Chapter 1

Nonstationary Times Series

1.1 Partial Sum Process and Functional Convergence

Let $y_t \in \mathbb{R}^m$ follow a random walk $y_t = y_{t-1} + e_t$, where $(e_t, \mathscr{F}_t)$ is a vector mds. Iterative substitution makes $y_t = y_0 + \sum_{i=1}^{t} e_i = y_0 + S_t$, where

$$S_t = \sum_{i=1}^{t} e_i$$

is the *partial sum.* We define the *standardized partial sum* as

$$S_n(r) = \frac{1}{\sqrt{n}} S_{\lfloor nr \rfloor} = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} e_t$$

for some real number $r \in [0, 1]$. For a finite n, $S_n(r)$ is a step function in r.

Recall **Convergence in distribution**: we say $\nu_n(r) \xrightarrow{d} \nu(r)$ if $E[f(\nu_n(r))] \to E[f(\nu(r))]$ for any bounded, continuous function $f : \nu \to \mathbb{R}$, where continuity is defined with respect to the uniform metric $\rho(\nu_1, \nu_2) = \sup_{0 \le r \le 1} \|\nu_1(r) - \nu_2(r)\|$. The definition of convergence in distribution is abstract and difficult to verify. It is easier to verify its equivalent conditions: (i) for any finite $r_1, ..., r_m$, we have $(\nu_n(r_1), ..., \nu_n(r_m)) \xrightarrow{d} (\nu(r_1), ..., \nu(r_m))$; (ii) $\nu_n(r)$ is stochastically equicontinuous.

As $n \to \infty$, asymptotically, the maximal jump size $\frac{1}{\sqrt{n}} \max_{i \le n} \|e_t\| = O_p(1)$, so jumps vanish and S_n is stochastically equicontinuous. Now we verify its finite joint distribution. For $S_n(r)$, we have

1. $S_n(0) = 0$
2. For any r, $S_n(r) \xrightarrow{d} N(0, r\Sigma)$
3. For $r_1 < r_2$, $S_n(r_1)$ and $S_n(r_2) - S_n(r_1)$ are asymptotically independent.

The second point holds as

$$S_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} e_t = \sqrt{\frac{\lfloor nr \rfloor}{n}} \frac{1}{\sqrt{\lfloor nr \rfloor}} \sum_{t=1}^{\lfloor nr \rfloor} e_t \xrightarrow{d} N(0, r\Sigma).$$

And the third point holds as

$$\begin{pmatrix} S_n(r_1) \\ S_n(r_2) - S_n(r_1) \end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix} r_1\Sigma & 0 \\ 0 & (r_2 - r_1)\Sigma \end{pmatrix}\right).$$

The above joint distribution are written for any two points $r_1, r_2 \in [0,1]$, and it is easy to see that the asymptotic normality can be extended to any $r_1, \ldots, r_m \in [0,1]$ with a finite m.

Notice that the initial value y_0 does not affect the asymptotic behavior, since $\frac{1}{\sqrt{n}} y_{\lfloor nr \rfloor} = S_n(r) + \frac{1}{\sqrt{n}} y_0$ with the initial value $\frac{1}{\sqrt{n}} y_0 = O_p\left(\frac{1}{\sqrt{n}}\right) = o_p(1)$. (For simplicity, we can simply assume $y_0 = 0$.)

Next, we introduce the Brownian motion.

Definition 1.1. A vector **Brownian motion** satisfies (of variance $var(B(1)) = \Sigma$)

1. $B(0) = 0$
2. $B(r) \sim N(0, r\Sigma)$
3. $B(r_1)$ is independent of $B(r_2) - B(r_1)$ for $r_1 < r_2$.

We find the limiting behavior of $S_n(r)$ in any finite coordinates coincides with the Brownian motion, and thus we have the following functional CLT.

Theorem 1.1. *if $(e_t, \mathscr{F}_t)$ is strictly stationary, ergodic mds with $\Sigma < \infty$, then $S_n(r) \xrightarrow{d} B(r)$*

1.2 Beveridge-Nelson Decomposition

So far we discussed mds innovation, which is a special case. In general, we want to allow the the innovations to be serially correlated. Let the innovation be $e_t = \Theta(L) u_t$, where u_t is mds and the polynomial $\Theta(z) = \theta_0 + \theta_1 z + \theta_2 z^2 + \theta_3 z^3 + \ldots$. Obviously,

$$e_t = \Theta(L) u_t = \Theta(1) u_t + (\Theta(L) - \Theta(1)) u_t.$$

Notice

$$\begin{aligned}
\Theta(1) - \Theta(z) &= \theta_0 + \theta_1 + \theta_2 + \theta_3 + \ldots - \left(\theta_0 + \theta_1 z + \theta_2 z^2 + \theta_3 z^3 + \ldots\right) \\
&= \theta_1 (1 - z) + \theta_2 \left(1 - z^2\right) + \theta_3 \left(1 - z^3\right) + \ldots \\
&= (1 - z) \left[\theta_1 + \theta_2 (1 + z) + \theta_3 \left(1 + z + z^2\right) + \ldots\right] \\
&= (1 - z) \Theta^*(z)
\end{aligned}$$

Replacing the dummy z by L, we write

$$\begin{aligned}
e_t &= \Theta(1) u_t + (1 - L) \left[-\Theta^*(L) u_t\right] \\
&= \Theta(1) u_t + (1 - L) \nu_t \\
&= \Theta(1) u_t + \nu_t - \nu_{t-1}
\end{aligned}$$

where $\nu_t = -\Theta^*(L) u_t$. As a result,

$$y_t = \sum_{s=1}^{t} e_s + y_0 = \Theta(1) \sum_{s=1}^{t} u_s + \nu_t + (y_0 - \nu_0)$$

where the first term is the permanent component, the second term the transitory component, and the third term in the parenthesis is the initial value.

The MA form of e_t ensures that it is stationary, with long-run variance

$$\begin{aligned} var\left(\frac{1}{\sqrt{n}}\sum_{s=1}^{n}e_s\right) &= var\left(\Theta\left(1\right)\frac{1}{\sqrt{n}}\sum_{s=1}^{t}u_s+\frac{\nu_t}{\sqrt{n}}-\frac{\nu_0}{\sqrt{n}}\right)\\ &= \Theta\left(1\right)\Sigma\Theta'\left(1\right)+o\left(1\right)\\ &\to \Theta\left(1\right)\Sigma\Theta'\left(1\right) \end{aligned}$$

where $\Sigma = var\left(u_s\right)$. In other word, the effect of the MA representation is multiply with the white noise variance by a factor $\Theta(1)$.

1.3 Functional CLT

Consider the representation

$$y_t = S_t + u_t + (y_0 - \nu_0).$$

Define $S_n\left(r\right) = \frac{1}{\sqrt{n}}S_{\lfloor nr\rfloor}$ and

$$z_n\left(r\right) = \frac{1}{\sqrt{n}}y_{\lfloor nr\rfloor} = S_n\left(r\right) + \frac{1}{\sqrt{n}}u_{\lfloor nr\rfloor} + \frac{1}{\sqrt{n}}(y_0-\nu_0).$$

If u_t is mds, we have

$$z_n\left(r\right) = S_n\left(r\right) + o_p\left(1\right) \xrightarrow{d} B\left(r\right)$$

where $B\left(1\right) \sim N\left(0, \Theta\left(1\right)\Sigma\Theta'\left(1\right)\right)$.

Linear projection ensures the innovations e_t in the Wold decomposition are white noise, but may not necessarily be mds. If u_t is not mds, we impose assumptions on the α-mixing coefficient so that we can still apply FCLT to conclude

$$z_n\left(r\right) \xrightarrow{d} B\left(r\right)$$

where $B(1) \sim \Omega$ with Ω being the long-run variance of Δy_t.

1.4 Orders of Integration

We say a time series y_t is $I\left(0\right)$ if y_t is weakly stationary with positive long-run variance. We say it is $I(d)$ if $\Delta^d y_t \sim I\left(0\right)$.

What happens if we "over differentiate" y_t? Suppose $y_t = \Theta\left(L\right)u_t$ in MA(∞) representation

$$\Delta y_t = \left(1-L\right)\Theta\left(L\right)u_t.$$

Consider $\left(1-L\right)\Theta\left(L\right)$ as an entity for the MA(∞) representation, and then the long-run $var\left(\Delta y_t\right) = \left(1-1\right)\Theta\left(L\right)var\left(u_t\right) = 0$.

1.5 Means

By the continuous mapping theorem, if $z_n\left(r\right) \xrightarrow{d} B\left(r\right)$, then $f\left(z_n\right) \xrightarrow{d} f\left(B\right)$ for continuous functional f. Notice $\frac{1}{\sqrt{n}}y_{\lfloor nr\rfloor} = z_n\left(r\right)$ is a step function.

$$\frac{1}{\sqrt{n}}\overline{y}_n = \frac{1}{n}\sum_{t=1}^{n}\frac{y_t}{\sqrt{n}} = \frac{1}{n}\sum_{r\in\left\{\frac{1}{n},\frac{2}{n},..,\frac{n-1}{n},1\right\}}z_n\left(r\right) = \int_0^1 z_n\left(r\right)\mathrm{d}r$$

for any finite n. We conclude

$$\frac{1}{\sqrt{n}}\overline{y}_n \xrightarrow{d} \int_0^1 B\left(r\right)\mathrm{d}r$$

is an average of a Brownian motions over $[0,1]$.

1.6 Regression with intercept and time trend

If we fit a unit root process y_t with a deterministic trend $y_t = \beta_0 + \beta_1 t + error_t$, we can denote the regressor as $X_t = \begin{pmatrix} 1 \\ t \end{pmatrix}$, and thus the OLS estimator is

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \left(X'X\right)^{-1} X'y.$$

As we have seen before, if we set $D_n = \begin{pmatrix} \sqrt{n} & 0 \\ 0 & n^{\frac{3}{2}} \end{pmatrix}$, then

$$\frac{1}{n}D_n \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = D_n \left(X'X\right)^{-1} D_n D_n^{-1} X'y = \left(D_n^{-1}\left(X'X\right)D_n^{-1}\right)^{-1}\frac{X'y}{n}.$$

The denominator

$$\begin{pmatrix} \frac{n}{n} & n^{-2}\sum t \\ n^{-2}\sum t & n^{-3}\sum t^2 \end{pmatrix} \to \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix}.$$

The numerator

$$D_n^{-1}\frac{1}{n}\sum_{t=1}^{n} X_t y_t = \frac{1}{n}\sum_{t=1}^{n}\begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{n} \end{pmatrix} X_t \frac{y_t}{\sqrt{n}} = \frac{1}{n}\sum_{r\in\left\{\frac{1}{n},\frac{2}{n},..,\frac{n-1}{n},1\right\}} X_n\left(r\right) z_n\left(r\right) = \int_0^1 X\left(r\right) z\left(r\right)\mathrm{d}r.$$

We conclude

$$\frac{1}{n}D_n \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} n^{-\frac{1}{2}}\hat{\beta}_0 \\ n^{-\frac{1}{2}}\hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix}^{-1} \begin{pmatrix} \int_0^1 B\left(r\right)\mathrm{d}r \\ \int_0^1 rB\left(r\right)\mathrm{d}r \end{pmatrix}.$$

The behavior of the OLS estimator is very different from our familiar iid cases. The intercept $\hat{\beta}_0 = O_p\left(\sqrt{n}\right)$ is explosive, whereas $\hat{\beta}_1 = O_p\left(n^{-\frac{1}{2}}\right)$. In particular, the trend coefficient matches the order of the two sides, but the estimated right-end of the trend is $n\hat{\beta}_1 = O_p\left(n^{-\frac{1}{2}}\right)$ is also explosive.

1.7 Demeaning and Detrending

When we witness a trend in a time series, one may attempt to detrend it. Have we investigate the consequence of demean and detrending if the true $\{y_t\}$ is a unit root process.

- demean: $y_t^* = y_t - \overline{y}_n$ is irrelevant of the initial value.

The standardized version

$$Z_n^*\left(r\right) = \frac{1}{\sqrt{n}}y_{\lfloor nr\rfloor} - \frac{1}{\sqrt{n}}\overline{y}_n = z_n\left(r\right) - \int_0^1 z\left(r\right)\mathrm{d}r \xrightarrow{d} B\left(r\right) - \int_0^1 B\left(r\right)\mathrm{d}r =: B^*\left(r\right)$$

demeaned B-motion

- detrending

$$\begin{aligned} Z_n^{**}(r) &= \frac{1}{\sqrt{n}} y_{\lfloor nr \rfloor} - \frac{1}{\sqrt{n}} X_{\lfloor nr \rfloor} \hat{\beta} \\ &= Z_n(r) - \frac{1}{\sqrt{n}} X'_{\lfloor nr \rfloor} n D_n^{-1} \frac{1}{n} D_n \hat{\beta} \\ &\xrightarrow{d} Z_n(r) - X'(r) \left(\int_0^1 XX' \right)^{-1} \left(\int_0^1 XB \right) =: B^{**}(r) \end{aligned}$$

detrended B-motion

- First difference

if $y_t = \beta_0 + \beta_1 t + z_t$, then $\Delta y_t = \beta_0 + \Delta z_t$

if β_1 is estemated by sample mean, then $\overline{\Delta y_n} = \frac{1}{n} \sum_{t=1}^{n} \Delta y_t = \frac{y_n - y_\infty}{n}$

And normalization $z_0 = 0$ gives $y_0 = \beta_0$

$$\widetilde{y}_t = y_t - y_0 - \frac{t}{n} (y_n - y_0)$$

this is the residual after (β_0, β_1) are estimated.

Standardization:

$$\widetilde{z}_n(r) = \frac{1}{\sqrt{n}} y_{\lfloor nr \rfloor} - \frac{y_0}{\sqrt{n}} - \frac{\lfloor nr \rfloor}{n} \frac{(y_n - y_0)}{\sqrt{n}} = \frac{1}{\sqrt{n}} y_{\lfloor nr \rfloor} - \frac{\lfloor nr \rfloor}{n} y_n + o_p(1) \xrightarrow{d} B(r) - rB(1) =: V(r)$$

Brownian bridge

1.8 Stochastic Integral

The Riemann-Stieltijes integral (deterministic) in $[0, 1]$ is defined as

$$\int_0^1 g(X) \, \mathrm{d}f(X) = \lim_{N \to \infty} \sum_{i=0}^{N-1} g\left(\frac{i}{n}\right) \left(f\left(\frac{i+1}{N}\right) - f\left(\frac{i}{N}\right) \right).$$

The key difference of the **stochastic integral** is that the measure for integration is a random:

$$\int_0^1 X \mathrm{d}z' = \int_0^1 X(r) \, \mathrm{d}z(r)' = \mathrm{plim}_{N \to \infty} \sum_{i=0}^{N-1} X\left(\frac{i}{n}\right) \left(z\left(\frac{i+1}{N}\right) - z\left(\frac{i}{N}\right) \right)$$

This RHS limit is a usually random variable, not a constant.

Consider (X_t, e_t), where e_t is a mds and X_t is non-stationary. If $X_n(r) = D_n^{-1} X_{\lfloor nr \rfloor}$ for some deterministic D_n and $X_n(r) \to X(r)$ then

$$\begin{aligned} \frac{1}{\sqrt{n}} D_n^{-1} \sum_{t=0}^{n-1} X_t e'_{t+1} &= \sum_{t=0}^{n-1} \left(D_n^{-1} X_t \right) \frac{e'_{t+1}}{\sqrt{n}} \\ &= \sum_{t=0}^{n-1} \left(D_n^{-1} X_t \right) \left(S_n\left(\frac{t+1}{N}\right) - S_n\left(\frac{t}{N}\right) \right) = \int_0^1 X_n \mathrm{d}S'_n \end{aligned}$$

Theorem 1.2. *If $(e_t, \mathscr{F}_t)$ is mds, $E(e_t e_t') = \sum < \infty$, $X_t \in \mathscr{F}_t$, and $(X_n(r), S_n(r)) \xrightarrow{d} X(r), B(r)$, then*

$$\int_0^1 X_n \mathrm{d}S_n' \xrightarrow{d} \int_0^1 X_n \mathrm{d}B'$$

Example 1.1. if $X_n(r) = S_n(r)$ and $S_t = \sum_{i=0}^{t} e_t$, where e_t is mds, then

$$\frac{1}{n}\sum_{t=0}^{n-1} S_t e_{t+1}' = \sum_{t=0}^{n-1} \frac{S_t}{\sqrt{n}} \frac{e_{t+1}'}{\sqrt{n}} \xrightarrow{d} \int B \mathrm{d}B'$$

If e_t is serially correlated, then

$$\frac{1}{n}\sum_{t=0}^{n-1} S_t e_{t+1}' \xrightarrow{d} \int B \mathrm{d}B' + \Lambda$$

where $\Lambda = \sum_{j=1}^{\infty} [z_{t-j} z_t']$

proof : use BN-decomposition for $e_t = \zeta_t + u_t - u_{t-1}$

1.9 AR(1) Regression

Let us start with the simplest model, an AR(1) regression with no intercept:

$$y_t = \alpha y_{t-1} + e_t$$

where e_t is a homoskedastic mds. Obviously, the OLS estimator satisfies

$$\hat{\alpha} - \alpha = \left(\sum_{t=0}^{n-1} y_t^2\right)^{-1} \sum_{t=0}^{n-1} y_t e_{t+1}$$

and proper scaling yields

$$n(\hat{\alpha} - \alpha) = \frac{1}{n}\sum_{t=0}^{n-1} y_t e_{t+1} / \frac{1}{n^2}\sum_{t=0}^{n-1} y_t^2.$$

The numerator in the last expression is

$$\begin{aligned}\sum_{t=0}^{n-1} \frac{y_t}{\sqrt{n}} \frac{y_{t+1} - y_t}{\sqrt{n}} &= \sum S_n(r)\left(S_n\left(r + \frac{1}{N}\right) - S(r)\right) \\ &= \int S_n(r)\, \mathrm{d}S_n(r) \xrightarrow{d} \int_0^1 B \mathrm{d}B = \sigma^2 \int W \mathrm{d}W\end{aligned}$$

and the denominator is

$$\frac{1}{n}\sum_{t=0}^{n-1} \left(\frac{y_t}{\sqrt{n}}\right)^2 = \sum_{t=0}^{n-1} \frac{1}{n} S_n^2(r) \xrightarrow{d} \int_0^1 B^2 = \sigma^2 \int_0^1 W^2.$$

Theorem 1.3. *if $(e_t, \mathscr{F}_t)$ is stationary, ergodic mds, then*

$$n(\hat{\alpha} - 1) \xrightarrow{d} \int_0^1 W \mathrm{d}W / \int_0^1 W^2$$

This estimator is super-consistent, in the sense that its rate of convergence is n, instead of $\sqrt{n}$ as in the iid case.

The stochastic integral $\int_0^1 W \mathrm{d}W = \frac{1}{2}\left(W^2(1) - 1\right)$ is an Ito integral. "-1" is present because $W_n(r)\left[W_n\left(r + \frac{1}{N}\right) - W_n(r)\right]$ is a mds.

Next, we usually use the t-statistic to infer the slope coefficient. Notice that the residual $\hat{e}_t = y_t - \hat{\alpha} y_{t-1}$ gives

$$\hat{\sigma}^2 = \frac{1}{n}\sum \hat{e}_t^2 = \frac{1}{n}\sum \hat{e}_t^2 + o_p(1) \xrightarrow{d} \sigma^2$$

Assume $(e_t, \mathscr{F}_t)$ homoskedastic mds. We have $\hat{var}(\hat{\alpha}) = \frac{\hat{\sigma}^2}{\sum y_t^2}$. The t-statistic is

$$\begin{aligned} t = \frac{\hat{\alpha} - 1}{s.e.(\hat{\alpha})} &= \frac{\left(\sum_{t=0}^{n-1} y_t^2\right)^{-1} \sum_{t=0}^{n-1} y_t e_{t+1}}{\hat{\sigma}/\sqrt{\sum y_t^2}} = \frac{\sum_{t=0}^{n-1} y_t e_{t+1}/\hat{\sigma}}{\sqrt{\sum y_t^2}} \\ &\xrightarrow{d} \frac{\sigma \int_0^1 W \mathrm{d}W/\sigma}{\sqrt{\int_0^1 W^2}} = \int_0^1 W \mathrm{d}W \Big/ \sqrt{\int_0^1 W^2} \end{aligned}$$

The above calculation is demonstrated by a regression with no intercept. For the regression with an intercept, $y_t = \mu + \alpha y_{t-1} + e_t$, by the Frisch-Waugh-Lovell Theorem the slope coefficient will be numerically equivalent to running OLS with $y_t = \alpha(y_t - \overline{y}_n) + e_t$, and thus

$$n(\hat{\alpha} - 1) \xrightarrow{d} \int_0^1 W^* \mathrm{d}W \Big/ \int_0^1 W^{*2}$$

where W^* is the demeaned Brownian motion. Similarly, if the regression has both an intercept and a time trend, then

$$n(\hat{\alpha} - 1) \xrightarrow{d} \int_0^1 W^{**} \mathrm{d}W \Big/ \int_0^1 W^{**2}$$

where W^{**} is the demeaned-and-detrended Brownian motion.

1.10 AR(p) Models with a Unit Root

If the true DGP is $e_t = a(L)\Delta y_t = a(L)(1-L)y_t$, then

$$\begin{aligned} y_t &= a_1 y_{t-1} + a_2 y_{t-2} + ... + a_p y_{t-p} + e_t \\ &= (a_1, a_2, ..., a_p)(y_{t-1}, y_{t-2}, ..., y_{t-p})' + e_t \\ &= (a_1, a_2, ..., a_p) A A^{-1} (y_{t-1}, y_{t-2} ..., y_{t-p})' + e_t \\ &= (\rho, \beta_1 ... \beta_{p-1})(y_{t-1}, \Delta y_{t-1}, ..., \Delta y_{t-p-1})' + e_t \end{aligned}$$

where

$$A = \begin{bmatrix} 1 & & & \cdots & 0 \\ 1 & -1 & & & \vdots \\ \vdots & \vdots & -1 & & \\ & & & \ddots & \\ 1 & -1 & & \cdots & -1 \end{bmatrix}, \text{ and } A^{-1} = \begin{bmatrix} 1 & & & \cdots & 0 \\ 1 & -1 & & & \vdots \\ & 1 & -1 & & \\ \vdots & & \ddots & \ddots & \\ 0 & \cdots & & 1 & -1 \end{bmatrix}$$

transforms keeps only one level variable y_{t-1} while transforms all other further lagged level variables $(y_{t-2}, ..., y_{t-p})$ into differenced variables $X_{t-1} = (\Delta y_{t-1}, ..., \Delta y_{t-p-1})$. If y_t is unit root, we have $a(1) = a_1 + ... + a_p = 1$.

The transformation separates the regressors into two types: one nonstationary variable and the other stationary variables. The OLS estimator of the transformed equation satisfies

$$\begin{pmatrix} n(\hat{\rho}-1) \\ \sqrt{n}\left(\hat{\beta}-\beta\right) \end{pmatrix} = \begin{pmatrix} \frac{1}{n^2}\sum_{t=p+1}^n y_{t-1}^2 & \frac{1}{n^{3/2}}\sum_{t=p+1}^n y_{t-1}X'_{t-1} \\ \frac{1}{n^{3/2}}\sum_{t=p+1}^n y_{t-1}X'_{t-1} & \frac{1}{n}\sum_{t=p+1}^n X_{t-1}X'_{t-1} \end{pmatrix} \begin{pmatrix} \frac{1}{n}\sum_{t=p+1}^n y_{t-1}e_t \\ \frac{1}{\sqrt{n}}\sum_{t=p+1}^n X_{t-1}e_t \end{pmatrix}.$$

notice

$$\frac{1}{n^{3/2}}\sum_{t=p+1}^n y_{t-1}X'_{t-1} = \frac{1}{n}\sum_{t=p+1}^n \frac{y_{t-1}}{\sqrt{n}}X'_{t-1} = \frac{1}{n}\sum_{t=p+1}^n S_n(r)\, X'_{t-1} \xrightarrow{p} 0$$

as $E[X_{t-1}] = 0$

- Alternatively, we understand it as $\frac{y_{t-1}}{\sqrt{n}} = \frac{y_{t-p}+y_{t-p+1}+...+y_{t-1}}{\sqrt{n}}$.

The denominator

$$\begin{pmatrix} \frac{1}{n^2}\sum_{t=p+1}^n y_{t-1}^2 & \frac{1}{n^{3/2}}\sum_{t=p+1}^n y_{t-1}X'_{t-1} \\ \frac{1}{n^{3/2}}\sum_{t=p+1}^n y_{t-1}X'_{t-1} & \frac{1}{n}\sum_{t=p+1}^n X_{t-1}X'_{t-1} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \omega^2\int_0^1 W^2(r) & 0 \\ 0 & Q \end{pmatrix}$$

The numerator

$$\begin{pmatrix} \frac{1}{n}\sum_{t=p+1}^n y_{t-1}e_t \\ \frac{1}{\sqrt{n}}\sum_{t=p+1}^n X_{t-1}e_t \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \omega\sigma\int_0^1 W\mathrm{d}W \\ N(0,\Omega) \end{pmatrix}$$

where ω is the long-run variance of Δy_t, $Q = E\left[X_{t-1}X'_{t-1}\right]$, Ω is the variance of $X_{t-1}e_t$

1.11 Test for Unit Root: ADF Test

When $(e_t, \mathscr{F}_t)$ is mds, we have $\omega = \sigma$. If we are interested in the null hypothesis that y_t is a unit root process, we have the celebrated Dicky-Fuller test.

Theorem 1.4. *Assume $a(L)\Delta y_t = e_t$, where $a(z)$ is p-1 order with $a_1 + ... + a_p = 1$. $(e_t, \mathscr{F}_t)$ is stationary mds with finite constant variance σ^2. Then*

$$\mathrm{ADF} = \frac{\hat{\alpha}-\alpha}{s.e.(\hat{\alpha})} \xrightarrow{d} \frac{\int_0^1 u\mathrm{d}W}{\sqrt{\int_0^1 u^2}},$$

where u depends on the specification of the deterministic part.

1.12 Test for a Unit Root: KPSS Stationarity Test

Kwiatkowski, Phillips, Schmidt, and Shin (1992) is an alternative test for nonstationarity. Its null hypothesis is that y_t is a stationary time series. Consider the model

$$y_t = \mu + S_t + e_t,$$

where $S_t = \sum_{s=1}^t u_t$. If $\sigma_u^2 = 0$, then S_t drops out and y_t is stationary as $y_t = \mu + e_t$.

The null hypothesis $H_0 : \sigma_u^2 = 0$ vs. $H_1 : \sigma_u^2 > 0$: we have the KPSS test statistic defined as

$$\mathrm{KPSS} = \frac{1}{n^2\hat{\omega}^2}\sum_{i=1}^{n}\sum_{t=1}^{i}\hat{e}_t^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{t=1}^{i}\frac{\hat{e}_t}{\sqrt{n}\hat{\omega}}\right]^2$$

It is a sample average of the square of the standardized paritial sum $\sum_{t=1}^{\lfloor nr\rfloor}\frac{\hat{e}_t}{\sqrt{n}\hat{\omega}} \xrightarrow{d} W(r) - rW(1) = V(r)$ is a Brownian Bridge.

To see this point, consider the simple case when e_t is mds so $\sigma = \omega$

$$\sum_{t=1}^{\lfloor nr\rfloor}\frac{\hat{e}_t}{\sqrt{n}\sigma} = \sum_{t=1}^{\lfloor nr\rfloor}\frac{t - \frac{1}{n}\sum_{t=1}^{n}e_t}{\sqrt{n}\sigma} = \sum_{t=1}^{\lfloor nr\rfloor}\frac{e_t}{\sqrt{n}\sigma} - \frac{\lfloor nr\rfloor}{n}\sum_{t=1}^{n}\frac{e_t}{\sqrt{n}\sigma} = S_n(r) - rS(1)$$

as $\hat{e}_t = y_t - \overline{y}_n$. Thus $\mathrm{KPSS} \xrightarrow{d} \int_0^1 V(r)\,\mathrm{d}r$.

If a trend is added in the form $y_t = \mu + \theta S_t + e_t$, then

$$\mathrm{KPSS} \xrightarrow{d} \int_0^1 V_2(r)\,\mathrm{d}r$$

where $V_2(r)$ is a 2nd-type Brownian bridge.

$$V_2(r) = W(r) - \left(\int_0^r X(S)\,\mathrm{d}S\right)'\left(\int_0^1 XX'\right)^{-1}\int_0^1 X\mathrm{d}W$$

where $X(S) = \begin{pmatrix} 1 \\ S \end{pmatrix}$.

`Zhentao Shi. Apr 11, 2023. Transcribed by Shu Shen.`