloor Tr \rfloor + 1} \cdot u_{\lfloor Tr \rfloor + 1} \right)$$

for any $r \in [0,1]$, the FCLT results

$$A_T \xrightarrow{d} \sigma Q_x^{\frac{1}{2}} \cdot W^p$$
$$B_T \xrightarrow{d} \sigma Q_z^{\frac{1}{2}} \cdot W^k$$

hold by the continuous mapping theorem, where $Q_x^{\frac{1}{2}}$ and $Q_z^{\frac{1}{2}}$ are the Cholesky factors of Q_x and Q_z.

7.1.3 The Asymptotic Distribution of the Test Statistic

In this section we derive the asymptotic distribution of the sup F test statistic defined above. To this end, we show that $F_T(\lambda_1, \cdots, \lambda_m; k)$ converges to some distribution that is a function of $(\lambda_1, \cdots, \lambda_m)$ uniformly in Λ_ε, and as such that the supremum statistic converges to the supremum of those distributions over Λ_ε.

First fix some $(\lambda_1, \cdots, \lambda_m) \in \Lambda_\varepsilon$. We will investigate the asymptotic properties of $F(\lambda_1, \cdots, \lambda_m; k)$. For the sake of completeness, we reiterate the definition of $F_T(\lambda_1, \cdots, \lambda_m; k)$:

$$F_T(\lambda_1, \cdots, \lambda_m; k) = \frac{(SSR_0 - SSR_m)/km}{SSR_m/(T-(m+1)k-p)}$$

Asymptotic Properties of the Unrestricted Model

We have already derived the least squares estimators $\hat{\beta}$ and $\hat{\delta}$ of β and δ under the unrestricted model as

$$\hat{\beta} = \beta + \left(X' M_{\bar{Z}} X\right)^{-1} X' M_{\bar{Z}} U$$
$$\hat{\delta} = \delta + \left(\bar{Z}' M_X \bar{Z}\right)^{-1} \bar{Z}' M_X U.$$

Denoting $\hat{\gamma} = (\hat{\beta}', \hat{\delta}')'$ and $\gamma = (\beta', \delta')'$,

$$\hat{\gamma} = \left(\bar{W}'\bar{W}\right)^{-1} \bar{W}'Y = \gamma + \left(\bar{W}'\bar{W}\right)^{-1} \bar{W}'U,$$

where we defined $\bar{W}$ earlier.
Define $D = \mathrm{diag}(\lambda_1, \lambda_2 - \lambda_1, \cdots, 1-\lambda_m) \in \mathbb{R}^{m\times m}$. Since

$$\frac{1}{T}\bar{W}'\bar{W} = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} x_t x_t' & \frac{1}{T}\sum_{t=1}^{T_1} x_t z_t' & \vdots & \frac{1}{T}\sum_{t=T_m+1}^{T} x_t z_t' \\ \frac{1}{T}\sum_{t=1}^{T_1} z_t x_t' & \frac{1}{T}\sum_{t=1}^{T_1} z_t z_t' & \vdots & O \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{T}\sum_{t=T_m+1}^{T} z_t x_t' & O & \cdots & \frac{1}{T}\sum_{t=T_m+1}^{T} z_t z_t' \end{pmatrix},$$

by assumption (1)

$$\frac{1}{T}\bar{W}'\bar{W} \xrightarrow{p} \begin{pmatrix} Q_x & \lambda_1 \cdot Q_{xz} & \vdots & (1-\lambda_m)\cdot Q_{xz} \\ \lambda_1 \cdot Q_{zx} & \lambda_1 \cdot Q_z & \vdots & O \\ \vdots & \vdots & \ddots & \vdots \\ (1-\lambda_m)\cdot Q_{zx} & O & \cdots & (1-\lambda_m)\cdot Q_z \end{pmatrix} = \begin{pmatrix} Q_x & (\iota_m' D) \bigotimes Q_{xz} \\ (D\iota_m) \bigotimes Q_{zx} & D \bigotimes Q_z \end{pmatrix} = \bar{Q},$$

where ι_m is the m-dimensional vector comprised of 1s, and $\bar{Q}$ is positive definite. Furthermore,

$$\frac{1}{\sqrt{T}}\bar{W}'U = \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t u_t \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T_1} z_t u_t \\ \vdots \\ \frac{1}{\sqrt{T}}\sum_{t=T_m+1}^{T} z_t u_t \end{pmatrix} = \begin{pmatrix} A_T(1) \\ B_T(\lambda_1) \\ \vdots \\ B_T(1) - B_T(\lambda_m) \end{pmatrix},$$

so by assumption (3),

$$\frac{1}{\sqrt{T}}\bar{W}'U \xrightarrow{d} \begin{pmatrix} \sigma Q_x^{\frac{1}{2}} \cdot W^p(1) \\ \sigma Q_z^{\frac{1}{2}} \cdot W^k(\lambda_1) \\ \vdots \\ \sigma Q_z^{\frac{1}{2}}\left(W^k(1) - W^k(\lambda_m)\right) \end{pmatrix} = \sigma\left(Q_x^{\frac{1}{2}} \cdot W^p(1)\right) = \begin{pmatrix} \sigma Q_x^{\frac{1}{2}} \cdot W^p(1) \\ B^* \end{pmatrix}$$

as $T \to \infty$. By the continuous mapping theorem and Slutsky's theorem, we now have

$$\sqrt{T}(\hat{\gamma} - \gamma) = \left(\frac{1}{T}\bar{W}'\bar{W}\right)^{-1}\frac{1}{\sqrt{T}}\bar{W}'U \xrightarrow{d} \bar{Q}^{-1}\begin{pmatrix} \sigma Q_x^{\frac{1}{2}} \cdot W^p(1) \\ B^* \end{pmatrix},$$

or that

$$\sqrt{T}(\hat{\gamma} - \gamma) = O_p(1).$$

By implication, $\hat{\gamma} - \gamma = o_p(1)$, that is, $\hat{\gamma}$ is consistent for γ.

The scaled sum of squared residuals SSR_m is decomposed as

$$\begin{aligned} \frac{1}{T}SSR_m &= \frac{1}{T}\left(Y - \bar{W}\hat{\gamma}\right)'\left(Y - \bar{W}\hat{\gamma}\right) \\ &= \frac{1}{T}\left(\bar{W}(\gamma - \hat{\gamma}) + U\right)'\left(\bar{W}(\gamma - \hat{\gamma}) + U\right) \\ &= (\gamma - \hat{\gamma})'\left(\frac{1}{T}\bar{W}'\bar{W}\right)(\gamma - \hat{\gamma}) + 2(\gamma - \hat{\gamma})'\left(\frac{1}{T}\bar{W}'U\right) + \frac{1}{T}U'U. \end{aligned}$$

We saw above that $\frac{1}{T}\bar{W}'\bar{W}$ is $O_p(1)$ and $\frac{1}{T}\bar{W}'U$ and $\hat{\gamma} - \gamma$ are $o_p(1)$. In addition, because $\{u_t^2 - \sigma^2\}_{t\in\mathbb{Z}}$ is an MDS with respect to $\mathcal{F}$, it is an uncorrelated sequence, which means that

$$\begin{aligned} \mathbb{E}\left|\frac{1}{T}U'U - \sigma^2\right|^2 &= \mathbb{E}\left|\frac{1}{T}\sum_{t=1}^{T}(u_t^2 - \sigma^2)\right|^2 \\ &= \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}\left[(u_t^2 - \sigma^2)\right]^2 \\ &\leq \frac{1}{T}\left(\sup_{t\in\mathbb{Z}}\mathbb{E}|u_t|^4\right). \end{aligned}$$

Since $\sup_{t\in\mathbb{Z}} \mathbb{E}|u_t|^4 < +\infty$ by assumption, we can conclude that

$$\mathbb{E}\left|\frac{1}{T}U'U - \sigma^2\right|^2 \to 0$$

as $T \to \infty$ and thus that

$$\frac{1}{T}U'U \xrightarrow{p} \sigma^2.$$

Therefore,

$$\frac{1}{T}SSR_m \xrightarrow{p} \sigma^2,$$

and likewise,

$$\frac{1}{T-(m+1)k-p}SSR_m \xrightarrow{p} \sigma^2,$$

where the left hand side is the denominator of $F(\lambda_1,\cdots,\lambda_m;k)$.

Decomposing the Difference of Sum of Squared Residuals

To investigate the asymptotic properties of

$$F_T^* = SSR_0 - SSR_m,$$

we introduce a convenient decomposition of the sum of squared residuals.

i) **The Unrestricted Model**

As usual, we start with the sum of squared residuals under the unrestricted model. Note that SSR_m can be written as

$$\begin{aligned} SSR_m &= \left(Y - X\hat{\beta} - \bar{Z}\hat{\delta}\right)' \left(Y - X\hat{\beta} - \bar{Z}\hat{\delta}\right) \\ &= \sum_{j=1}^{m+1} \sum_{t=T_{j-1}+1}^{T_j} (y_t - x_t'\hat{\beta} - z_j'\hat{\delta}_j)^2, \end{aligned}$$

where $\hat{\delta}_j$ is the part of $\hat{\delta}$ estimating δ_j. Defining

$$D^U(j,j) = \sum_{t=T_{j-1}+1}^{T_j} (y_t - x_t'\hat{\beta} - z_j'\hat{\delta}_j)^2$$

for $1 \le j \le m+1$, which is the sum of squared residuals under the unrestricted model computed using only the data from the jth regime, we can write

$$SSR_m = \sum_{j=1}^{m+1} D^U(j,j).$$

It remains to be seen ho to compute each $D^U(j,j)$.

Define $X_1, \cdots, X_{m+1}$, $Y_1, \cdots, Y_{m+1}$ and $U_1, \cdots, U_{m+1}$ as

$$X_j = \begin{pmatrix} x'_{T_{j-1}+1} \\ \vdots \\ x'_{T_j} \end{pmatrix}, \quad Y_j = \begin{pmatrix} y_{T_{j-1}+1} \\ \vdots \\ y_{T_j} \end{pmatrix}, \quad \text{and} \quad U_j = \begin{pmatrix} u_{T_{j-1}+1} \\ \vdots \\ u_{T_j} \end{pmatrix}$$

for $1 \le j \le m+1$. Then,

$$\bar{Z}'X = \begin{pmatrix} Z_1' & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & Z_{m+1}' \end{pmatrix} X = \begin{pmatrix} \sum_{t=1}^{T_1} z_t x_t' \\ \vdots \\ \sum_{t=T_m+1}^{T} z_t x_t' \end{pmatrix} := \begin{pmatrix} Z_1' X_1 \\ \vdots \\ Z_{m+1}' X_{m+1} \end{pmatrix},$$

so that

$$\begin{pmatrix} X'X & X'\bar{Z} \\ \bar{Z}'X & \bar{Z}'\bar{Z} \end{pmatrix} = \begin{pmatrix} X'X & X_1'Z_1 & \cdots & X_{m+1}'Z_{m+1} \\ Z_1'X_1 & Z_1'Z_1 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ Z_{m+1}'X_{m+1} & O & \cdots & Z_{m+1}'Z_{m+1} \end{pmatrix},$$

and

$$\begin{pmatrix} X'Y \\ \bar{Z}'Y \end{pmatrix} = \begin{pmatrix} X'Y \\ Z_1'Y_1 \\ \vdots \\ Z_{m+1}'Y_{m+1} \end{pmatrix}.$$

By definition, the least squares estimators $\hat{\beta}$ and $\hat{\delta}$ of β and δ satisfy

$$\begin{pmatrix} X'X & X'\bar{Z} \\ \bar{Z}'X & \bar{Z}'\bar{Z} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{\delta}_1 \\ \vdots \\ \hat{\delta}_{m+1} \end{pmatrix} = \begin{pmatrix} X'Y \\ \bar{Z}'Y \end{pmatrix},$$

so for any $1 \leq j \leq m+1$, we have

$$\begin{pmatrix} Z_j'X_j & Z_j'Z_j \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{\delta}_j \end{pmatrix} = Z_j'Y_j,$$

or

$$\hat{\delta}_j = (Z_j'Z_j)^{-1}Z_j'\left(Y_j - X_j\hat{\beta}\right).$$

For any $1 \leq j \leq m+1$,

$$Y_j = X_j\beta + Z_j\delta_j + U_j$$

where

$$\delta_j = \delta_1$$

under the null hypothesis. Since

$$\hat{\beta} = \beta + (X'M_{\bar{Z}}X)^{-1}X'M_{\bar{Z}}U,$$

it follows that

$$\begin{aligned}\hat{\delta}_j &= (Z_j'Z_j)^{-1}Z_j'\left(Y_j - X_j\hat{\beta}\right)\\ &= \delta_j + (Z_j'Z_j)^{-1}Z_j'X_j(\beta-\hat{\beta}) + (Z_j'Z_j)^{-1}Z_j'U_j\end{aligned}$$

and thus

$$\begin{aligned}D^U(j,j) &= \sum_{t=T_{j-1}+1}^{T_j}(y_t - x_t'\hat{\beta} - z_t'\hat{\delta}_j)^2\\ &= \left(Y_j - X_j\hat{\beta} - Z_j'\hat{\delta}_j\right)'\left(Y_j - X_j\hat{\beta} - Z_j'\hat{\delta}_j\right)\\ &= \left(X_j(\beta-\hat{\beta}) + Z_j'(\delta_j - \hat{\delta}_j) + U_j\right)'\left(X_j(\beta-\hat{\beta}) + Z_j'(\delta_j - \hat{\delta}_j) + U_j\right)\\ &= \left|\left(I_{T_j - T_{j-1}} - Z_j(Z_j'Z_j)^{-1}Z_j'\right)X_j(\beta-\hat{\beta}) + \left(I_{T_j-T_{j-1}} - Z_j'(Z_j'Z_j)^{-1}Z_j'\right)U_j\right|^2\\ &= \left|M_{Z_j}\left(U_j - X_j(X'M_{\bar{Z}}X)^{-1}X'M_{\bar{Z}}U\right)\right|^2.\end{aligned}$$

ii) **The Restricted Model**

We now decompose SSR_0 so that each piece has a similar representation as $D^U(j,j)$. Define $X_{1,j}$, $Z_{1,j}$, $Y_{1,j}$, and $U_{1,j}$ as

$$X_{1,j} = \begin{pmatrix} x_1' \\ \vdots \\ x_{T_j}' \end{pmatrix}, \quad Z_{1,j} = \begin{pmatrix} z_1' \\ \vdots \\ z_{T_j}' \end{pmatrix}, \quad Y_{1,j} = \begin{pmatrix} y_1 \\ \vdots \\ y_{T_j} \end{pmatrix} \quad \text{and} \quad U_{1,j} = \begin{pmatrix} u_1 \\ \vdots \\ u_{T_j} \end{pmatrix}$$

for $1 \le j \le m+1$. That is, $X_{1,j}$, $Z_{1,j}$, $Y_{1,j}$, and $U_{1,j}$ collect the observations of the respective variables up to regime j.

Recalling that $\tilde{\beta}$ and $\tilde{\delta}$ are the least squares estimators of β and δ under the restricted model,

$$\tilde{\delta}_j = \tilde{\delta}_1$$

for any $1 \le j \le m+1$ due to the restriction $R\tilde{\delta} = \mathbf{0}$, and the sum of squared residuals under the restricted model is given by

$$SSR_0 = \sum_{t=1}^{T} (y_t - x_t'\tilde{\beta} - z_t'\tilde{\delta}_1)^2 = \left(Y - X\tilde{\beta} - Z\tilde{\delta}_1\right)' \left(Y - X\tilde{\beta} - Z\tilde{\delta}_1\right).$$

We now derive an expression for SSR_0.

By definition, $\tilde{\beta}$ and $\tilde{\delta}$ are solutions to the minimization problem

$$\min_{\beta\in\mathbb{R}^p, \delta\in\mathbb{R}^{(m+1)k}} \quad \frac{1}{2}\left(Y - X\beta - \bar{Z}\delta\right)' \left(Y - X\beta - \bar{Z}\delta\right) = \frac{1}{2}\sum_{j=1}^{m+1} \sum_{t=T_{j-1}+1}^{T_j} (y_t - x_t'\beta - z_t'\delta_j)^2$$

$$\text{subject to} \quad \delta_j = \delta_1 \text{ for any } 1 \le j \le m+1.$$

Substituting the constraints into the objective function, our goal reduces to solving the minimization problem

$$\min_{\beta\in\mathbb{R}^p, \delta_1\in\mathbb{R}^k} \quad \frac{1}{2}\sum_{t=1}^{T} (y_t - x_t'\beta - z_t'\delta_1)^2 = \frac{1}{2}(Y - X\beta - Z\delta_1)'(Y - X\beta - Z\delta_1).$$

Therefore,

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z \end{pmatrix} \begin{pmatrix} \tilde{\beta} \\ \tilde{\delta}_1 \end{pmatrix} = \begin{pmatrix} X'Y \\ Z'Y \end{pmatrix}.$$

It follows that

$$\tilde{\beta} = (X'M_Z X)^{-1} X'M_Z Y = \beta + (X'M_Z X)^{-1} X'M_Z U$$

and

$$\tilde{\delta}_1 = (Z'Z)^{-1}Z'(Y - X\tilde{\beta}),$$

and because

$$Y = X\beta + Z\delta_1 + U$$

under the null hypothesis, we have

$$\tilde{\delta}_1 = \delta_1 + (Z'Z)^{-1}Z'X(\beta - \tilde{\beta}) + (Z'Z)^{-1}Z'U.$$

Substituting this into the formula for SSR_0, we have

$$\begin{aligned} SSR_0 &= \left(Y - X\tilde{\beta} - Z\tilde{\delta}_1\right)' \left(Y - X\tilde{\beta} - Z\tilde{\delta}_1\right) \\ &= \left(X(\beta - \tilde{\beta}) + Z(\delta_1 - \tilde{\delta}_1) + U\right)' \left(X(\beta - \tilde{\beta}) + Z(\delta_1 - \tilde{\delta}_1) + U\right) \\ &= \left|\left(I_T - Z(Z'Z)^{-1}Z'\right) X(\beta - \tilde{\beta}) + \left(I_T - Z(Z'Z)^{-1}Z'\right) U\right|^2 \\ &= \left|M_Z \left(U - X(X'M_Z X)^{-1}X'M_Z U\right)\right|^2. \end{aligned}$$

Analogously, define

$$D^R(1,j) = \left|M_{Z_{1,j}} \left(U_{1,j} - X_{1,j}(X'M_Z X)^{-1}X'M_Z U\right)\right|^2$$

for any $1 \le j \le m+1$. Then, $SSR_0 = D^R(1,m+1)$, so that

$$SSR_0 = \sum_{j=1}^{m} \left(D^R(1,j+1) - D^R(1,j)\right) + D^R(1,1).$$

Therefore, we can see that

$$\begin{aligned}
F_T^* &= SSR_0 - SSR_m \\
&= \sum_{j=1}^{m} \Big(D^R(1, j+1) - D^R(1, j) \Big) + D^R(1,1) - \sum_{j=1}^{m+1} D^U(j, j) \\
&= \sum_{j=1}^{m} \Big(D^R(1, j+1) - D^R(1, j) - D^U(j+1, j+1) \Big) + D^R(1,1) - D^U(1,1).
\end{aligned}$$

Defining

$$F_{T,j}^* = D^R(1, j+1) - D^R(1, j) - D^U(j+1, j+1)$$

and

$$F_{T,0}^* = D^R(1,1) - D^U(1,1)$$

for $1 \leq j \leq m$, we will now study the asymptotic properties of $F_{T,j}^*$ and $F_{T,0}^*$.

The Convergence Properties of $F^*_{T,j}$ for $1 \le j \le m$

For this part, fix $1 \le j \le m$, and define

$$V_T = (X' M_Z X)^{-1} X' M_Z U \quad \text{and} \quad \bar{V}_T = (X' M_{\bar{Z}} X)^{-1} X' M_{\bar{Z}} U.$$

Noting that

$$\begin{aligned}
U'_{1,j+1} U_{1,j+1} &= U'_{1,j} U_{1,j} + U'_{j+1} U_{j+1} \\
X'_{1,j+1} X_{1,j+1} &= X'_{1,j} X_{1,j} + X'_{j+1} X_{j+1} \\
Z'_{1,j+1} Z_{1,j+1} &= Z'_{1,j} Z_{1,j} + Z'_{j+1} Z_{j+1} \\
X'_{1,j+1} Z_{1,j+1} &= X'_{1,j} Z_{1,j} + X'_{j+1} Z_{j+1} \\
X'_{1,j+1} U_{1,j+1} &= X'_{1,j} U_{1,j} + X'_{j+1} U_{j+1} \\
Z'_{1,j+1} U_{1,j+1} &= Z'_{1,j} U_{1,j} + Z'_{j+1} U_{j+1},
\end{aligned}$$

and defining

$$\begin{aligned}
S_j &= Z'_{1,j} U_{1,j} = \sum_{t=1}^{T_j} z_t u_t \\
H_j &= Z'_{1,j} Z_{1,j} = \sum_{t=1}^{T_j} z_t z'_t \\
K_j &= Z'_{1,j} X_{1,j} = \sum_{t=1}^{T_j} z_t x'_t \\
L_j &= X'_{1,j} X_{1,j} = \sum_{t=1}^{T_j} x_t x'_t \\
M_j &= X'_{1,j} U_{1,j} = \sum_{t=1}^{T_j} x_t u_t,
\end{aligned}$$

we can see that

$$\begin{aligned}
F^*_{T,j} &= D^R(1, j+1) - D^R(1, j) - D^U(j+1, j+1) \\
&= \left| M_{Z_{1,j+1}} \left(U_{1,j+1} - X_{1,j+1} V_T \right) \right|^2 - \left| M_{Z_{1,j}} \left(U_{1,j} - X_{1,j} V_T \right) \right|^2 - \left| M_{Z_j} \left(U_j - X_j \, \delta V_T \right) \right|^2 \\
&= \left(U_{1,j+1} - X_{1,j+1} V_T \right)' M_{Z_{1,j+1}} \left(U_{1,j+1} - X_{1,j+1} V_T \right) - \left(U_{1,j} - X_{1,j} V_T \right)' M_{Z_{1,j}} \left(U_{1,j} - X_{1,j} V_T \right) \\
&\quad - \left(U_{j+1} - X_{j+1} V_T \right)' M_{Z_j} \left(U_{j+1} - X_{j+1} V_T \right) \\
&= U'_{1,j+1} M_{Z_{1,j+1}} U'_{1,j+1} - 2 \cdot V'_T X'_{1,j+1} M_{Z_{1,j+1}} U_{1,j+1} + V'_T X'_{1,j+1} M_{Z_{1,j+1}} X_{1,j+1} V_T \\
&\quad - U'_{1,j} M_{Z_{1,j}} U'_{1,j} + 2 \cdot V'_T X'_{1,j} M_{Z_{1,j}} U_{1,j} - V'_T X'_{1,j} M_{Z_{1,j}} X_{1,j} V_T \\
&\quad - U'_{j+1} M_{Z_{j+1}} U'_{j+1} + 2 \cdot \bar{V}'_T X'_{j+1} M_{Z_{j+1}} U_{j+1} - \bar{V}'_T X'_{j+1} M_{Z_{j+1}} X_{j+1} \bar{V}_T \\
&= U'_{1,j+1} U'_{1,j+1} - 2 \cdot V'_T X'_{1,j+1} U_{1,j+1} + V'_T X'_{1,j+1} X_{1,j+1} V_T
\end{aligned}$$

$$
\begin{aligned}
&- U_{1,j}' U_{1,j}' + 2 \cdot V_T' X_{1,j}' U_{1,j} - V_T' X_{1,j}' X_{1,j} V_T \\
&- U_{j+1}' U_{j+1}' + 2 \cdot \bar{V}_T' X_{j+1}' U_{j+1} - \bar{V}_T' X_{j+1}' X_{j+1} \bar{V}_T \\
&- S_{j+1}' H_{j+1}^{-1} S_{j+1} + 2 \cdot V_T' K_{j+1}' H_{j+1}^{-1} S_{j+1} - V_T' K_{j+1}' H_{j+1}^{-1} K_{j+1} V_T \\
&+ S_j' H_j^{-1} S_j - 2 \cdot V_T' K_j' H_j^{-1} S_j + V_T' K_j' H_j^{-1} K_j V_T \\
&+ (S_{j+1} - S_j)' (H_{j+1} - H_j)^{-1} (S_{j+1} - S_j) - 2 \cdot \bar{V}_T' (K_{j+1} - K_j)' (H_{j+1} - H_j)^{-1} (S_{j+1} - S_j) \\
&+ \bar{V}_T' (K_{j+1} - K_j)' (H_{j+1} - H_j)^{-1} (K_{j+1} - K_j) \bar{V}_T \\
= {} & 2 \cdot \left(\bar{V}_T - V_T \right)' (M_{j+1} - M_j) + V_T' (L_{j+1} - L_j) V_T - \bar{V}_T' (L_{j+1} - L_j) \bar{V}_T \\
&- S_{j+1}' H_{j+1}^{-1} S_{j+1} + 2 \cdot V_T' K_{j+1}' H_{j+1}^{-1} S_{j+1} - V_T' K_{j+1}' H_{j+1}^{-1} K_{j+1} V_T \\
&+ S_j' H_j^{-1} S_j - 2 \cdot V_T' K_j' H_j^{-1} S_j + V_T' K_j' H_j^{-1} K_j V_T \\
&+ (S_{j+1} - S_j)' (H_{j+1} - H_j)^{-1} (S_{j+1} - S_j) - 2 \cdot \bar{V}_T' (K_{j+1} - K_j)' (H_{j+1} - H_j)^{-1} (S_{j+1} - S_j) \\
&+ \bar{V}_T' (K_{j+1} - K_j)' (H_{j+1} - H_j)^{-1} (K_{j+1} - K_j) \bar{V}_T.
\end{aligned}
$$

There are a total of 12 terms in the above expression; we study each in detail.

For $T \geq d$, $T\lambda_j$ is an integer, so by assumption (3),

$$\begin{pmatrix} \frac{1}{\sqrt{T}} M_j \\ \frac{1}{\sqrt{T}} S_j \end{pmatrix} = \frac{1}{\sqrt{T}} \begin{pmatrix} X'_{1,j} \\ Z'_{1,j} \end{pmatrix} U_{1,j} = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\lambda_j \rfloor} v_t = S_T(\lambda_j)$$
$$\xrightarrow{d} \sigma Q^{\frac{1}{2}} \cdot W^{p+k}(\lambda_j),$$

and by assumption (1), we have

$$\begin{pmatrix} \frac{1}{T} L_j & \frac{1}{T} K'_j \\ \frac{1}{T} K_j & \frac{1}{T} H_j \end{pmatrix} = \frac{1}{T} \begin{pmatrix} \sum_{t=1}^{T_j} x_t x'_t & \sum_{t=1}^{T_j} x_t z'_t \\ \sum_{t=1}^{T_j} z_t x'_t & \sum_{t=1}^{T_j} z_t z'_t \end{pmatrix} = \frac{1}{T} \sum_{t=1}^{\lfloor T\lambda_j \rfloor} w_t w'_t \xrightarrow{p} \lambda_j \cdot Q$$

and

$$\begin{pmatrix} \frac{1}{T}(L_{j+1} - L_j) & \frac{1}{T}(K_{j+1} - K_j)' \\ \frac{1}{T}(K_{j+1} - K_j) & \frac{1}{T}(H_{j+1} - H_j) \end{pmatrix} = \frac{1}{T} \sum_{t=\lfloor T\lambda_j \rfloor + 1}^{\lfloor T\lambda_{j+1} \rfloor} w_t w'_t \xrightarrow{p} (\lambda_{j+1} - \lambda_j) \cdot Q.$$

Furthermore,

$$\begin{aligned} \sqrt{T} V_T &= \left(\frac{1}{T} X' M_Z X\right)^{-1} \frac{1}{\sqrt{T}} X' M_Z U \\ &= \left[\frac{1}{T}\sum_{t=1}^{T} x_t x'_t - \left(\frac{1}{T}\sum_{t=1}^{T} x_t z'_t\right)\left(\frac{1}{T}\sum_{t=1}^{T} z_t z'_t\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} x_t z'_t\right)'\right]^{-1} \\ &\quad \times \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t u_t - \left(\frac{1}{T}\sum_{t=1}^{T} x_t z'_t\right)\left(\frac{1}{T}\sum_{t=1}^{T} z_t z'_t\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_t u_t\right)\right] \\ &\xrightarrow{d} \sigma \cdot \left(Q_x - Q_{xz} Q_z^{-1} Q_{zx}\right)^{-1} \left(Q_x^{\frac{1}{2}} \cdot W^p(1) - Q_{xz} Q_z^{-\frac{1}{2}\prime} \cdot W^k(1)\right) = A^* \end{aligned}$$

as $T \to \infty$ by assumptions (1) and (3), where

$$Q_x - Q_{xz} Q_z^{-1} Q_{zx}$$

is nonsingular because Q is positive definite.
Similarly,

$$\begin{aligned} \sqrt{T} \bar{V}_T &= \left(\frac{1}{T} X' M_{\bar{Z}} X\right)^{-1} \frac{1}{\sqrt{T}} X' M_{\bar{Z}} U \\ &= \left[\frac{1}{T}\sum_{t=1}^{T} x_t x'_t - \left(\frac{1}{T} X' \bar{Z}\right)\left(\frac{1}{T} \bar{Z}' \bar{Z}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} \bar{Z}' X\right)\right]^{-1} \\ &\quad \times \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t u_t - \left(\frac{1}{T} X' \bar{Z}\right)\left(\frac{1}{T} \bar{Z}' \bar{Z}\right)^{-1}\left(\frac{1}{\sqrt{T}} \bar{Z}' U\right)\right] \\ &\xrightarrow{d} \left(Q_x^{-1} - (\iota'_m D \iota_m) \bigotimes Q_{xz} Q_z^{-1} Q_{zx}\right)^{-1} \left(\sigma Q_x^{\frac{1}{2}} \cdot W^p(1) - \left(\iota'_m \bigotimes Q_{xz} Q_z^{-1}\right) \cdot B^*\right). \end{aligned}$$

Since

$$\iota_m' D \iota_m = \lambda_1 + (\lambda_2 - \lambda_1) + \cdots + (1 - \lambda_m) = 1$$

and

$$\begin{aligned}\left(\iota_m' \bigotimes Q_{zx} Q_z^{-1}\right) B^* &= \sigma \cdot Q_{zx} Q_z^{-\frac{1}{2}\prime} \left(W^k(\lambda_1) + (W^k(\lambda_2) - W^k(\lambda_1)) + \cdots + (W^k(1) - W^k(\lambda_m))\right) \\ &= \sigma \cdot Q_{zx} Q_z^{-\frac{1}{2}\prime} W^k(1),\end{aligned}$$

the weak limit above can be expressed as

$$\sqrt{T}\bar{V}_T \xrightarrow{d} \sigma \cdot \left(Q_x^{-1} - Q_{xz} Q_z^{-1} Q_{zx}\right)^{-1} \left(Q_x^{\frac{1}{2}} \cdot W^p(1) - Q_{xz} Q_z^{-\frac{1}{2}\prime} \cdot W^k(1)\right) = A^*.$$

Thus, $\sqrt{T}V_T$ and $\sqrt{T}\bar{V}_T$ converge weakly to the same limit.
We can then write F_T^* as

$$\begin{aligned}F_{T,j}^* = {} & 2 \cdot \sqrt{T}\left(\bar{V}_T - V_T\right)' (A_T(\lambda_{j+1}) - A_T(\lambda_j)) + \left(\sqrt{T}V_T\right)' \left(\frac{1}{T} \sum_{t=\lfloor T\lambda_j \rfloor + 1}^{\lfloor T\lambda_{j+1} \rfloor} x_t x_t'\right) \left(\sqrt{T}V_T\right) \\ & - \left(\sqrt{T}\bar{V}_T\right)' \left(\frac{1}{T} \sum_{t=\lfloor T\lambda_j \rfloor + 1}^{\lfloor T\lambda_{j+1} \rfloor} x_t x_t'\right) \left(\sqrt{T}\bar{V}_T\right)' \\ & - B_T(\lambda_{j+1})' \left(\frac{1}{T} \sum_{t=1}^{\lfloor T\lambda_{j+1} \rfloor} z_t z_t'\right)^{-1} B_T(\lambda_{j+1}) \\ & + \left(\sqrt{T}V_T\right)' \cdot \left(\frac{1}{T} \sum_{t=1}^{\lfloor T\lambda_{j+1} \rfloor} x_t z_t'\right) \left(\frac{1}{T} \sum_{t=1}^{\lfloor T\lambda_{j+1} \rfloor} z_t z_t'\right)^{-1} \left[2 \cdot B_T(\lambda_{j+1}) - \left(\frac{1}{T} \sum_{t=1}^{\lfloor T\lambda_{j+1} \rfloor} x_t z_t'\right)' \left(\sqrt{T}V_T\right)\right] \\ & + B_T(\lambda_j)' \left(\frac{1}{T} \sum_{t=1}^{\lfloor T\lambda_j \rfloor} z_t z_t'\right)^{-1} B_T(\lambda_j) \\ & - \left(\sqrt{T}V_T\right)' \cdot \left(\frac{1}{T} \sum_{t=1}^{\lfloor T\lambda_j \rfloor} x_t z_t'\right) \left(\frac{1}{T} \sum_{t=1}^{\lfloor T\lambda_j \rfloor} z_t z_t'\right)^{-1} \left[2 \cdot B_T(\lambda_j) - \left(\frac{1}{T} \sum_{t=1}^{\lfloor T\lambda_j \rfloor} x_t z_t'\right)' \left(\sqrt{T}V_T\right)\right] \\ & + (B_T(\lambda_{j+1}) - B_T(\lambda_j))' \left(\frac{1}{T} \sum_{t=\lfloor T\lambda_j \rfloor + 1}^{\lfloor T\lambda_{j+1} \rfloor} z_t z_t'\right)^{-1} (B_T(\lambda_{j+1}) - B_T(\lambda_j)) \\ & - 2 \cdot \left(\sqrt{T}\bar{V}_T\right)' \left(\frac{1}{T} \sum_{t=\lfloor T\lambda_j \rfloor + 1}^{\lfloor T\lambda_{j+1} \rfloor} x_t z_t'\right) \left(\frac{1}{T} \sum_{t=\lfloor T\lambda_j \rfloor + 1}^{\lfloor T\lambda_{j+1} \rfloor} z_t z_t'\right)^{-1} (B_T(\lambda_{j+1}) - B_T(\lambda_j)) \\ & + \left(\sqrt{T}\bar{V}_T\right)' \left(\frac{1}{T} \sum_{t=\lfloor T\lambda_j \rfloor + 1}^{\lfloor T\lambda_{j+1} \rfloor} x_t z_t'\right) \left(\frac{1}{T} \sum_{t=\lfloor T\lambda_j \rfloor + 1}^{\lfloor T\lambda_{j+1} \rfloor} z_t z_t'\right)^{-1} \left(\frac{1}{T} \sum_{t=\lfloor T\lambda_j \rfloor + 1}^{\lfloor T\lambda_{j+1} \rfloor} x_t z_t'\right)' \left(\sqrt{T}\bar{V}_T\right)\end{aligned}$$

which converges in distribution to

$$
\begin{aligned}
F_j^* = &-\sigma^2 \lambda_{j+1}^{-1} \cdot W^k(\lambda_{j+1})' W^k(\lambda_{j+1}) + A^{*\prime} Q_{xz} Q_z^{-1} \left[2\sigma Q_z^{\frac{1}{2}} \cdot W^k(\lambda_{j+1}) - \lambda_{j+1} \cdot Q_{zx} A^* \right] \\
&+ \sigma^2 \lambda_j^{-1} \cdot W^k(\lambda_j)' W^k(\lambda_j) - A^{*\prime} Q_{xz} Q_z^{-1} \left[2\sigma Q_z^{\frac{1}{2}} \cdot W^k(\lambda_j) - \lambda_j \cdot Q_{zx} A^* \right] \\
&+ \sigma^2 (\lambda_{j+1} - \lambda_j)^{-1} \cdot \left(W^k(\lambda_{j+1}) - W^k(\lambda_j) \right)' \left(W^k(\lambda_{j+1}) - W^k(\lambda_j) \right) \\
&- 2 \cdot A^{*\prime} Q_{xz} Q_z^{-1} \left(\sigma Q_z^{\frac{1}{2}} \left(W^k(\lambda_{j+1}) - W^k(\lambda_j) \right) \right) \\
&+ (\lambda_{j+1} - \lambda_j) \cdot A^{*\prime} Q_{xz} Q_z^{-1} Q_{zx} A^* \\
= &-\sigma^2 \lambda_{j+1}^{-1} \cdot W^k(\lambda_{j+1})' W^k(\lambda_{j+1}) + \sigma^2 \lambda_j^{-1} \cdot W^k(\lambda_j)' W^k(\lambda_j) \\
&+ \sigma^2 (\lambda_{j+1} - \lambda_j)^{-1} \left(W^k(\lambda_{j+1}) - W^k(\lambda_j) \right)' \left(W^k(\lambda_{j+1}) - W^k(\lambda_j) \right) \\
= &\frac{\sigma^2}{\lambda_{j+1}\lambda_j(\lambda_{j+1} - \lambda_j)} [-\lambda_j(\lambda_{j+1} - \lambda_j) \cdot W^k(\lambda_{j+1})' W^k(\lambda_{j+1}) + \lambda_{j+1}(\lambda_{j+1} - \lambda_j) \cdot W^k(\lambda_j)' W^k(\lambda_j) \\
&+ \lambda_{j+1}\lambda_j \cdot \left(W^k(\lambda_{j+1}) - W^k(\lambda_j) \right)' \left(W^k(\lambda_{j+1}) - W^k(\lambda_j) \right)] \\
= &\frac{\sigma^2}{\lambda_{j+1}\lambda_j(\lambda_{j+1} - \lambda_j)} \left[\lambda_j^2 \cdot W^k(\lambda_{j+1})' W^k(\lambda_{j+1}) - 2 \cdot \lambda_{j+1}\lambda_j \cdot W^k(\lambda_{j+1})' W^k(\lambda_j) + \lambda_{j+1}^2 \cdot W^k(\lambda_j)' W^k(\lambda_j) \right] \\
= &\frac{\sigma^2}{\lambda_{j+1}\lambda_j(\lambda_{j+1} - \lambda_j)} \cdot \left| \lambda_j \cdot W^k(\lambda_{j+1}) - \lambda_{j+1} \cdot W^k(\lambda_j) \right|^2.
\end{aligned}
$$

The Convergence Properties of $F^*_{T,0}$

Now we turn to the asymptotic properties of the term

$$F^*_{T,0} = D^R(1,1) - D^U(1,1).$$

In this case, since

$$Y_{1,1} = Y_1, \quad X_{1,1} = X_1, \quad Z_{1,1} = Z_1, \quad \text{and} \quad U_{1,1} = U_1,$$

we can see that

$$\begin{aligned}
F^*_{T,0} &= \left| M_{Z_{1,1}} \left(U_{1,1} - X_{1,1} V_T \right) \right|^2 - \left| M_{Z_1} \left(U_1 - X_1 \bar{V}_T \right) \right|^2 \\
&= \left(U_1 - X_1 V_T \right)' M_{Z_1} \left(U_1 - X_1 V_T \right) - \left(U_1 - X_1 \bar{V}_T \right)' M_{Z_1} \left(U_1 - X_1 \bar{V}_T \right) \\
&= (\bar{V}_T - V_T)' X_1' M_{Z_1} X_1 (\bar{V}_T - V_T) + 2 \cdot \left(U_1 - X_1 V_T \right)' M_{Z_1} X_1 (\bar{V}_T - V_T) \\
&= \left[\sqrt{T} (\bar{V}_T - V_T) \right]' \left[\frac{1}{T} X_1' X_1 - \left(\frac{1}{T} X_1' Z_1 \right) \left(\frac{1}{T} Z_1' Z_1 \right)^{-1} \left(\frac{1}{T} X_1' Z_1 \right)' \right] \left[\sqrt{T} (\bar{V}_T - V_T) \right] \\
&\quad + 2 \cdot \left[\frac{1}{\sqrt{T}} U_1 - \left(\frac{1}{T} X_1 \right) \left(\sqrt{T} V_T \right) \right]' M_{Z_1} \left[\sqrt{T} (\bar{V}_T - V_T) \right] \\
&\leq \left| \sqrt{T} \left(\hat{V}_T - V_T \right) \right|^2 \cdot \left\| \frac{1}{T} X_1' X_1 - \left(\frac{1}{T} X_1' Z_1 \right) \left(\frac{1}{T} Z_1' Z_1 \right)^{-1} \left(\frac{1}{T} X_1' Z_1 \right)' \right\| \\
&\quad + 2 \cdot \left| \frac{1}{\sqrt{T}} U_1 - \left(\frac{1}{T} X_1 \right) \left(\sqrt{T} V_T \right) \right| \cdot \left| \sqrt{T} (\bar{V}_T - V_T) \right|.
\end{aligned}$$

Since

$$\frac{1}{T} X_1' X_1 - \left(\frac{1}{T} X_1' Z_1 \right) \left(\frac{1}{T} Z_1' Z_1 \right)^{-1} \left(\frac{1}{T} X_1' Z_1 \right)' \xrightarrow{p} Q_x - Q_{xz} Q_z^{-1} Q_{zx}$$

and

$$\left| \frac{1}{\sqrt{T}} U_1 - \left(\frac{1}{T} X_1 \right) \left(\sqrt{T} V_T \right) \right|^2 \leq 2 \cdot \frac{1}{T} U_1' U_1 + 2 \cdot \left(\frac{1}{T} \sum_{t=1}^{T_1} x_t' x_t \right) \cdot \left| \sqrt{T} V_T \right|^2,$$

where

$$2 \cdot \frac{1}{T} U_1' U_1 + 2 \cdot \left(\frac{1}{T} \sum_{t=1}^{T_1} x_t' x_t \right) \cdot \left| \sqrt{T} V_T \right|^2 \xrightarrow{p} 2\lambda_1 \cdot \left(\sigma^2 + \operatorname{tr}(Q_x) \left| A^* \right|^2 \right),$$

it follows that all the terms aside from $\sqrt{T} \left(\bar{V}_T - V_T \right)$ are $O_p(1)$. Since $\sqrt{T} \left(\bar{V}_T - V_T \right) = o_p(1)$, we can see that

$$F^*_{T,0} = o_p(1).$$

The Weak Limit of $F_T(\lambda_1, \cdots, \lambda_m; k)$

We have seen that

$$F_{T,j}^* \xrightarrow{d} F_j^*$$

for any $1 \le j \le m$ and that $F_{T,0}^* = o_p(1)$ under the null. Therefore,

$$\begin{aligned}
F_T^* &= \sum_{j=1}^{m} \Big(D^R(1, j+1) - D^R(1, j) - D^U(j+1, j+1) \Big) + D^R(1,1) - D^U(1,1) \\
&= \sum_{j=1}^{m} F_{T,j}^* + F_{T,0}^* \\
&\xrightarrow{d} \sum_{j=1}^{m} F_j^* = \sigma^2 \sum_{j=1}^{m} \frac{\left| \lambda_j \cdot W^k(\lambda_{j+1}) - \lambda_{j+1} \cdot W^k(\lambda_j) \right|^2}{\lambda_{j+1}\lambda_j(\lambda_{j+1} - \lambda_j)}
\end{aligned}$$

under H_0. Putting this together with the asymptotic properties of SSR_m, we can see that, under H_0,

$$\begin{aligned}
F_T(\lambda_1, \cdots, \lambda_m; k) &= \frac{1}{km} \cdot \frac{F_T^*}{SSR_m/(T - (m+1)k - p)} \\
&\xrightarrow{d} \frac{1}{km} \sum_{j=1}^{m} \frac{\left| \lambda_j \cdot W^k(\lambda_{j+1}) - \lambda_{j+1} \cdot W^k(\lambda_j) \right|^2}{\lambda_{j+1}\lambda_j(\lambda_{j+1} - \lambda_j)}.
\end{aligned}$$

For any $1 \le j \le m$, define

$$\bar{F}_j = \frac{\left| \lambda_j \cdot W^k(\lambda_{j+1}) - \lambda_{j+1} \cdot W^k(\lambda_j) \right|^2}{\lambda_{j+1}\lambda_j(\lambda_{j+1} - \lambda_j)}.$$

Note that

$$\begin{aligned}
\lambda_j \cdot W^k(\lambda_{j+1}) - \lambda_{j+1} \cdot W^k(\lambda_j) &= \lambda_j \left(W^k(\lambda_{j+1}) - W^k(\lambda_j) \right) + (\lambda_j - \lambda_{j+1}) \cdot W^k(\lambda_j) \\
&\sim N\left[\mathbf{0}, \left(\lambda_j^2 \cdot (\lambda_{j+1} - \lambda_j) + (\lambda_{j+1} - \lambda_j)^2 \cdot \lambda_j \right) \cdot I_k \right] \\
&= N\left[\mathbf{0}, \lambda_{j+1}\lambda_j(\lambda_{j+1} - \lambda_j) \cdot I_k \right]
\end{aligned}$$

since $W^k(\lambda_{j+1}) - W^k(\lambda_j)$ and $W^k(\lambda_j)$ are independent Gaussian random vectors with mean zero and covariance matrix $(\lambda_{j+1} - \lambda_j) \cdot I_k$ and $\lambda_j \cdot I_k$. Therefore,

$$\bar{F}_j = \left(\lambda_j \cdot W^k(\lambda_{j+1}) - \lambda_{j+1} \cdot W^k(\lambda_j) \right)' (\lambda_{j+1}\lambda_j(\lambda_{j+1} - \lambda_j) \cdot I_k)^{-1} \left(\lambda_j \cdot W^k(\lambda_{j+1}) - \lambda_{j+1} \cdot W^k(\lambda_j) \right)$$

is a chi squared random variable with k degrees of freedom.
As such, the weak limit $\sum_{j=1}^{m} \frac{\bar{F}_j}{m}$ of $k \cdot F_T(\lambda_1, \cdots, \lambda_m; k)$ is the sum of m dependent chi squared variables with k degrees of freedom, each divided by the number of breaks m.

The Weak Limit of the Supremum Statistic

Now we are in a position to derive the limit of the test statistic under the null. For any $(\lambda_1, \cdots, \lambda_m) \in \Lambda_\varepsilon$, we saw that each $F_T(\lambda_1, \cdots, \lambda_m; k)$ can be expressed as a continuous function of $V_T, \bar{V}_T, A_T, B_T$ and the mapping

$$r \mapsto \frac{1}{T} \sum_{t=1}^{\lfloor Tr \rfloor} w_t w_t',$$

all of which converge uniformly in $[0,1]$. This indicates that

$$\begin{aligned}\sup F_T(m;k) &= \sup_{(\lambda_1,\cdots,\lambda_m)\in\Lambda_\varepsilon} F_T(\lambda_1, \cdots, \lambda_m; k) \\ &\xrightarrow{d} \sup_{(\lambda_1,\cdots,\lambda_m)\in\Lambda_\varepsilon} \left(\frac{1}{km} \sum_{j=1}^{m} \frac{\left|\lambda_j \cdot W^k(\lambda_{j+1}) - \lambda_{j+1} \cdot W^k(\lambda_j)\right|^2}{\lambda_{j+1}\lambda_j(\lambda_{j+1} - \lambda_j)} \right)\end{aligned}$$

under H_0.

We only provide a heuristic reasoning for the consistency of the above test statistic.
If the alternative is true, that is, if there truly are m breaks in the model, then at the true break fractions $\lambda_1, \cdots, \lambda_m$, the sum of squared residuals SSR_0 subject to the restriction that there are no breaks will be much larger than SSR_m. Therefore, the supremum of the F statistics over the collection Λ_ε will also be much greater than 0, provided that the true fractions are contained in Λ_ε. This means that the sup F statistic will then be large and thus more likely to reject the null.

7.1.4 Double Maximum Tests

So far, we have studied the asymptotic distribution of the sup F test statistic under the null, where the alternative is that there are m breaks. However, we may wish not to specify the number of breaks under the alternative. In this case, we can choose a large enough $M \in N_+$ to serve as the maximum possible number of breaks and test the null hypothesis

$$H_0: \text{ There are no breaks}$$

against the alternative

$$H_1: \text{ There are at most } M \text{ breaks}$$

using the double maximum statistic

$$\begin{aligned}\text{Dmax } F_T(M, a_1, \cdots, a_M; k) &= \max_{1 \le m \le M} a_m \cdot (\sup F_T(m;k)) \\ &= \max_{1 \le m \le M} a_m \cdot \left(\sup_{(\lambda_1, \cdots, \lambda_m) \in \Lambda_\varepsilon} F_T(\lambda_1, \cdots, \lambda_m; k) \right),\end{aligned}$$

where $a_1, \cdots, a_M > 0$ are exogenous weights assigned to each possible number of breaks. If $a_1 = \cdots = a_M = 1$, then we call the above statistic the uniform double maximum test statistic. Under H_0, the asymptotic distribution of the test statistic follows directly from the above derivation; since

$$\sup_{(\lambda_1, \cdots, \lambda_m) \in \Lambda_\varepsilon} F_T(\lambda_1, \cdots, \lambda_m; k) \xrightarrow{d} \sup_{(\lambda_1, \cdots, \lambda_m) \in \Lambda_\varepsilon} \left(\frac{1}{km} \sum_{j=1}^{m} \frac{\left| \lambda_j \cdot W^k(\lambda_{j+1}) - \lambda_{j+1} \cdot W^k(\lambda_j) \right|^2}{\lambda_{j+1} \lambda_j (\lambda_{j+1} - \lambda_j)} \right)$$

for any $1 \le m \le M$ under the null of no breaks, and because the maximum over a finite set is a continuous function, by the continuous mapping theorem

$$\text{Dmax } F_T(M, a_1, \cdots, a_M; k) \xrightarrow{d} \max_{1 \le m \le M} a_m \cdot \left[\sup_{(\lambda_1, \cdots, \lambda_m) \in \Lambda_\varepsilon} \left(\frac{1}{km} \sum_{j=1}^{m} \frac{\left| \lambda_j \cdot W^k(\lambda_{j+1}) - \lambda_{j+1} \cdot W^k(\lambda_j) \right|^2}{\lambda_{j+1} \lambda_j (\lambda_{j+1} - \lambda_j)} \right) \right]$$

under H_0.

Heuristically, the consistency of the test follows because, if the true model has $1 \le m \le M$ breaks, then the sup F statistic corresponding to the alternative of m breaks will be large, and as such the maximum of the sup F statistics corresponding to each possible number of breaks will also be large, making it likely that the null will be rejected.

7.2 Sequentially Testing for Breaks

In the previous section, we studied testing the null of no breaks against the alternative of m breaks, or at most M breaks. Here we study how to test sequentially for m breaks against $m+1$ breaks for $m \geq 0$. To test for no breaks against the alternative of a single break, we can simply use the test devised above, so we focus on the case where $m \geq 1$.

The intuition behind the test is to test for a break in each of the $m+1$ regimes under the null hypothesis, or the assumption that there are m breaks. Taking the maximum difference in sum of squared residuals from putting an additional break in each of the $m+1$ regimes now becomes our test statistic, since if the maximum is large, then it means that a huge reduction in SSR can be achieved by placing an additional break in between two extant breaks.

Formally, the test statistic given the hypotheses

$$H_0: \text{ There are } m \text{ breaks} \quad \text{v.s.} \quad H_1: \text{ There are } m+1 \text{ breaks}$$

is given by

$$\begin{aligned} F_T(m+1 \mid m) &= \frac{1}{\hat{\sigma}^2}\left(SSR(\hat{T}_1,\cdots,\hat{T}_m) - \min_{1\leq j\leq m+1} \inf_{\lambda\in\Lambda_{j,\eta}} SSR(\hat{T}_1,\cdots,\hat{T}_{j-1},\lfloor T\lambda\rfloor,\hat{T}_j,\cdots,\hat{T}_m)\right) \\ &= \frac{1}{\hat{\sigma}^2}\max_{1\leq j\leq m+1}\sup_{\tau\in\Lambda_{j,\eta}}\Big(SSR(\hat{T}_1,\cdots,\hat{T}_m) - SSR(\hat{T}_1,\cdots,\hat{T}_{j-1},\tau,\hat{T}_j,\cdots,\hat{T}_m)\Big), \end{aligned}$$

where

$$\Lambda_{j,\eta} = \left\{\lambda \in [0,1] \mid \frac{\lambda - \hat{\lambda}_{j-1}}{\hat{\lambda}_j - \hat{\lambda}_{j-1}} \in [\eta, 1-\eta]\right\}$$

for some $\eta > 0$, so that the additional break $\tau = \lfloor T\lambda \rfloor$ is not chosen too close to some existing break, $\hat{\sigma}^2$ is a consistent estimator of the error variance σ^2, and $\hat{T}_1,\cdots,\hat{T}_m$ are the break dates estimated via least squares.

To derive the asymptotic distribution of the above test statistic, we assume that the difference between the estimated break dates and the true break dates do not diverge, that is,

$$\hat{T}_j - T_j^0 = O_p(1)$$

for any $1 \leq j \leq m$ under H_0. This simply states that the estimated break fractions converge to the true break fractions at rate T.
We assume the above in addition to assumptions (1) to (3) above, and, for simplicity, that the model is one of a pure structural break.

7.2.1 The Asymptotic Distribution of the Test Statistic

The basic idea of the proof is the same as that of the test for no breaks against m breaks, that is, to decompose the sum of squared residuals into the sum of SSRs in each regime.
The pure structural break model assumption above implies that the model is given by

$$y_t = z_t'\delta_j^0 + u_t \qquad \text{if } T_{j-1}^0 + 1 \leq t \leq T_j^0$$

for the true parameter $\delta^0 = (\delta_1^{0\prime}, \cdots, \delta_m^{0\prime})'$ under the null hypothesis H_0.
We again proceed in steps.

The SSR under m breaks

Letting $\hat{T}_1, \cdots, \hat{T}_m$ be the break dates estimated via least squares, the estimated coefficients $\hat{\delta}$ are given by

$$\hat{\delta} = \left(\bar{Z}'\bar{Z}\right)^{-1} \bar{Z}'Y$$

and the estimator of each $\hat{\delta}_j$ by

$$\begin{aligned}\hat{\delta}_j &= (Z_j'Z_j)^{-1} Z_j'Y_j \\ &= \left(\sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} z_t z_t'\right)^{-1} \sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} z_t y_t,\end{aligned}$$

where

$$\bar{Z} = \operatorname{diag}(Z_1, \cdots, Z_{m+1})$$

is the diagonal partition of Z according to the breaks $\hat{T}_1, \cdots, \hat{T}_m$, and $Y = (Y_1', \cdots, Y_{m+1}')'$ is a similar partition for Y.
The sum of squared residuals under m breaks is then given by

$$\begin{aligned}SSR(\hat{T}_1, \cdots, \hat{T}_m) &= \left(Y - \bar{Z}\hat{\delta}\right)' \left(Y - \bar{Z}\hat{\delta}\right) \\ &= Y'M_{\bar{Z}}Y.\end{aligned}$$

Because $\hat{\lambda}_j \xrightarrow{p} \lambda_j^0$ for any $1 \leq j \leq m$, we can assume that, under large T,

$$\left|\hat{\lambda}_j - \lambda_j^0\right| < \frac{\varepsilon}{2}$$

with large probability for every $1 \leq j \leq m$, so that each $\hat{\lambda}_j$ lies between λ_{j-1}^0 and λ_{j+1}^0 (the

distance between each true break fraction is larger than ε).
Letting $\bar{Z}^0$ be the diagonal partition of Z according to the true break points $T_1^0, \cdots, T_m^0$, we have

$$Y = \bar{Z}^0 \delta^0 + U,$$

and as such

$$\begin{aligned} SSR(\hat{T}_1, \cdots, \hat{T}_m) - U' M_{\bar{Z}} U &= Y' M_{\bar{Z}} Y - U' M_{\bar{Z}} U \\ &= \delta^{0\prime} \bar{Z}^{0\prime} M_{\bar{Z}} \bar{Z}^0 \delta^0 + 2 \cdot \delta^{0\prime} \bar{Z}^{0\prime} M_{\bar{Z}} U \\ &= \delta^{0\prime} \left(\bar{Z}^0 - \bar{Z} \right)' M_{\bar{Z}} \left(\bar{Z}^0 - \bar{Z} \right) \delta^0 + 2 \cdot \delta^{0\prime} \left(\bar{Z}^0 - \bar{Z} \right)' M_{\bar{Z}} U. \end{aligned}$$

Note that, for large T,

$$\left\| \frac{1}{T} \left(\bar{Z}^0 - \bar{Z} \right) \right\|^2 \leq \frac{1}{T} \sum_{j=1}^{m+1} \left| \hat{T}_j - T_j^0 \right| \left(\max_{1 \leq t \leq T} |z_t|^2 \right)$$

with large probability. For any $\delta > 0$, because $\hat{T}_j - T_j^0 = O_p(1)$ for any $1 \leq j \leq m$,

$$\sum_{j=1}^{m+1} \left| \hat{T}_j - T_j^0 \right| = O_p(1)$$

and there exists a $C > 0$ such that, for large T,

$$\mathbb{P} \left(\sum_{j=1}^{m+1} \left| \hat{T}_j - T_j^0 \right| > C \right)$$

is small for large T. Additionally, by assumption

$$M = \sup_{t \in \mathbb{Z}} \mathbb{E} |z_t|^4 < +\infty.$$

Then, we can see that

$$\begin{aligned} \mathbb{P} \left(\frac{1}{T} \max_{1 \leq t \leq T} |z_t|^2 > \frac{\delta}{C} \right) &\leq \sum_{t=1}^{T} \mathbb{P} \left(\frac{1}{T} |z_t|^2 > \frac{\delta}{C} \right) \\ &\leq \sum_{t=1}^{T} \frac{C^2}{\delta^2} \cdot \frac{1}{T^2} \mathbb{E} |z_t|^4 \\ &\leq \left(M \cdot \frac{C^2}{\delta^2} \right) \frac{1}{T}, \end{aligned}$$

and as such

$$\mathbb{P} \left(\left\| \frac{1}{T} \left(\bar{Z}^0 - \bar{Z} \right) \right\|^2 > \delta \right) \leq \mathbb{P} \left(\sum_{j=1}^{m+1} \left| \hat{T}_j - T_j^0 \right| > C \right) + \mathbb{P} \left(\frac{1}{T} \left(\max_{1 \leq t \leq T} |z_t|^2 \right) > \frac{\delta}{C} \right)$$

$$\leq \mathbb{P}\left(\sum_{j=1}^{m+1}\left|\hat{T}_j - T_j^0\right| > C\right) + \left(M \cdot \frac{C^2}{\delta^2}\right)\frac{1}{T}$$

is small for large T. Thus, we can conclude that

$$\lim_{T\to\infty} \mathbb{P}\left(\left\|\frac{1}{T}\left(\bar{Z}^0 - \bar{Z}\right)\right\|^2 > \delta\right) = 0.$$

This holds for any $\delta > 0$, so

$$\left\|\frac{1}{T}\left(\bar{Z}^0 - \bar{Z}\right)\right\|^2 \xrightarrow{p} 0.$$

We saw previously that

$$\frac{1}{\sqrt{T}}U = O_p(1),$$

so

$$SSR(\hat{T}_1, \cdots, \hat{T}_m) - U' M_{\bar{Z}} U = o_p(1).$$

The SSR with an aditional break

Choose any $1 \le i \le m$ and $\lambda_\tau \in \Lambda_{i,\eta}$. Denote $\tau = \lfloor T\lambda_\tau \rfloor$.
Because $\hat{\lambda}_j \xrightarrow{p} \lambda_j^0$ for any $1 \le j \le m$, we can assume that, under large T,

$$\left|\hat{\lambda}_j - \lambda_j^0\right| < \varepsilon\eta$$

with large probability for any $1 \le j \le m$. Since

$$\hat{\lambda}_{i-1} + \left(\hat{\lambda}_i - \hat{\lambda}_{i-1}\right)\eta \le \lambda_\tau \le \hat{\lambda}_i - \left(\hat{\lambda}_i - \hat{\lambda}_{i-1}\right)\eta,$$

this means that

$$\begin{aligned}\lambda_\tau - \lambda_{i-1}^0 &\ge \left(\hat{\lambda}_i - \hat{\lambda}_{i-1}\right)\eta + \hat{\lambda}_{i-1} - \lambda_{i-1}^0 \\ &> \varepsilon\eta - \varepsilon\eta = 0,\end{aligned}$$

so that $\lambda_{i-1}^0 < \lambda_\tau$ with large probability for large T. Similarly, $\lambda_\tau < \lambda_i^0$ with large probability for large T.

Letting

$$\bar{Z}^* = \text{diag}(Z_1^*, \cdots, Z_{i-1}^*, Z_\Delta^*, Z_i^*, \cdots, Z_{m+1}^*)$$

be the diagonal partition of Z under the break points $\hat{T}_1, \cdots, \hat{T}_{i-1}, \tau, \hat{T}_i, \cdots, \hat{T}_m$ and

$$Y = (Y_1^{*\prime}, \cdots, Y_{i-1}^{*\prime}, Y_\Delta^{*\prime}, Y_i^{*\prime}, \cdots, Y_{m+1}^{*\prime})$$

a similar partition of Y according to the above break points, we can see that

$$Z_j = Z_j^* \quad \text{and} \quad Y_j = Y_j^*$$

for any $1 \le j \le i-1$ and $i+1 \le j \le m$, while

$$Z_i = \begin{pmatrix} Z_\Delta^* \\ Z_i^* \end{pmatrix} \quad \text{and} \quad Y_i = \begin{pmatrix} Y_\Delta^* \\ Y_i^* \end{pmatrix}.$$

The sum of squared residuals under the above break points is given by

$$SSR(\hat{T}_1, \cdots, \hat{T}_{i-1}, \tau, \hat{T}_i, \cdots, \hat{T}_m) = Y' M_{\bar{Z}^*} Y.$$

Letting $\bar{Z}^{0*}$ be the partition of Z under the "true" break points $T_1^0, \cdots, T_{i-1}^0, \tau, T_i^0, \cdots, T_m^0$, and defining

$$\delta^{0*} = \begin{pmatrix} \delta_1^0 \\ \vdots \\ \delta_{i-1}^0 \\ \delta_\Delta^0 \\ \delta_i^0 \\ \vdots \\ \delta_m^0 \end{pmatrix},$$

where $\delta_\Delta^0 = \delta_i^0$, we have

$$Y = \bar{Z}^{0*}\delta^{0*} + U$$

and as such

$$\begin{aligned} SSR(\hat{T}_1, \cdots, \hat{T}_{i-1}, \tau, \hat{T}_i, \cdots, \hat{T}_m) - U' M_{\bar{Z}^*} U &= \delta^{0*\prime} \bar{Z}^{0*\prime} M_{\bar{Z}^*} \bar{Z}^{0*} \delta^{0*} + 2 \cdot \delta^{0*\prime} \bar{Z}^{0*\prime} M_{\bar{Z}^*} U \\ &= \delta^{0*\prime} \left(\bar{Z}^{0*} - \bar{Z}^*\right)' M_{\bar{Z}^*} \left(\bar{Z}^{0*} - \bar{Z}^*\right) \delta^{0*} + 2 \cdot \delta^{0*\prime} \left(\bar{Z}^{0*} - \bar{Z}^*\right)' M_{\bar{Z}^*} U. \end{aligned}$$

Since the squared norm of the difference $\bar{Z}^{0*} - \bar{Z}^*$ is, as before, a sum of

$$\sum_{j=1}^{m} \left|\hat{T}_j - T_j^0\right|$$

elements, $\frac{1}{T}\left(\bar{Z}^{0*} - \bar{Z}^*\right) = o_p(1)$ and

$$SSR(\hat{T}_1, \cdots, \hat{T}_{i-1}, \tau, \hat{T}_i, \cdots, \hat{T}_m) - U' M_{\bar{Z}^*} U = o_p(1).$$

Decomposing the Difference of Sum of Squared Residuals

Define

$$F_T^* = U' M_{\bar{Z}} U - U' M_{\bar{Z}^*} U.$$

We have

$$U' M_{\bar{Z}} U = \sum_{j=1}^{m} U_j' Z_j (Z_j' Z_j)^{-1} Z_j' U_j,$$

where

$$U_j = \begin{pmatrix} u_{\hat{T}_{j-1}+1} \\ \vdots \\ u_{\hat{T}_j} \end{pmatrix}$$

for $1 \le j \le m$, and likewise,

$$\begin{aligned} U' M_{\bar{Z}^*} U = &\sum_{j=1}^{i-1} U_j' M_{Z_j} U_j + \sum_{j=i+1}^{m} U_j' M_{Z_j} U_j \\ &+ U_\Delta^{*\prime} M_{Z_\Delta^*} U_\Delta^* + U_i^{*\prime} M_{Z_i^*} U_i^*, \end{aligned}$$

where

$$U_\Delta^* = \begin{pmatrix} u_{\hat{T}_{i-1}+1} \\ \vdots \\ u_\tau \end{pmatrix} \quad \text{and} \quad U_i^* = \begin{pmatrix} u_{\tau+1} \\ \vdots \\ u_{\hat{T}_i} \end{pmatrix}.$$

Seeing as how

$$U_i = \begin{pmatrix} U_\Delta^* \\ U_i^* \end{pmatrix},$$

we have

$$\begin{aligned} F_T^* &= U_i' M_{Z_i} U_i - U_\Delta^{*\prime} M_{Z_\Delta^*} U_\Delta^* - U_i^{*\prime} M_{Z_i^*} U_i^* \\ &= U_i' U_i - U_i' Z_i^0 \left(Z_i' Z_i\right)^{-1} Z_i' U_i - U_\Delta^{*\prime} U_\Delta^* + U_\Delta^{*\prime} Z_\Delta^* \left(Z_\Delta^{*\prime} Z_\Delta^*\right)^{-1} Z_\Delta^{*\prime} U_\Delta^{0*} \\ &\quad - U_i^{*\prime} U_i^* + U_i^{*\prime} Z_i^* \left(Z_i^{*\prime} Z_i^*\right)^{-1} Z_i^{*\prime} U_i^* \\ &= -U_i' Z_i \left(Z_i' Z_i\right)^{-1} Z_i' U_i + U_\Delta^{0*\prime} Z_\Delta^* \left(Z_\Delta^{*\prime} Z_\Delta^*\right)^{-1} Z_\Delta^{*\prime} U_\Delta^* \\ &\quad + U_i^{*\prime} Z_i^* \left(Z_i^{*\prime} Z_i^*\right)^{-1} Z_i^{*\prime} U_i^* \\ &= -\Big(S_T(\hat{\lambda}_i) - S_T(\hat{\lambda}_{i-1})\Big)' \Big(M_T(\hat{\lambda}_i) - M_T(\hat{\lambda}_{i-1})\Big)^{-1} \Big(S_T(\hat{\lambda}_i) - S_T(\hat{\lambda}_{i-1})\Big) \\ &\quad + \Big(S_T(\lambda_\tau) - S_T(\hat{\lambda}_{i-1})\Big)' \Big(M_T(\lambda_\tau) - M_T(\hat{\lambda}_{i-1})\Big)^{-1} \Big(S_T(\lambda_\tau) - S_T(\hat{\lambda}_{i-1})\Big) \end{aligned}$$

$$+\left(S_T(\hat{\lambda}_i)-S_T(\lambda_\tau)\right)'\left(M_T(\hat{\lambda}_i)-M_T(\lambda_\tau)\right)^{-1}\left(S_T(\hat{\lambda}_i)-S_T(\lambda_\tau)\right).$$

Define

$$M_T(\lambda)=\frac{1}{T}\sum_{t=1}^{\lfloor T\lambda\rfloor} z_t z_t'$$

for any $\lambda\in[0,1]$.

Note that

$$\left|S_T(\hat{\lambda}_i)-S_T(\lambda_i^0)\right|\leq\frac{1}{\sqrt{T}}\left|\hat{T}_i-T_i^0\right|\left(\max_{1\leq t\leq T}|z_t|\right).$$

For any $\delta>0$ and $\zeta>0$, because $\hat{T}_i-T_i^0=O_p(1)$, there exists a $C>0$ such that, for large T,

$$\mathbb{P}\left(\left|\hat{T}_i-T_i^0\right|>C\right)<\frac{\zeta}{2}.$$

Additionally, by assumption

$$M=\sup_{t\in\mathbb{Z}}\mathbb{E}|z_t|^4<+\infty.$$

Then, we can see that

$$\begin{aligned}\mathbb{P}\left(\frac{1}{\sqrt{T}}\max_{1\leq t\leq T}|z_t|>\frac{\delta}{C}\right)&\leq\sum_{t=1}^{T}\mathbb{P}\left(\frac{1}{\sqrt{T}}|z_t|>\frac{\delta}{C}\right)\\&\leq\sum_{t=1}^{T}\frac{C^4}{\delta^4}\cdot\frac{1}{T^2}\mathbb{E}|z_t|^4\\&\leq\left(M\cdot\frac{C^4}{\delta^4}\right)\frac{1}{T},\end{aligned}$$

and as such

$$\begin{aligned}\mathbb{P}\left(\left|S_T(\hat{\lambda}_i)-S_T(\lambda_i^0)\right|>\delta\right)&\leq\mathbb{P}\left(\left|\hat{T}_i-T_i^0\right|>C\right)+\mathbb{P}\left(\frac{1}{\sqrt{T}}\left(\max_{1\leq t\leq T}|z_t|\right)>\frac{\delta}{C}\right)\\&\leq\frac{\zeta}{2}+\left(M\cdot\frac{C^4}{\delta^4}\right)\frac{1}{T}\\&<\zeta\end{aligned}$$

for large T. Thus, we can conclude that

$$\mathbb{P}\left(\left|S_T(\hat{\lambda}_i)-S_T(\lambda_i^0)\right|>\delta\right)<\zeta$$

for large T that depends only on δ and ζ, so by definition

$$\lim_{T\to\infty} \mathbb{P}\left(\left|S_T(\hat{\lambda}_i) - S_T(\lambda_i^0)\right| > \delta\right) = 0.$$

This holds for any $\delta > 0$, so

$$\left|S_T(\hat{\lambda}_i) - S_T(\lambda_i^0)\right| \xrightarrow{p} 0.$$

It follows then that

$$\begin{aligned} S_T(\hat{\lambda}_i) - S_T(\hat{\lambda}_{i-1}) &= \left(S_T(\hat{\lambda}_i) - S_T(\lambda_i^0)\right) + \left(S_T(\lambda_i^0) - S_T(\lambda_{i-1}^0)\right) + \left(S_T(\hat{\lambda}_{i-1}) - S_T(\lambda_{i-1}^0)\right) \\ &\xrightarrow{d} \sigma Q^{\frac{1}{2}} \cdot \left(W^k(\lambda_i^0) - W^k(\lambda_{i-1}^0)\right) \end{aligned}$$

by assumption (3). Similar arguments establish that

$$S_T(\lambda_\tau) - S_T(\hat{\lambda}_{i-1}) \xrightarrow{d} \sigma Q^{\frac{1}{2}} \cdot \left(W^k(\lambda_\tau) - W^k(\lambda_{i-1}^0)\right)$$

and

$$S_T(\hat{\lambda}_i) - S_T(\lambda_\tau) \xrightarrow{d} \sigma Q^{\frac{1}{2}} \cdot \left(W^k(\lambda_i^0) - W^k(\lambda_\tau)\right).$$

Similarly,

$$\left\|M_T(\hat{\lambda}_i) - M_T(\lambda_i^0)\right\| \le \frac{1}{T}\left|\hat{T}_i - T_i^0\right| \left(\max_{1\le t\le T} |z_t|^2\right),$$

and by the same line of reasoning as above, we have

$$\left\|M_T(\hat{\lambda}_i) - M_T(\lambda_i^0)\right\| \xrightarrow{p} 0.$$

It follows then that,

$$\begin{aligned} M_T(\hat{\lambda}_i) - M_T(\hat{\lambda}_{i-1}) &= \left(M_T(\hat{\lambda}_i) - M_T(\lambda_i^0)\right) + \left(M_T(\lambda_i^0) - M_T(\lambda_{i-1}^0)\right) + \left(M_T(\hat{\lambda}_{i-1}) - M_T(\lambda_{i-1}^0)\right) \\ &\xrightarrow{p} (\lambda_i^0 - \lambda_{i-1}^0) \cdot Q \end{aligned}$$

by assumption (1), which tells us that

$$\left(M_T(\hat{\lambda}_i) - M_T(\hat{\lambda}_{i-1})\right)^{-1} \xrightarrow{p} \frac{1}{\lambda_i^0 - \lambda_{i-1}^0} Q^{-1}.$$

Similar arguments allow us to establish that

$$\left(M_T(\hat{\lambda}_i) - M_T(\lambda_\tau)\right)^{-1} \xrightarrow{p} \frac{1}{\lambda_i^0 - \lambda_\tau} Q^{-1}$$

and

$$\left(M_T(\lambda_\tau) - M_T(\hat{\lambda}_{i-1})\right)^{-1} \xrightarrow{p} \frac{1}{\lambda_\tau - \lambda_{i-1}^0} Q^{-1}.$$

Therefore, F_T^* converges in distribution to

$$\sigma^2 \left[\frac{\left|W^k(\lambda_i^0) - W^k(\lambda_\tau)\right|^2}{\lambda_i^0 - \lambda_\tau} + \frac{\left|W^k(\lambda_\tau) - W^k(\lambda_{i-1}^0)\right|^2}{\lambda_\tau - \lambda_{i-1}^0} - \frac{\left|W^k(\lambda_i^0) - W^k(\lambda_{i-1}^0)\right|^2}{\lambda_i^0 - \lambda_{i-1}^0} \right],$$

which is identically distirbuted to

$$F^*(\lambda_\tau) = \sigma^2 \left[\frac{\left|W^k(\lambda_i^0 - \lambda_\tau)\right|^2}{\lambda_i^0 - \lambda_\tau} + \frac{\left|W^k(\lambda_\tau - \lambda_{i-1}^0)\right|^2}{\lambda_\tau - \lambda_{i-1}^0} - \frac{\left|W^k(\lambda_i^0 - \lambda_{i-1}^0) - W^k(\lambda_\tau - \lambda_{i-1}^2)\right|^2}{\lambda_i^0 - \lambda_{i-1}^0} \right].$$

Noting that

$$\frac{1}{\sqrt{\lambda_i^0 - \lambda_{i-1}^0}} W^k(\lambda_i^0 - \lambda_\tau) \sim W^k(1),$$

and, defining

$$\mu = \frac{\lambda_\tau - \lambda_{i-1}^0}{\lambda_i^0 - \lambda_{i-1}^0}.$$

that

$$\frac{1}{\sqrt{\lambda_\tau - \lambda_{i-1}^0}} W^k(\lambda_\tau - \lambda_{i-1}^0) \sim \mu^{-\frac{1}{2}} \cdot W^k(\mu)$$

and

$$\frac{1}{\lambda_i^0 - \lambda_\tau} \left(W^k(\lambda_i^0 - \lambda_{i-1}^0) - W^k(\lambda_\tau - \lambda_{i-1}^2)\right) \sim (1-\mu)^{-\frac{1}{2}} \cdot \left(W^k(1) - W^k(\mu)\right)$$

the weak limit can be rewritten as

$$\begin{aligned} F^*(\lambda_\tau) &\sim \sigma^2 \left[\frac{\left|W^k(1) - W^k(\mu)\right|^2}{1-\mu} + \frac{\left|W^k(\mu)\right|^2}{\mu} - W^k(1)^2 \right] \\ &= \frac{\sigma^2}{\mu(1-\mu)} \left[\mu \cdot \left(W^k(1) - W^k(\mu)\right)' \left(W^k(1) - W^k(\mu)\right) + (1-\mu) \cdot W^k(\mu)' W^k(\mu) - \mu(1-\mu) W^k(1)' W^k(1)\right] \\ &= \frac{\sigma^2}{\mu(1-\mu)} \left[W^k(\mu)' W^k(\mu) - 2\mu W^k(1)' W^k(\mu) + \mu^2 \cdot W^k(1)' W^k(1)\right] \end{aligned}$$

$$= \sigma^2 \cdot \frac{\left| W^k(\mu) - \mu \cdot W^k(1) \right|^2}{\mu(1-\mu)}.$$

The Weak Limit of the Difference of SSRs

We have seen above that

$$\left[SSR(\hat{T}_1,\cdots,\hat{T}_m) - SSR(\hat{T}_1,\cdots,\hat{T}_{i-1},\tau,\hat{T}_i,\cdots,\hat{T}_m)\right] - \left[U'M_{\bar{Z}^*}U - U'M_{\bar{Z}^*}U\right] = o_p(1),$$

and that

$$F_T^* = U'M_{\bar{Z}^*}U - U'M_{\bar{Z}^*}U \xrightarrow{d} F^*(\lambda_\tau).$$

Therefore,

$$F_{T,i}(\lambda_\tau) = SSR(\hat{T}_1,\cdots,\hat{T}_m) - SSR(\hat{T}_1,\cdots,\hat{T}_{i-1},\tau,\hat{T}_i,\cdots,\hat{T}_m) \xrightarrow{d} F^*(\lambda_\tau).$$

This holds uniformly for any $\lambda_\tau \in \Lambda_{i,\eta}$, so

$$\sup_{\lambda\in\Lambda_{i,\eta}} |F_{T,i}(\lambda) - F^*(\lambda)| \xrightarrow{d} 0.$$

as $T \to \infty$. Given any outcome, for any $T \in N_+$, because $\Lambda_{i,\eta}$ is a compact set and $F_{T,i}$ and F^* have finitely many discontinuities,

$$\left|\sup_{\lambda\in\Lambda_{i,\eta}} F_{T,i}(\lambda) - \sup_{\lambda\in\Lambda_{i,\eta}} F^*(\lambda)\right| \le \sup_{\lambda\in\Lambda_{i,\eta}} |F_{T,i}(\lambda) - F^*(\lambda)|,$$

which implies that

$$\sup_{\lambda\in\Lambda_{i,\eta}} F_{T,i}(\lambda) - \sup_{\lambda\in\Lambda_{i,\eta}} F^*(\lambda) \xrightarrow{d} 0$$

as well.

Because

$$\hat{\lambda}_j \xrightarrow{p} \lambda_j^0$$

for any $1 \le j \le m$, we can see that

$$\mathbb{P}\left(\Lambda_{i,\eta} \ne \Lambda_{i,\eta}^0\right) \to 0$$

as $T \to \infty$, where

$$\Lambda_{i,\eta}^0 = \left\{\lambda \in [0,1] \mid \frac{\lambda - \lambda_{i-1}^0}{\lambda_i^0 - \lambda_{i-1}^0} \in [\eta, 1-\eta]\right\}.$$

For any outcome $\omega \in \Omega$, F^* is a continuous function on the compact set $[0,1]$ due to the fact that the Wiener process has continuous paths. If $\Lambda_{i,\eta} = \Lambda^0_{i,\eta}$, then

$$\sup_{\lambda \in \Lambda_{i,\eta}} F^*(\lambda) = \sup_{\lambda \in \Lambda^0_{i,\eta}} F^*(\lambda),$$

which tells us that, for any $\delta > 0$,

$$\mathbb{P}\left(\left|\sup_{\lambda \in \Lambda_{i,\eta}} F^*(\lambda) - \sup_{\lambda \in \Lambda^0_{i,\eta}} F^*(\lambda)\right| > \delta\right) \leq \mathbb{P}\left(\Lambda_{i,\eta} \neq \Lambda^0_{i,\eta}\right).$$

The latter goes to 0 as $T \to \infty$, so that

$$\lim_{T \to \infty} \mathbb{P}\left(\left|\sup_{\lambda \in \Lambda_{i,\eta}} F^*(\lambda) - \sup_{\lambda \in \Lambda^0_{i,\eta}} F^*(\lambda)\right| > \delta\right) = 0.$$

This holds for any $\delta > 0$, so

$$\sup_{\lambda \in \Lambda_{i,\eta}} F^*(\lambda) \xrightarrow{p} \sup_{\lambda \in \Lambda^0_{i,\eta}} F^*(\lambda).$$

By implication,

$$\sup_{\lambda \in \Lambda_{i,\eta}} F_{T,i}(\lambda) \xrightarrow{d} \sup_{\lambda \in \Lambda^0_{i,\eta}} F^*(\lambda).$$

Since

$$\sup_{\lambda \in \Lambda^0_{i,\eta}} F^*(\lambda) \sim \sigma^2 \cdot \sup_{\eta \leq \mu \leq 1-\eta} \frac{\left|W^k(\mu) - \mu \cdot W^k(1)\right|^2}{\mu(1-\mu)},$$

we finally have

$$\sup_{\lambda \in \Lambda_{i,\eta}} F_{T,i}(\lambda) \xrightarrow{d} \sigma^2 \cdot \sup_{\eta \leq \mu \leq 1-\eta} \frac{\left|W^k(\mu) - \mu \cdot W^k(1)\right|^2}{\mu(1-\mu)}.$$

Finally, we can see that

$$\begin{aligned} F_T(m+1 \mid m) = \max_{1 \leq i \leq m+1} \sup_{\tau \in \Lambda_{i,\eta}} \left[SSR(\hat{T}_1, \cdots, \hat{T}_m) - SSR(\hat{T}_1, \cdots, \hat{T}_{i-1}, \tau, \hat{T}_i, \cdots, \hat{T}_m)\right] \\ \xrightarrow{d} \max_{1 \leq i \leq m+1} \sup_{\eta \leq \mu \leq 1-\eta} \frac{\left|W^k(\mu) - \mu \cdot W^k(1)\right|^2}{\mu(1-\mu)}, \end{aligned}$$

where the right hand side is the maximum of $m+1$ independent supremums over chi-squared random variables with k degrees of freedom.