Introduction

Consider the simple AR(1) model for $t = 1, \ldots, T$

$$y_t = \phi y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim \mathsf{WN}(0, \sigma^2)$$

If $|\phi| < 1$, then $y_t \sim I(0)$ and

$$y_t = \psi(L)\varepsilon_t,\ \psi(L) = \sum_{k=0}^{\infty} \psi_k L^k,\ \psi_k = \phi^k$$

such that

$$\sum_{k=0}^{\infty} k|\psi_k| \ < \ \infty$$

$$\mathsf{LRV} = \sigma^2 \psi(1)^2 = \sigma^2 (1 - \phi)^{-2} < \infty$$

Furthermore, by the LLN and the CLT

$$T^{-1}\sum_{t=1}^{T} y_t \xrightarrow{p} E[y_t] = 0$$

$$T^{-1/2}\sum_{t=1}^{T} y_t \xrightarrow{d} N(0, \sigma^2\psi(1)^2) = N(0, \sigma^2(1-\phi)^{-2})$$

$$T^{1/2}(\hat{\phi} - \phi) \xrightarrow{d} N(0, (1-\phi^2))$$

where $\hat{\phi} = \left(\sum_{t=1}^{T} y_{t-1}^2\right)^{-1}\sum_{t=1}^{T} y_{t-1}y_t$ is the least squares estimate of ϕ.

If $\phi = 1$, then $y_t \sim I(1)$ and

$$\begin{aligned} \psi_k &= 1 \\ \sum_{k=0}^{\infty} k|\psi_k| &= \infty \\ \sigma^2 \psi(1)^2 &= \infty \end{aligned}$$

Furthermore,

$$T^{-1} \sum_{t=1}^{T} y_t \to \infty \text{ as } T \to \infty$$

$$T^{-1/2} \sum_{t=1}^{T} y_t \to \infty \text{ as } T \to \infty$$

$$T^{1/2}(\hat{\phi} - 1) \xrightarrow{p} 0$$

Clearly, the asymptotic results for $I(0)$ processes are not applicable.

Sample Moments of $I(1)$ Processes

When $\phi = 1$

$$\begin{aligned} y_t &= y_{t-1} + \varepsilon_t \\ &= y_0 + \sum_{j=1}^{t} \varepsilon_j \\ &= \sum_{j=1}^{t} \varepsilon_j \text{ if } y_0 = 0 \end{aligned}$$

Now, consider the sample mean of y_t when $y_0 = 0$:

$$\bar{y} = T^{-1} \sum_{t=1}^{T} y_t = T^{-1} \sum_{t=1}^{T} \left(\sum_{j=1}^{t} \varepsilon_j \right)$$

Notice that the sample mean is a normalized sum of partial sums of the white noise error term ε_t. As such, it exhibits very different probabilistic behavior than the sum of stationary and ergodic errors. It turns out that the limit behavior of y when $\phi = 1$ is described by simple functionals of Brownian motion.

Brownian Motion

Standard Brownian motion (Wiener process) is a continuous-time process $W(\cdot)$ associating each date $r \in [0, 1]$ the scalar random ariable $W(r)$ such that

1. $W(0) = 0$

2. For any dates $0 \le r_1 < r_2 < \cdots < r_k \le k$, the random increments

   $W(r_2) - W(r_1), W(r_3) - W(r_2), \ldots, W(r_k) - W(r_{k-1})$

   are independent Gaussian random variables with

   $$W(t) - W(s) \sim N(0, t - s)$$

3. For any given realization, $W(r)$ is continuous at r with probability 1. That is, $W(r) \in C[0, 1] =$ space of continuous real valued functions on $[0, 1]$.

The standard Brownian motion, or Wiener process, may be intuitively thought of as the continuous-time limit of a random walk process in which the integer time index $t = 1, 2, \ldots, \infty$ has been rescaled to the continuous time index $r = 0, \ldots, 1$. The Wiener process may be shown to have the following properties:

1. $W(r) \sim N(0, r)$
2. $\sigma W(r) = B(r) \sim N(0, \sigma^2 r)$
3. $W(r)^2 \sim r \cdot \chi^2(1)$
4. $W(r)$ is not differentiable and exhibits unbounded variation.

Partial Sum Processes and the Functional Central Limit Theorem

Let $\varepsilon_t \sim WN(0, \sigma^2)$. For $r \in [0, 1]$, define the *partial sum process*

$$
\begin{aligned}
X_T(r) &= T^{-1} \sum_{t=1}^{[Tr]} \varepsilon_t \\
[Tr] &= \text{integer part of } T \cdot r
\end{aligned}
$$

For example, let $T = 10$ and consider $X_T(r)$ for $r = 0, 0.01, 0.1, 0.2$:

$$\begin{aligned}
r &= 0, [10 \cdot 0] = 0 : X_{10}(0) = \frac{1}{10} \sum_{t=1}^{[10 \cdot 0]} \varepsilon_t = 0 \\
r &= 0.01, [10 \cdot 0.01] = 0 : X_{10}(0.01) = \frac{1}{10} \sum_{t=1}^{[10 \cdot 0.01]} \varepsilon_t = 0 \\
r &= 0.1, [10 \cdot 0.1] = 1 : X_{10}(0.1) = \frac{1}{10} \sum_{t=1}^{[10 \cdot 0.1]} \varepsilon_t = \frac{\varepsilon_1}{10} \\
r &= 0.2, [10 \cdot 0.2] = 2 : X_{10}(0.2) = \frac{1}{10} \sum_{t=1}^{[10 \cdot 0.2]} \varepsilon_t = \frac{\varepsilon_1 + \varepsilon_2}{10}
\end{aligned}$$

In general,

$$X_{10}(r) = \frac{\varepsilon_1 + \cdots + \varepsilon_j}{10}, \ \frac{j}{10} \le r < \frac{j+1}{10}$$

For a sequence of errors $\varepsilon_1, \ldots, \varepsilon_T$:

1. the function $X_T(r)$ is a random step function defined on $[0, 1]$.

2. As T gets bigger the spaces between the steps gets smaller and the random step function begins to look more and more like a Wiener process.

The Functional Central Limit Theorem

For any fixed $r \in [0, 1]$, consider

$$
\begin{aligned}
\sqrt{T} X_T(r) &= \sqrt{T} \left(T^{-1} \sum_{t=1}^{[Tr]} \varepsilon_t \right) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \varepsilon_t \\
&= \left(\frac{\sqrt{[Tr]}}{\sqrt{T}} \right) \left(\frac{1}{\sqrt{[Tr]}} \sum_{t=1}^{[Tr]} \varepsilon_t \right)
\end{aligned}
$$

Now, as $T \to \infty$

$$
\frac{\sqrt{[Tr]}}{\sqrt{T}} \to \sqrt{r}
$$

$$
\frac{1}{\sqrt{[Tr]}} \sum_{t=1}^{[Tr]} \varepsilon_t \xrightarrow{d} N(0, \sigma^2)
$$

It follows from Slutsky's theorem that

$$\sqrt{T}X_T(r) \xrightarrow{d} N(0, r \cdot \sigma^2) \equiv \sigma \cdot W(r)$$

or

$$\sqrt{T}X_T(r)/\sigma \xrightarrow{d} N(0, r) \equiv W(r)$$

Notice that when $r = 1$, we have the usual result

$$\sqrt{T}X_T(1)/\sigma = \frac{1}{\sigma\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t \xrightarrow{d} N(0,1) \equiv W(1)$$

Since the above result holds for any $r \in [0, 1]$, one might expect that the result holds uniformly for $r \in [0, 1]$. In fact, the probability distribution of the sequence of stochastic step functions

$$\{\sqrt{T} X_T(\cdot)/\sigma\}_{T=1}^{\infty}$$

defined on $[0, 1]$ converges asymptotically to that of standard Brownian motion $W(\cdot)$.

This convergence result, know as *Donsker's Theorem for Partial Sums* or the *Functional Central Limit Theorem* (FCLT), is often represented as

$$\sqrt{T} X_T(\cdot)/\sigma \Rightarrow W(\cdot)$$

The symbol "$\Rightarrow$" denotes convergence in distribution for random functions.

The Continuous Mapping Theorem

Recall, if X_T is a sequence of random variables such that $X_T \stackrel{d}{\to} X$ and $g(\cdot)$ is a continuous function then $g(X_T) \stackrel{d}{\to} g(X)$. A similar result holds for random functions and is called the *Continuous Mapping Theorem* (CMT).

Let $\{S_T(\cdot)\}_{T=1}^{\infty}$ be a sequence of random functions such that

$$\begin{aligned} S_T(\cdot) &\Rightarrow S(\cdot) \\ g(\cdot) &= \text{continuous functional} \end{aligned}$$

Then the CMT states that

$$g(S_T(\cdot)) \Rightarrow g(S(\cdot))$$

Example 1

Suppose $S_T(\cdot) = \sqrt{T}X_T(\cdot)/\sigma$ so that $S(\cdot) = W(\cdot)$ by the FCLT. Let $g(S_T(\cdot)) = \sigma \cdot S_T(\cdot)$. Then

$$g(S_T(\cdot)) \Rightarrow g(W(\cdot)) = \sigma W(\cdot)$$

Example 2

Let $g(S_T(\cdot)) = \int_0^1 S_T(r)dr$. Then

$$g(S_T(\cdot)) \Rightarrow g(W(\cdot)) = \int_0^1 W(r)dr$$

Convergence of Sample Moments of I(1) Processes

Let y_t be the $I(1)$ process

$$y_t = y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$$

For $r \in [0, 1]$, define the partial sum process

$$X_T(r) = T^{-1} \sum_{t=1}^{[Tr]} \varepsilon_t$$

such that $\sqrt{T} X_T(\cdot) \Rightarrow \sigma W(\cdot)$. The FCLT and the CMT may be used to deduce the following results:

$$\begin{aligned}
T^{-3/2} \sum_{t=1}^{T} y_{t-1} &\Rightarrow \sigma \int_0^1 W(r) dr \\
T^{-2} \sum_{t=1}^{T} y_{t-1}^2 &\Rightarrow \sigma^2 \int_0^1 W(r)^2 dr \\
T^{-1} \sum_{t=1}^{T} y_{t-1} \varepsilon_t &\Rightarrow \sigma^2 \int_0^1 W(r) dW(r) \\
&= \sigma^2 \left(W(1)^2 - 1 \right) / 2
\end{aligned}$$

For example, it can be shown that

$$T^{-3/2} \sum_{t=1}^{T} y_{t-1} = \int_0^1 \sqrt{T} X_T(r) dr \Rightarrow \sigma \int_0^1 W(r) dr$$

using the FCLT and the CMT. The details are given in chapter 17 of Hamilton.

Application: Unit Root Tests

To illustrate the convergence of sample moments of $I(1)$ processes, consider the AR(1) regression

$$y_t = \phi y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim \text{WN}(0, \sigma^2)$$

If $\phi = 1$ then $y_t \sim I(1)$; if $|\phi| < 1$ then $y_t \sim I(0)$. A test of $y_t \sim I(1)$ against the alternative that $y_t \sim I(0)$ may therefore be formulated as

$$H_0 : \phi = 1 \text{ vs. } H_1 : |\phi| < 1$$

A natural test statistic is the t-statistic

$$t_{\phi=1} = \frac{\hat{\phi} - 1}{\text{SE}(\hat{\phi})}$$

where

$$\begin{aligned}
\hat{\phi} &= \left(\sum_{t=1}^{T} y_{t-1}^2\right)^{-1} \sum_{t=1}^{T} y_{t-1} y_t \\
\mathsf{SE}(\hat{\phi}) &= \left(\hat{\sigma}^2 \left(\sum_{t=1}^{T} y_{t-1}^2\right)^{-1}\right)^{1/2} \\
\hat{\sigma}^2 &= T^{-1} \sum_{t=1}^{T} (y_t - \hat{\phi} y_{t-1})^2
\end{aligned}$$

Consistency of $\hat{\phi}$ under $H_0 : \phi = 1$

Under $H_0 : \phi = 1$

$$
\begin{aligned}
\hat{\phi} - 1 &= \left(\sum_{t=1}^{T} y_{t-1}^2 \right)^{-1} \sum_{t=1}^{T} y_{t-1} \varepsilon_t \\
& \left(T^{-2} \sum_{t=1}^{T} y_{t-1}^2 \right)^{-1} T^{-2} \sum_{t=1}^{T} y_{t-1} \varepsilon_t
\end{aligned}
$$

Using the results

$$
\begin{aligned}
T^{-2} \sum_{t=1}^{T} y_{t-1}^2 &\Rightarrow \sigma^2 \int_0^1 W(r)^2 dr \\
T^{-1} \sum_{t=1}^{T} y_{t-1} \varepsilon_t &\Rightarrow \sigma^2 \int_0^1 W(r) dW(r)
\end{aligned}
$$

and the CMT, it follows that

$$
\hat{\phi} - 1 \xrightarrow{p} \left(\sigma^2 \int_0^1 W(r)^2 dr \right)^{-1} \times 0 = 0
$$

so that $\hat{\phi} \xrightarrow{p} 1$.

DF Test with Intercept

$$
\begin{aligned}
y_t &= c + \phi y_{t-1} + \varepsilon_t \\
&= \mathbf{x}_t' \boldsymbol{\beta} + \varepsilon_t \\
\mathbf{x}_t &= (1, y_{t-1})', \ \boldsymbol{\beta} = (c, \phi)'
\end{aligned}
$$

OLS gives

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \sum_{t=1}^{T} \mathbf{x}_t y_{t-1} \\
\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' &= \begin{pmatrix} T & \sum_{t=1}^{T} y_{t-1} \\ \sum_{t=1}^{T} y_{t-1} & \sum_{t=1}^{T} y_{t-1}^2 \end{pmatrix} \\
\sum_{t=1}^{T} \mathbf{x}_t y_{t-1} &= \begin{pmatrix} \sum_{t=1}^{T} y_{t-1} \\ \sum_{t=1}^{T} y_t y_{t-1} \end{pmatrix}
\end{aligned}
$$

Now, under $H_0 : \phi = 1$ and $c = 0$

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \begin{pmatrix} \hat{c} - 0 \\ \hat{\phi} - 1 \end{pmatrix} = \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \sum_{t=1}^{T} \mathbf{x}_t \varepsilon_t$$

$$\begin{pmatrix} T & \sum_{t=1}^{T} y_{t-1} \\ \sum_{t=1}^{T} y_{t-1} & \sum_{t=1}^{T} y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{T} \varepsilon_t \\ \sum_{t=1}^{T} y_{t-1} \varepsilon_t \end{pmatrix}$$

Problem: Elements of $\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'$ and $\sum_{t=1}^{T} \mathbf{x}_t \varepsilon_t$ converge at different rates!

$$\begin{pmatrix} T & \sum_{t=1}^{T} y_{t-1} \\ \sum_{t=1}^{T} y_{t-1} & \sum_{t=1}^{T} y_{t-1}^2 \end{pmatrix} = \begin{pmatrix} O(T) & O_p(T^{3/2}) \\ O_p(T^{3/2}) & O_p(T^2) \end{pmatrix}$$

$$\begin{pmatrix} \sum_{t=1}^{T} \varepsilon_t \\ \sum_{t=1}^{T} y_{t-1} \varepsilon_t \end{pmatrix} = \begin{pmatrix} O_p(T^{1/2}) \\ O_p(T) \end{pmatrix}$$

Implication: Cannot get sensible convergence results using traditional scaling

$$
\begin{aligned}
T\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) &= \left(T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1} T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_t \\
&= \begin{pmatrix} T^{-1} & T^{-2}\sum_{t=1}^{T} y_{t-1} \\ T^{-1}\sum_{t=1}^{T} y_{t-1} & T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \end{pmatrix}^{-1} \\
&\quad \times \begin{pmatrix} T^{-1}\sum_{t=1}^{T}\varepsilon_t \\ T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \end{pmatrix} \\
&\Rightarrow \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2\int_0^1 W(r)^2 dr \end{pmatrix}^{-1} \\
&\quad \times \begin{pmatrix} 0 \\ \sigma^2\int_0^1 W(r)dW(r) \end{pmatrix}
\end{aligned}
$$

which is not well defined.

Sims-Stock-Watson Trick

Define the diagonal and invertible scaling matrix

$$\mathbf{D}_T = \begin{pmatrix} T^{1/2} & 0 \\ 0 & T \end{pmatrix}$$

Then write

$$\begin{aligned} \mathbf{D}_T\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) &= \mathbf{D}_T\left(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\mathbf{D}_T\mathbf{D}_T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_t \\ &= \left(\mathbf{D}_T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\mathbf{D}_T^{-1}\right)^{-1}\mathbf{D}_T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_t \end{aligned}$$

where

$$\mathbf{D}_T\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)=\begin{pmatrix} T^{1/2}\hat{c} \\ T(\hat{\phi}-1) \end{pmatrix}$$

$$\mathbf{D}_T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\mathbf{D}_T^{-1}$$

$$=\begin{pmatrix} 1 & T^{-3/2}\sum_{t=1}^{T}y_{t-1} \\ T^{-3/2}\sum_{t=1}^{T}y_{t-1} & T^{-2}\sum_{t=1}^{T}y_{t-1}^2 \end{pmatrix}$$

$$\mathbf{D}_T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_t=\begin{pmatrix} T^{-1/2}\sum_{t=1}^{T}\varepsilon_t \\ T^{-1}\sum_{t=1}^{T}y_{t-1}\varepsilon_t \end{pmatrix}$$

Therefore,

$$\mathbf{D}_T\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) \Rightarrow \begin{pmatrix} 1 & \sigma\int_0^1 W(r) \\ \sigma\int_0^1 W(r) & \sigma^2\int_0^1 W(r)^2 dr \end{pmatrix}^{-1} \times \begin{pmatrix} N(0,\sigma^2) \\ \sigma^2\int_0^1 W(r)dW(r) \end{pmatrix}$$

Straightforward algebra shows that

$$T^{1/2}\hat{c} \xrightarrow{d} N(0,\sigma^2)$$

$$T(\hat{\phi}-1) \Rightarrow \left(\int_0^1 W^{\mu}(r)^2 dr\right)^{-1}\int_0^1 W^{\mu}(r)dW(r)$$

$$W^{\mu}(r) = W(r)-\int_0^1 W(r)$$

Convergence of Sample Moments with General Serial Correlation

$$
\begin{aligned}
y_t &= y_{t-1} + \psi^*(L)\varepsilon_t,\ \varepsilon_t \sim \mathsf{WN}(0, \sigma^2) \\
&= y_{t-1} + u_t \\
& \quad \psi^*(L) \text{ is 1-summable} \\
\mathsf{LRV} &= \sigma^2 \psi^*(1)^2 = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j \\
\gamma_j &= \mathsf{cov}(u_t, u_{t-j})
\end{aligned}
$$

FCLT

$$
\sqrt{T} X_T(\cdot) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\cdot]} u_t \Rightarrow \mathsf{LRV} \times W(\cdot)
$$

1. $T^{-3/2}\sum_{t=1}^{T} y_{t-1} \Rightarrow \sqrt{\text{LRV}}\int_0^1 W(r)dr$

2. $T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \Rightarrow \text{LRV}\int_0^1 W(r)^2 dr$

3. $T^{-1}\sum_{t=1}^{T} y_{t-1}u_t \Rightarrow \text{LRV}\int_0^1 W(r)dW(r)+\omega, \omega = \frac{1}{2}(\text{LRV}-\gamma_0)$

4. $T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \Rightarrow \sqrt{\sigma^2\text{LRV}}\int_0^1 W(r)dW(r)$

5. $T^{-1}\sum_{t=1}^{T} y_{t-1}u_{t-1} = \text{LRV}\int_0^1 W(r)dW(r) + \omega + \gamma_0$

Application: Asymptotic Distribution of ADF test

Assume y_t is $I(1)$ and that $\Delta y_t \sim AR(1)$

$$\begin{aligned} \Delta y_t &= \xi \Delta y_{t-1} + \varepsilon_t, \varepsilon_t \sim \text{WN}(0, \sigma^2) \\ |\xi| &< 1 \end{aligned}$$

Therefore, Δy_t has Wold representation

$$\begin{aligned} \Delta y_t &= \psi^*(L)\varepsilon_t = u_t \\ \psi^*(L) &= (1 - \xi L)^{-1} = \sum_{j=0}^{\infty} \psi_j^* L^j,\ \psi_j^* = \xi^j \\ \text{LRV} &= \sigma^2 \psi^*(L) = \sigma^2 (1-\xi)^{-1} \end{aligned}$$

The ADF test regression is

$$\begin{aligned} y_t &= \phi y_{t-1} + \xi \Delta y_{t-1} + \varepsilon_t \\ & \mathbf{x}_t' \boldsymbol{\beta} + \varepsilon_t \\ \mathbf{x}_t' &= (y_{t-1}, \Delta y_{t-1})',\ \boldsymbol{\beta} = (\phi, \xi)' \end{aligned}$$

Notice that

$$x_t = \begin{pmatrix} y_{t-1} \\ \Delta y_{t-1} \end{pmatrix} \begin{matrix} \sim I(1) \\ \sim I(0) \end{matrix}$$

OLS on the ADF test regression gives

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \sum_{t=1}^{T} \mathbf{x}_t \varepsilon_t$$

where

$$\begin{aligned} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' &= \begin{pmatrix} \sum_{t=1}^{T} y_{t-1}^2 & \sum_{t=1}^{T} y_{t-1} \Delta y_{t-1} \\ \sum_{t=1}^{T} \Delta y_{t-1} y_{t-1} & \sum_{t=1}^{T} \Delta y_{t-1}^2 \end{pmatrix} \\ &= \begin{pmatrix} O_p(T^2) & O_p(T) \\ O_p(T) & O_p(T^{1/2}) \end{pmatrix} \\ \sum_{t=1}^{T} \mathbf{x}_t \varepsilon_t &= \begin{pmatrix} \sum_{t=1}^{T} y_{t-1} \varepsilon_t \\ \sum_{t=1}^{T} \Delta y_{t-1} \varepsilon_t \end{pmatrix} \\ &= \begin{pmatrix} O_p(T) \\ O_p(T^{1/2}) \end{pmatrix} \end{aligned}$$

Use Sims-Stock-Watson trick and define the scaling matrix

$$\mathbf{D}_T = \begin{pmatrix} T & 0 \\ 0 & T^{1/2} \end{pmatrix}$$

Then write

$$\begin{aligned} \mathbf{D}_T\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) &= \mathbf{D}_T\left(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\mathbf{D}_T\mathbf{D}_T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_t \\ &= \left(\mathbf{D}_T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\mathbf{D}_T^{-1}\right)^{-1}\mathbf{D}_T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_t \end{aligned}$$

where

$$\mathbf{D}_T\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = \begin{pmatrix} T(\hat{\phi} - 1) \\ T^{1/2}\left(\hat{\xi} - \xi\right) \end{pmatrix}$$

$$\mathbf{D}_T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\mathbf{D}_T^{-1}$$

$$= \begin{pmatrix} T^{-2}\sum_{t=1}^{T} y_{t-1}^2 & T^{-3/2}\sum_{t=1}^{T} y_{t-1}\Delta y_{t-1} \\ T^{-3/2}\sum_{t=1}^{T}\Delta y_{t-1}y_{t-1} & T^{-1}\sum_{t=1}^{T}\Delta y_{t-1}^2 \end{pmatrix}$$

and

$$\mathbf{D}_T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\varepsilon_t = \begin{pmatrix} T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \\ T^{-1/2}\sum_{t=1}^{T}\Delta y_{t-1}\varepsilon_t \end{pmatrix}$$

Note: $\Delta y_{t-1}\varepsilon_t = u_{t-1}\varepsilon_t$ is a stationary and ergodic MDS with

$$\begin{aligned} E[(u_{t-1}\varepsilon_t)^2] &= E[E\,(u_{t-1}\varepsilon_t)^2\,|I_{t-1}] \\ &= E[u_{t-1}^2 E[\varepsilon_t^2]] = \sigma^2\gamma_0 \end{aligned}$$

Therefore, by the appropriate CLT

$$T^{-1/2}\sum_{t=1}^{T}\Delta y_{t-1}\varepsilon_t \to N(0,\sigma^2\gamma_0)$$

Using the convergence results for the sample moments of serially correlated I(1) process, the above result, and the CMT gives

$$T(\hat{\phi}-1) \Rightarrow \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr}$$

$$T^{1/2}\left(\hat{\xi}-\xi\right) \xrightarrow{d} N(0,\sigma^2\gamma_0)$$

Furthermore, $\hat{\phi}$ and $\hat{\xi}$ are asymptotically independent.