Unit root processes and functional central limit theorem

A. Small-sample estimation properties for stationary AR(1)

Consider next an AR(1) process:

$$y_t = \rho y_{t-1} + \varepsilon_t$$

$\varepsilon_t \sim$ i.i.d. $(0, \sigma^2)$

$|\rho| < 1$

Then

$$E(y_t) = 0$$

Var$(y_t) = \sigma^2/(1-\rho^2)$

Estimate by OLS regression of y_t on y_{t-1}

$$\hat{\rho} = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

Substitute $y_t = \rho y_{t-1} + \varepsilon_t$

$$\hat{\rho} = \rho + \frac{\sum_{t=1}^{T} y_{t-1} \varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

$$\sqrt{T}(\hat{\rho}_T - \rho) = \frac{\sqrt{T}\, T^{-1} \sum_{t=1}^{T} y_{t-1} \varepsilon_t}{T^{-1} \sum_{t=1}^{T} y_{t-1}^2}$$

Numerator of $\sqrt{T}\,(\hat{\rho}_T - \rho)$

$$\sqrt{T}\,T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t$$

Notice $y_{t-1}\varepsilon_t$ is martingale difference

$$E(y_{t-1}\varepsilon_t y_{t-s}\varepsilon_{t-s}) = 0 \quad \text{for } s > 0$$

with mean $E(y_{t-1}\varepsilon_t) = 0$ and variance

$$E(y_{t-1}^2\varepsilon_t^2) = \sigma^4/(1-\rho^2)$$

So by central limit theorem,

$$\sqrt{T}\,T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \xrightarrow{L} N(0, \sigma^4/(1-\rho^2))$$

Denominator of $\sqrt{T}\,(\hat{\rho}_T - \rho)$

$$T^{-1}\sum_{t=1}^{T} y_{t-1}^2$$

is the sample mean of y_{t-1}^2 which by law of large numbers converges in probability to its population mean,

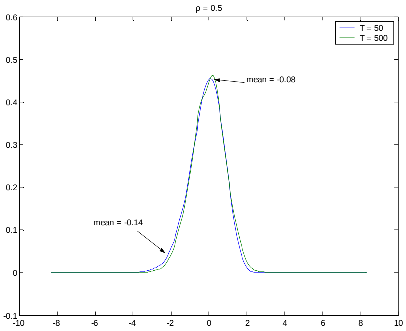
$$T^{-1}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{p} \sigma^2/(1-\rho^2)$$

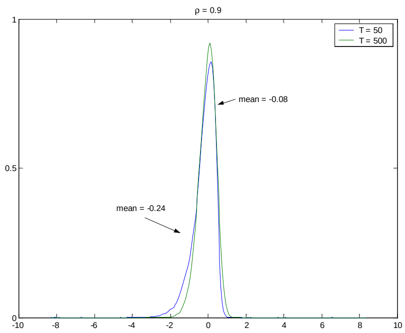
Conclusion:

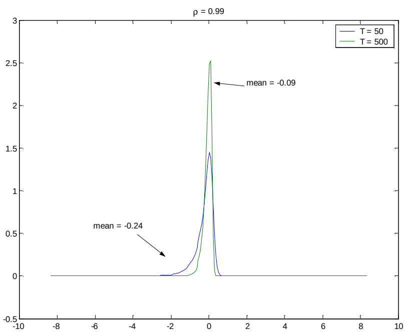
$$\sqrt{T}\,(\hat{\rho}_T - \rho) = \frac{\sqrt{T}\,T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{T^{-1}\sum_{t=1}^{T} y_{t-1}^2}$$

is asymptotically a $N(0, \sigma^4/(1-\rho^2))$ variable divided by the constant $\sigma^2/(1-\rho^2)$, or

$$\sqrt{T}\,(\hat{\rho}_T - \rho) \xrightarrow{L} N(0, 1-\rho^2)$$


ρ = 0.5
T = 50
T = 500
mean = -0.08
mean = -0.14


ρ = 0.9
T = 50
T = 500
mean = -0.08
mean = -0.24


ρ = 0.99
T = 50
T = 500
mean = -0.09
mean = -0.24

Observations

$\hat{\rho}$ is downward-biased in small samples

This bias is more severe as ρ gets bigger

Normal approximation gets better as T gets bigger

Variance of $\hat{\rho}$ around ρ gets smaller as ρ gets bigger

$$\sqrt{T}\,(\hat{\rho}_T - \rho) \xrightarrow{L} N(0, 1-\rho^2)$$

Next consider the OLS t-statistic

$$\tau_\rho = \frac{\hat{\rho}-\rho}{\hat{\sigma}_{\hat{\rho}}}$$

$$\hat{\sigma}^2_{\hat{\rho}} = \frac{\hat{\sigma}^2}{\sum_{t=1}^{T} y^2_{t-1}}$$

$$\tau_\rho = \frac{\sqrt{T}\,(\hat{\rho}_T - \rho)}{\sqrt{\hat{\sigma}^2_T/(T^{-1}\sum y^2_{t-1})}}$$

Numerator of τ_ρ

$$\sqrt{T}\,(\hat{\rho}_T - \rho) \xrightarrow{L} N(0, 1 - \rho^2)$$

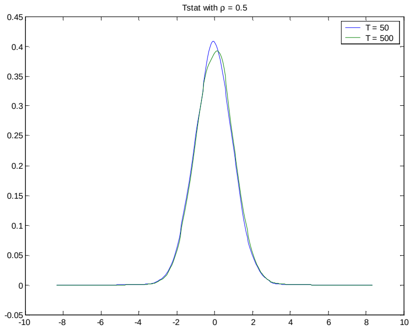
Denominator of τ_ρ

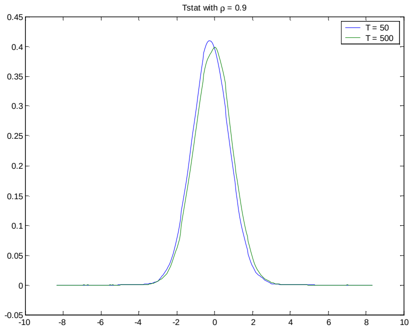
$$\sqrt{\hat{\sigma}_T^2 / (T^{-1} \sum y_{t-1}^2)} \xrightarrow{p} \sqrt{\sigma^2 / [\sigma^2 / (1 - \rho^2)]}$$

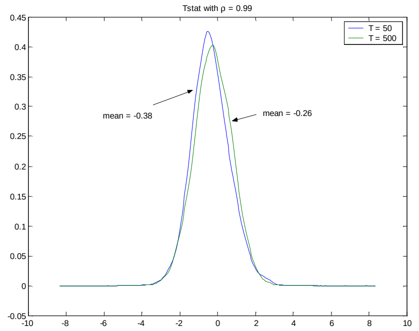
$$= \sqrt{1 - \rho^2}$$

Conclusion

$$\tau_\rho \xrightarrow{L} N(0,1)$$







Observations

τ_ρ is negatively skewed relative to the $N(0,1)$ in small samples

This skew is more severe as ρ gets bigger

Normal approximation gets better as T gets bigger

Unit root processes and functional central limit theorem

A. Small-sample estimation properties for stationary AR(1)

B. Properties of OLS estimate of ρ when true value is unity

Summary of asymptotic results:

$$\sqrt{T}\,(\hat{\rho}_T - \rho) \xrightarrow{L} N(0, 1-\rho^2)$$

$$\tau_\rho \xrightarrow{L} N(0,1)$$

Conjecture: when $\rho = 1$,

$$\sqrt{T}\,(\hat{\rho}_T - \rho) \xrightarrow{p} 0$$

$\tau_\rho \xrightarrow{L}$ something nonstandard

$$y_t = \rho y_{t-1} + \varepsilon_t$$

true $\rho_0 = 1$

$$\hat{\rho} = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

$$= \rho_0 + \frac{\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

$$\sqrt{T}\,(\hat{\rho}_T - \rho_0) \xrightarrow{p} 0$$

$$T(\hat{\rho}_T - \rho_0) = \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2}$$

Numerator of $T(\hat{\rho}_T - \rho_0) = T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t$

Notice under $H_0 : \rho_0 = 1$

$$y_t = y_{t-1} + \varepsilon_t$$

$$y_t^2 = y_{t-1}^2 + 2y_{t-1}\varepsilon_t + \varepsilon_t^2$$

$$y_{t-1}\varepsilon_t = (1/2)(y_t^2 - y_{t-1}^2 - \varepsilon_t^2)$$

$$T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t =$$

$$T^{-1}(1/2)\left(y_T^2 - y_0^2 - \sum_{t=1}^{T} \varepsilon_t^2\right)$$

Numerator of $T(\hat{\rho}_T - \rho_0)$

$$= T^{-1}(1/2)\left(y_T^2 - y_0^2 - \sum_{t=1}^{T} \varepsilon_t^2\right)$$

Suppose $y_0 = 0$ (asymptotically same)

$T^{-1} \sum_{t=1}^{T} \varepsilon_t^2 \xrightarrow{p} \sigma^2$

by law of large numbers

$$T^{-1} y_T^2 = T^{-1}(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_T)^2$$

$$= \left(\sqrt{T}\, T^{-1} \sum_{t=1}^{T} \varepsilon_t\right)^2$$

which by central limit theorem

converges to the square of a $N(0, \sigma^2)$

$= \sigma^2 \chi^2(1)$

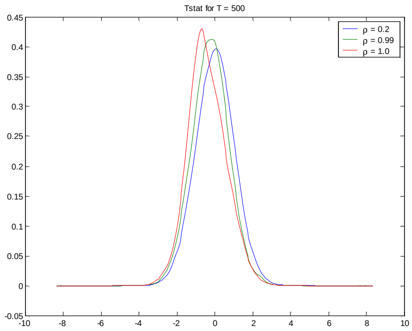
Conclusion: numerator of $T(\hat{\rho}_T - \rho_0)$

$$= T^{-1}(1/2)\left(y_T^2 - y_0^2 - \sum_{t=1}^{T} \varepsilon_t^2\right)$$

converges in distribution to

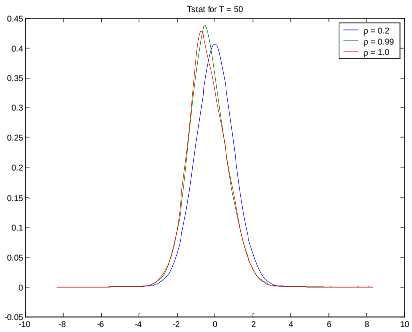
$\sigma^2 (1/2)[\chi^2(1) - 1]$

Denominator of $T(\hat{\rho}_T - \rho_0)$ also has a nonstandard asymptotic distribution when $\rho_0 = 1$, and t-statistic τ_ρ also has nonstandard distribution



Probability that $\tau_\rho < -1.96 = 0.05$

Probability that $N(0,1) < -1.96 = 0.025$



A more elegant characterization of these “nonstandard” distributions

Consider $u_t \sim$ i.i.d. $(0, 1)$

Define $X_T(r)$ for $r \in [0, 1]$ to be T^{-1} times the sum of the first rth fraction of a sample of size T

$$X_T(r) = T^{-1} \sum_{t:(t/T)\leq r} u_t$$

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

where

$[Tr]^* =$ greatest integer less than or equal to Tr

$X_T(1) =$ sample mean

$X_T(0) = 0$

Notice

$$\sqrt{T}\,X_T(r) = \sqrt{\frac{[Tr]^*}{T}}\sqrt{\frac{1}{[Tr]^*}}\sum_{t=1}^{[Tr]^*} u_t$$
$$\overset{L}{\to} N(0,r)$$

by the CLT

$$\sqrt{T}\,[X_T(r_2) - X_T(r_1)] \overset{L}{\to} N(0, r_2 - r_1)$$

and is independent of $X_T(r_1)$

for $r_2 > r_1$

Definition: Standard Brownian motion

$W(r)$ is a random process

$$W : r \in [0,1] \to \Re^1$$

such that

(a) $W(0) = 0$

(b) for any dates $0 \leq r_1 < r_2 < \cdots < r_k \leq 1$, the values $W(r_2) - W(r_1),\ W(r_3) - W(r_2),\ \ldots, W(r_k) - W(r_{k-1})$ are independent multivariate Gaussian with $W(s) - W(t) \sim N(0, s-t)$

(c) for any realization, $W(r)$ is continuous in r with probability 1

Can think of $W(r)$ as the limiting distribution of

$$\sqrt{T}\, X_T(r) = T^{-1/2} \sum_{t=1}^{[Tr]^*} u_t$$

$$\sqrt{T}\, X_T(.) \xrightarrow{L} W(.)$$

[functional central limit theorem]

Continuous mapping theorem:
If $S_T(.) \overset{L}{\rightarrow} S(.)$ and $g(.)$ is a continuous functional, then
$g[S_T(.)] \overset{L}{\rightarrow} g[S(.)]$

Example 1: if we multiply u_t by σ (so have white noise with variance σ^2), then multiply $W(.)$ by σ, that is, for $\varepsilon_t \sim$ i.i.d. $(0, \sigma^2)$,

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

$$= \sigma T^{-1} \sum_{t=1}^{[Tr]^*} u_t$$

$$\sqrt{T} X_T(.) \overset{L}{\rightarrow} \sigma W(.)$$

Example 2:

$$\sqrt{T}\int_0^1 X_T(r)\,dr \xrightarrow{L} \sigma\int_0^1 W(r)\,dr$$

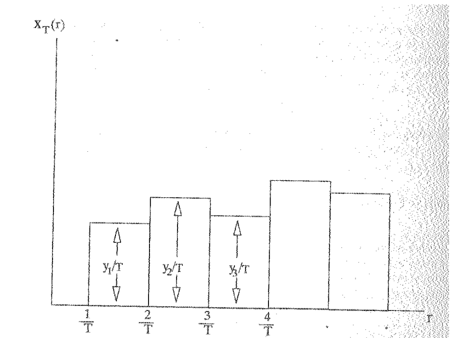


FIGURE 17.1 Plot of $X_T(r)$ as a function of r.

Usefulness of Example 2. Define

$\xi_t = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t$, $\xi_0 = 0$

Then $X_T(r) = T^{-1}\xi_{[Tr]^*}$

$$\int_0^1 X_T(r)\,dr = T^{-1}\int_0^1 \xi_{[Tr]^*}\,dr$$

$$= T^{-1}[\xi_1/T + \xi_2/T + \cdots + \xi_{T-1}/T]$$

$$\sqrt{T}\int_0^1 X_T(r)\,dr = T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}$$

Thus since

$$\sqrt{T}\int_0^1 X_T(r)\,dr \xrightarrow{L} \sigma\int_0^1 W(r)\,dr$$

it follows that

$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1} \xrightarrow{L} \sigma\int_0^1 W(r)\,dr$$

But $\xi_t = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t$

was a random walk

$$T^{-1/2}\sum_{t=1}^{T}\varepsilon_t \xrightarrow{L} \sigma W(1)$$

$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1} \xrightarrow{L} \sigma\int_0^1 W(r)\,dr$$

Example 3:

$$\begin{aligned} S_T(r) &= \left[\sqrt{T}X_T(r)\right]^2 \\ &\xrightarrow{L} \sigma^2[W(r)]^2 \\ &\sim \sigma^2 r^2 \cdot \chi^2(1) \end{aligned}$$

Example 4:

$$\int_0^1 S_T(r)\,dr = \frac{\xi_1^2}{T^2} + \frac{\xi_2^2}{T^2} + \cdots \frac{\xi_{T-1}^2}{T^2}$$

$$= T^{-2}\sum_{t=1}^{T} \xi_{t-1}^2$$

$$\xrightarrow{L} \sigma^2 \int_0^1 [W(r)]^2\,dr$$

So return to our result that for

$y_t = \rho y_{t-1} + \varepsilon_t$

true $\rho_0 = 1$

$$T(\hat{\rho}_T - \rho_0) = \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2}$$

We earlier saw that for numerator,

$$T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t$$

$$\xrightarrow{L} (1/2)\sigma^2[\chi^2(1) - 1]$$

$$= (1/2)\sigma^2\{W(1)^2 - 1\}$$

Denominator:

$$T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{L} \sigma^2 \int_0^1 [W(r)]^2\, dr$$

Conclusion:

$$T(\hat{\rho}_T - \rho_0) \xrightarrow{L} \frac{(1/2)\{W(1)^2 - 1\}}{\int_0^1 [W(r)]^2\, dr}$$

Can also show that, for the usual OLS t statistic, when $\rho_0 = 1$,

$$\frac{\hat{\rho}_T - \rho_0}{\hat{\sigma}_{\hat{\rho}}} \xrightarrow{L} \frac{(1/2)\{W(1)^2 - 1\}}{\left[\int_0^1 [W(r)]^2\, dr\right]^{1/2}}.$$

This is called the Dickey-Fuller (Case 1) distribution.

Using same methods it can be shown that if we include a constant term in the estimated regression,

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

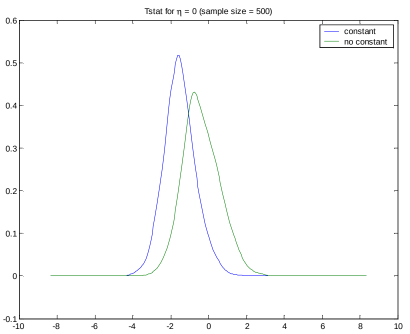
true $\alpha_0 = 0, \rho_0 = 1$

then the OLS t test is characterized by

$$\frac{\hat{\rho}_T - \rho_0}{\hat{\sigma}_{\hat{\rho}}} \xrightarrow{L}$$

$$\frac{(1/2)\{W(1)^2 - 1\} - W(1)\int_0^1 W(r)\,dr}{\left\{\int_0^1 [W(r)]^2\,dr - \left[\int_0^1 W(r)\,dr\right]^2\right\}^{1/2}}$$

which is called the Dickey-Fuller (Case 2) distribution



Unit root processes and functional central limit theorem

A. Small-sample estimation properties for stationary AR(1)
B. Properties of OLS estimate of ρ when true value is unity
C. Augmented Dickey-Fuller test

Suppose true process is

$\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \cdots + \phi_p \Delta y_{t-p} + \varepsilon_t$

ε_t ~ i.i.d. $(0, \sigma^2)$

$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$

implies $|z| > 1$

Then y_t ~ I(1).

What would happen if we estimated the following regession

$$\Delta y_t = \eta y_{t-1} + \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \cdots + \phi_p \Delta y_{t-p} + \varepsilon_t$$

(true $\eta = 0$).

Note variance of Δy_{t-j} is finite, but variance of y_{t-1} grows with t.

This suggests OLS estimate of $\hat{\eta}$ converges faster than the estimates of $\hat{\phi}_j$

Let $\mathbf{z}_t = (\Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-p})'$

$\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_p)'$

$\Delta y_t = \eta y_{t-1} + \mathbf{z}_t'\boldsymbol{\phi} + \varepsilon_t$

$$\begin{bmatrix} \hat{\eta} \\ \hat{\boldsymbol{\phi}} \end{bmatrix} = \begin{bmatrix} \sum y_{t-1}^2 & \sum y_{t-1}\mathbf{z}_t' \\ \sum \mathbf{z}_t y_{t-1} & \sum \mathbf{z}_t \mathbf{z}_t' \end{bmatrix}^{-1} \begin{bmatrix} \sum y_{t-1} y_t \\ \sum \mathbf{z}_t y_t \end{bmatrix}$$

$$\begin{bmatrix} \hat{\eta} \\ \hat{\boldsymbol{\phi}} \end{bmatrix} = \begin{bmatrix} \sum y_{t-1}^2 & \sum y_{t-1}\mathbf{z}_t' \\ \sum \mathbf{z}_t y_{t-1} & \sum \mathbf{z}_t \mathbf{z}_t' \end{bmatrix}^{-1} \begin{bmatrix} \sum y_{t-1}(\eta_0 y_{t-1} + \mathbf{z}_t'\boldsymbol{\phi}_0 + \varepsilon_t) \\ \sum \mathbf{z}_t(\eta_0 y_{t-1} + \mathbf{z}_t'\boldsymbol{\phi}_0 + \varepsilon_t) \end{bmatrix}$$

$$\begin{bmatrix} \hat{\eta} - \eta_0 \\ \hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0 \end{bmatrix} = \begin{bmatrix} \sum y_{t-1}^2 & \sum y_{t-1}\mathbf{z}_t' \\ \sum \mathbf{z}_t y_{t-1} & \sum \mathbf{z}_t \mathbf{z}_t' \end{bmatrix}^{-1} \begin{bmatrix} \sum y_{t-1} \varepsilon_t \\ \sum \mathbf{z}_t \varepsilon_t \end{bmatrix}$$

If $\hat{\eta}_T$ converges at rate T and $\hat{\phi}_T$ converges at rate $\sqrt{T}$, then we want to premultiply by

$$\Upsilon_T = \begin{bmatrix} T & 0 \\ 0 & T^{1/2}\mathbf{I}_{p-1} \end{bmatrix}$$

$$\Upsilon_T \begin{bmatrix} \hat{\eta}_T - \eta_0 \\ \hat{\phi}_T - \phi_0 \end{bmatrix} = \Upsilon_T \begin{bmatrix} \sum y_{t-1}^2 & \sum y_{t-1}\mathbf{z}_t' \\ \sum \mathbf{z}_t y_{t-1} & \sum \mathbf{z}_t\mathbf{z}_t' \end{bmatrix}^{-1} \Upsilon_T \Upsilon_T^{-1} \begin{bmatrix} \sum y_{t-1}\varepsilon_t \\ \sum \mathbf{z}_t\varepsilon_t \end{bmatrix}$$

$$\begin{bmatrix} T(\hat{\eta}_T - \eta_0) \\ T^{1/2}(\hat{\phi}_T - \phi_0) \end{bmatrix} = \left\{ \Upsilon_T^{-1} \begin{bmatrix} \sum y_{t-1}^2 & \sum y_{t-1}\mathbf{z}_t' \\ \sum \mathbf{z}_t y_{t-1} & \sum \mathbf{z}_t\mathbf{z}_t' \end{bmatrix} \Upsilon_T^{-1} \right\}^{-1} \Upsilon_T^{-1} \begin{bmatrix} \sum y_{t-1}\varepsilon_t \\ \sum \mathbf{z}_t\varepsilon_t \end{bmatrix}$$

$$\begin{bmatrix} T(\hat{\eta}_T - \eta_0) \\ T^{1/2}(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0) \end{bmatrix} = \begin{bmatrix} T^{-2}\sum y_{t-1}^2 & T^{-3/2}\sum y_{t-1}\mathbf{z}_t' \\ T^{-3/2}\sum \mathbf{z}_t y_{t-1} & T^{-1}\sum \mathbf{z}_t\mathbf{z}_t' \end{bmatrix}^{-1} \begin{bmatrix} T^{-1}\sum y_{t-1}\varepsilon_t \\ T^{-1/2}\sum \mathbf{z}_t\varepsilon_t \end{bmatrix}$$

Recall that we showed that when y_t follows a random walk $(\Delta y_t = \varepsilon_t)$,

$$T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \xrightarrow{L} (1/2)\sigma^2[\chi^2(1) - 1]$$

More generally, if

$$\Delta y_t = \psi_0\varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots$$

with $\sum_{j=0}^{\infty} |\psi_j| < \infty$, then

$$T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \xrightarrow{L} (1/2)\sigma^2\left(\sum_{j=0}^{\infty}\psi_j\right)[\chi^2(1) - 1]$$

Furthermore,

$$T^{-3/2}\sum_{t=1}^{T} y_{t-1}\Delta y_{t-j} \xrightarrow{p} 0$$

$$\begin{bmatrix} T(\hat{\eta}_T - \eta_0) \\ T^{1/2}(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0) \end{bmatrix} = \begin{bmatrix} T^{-2}\sum y_{t-1}^2 & T^{-3/2}\sum y_{t-1}\mathbf{z}_t' \\ T^{-3/2}\sum \mathbf{z}_t y_{t-1} & T^{-1}\sum \mathbf{z}_t\mathbf{z}_t' \end{bmatrix}^{-1} \begin{bmatrix} T^{-1}\sum y_{t-1}\varepsilon_t \\ T^{-1/2}\sum \mathbf{z}_t\varepsilon_t \end{bmatrix}$$

$$\begin{bmatrix} T(\hat{\eta}_T - \eta_0) \\ T^{1/2}(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \dfrac{(1/2)\sigma^2\left(\sum_{j=0}^{\infty}\psi_j\right)[\chi^2(1)-1]}{T^{-2}\sum y_{t-1}^2} \\ \left(T^{-1}\sum \mathbf{z}_t\mathbf{z}_t'\right)^{-1}(T^{-1/2}\sum \mathbf{z}_t\varepsilon_t) \end{bmatrix}$$

$$\left(T^{-1}\sum \mathbf{z}_t\mathbf{z}_t'\right)^{-1}(T^{-1/2}\sum \mathbf{z}_t\varepsilon_t) \xrightarrow{L}$$

$$\mathbf{q} \sim N(\mathbf{0}, \sigma^2\boldsymbol{\Gamma}_0^{-1})$$

$$\boldsymbol{\Gamma}_0 = E(\mathbf{z}_t\mathbf{z}_t') = \operatorname*{plim}_{T\to\infty} T^{-1}\sum \mathbf{z}_t\mathbf{z}_t'$$

Conclusion:

Asymptotic distribution of $T^{1/2}(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0)$ is same as standard case, in fact, identical to that if we hadn't estimated η at all!

All t and F tests involving $\boldsymbol{\phi}$ are asymptotically valid.

Asymptotic distribution of $\hat{\eta}$ is nonstandard, in fact, asymptotic distribution of t test of $\eta = 0$ is same (Dickey-Fuller case 1) as that of $\rho = 1$ in the regression

$$y_t = \rho y_{t-1} + \varepsilon_t$$

$T(\hat{\eta}_T - \eta_0) \xrightarrow{L}$ nonstandard

$$T^{1/2}(\hat{\eta}_T - \eta_0) \xrightarrow{p} 0$$

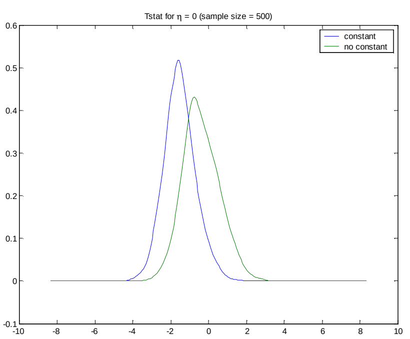
When constant is included in estimated regression,

$$\Delta y_t = c + \eta y_{t-1} + \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \cdots + \phi_p \Delta y_{t-p} + \varepsilon_t,$$

but true process is AR(p) in first-differences without constant or drift

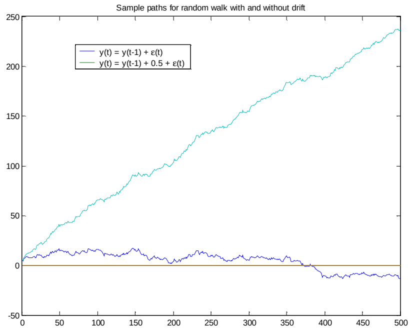
$$\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \cdots + \phi_p \Delta y_{t-p} + \varepsilon_t$$

then t-test of $\eta = 0$ has the Dickey-Fuller Case 2 distribution.



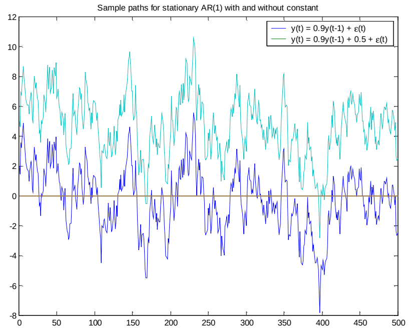
Why include constant in regression when true model doesn't have it and it makes distribution more skewed?

Answer: if true $\eta = 0$, it's obvious that c must be near zero

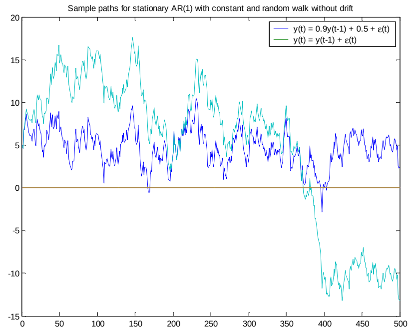


$$y_t = c + y_{t-1} + \varepsilon_t$$
$$\Rightarrow y_t = ct + y_0 + \sum_{\tau=1}^{t} \varepsilon_\tau$$
$$\Rightarrow y_t = ct + y_0 \pm 2\sigma\sqrt{t}$$

If true $\eta = 0$, it's obvious
that c must be near zero.
If true $\eta < 0$, it's obvious
that c must be greater than zero.



$H_0 : \eta = 0$ (only sensible if $c \simeq 0$)
$H_A : \eta < 0$ (only sensible if $c > 0$)



Conclusion: to test for unit root $(\eta = 0)$ estimate

$$\Delta y_t = c + \eta y_{t-1} + \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \cdots + \phi_p \Delta y_{t-p} + \varepsilon_t$$

Calculate distribution assuming true $c = 0,\ \eta = 0$ (i.e., compare t test of $\eta = 0$ with Case 2 in Table B.6)

What if data do exhibit a strong trend? Estimate

$$\Delta y_t = c + \delta t + \eta y_{t-1} + \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \cdots + \phi_p \Delta y_{t-p} + \varepsilon_t$$

$H_0 : \eta = 0$

(only sensible if $c > 0$ and $\delta = 0$)

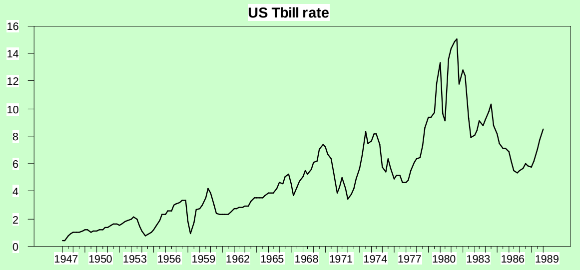
$H_A : \eta < 0$

(only sensible if $c > 0$ and $\delta > 0$)

Compare OLS t test of $\eta = 0$ with Case 4 in Table B.6)

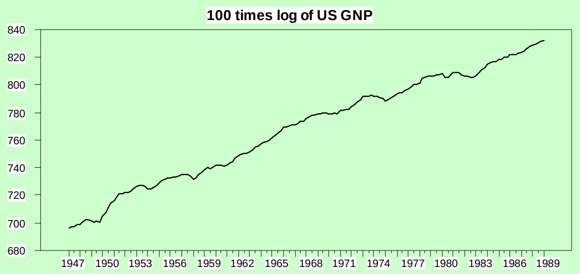
Example 17.8 in text

U.S. 3-month Treasury bill rate, 1947:Q2 to 1989:Q1



Sensible H_0 excludes a constant
Sensible H_A includes a constant

Use Case 2



Sensible H_0 includes a constant, excludes trend
Sensible H_A includes a constant and trend

Use Case 4

Caution on interpreting ADF tests

The ADF test assumes

$$y_t = \alpha + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

and tests

$$H_0 : \phi_1 + \cdots + \phi_p = 1$$

$$H_A : \phi_1 + \cdots + \phi_p < 1$$

If we reject H_0 that does not mean we reject the hypothesis that $y_t \sim I(1)$

Example:

a_t and ε_t uncorrelated white noise each with unit variance

$$\xi_t = \xi_{t-1} + a_t$$

$$y_t = \varepsilon_t + 10^{-6} \times \xi_t$$

$$(1-L)y_t = 10^{-6} a_t + (1-L)\varepsilon_t$$

$$s_{\Delta y}(0) = 10^{-12} > 0$$

So y_t is a unit root process, but is impossible to distinguish from white noise if $T < 10^6$.

Could not claim to have rejected this possibility based on observed data.

Issue: this $y_t \sim ARMA(1,1)$ for which $AR(p)$ is poor approximation.

Correct interpretation of ADF:

Given that I approximate with $AR(p)$, is $\phi_1 + \cdots + \phi_p = 1$ consistent with the data?

As T gets larger, could increase p so the set of considered models gets closer to set of all $I(1)$ processes.

Unit root processes and functional central limit theorem

A. Small-sample estimation properties for stationary AR(1)
B. Properties of OLS estimate of ρ when true value is unity
C. Augmented Dickey-Fuller test
D. Elliott, Rothenberg and Stock test

Note the ADF test conditions on initial p values for y_t.

ERS find that these values have information about constant and trend that could produce a more powerful test.

Uniformly most powerful test (against all $\rho = 1 + \eta$) does not exist. But choosing a test with strong power for $\rho = 1 + c/T$ with $c = -7.0$ (case 2) or $c = -13.5$ (case 4) works well.

Elliott-Rothenberg-Stock GLS augmented Dickey-Fuller test

Case 2: want to include constant under alternative but not under the null.
(1) Set $\rho = 1 - 7/T$ (e.g, $\rho = 0.93$ when $T = 100$.

(2) Calculate

$$\underset{(T\times 1)}{\mathbf{y}^*} = \begin{bmatrix} y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \vdots \\ y_T - \rho y_{T-1} \end{bmatrix} \qquad \underset{(T\times 1)}{\mathbf{x}^*} = \begin{bmatrix} 1 \\ 1-\rho \\ 1-\rho \\ \vdots \\ 1-\rho \end{bmatrix}$$

(3) Regress $\mathbf{y}^*$ on $\mathbf{x}^*$

$$\hat{c} = (\mathbf{x}^{*\prime}\mathbf{x}^*)^{-1}(\mathbf{x}^{*\prime}\mathbf{y}^*)$$

(4) Calculate $\tilde{y}_t = y_t - \hat{c}$

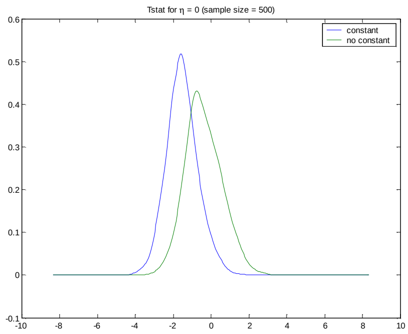
(5) Do ADF using $\tilde{y}_t$ with no constant term:

$$\Delta\tilde{y}_t = \eta\tilde{y}_{t-1} + \phi_1\Delta\tilde{y}_{t-1} + \phi_2\Delta\tilde{y}_{t-2} \cdots + \phi_p\Delta\tilde{y}_{t-p} + \varepsilon_t$$

$$t = p+1, p+2, \ldots, T$$

Compare OLS t stat for $\eta = 0$ with Case 1 in Table B.6.

Difference between case 1 and case 2 DF distributions



Elliott-Rothenberg-Stock GLS augmented Dickey-Fuller test

Case 4: want to include constant and time trend under alternative but not under the null.

(1) Set $\rho = 1 - 13.5/T$ (e.g, $\rho = 0.865$ when $T = 100$.

(2) Calculate

$$\underset{(T\times 1)}{\mathbf{y}^*} = \begin{bmatrix} y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \vdots \\ y_T - \rho y_{T-1} \end{bmatrix}$$

$$\underset{(T\times 2)}{\mathbf{x}^*} = \begin{bmatrix} 1 & 1 \\ 1-\rho & 2-\rho 1 \\ 1-\rho & 3-\rho 2 \\ \vdots & \vdots \\ 1-\rho & T-\rho(T-1) \end{bmatrix}$$

(3) Regress $\mathbf{y}^*$ on $\mathbf{x}^*$

$$\begin{bmatrix} \hat{c} \\ \hat{\delta} \end{bmatrix} = (\mathbf{x}^{*\prime}\mathbf{x}^*)^{-1}(\mathbf{x}^{*\prime}\mathbf{y}^*)$$

(4) Calculate $\tilde{y}_t = y_t - \hat{c} - \hat{\delta} t$

(5) Do ADF using $\tilde{y}_t$ with no constant term or trend:

$$\Delta\tilde{y}_t = \eta\tilde{y}_{t-1} + \phi_1\Delta\tilde{y}_{t-1} + \phi_2\Delta\tilde{y}_{t-2}$$
$$\cdots + \phi_p\Delta\tilde{y}_{t-p} + \varepsilon_t$$
$$t = p+1, p+2, \ldots, T$$

Compare OLS t stat for $\eta = 0$ with Panel C in Table 1 of ERS.

Unit root processes and functional central limit theorem

A. Small-sample estimation properties for stationary AR(1)
B. Properties of OLS estimate of ρ when true value is unity
C. Augmented Dickey-Fuller test
D. Elliott, Rothenberg and Stock test
E. Consequences for nonstationary AR(p)

Above discussion considered:

$$\Delta y_t = \eta y_{t-1} + \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \cdots + \phi_p \Delta y_{t-p} + \varepsilon_t$$

Now consider OLS estimation of

$$\begin{aligned} y_t &= \rho_1 y_{t-1} + \rho_2 y_{t-2} + \cdots + \rho_{p+1} y_{t-p-1} + \varepsilon_t \\ &= (1 + \eta + \phi_1) y_{t-1} + (\phi_2 - \phi_1) y_{t-2} + \cdots + (-\phi_p) y_{t-p-1} + \varepsilon_t \end{aligned}$$

OLS: $\hat{\rho}_1 \equiv 1 + \hat{\eta} + \hat{\phi}_1$

Suppose we wanted to do a hypothesis test about the value of $\eta + \phi_1$

$$T^{1/2} \hat{\eta}_T \xrightarrow{p} \eta_0$$

$$T^{1/2} (\hat{\phi}_{1T} - \phi_{10}) \xrightarrow{L} N(0, v)$$

$$\Rightarrow T^{1/2} (\hat{\phi}_{1T} + \hat{\eta}_T - \phi_{10} - \eta_0) \xrightarrow{L} N(0, v)$$

OLS: $\hat{\rho}_1 \equiv 1 + \hat{\eta} + \hat{\phi}_1$

$$T^{1/2} (\hat{\rho}_{1T} - \rho_{10}) \xrightarrow{L} N(0, v)$$

Original regression:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t$$

$$\hat{\boldsymbol{\beta}} = \left(\sum \mathbf{x}_t\mathbf{x}_t'\right)^{-1}\left(\sum \mathbf{x}_t y_t\right)$$

Transformed regression:

$$y_t = \mathbf{x}_t^{*\prime}\boldsymbol{\beta}^* + \varepsilon_t$$

$$\mathbf{x}_t^* = \mathbf{H}\mathbf{x}_t$$

$$\hat{\boldsymbol{\beta}}^* = \left(\sum \mathbf{x}_t^*\mathbf{x}_t^{*\prime}\right)^{-1}\left(\sum \mathbf{x}_t^* y_t\right)$$

$$= \left(\mathbf{H}\sum \mathbf{x}_t\mathbf{x}_t'\mathbf{H}'\right)^{-1}\left(\mathbf{H}\sum \mathbf{x}_t y_t\right)$$

$$= \mathbf{H}'^{-1}\hat{\boldsymbol{\beta}}$$

Summary

True model:

$$\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \cdots + \phi_p \Delta y_{t-p} + \varepsilon_t$$

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

implies $|z| > 1$

Estimated model:

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \cdots + \rho_{p+1} y_{t-p-1} + \varepsilon_t$$

Each estimate $\hat{\rho}_j$ is asymptotically Normal.

A standard t test of $\hat{\rho}_j = \rho_{j0}$ is asymptotically valid for any j.

Most F tests on the ρ_j's are asymptotically valid.

The only problem comes from the fact that

$$\hat{\rho}_1 + \hat{\rho}_2 + \cdots + \hat{\rho}_{p+1} = 1 + \hat{\eta},$$

which has a nonstandard distribution (though Normal approximation to t statistic is not too bad).

To test null hypothesis that

$$\rho_1 + \rho_2 + \cdots + \rho_{p+1} = 1$$

(i.e., null hypothesis that there is a unit root), compare t test of this hypothesis with the distribution derived above for the case

$$y_t = \rho y_{t-1} + \varepsilon_t$$

What about the following?

True model:

$$\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \cdots + \phi_p \Delta y_{t-p} + \varepsilon_t$$

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

implies $|z| > 1$ (same as before)

Estimated model:

$$\Delta y_t = c + \eta y_{t-1} + \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \cdots + \phi_p \Delta y_{t-p} + \varepsilon_t$$

(added a constant term)

Results: individual coefficients $\hat{\phi}_j$ are asymptotically Normal, and t statistics for $\hat{\phi}_j = \phi_{j0}$ are asymptotically valid.

This implies same for t statistics on $\hat{\rho}_j$ in

$$y_t = c + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \cdots + \rho_{p+1} y_{t-p-1} + \varepsilon_t$$

$$\Delta y_t = c + \eta y_{t-1} + \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \cdots + \phi_p \Delta y_{t-p} + \varepsilon_t$$

Both $\hat{c}$ and $\hat{\eta}$ have nonstandard asymptotic distributions, and t test of $\eta = 0$ has different asymptotic distribution from case with no constant term (and more different from $N(0,1)$ distribution).

If we estimate in the form

$$y_t = c + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \cdots + \rho_{p+1} y_{t-p-1} + \varepsilon_t,$$

problematic hypothesis tests:

$$c = c_0$$

$$\rho_1 = \rho_2 = \cdots = \rho_{p+1} = 0$$

$$\rho_1 + \rho_2 + \cdots + \rho_{p+1} = 1$$

all other hypothesis tests, done in usual way, are asymptotically ok