UNIVERSITÀ DEGLI STUDI DI MILANO
Dipartimento di Economia, Management e Metodi Quantitativi

Academic Year 2019-2020

Time Series Econometics

Fabrizio Iacone

Chapter 10: Unit Root testing

Topics: Brownian motion, Functional central limit theorem, Limit properties of the sample mean of a random walk, Limit properties of the OLS estimate of the autoregressive parameter in a random walk, Limit properties of the t statistic associated to the OLS estimate of the autoregressive parameter in a random walk, The Dickey Fuller test for a unit root in a random walk: Case 1, The Dickey Fuller test for a unit root in a random walk: Case 2, The Dickey Fuller test for a unit root in a random walk with drift: Case 3, The Dickey Fuller test for a unit root in a random walk with drift: Case 4, Choice of the unit root test, Augmented Dickey Fuller test for a unit root when the disturbances have a stationary AR(p) structure: Case 1, Case 2, Case 3, Case 4, Choice of the order p in the ADF test, Phillips-Perron tests for a unit root in a generic I(1) process

We saw that

$$Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t, \;\; \varepsilon_t \; w.n.(0,\sigma^2), \text{ when } t > 0$$

$$Y_t = 0 \text{ when } t \leq 0$$

has different properties depending on whether $\rho = 1$ or $|\rho| < 1$.

We want a test to distinguish between the two cases.

Introduce

Brownian motion (heuristic)

A Browian motion $W(.)$ is a continuous time stochastic process that associates to each date $t \in [0,1]$ a value $W(t)$ such that

★ $W(0) = 0$

★ for any date $0 \leq t_1 < t_e < \ldots < t_k \leq 1$, the differences $W(t_2) - W(t_1)$, $W(t_3) - W(t_2)$, ..., $W(t_k) - W(t_{k-1})$ are normally independently distributed random variables such that, for s, $0 \leq t < s \leq 1$,

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

★ $W(t)$ is continous with probability 1

Introduce the operator $[.]^*$, such that $[x]^*$ returns the integer part of a number x. Introduce

$$X_T(r) = \frac{1}{T}\sum_{t=1}^{[rT]^*} \varepsilon_t,\ \varepsilon_t \text{ i.i.d.}(0,\sigma^2),\text{ for } r \in [0,1]$$

Functional Central Limit Theorem (heuristic)

$$\sqrt{T}X_T(.)/\sigma \to_d W(.)$$

(here and after, these limits are as $T \to \infty$)

(The FCLT links functions on $[0,1]$: we should define what convergence in distribution means, there. It turns out that the nature of the convergence, and even the notation, have to be generalised; however, we do not discuss this).

The Central Limit Theorem,

$$\sqrt{T}\,\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t \to_d N(0,\sigma^2)$$

is a byproduct of the FCLT:

$$\sqrt{T}X_T(1)/\sigma \to_d W(1).$$

(just set $r = 1$ in the FCLT)

Now, we can see what happens to the sample mean of an $I(1)$ process

$$Y_t = Y_{t-1} + \varepsilon_t,\ \ \varepsilon_t\ i.i.d.\ (0, \sigma^2),\ \text{ when } t > 0$$

$$Y_t = 0 \text{ when } t \leq 0$$

We can express $Y_1, \ldots, Y_t$ as a function of $X_T(r)$:

$$X_T(.) = \begin{cases} 0 \text{ for } 0 \leq r < 1/T \\ Y_1/T \text{ for } 1/T \leq r < 2/T \\ Y_2/T \text{ for } 2/T \leq r < 3/T \\ \cdots \\ Y_t/T \text{ for } t/T \leq r < (t+1)/T \\ \cdots \\ Y_{T-1}/T \text{ for } (T-1)/T \leq r < 1 \\ Y_T/T \text{ for } r = 1 \end{cases}$$

$X_T(.)$ is a step function:
for $t/T \leq r < (t+1)/T,\ \ X_T(.) = Y_t/T$.

For any constant c,

$$\int_{t/T}^{(t+1)/T} c dr = c|r|_{t/T}^{(t+1)/T} = c\frac{1}{T}.$$

In the same way, we can compute

$$\int_{t/T}^{(t+1)/T} X_T(r)dr = Y_t/T * 1/T = Y_t/T^2.$$

Then,

$$\begin{aligned} & Y_0/T^2 \qquad\qquad + \ldots + Y_t/T^2 \qquad\qquad + \ldots + Y_{T-1}/T^2 \\ = & \int_{0/T}^{1/T} X_T(r)dr + \ldots + \int_{t/T}^{(t+1)/T} X_T(r)dr + \ldots + \int_{(T-1)/T}^{T/T} X_T(r)dr \end{aligned}$$

ie.

$$\frac{1}{T^2}\sum_{t=1}^{T} Y_{t-1} = \int_0^1 X_T(r)dr$$

From the FCLT we know that

$$\sqrt{T}X_T(.)/\sigma \rightarrow_d W(.),$$

so

$$\sqrt{T}\frac{1}{T^2}\sum_{t=1}^{T} Y_{t-1}/\sigma = \int_0^1 \sqrt{T}X_T(r)/\sigma dr \rightarrow_d \int_0^1 W(r)dr$$

What is $\int_0^1 W(r)dr$? It is a random variable, obtained by reweighting and averaging normally distributed random variables.
In particular, $\int_0^1 W(r)dr$ is a $N(0, 1/3)$.

We can now conclude

$$\frac{1}{\sqrt{T}} \frac{1}{T} \sum_{t=1}^{T} Y_{t-1} \rightarrow_d \sigma \int_0^1 W(r)dr,$$

which is $N(0, 1/3\sigma^2)$.

Since

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

$$= \frac{1}{\sqrt{T}} \frac{1}{T} \sum_{t=0}^{T-1} Y_t + \frac{1}{\sqrt{T}} \frac{1}{T} Y_T - \frac{1}{\sqrt{T}} \frac{1}{T} Y_0$$

$$= \frac{1}{\sqrt{T}} \frac{1}{T} \sum_{t=1}^{T} Y_{t-1} + \frac{1}{\sqrt{T}} \frac{1}{T} Y_T - \frac{1}{\sqrt{T}} \frac{1}{T} Y_0,$$

notice that $Y_0 = 0$, and that $\frac{1}{\sqrt{T}} \frac{1}{T} Y_T \rightarrow_p 0$, so

$$\frac{1}{\sqrt{T}} \bar{Y} \rightarrow_d \sigma \int_0^1 W(r)dr$$

as well.

A test to check if Y_t is a random walk:

Estimate ρ via OLS in

$$Y_t = \rho Y_{t-1} + \varepsilon_t,\ \ \varepsilon_t\ i.i.d.\ (0, \sigma^2), \text{ when } t > 0$$

$$Y_t = 0 \text{ when } t \leq 0$$

When $\rho = 1$,

$$\widehat{\rho} = \frac{\sum_{t=2}^{T} Y_t Y_{t-1}}{\sum_{t=2}^{T} Y_{t-1}^2} = \frac{\sum_{t=2}^{T} (Y_{t-1} + \varepsilon_t) Y_{t-1}}{\sum_{t=2}^{T} Y_{t-1}^2}$$

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

In order to find out more about $\sum_{t=2}^{T} Y_{t-1}^2$,

$$X_T(.)^2 = \begin{cases} 0 \text{ for } 0 \leq r < 1/T \\ Y_1^2/T^2 \text{ for } 1/T \leq r < 2/T \\ Y_2^2/T^2 \text{ for } 2/T \leq r < 3/T \\ \cdots \\ Y_t^2/T^2 \text{ for } t/T \leq r < (t+1)/T \\ \cdots \\ Y_{T-1}^2/T^2 \text{ for } (T-1)/T \leq r < 1 \\ Y_T^2/T^2 \text{ for } r = 1 \end{cases}$$

$X_T(.)^2$ is a step function: for $t/T \leq r < (t+1)/T$, $X_T(.)^2 = Y_t^2/T^2$, so

$$\int_{t/T}^{(t+1)/T} X_T(r)^2 dr = Y_t^2/T^2 * 1/T = Y_t^2/T^3.$$

Then,

$$\begin{aligned} & Y_0^2/T^3 \quad + \ldots + Y_t^2/T^3 \quad + \ldots + Y_{T-1}^2/T^3 \\ = & \int_{0/T}^{1/T} X_T(r)^2 dr + \ldots + \int_{t/T}^{(t+1)/T} X_T(r)^2 dr + \ldots \\ & + \int_{(T-1)/T}^{T/T} X_T(r)^2 dr \end{aligned}$$

i.e.

$$\frac{1}{T^3} \sum_{t=1}^{T} Y_{t-1}^2 = \int_0^1 X_T(r)^2 dr$$

From the FCLT, we can immediately derive

$$T X_T(.)^2/\sigma^2 \rightarrow_d W(.)^2,$$

($W(r)^2$ is a well defined random variable, because $W(r)^2/r$ is a χ_1^2) so

$$T \frac{1}{T^3} \sum_{t=1}^{T} Y_{t-1}^2/\sigma^2 = \int_0^1 T X_T(r)^2/\sigma^2 dr \rightarrow_d \int_0^1 W(r)^2 dr,$$

so we can conclude

$$\frac{1}{T^2} \sum_{t=1}^{T} Y_{t-1}^2 = \int_0^1 T X_T(r)^2 dr \rightarrow_d \sigma^2 \int_0^1 W(r)^2 dr.$$

In order to find out more about $\sum_{t=2}^{T} \varepsilon_t Y_{t-1}$, consider

$$Y_t^2 = (Y_{t-1} + \varepsilon_t)^2 = Y_{t-1}^2 + \varepsilon_t^2 + 2Y_{t-1}\varepsilon_t$$

so, rearranging terms,

$$Y_t^2 - Y_{t-1}^2 - \varepsilon_t^2 = 2Y_{t-1}\varepsilon_t.$$

Summing over t, $t = 1, \ldots, T$,

$$\sum_{t=1}^{T} Y_t^2 - \sum_{t=1}^{T} Y_{t-1}^2 - \sum_{t=1}^{T} \varepsilon_t^2 = 2\sum_{t=1}^{T} Y_{t-1}\varepsilon_t$$

and

$$\begin{aligned} &\sum_{t=1}^{T} Y_t^2 - \sum_{t=1}^{T} Y_{t-1}^2 \\ &= (Y_1^2 + Y_2^2 + \ldots + Y_t^2 + \ldots + Y_{T-1}^2 + Y_T^2) \\ &\quad - (Y_0^2 + Y_1^2 + \ldots + Y_{t-1}^2 + \ldots + Y_{T-2}^2 + Y_{T-1}^2) \\ &= Y_T^2 - Y_0^2 = Y_T^2 \end{aligned}$$

because $Y_0 = 0$, so

$$\sum_{t=1}^{T} Y_{t-1}\varepsilon_t = \frac{1}{2}\left(Y_T^2 - \sum_{t=1}^{T} \varepsilon_t^2 \right).$$

Normalising by T,

$$\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t = \frac{1}{2}\left(\frac{1}{T}Y_T^2 - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 \right).$$

Since

$$\frac{1}{T}Y_T^2 = TX_T(1)^2 \to_d \sigma^2 W(1)^2$$

(by the CLT), and

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 \to_p \sigma^2$$

(by the law of large numbers) then

$$\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t \to_d \frac{1}{2}\sigma^2\left(W(1)^2 - 1\right).$$

Summarising,

$$T(\widehat{\rho} - 1) = \frac{\frac{1}{T}\sum_{t=2}^{T}\varepsilon_t Y_{t-1}}{\frac{1}{T^2}\sum_{t=2}^{T} Y_{t-1}^2} \to_d \frac{\frac{1}{2}\left(W(1)^2 - 1\right)}{\int_0^1 W(r)^2 dr}$$

★ $\widehat{\rho}$ is still consistent ($\widehat{\rho} \to_p 1$)

★ indeed, $\widehat{\rho}$ is "superconsistent" (see the rate T rather then the usual $\sqrt{T}$)

★ $\dfrac{\frac{1}{2}\left(W(1)^2-1\right)}{\int_0^1 W(r)^2 dr}$ is not a normal distribution

★ in small samples (and $\varepsilon_t\, Nid(0,\sigma^2)$), $\widehat{\rho}$ underestimates 1 (in a probabilistic sense)

★ $\dfrac{\frac{1}{2}\left(W(1)^2-1\right)}{\int_0^1 W(r)^2 dr}$ is skewed to the left

Testing

$$H_0 : \{\rho = 1\} \text{ vs } H_A : \{|\rho| < 1\}$$

in

$$Y_t = \rho Y_{t-1} + \varepsilon_t,\ \ \varepsilon_t\ \ i.i.d.(0,\sigma^2) \text{ when } t > 0$$

$$Y_t = 0 \text{ when } t \leq 0$$

the 5% critical value for the $T(\widehat{\rho} - 1)$ statistic is -8.1.

t –statistic:

$$t = \frac{(\hat{\rho} - \rho)}{\hat{\sigma}_{\hat{\rho}}}$$

$$\text{where } \hat{\sigma}^2_{\hat{\rho}} = \frac{s^2}{\sum_{t=2}^{T} Y^2_{t-1}}$$

$$\text{and } s^2 = \frac{1}{T-1} \sum_{t=2}^{T} (Y_t - \hat{\rho} Y_{t-1})^2$$

When $|\rho| = 1$,

rewrite

$$t = \frac{T(\hat{\rho} - \rho)}{T\hat{\sigma}_{\hat{\rho}}}.$$

Look at $T\hat{\sigma}_{\hat{\rho}}$ first.

again,

$$\hat{\rho} \to_p \rho, \text{ so } s^2 = \frac{1}{T-1} \sum_{t=2}^{T} (Y_t - \hat{\rho} Y_{t-1})^2 \to_p \sigma^2.$$

Since we already saw that

$$\frac{1}{T^2} \sum_{t=2}^{T} Y^2_{t-1} \to_d \sigma^2 \int_0^1 W(r)^2 dr,$$

then

$$T^2\widehat{\sigma}_{\widehat{\rho}}^2 = \frac{s^2}{\frac{1}{T^2}\sum_{t=2}^{T} Y_{t-1}^2}$$

$$\to_d \frac{\sigma^2}{\sigma^2 \int_0^1 W(r)^2 dr} = \frac{1}{\int_0^1 W(r)^2 dr}$$

and $T\widehat{\sigma}_{\widehat{\rho}} \to_d \dfrac{1}{\sqrt{\int_0^1 W(r)^2 dr}}$

As for the numerator,

$$T(\widehat{\rho} - 1) \to_d \frac{\frac{1}{2}\left(W(1)^2 - 1\right)}{\int_0^1 W(r)^2 dr}$$

summarising,

$$t = \frac{T(\widehat{\rho} - 1)}{T\widehat{\sigma}_{\widehat{\rho}}}, \quad t \to_d \frac{\frac{1}{2}\left(W(1)^2 - 1\right)}{\sqrt{\int_0^1 W(r)^2 dr}}.$$

★ $\frac{\frac{1}{2}\left(W(1)^2-1\right)}{\sqrt{\int_0^1 W(r)^2 dr}}$ is not normally distributed; it is skewed to the left.

Testing $H_0 : \{\rho = 1\}$ vs. $H_A : \{|\rho| < 1\}$ with a t statistic using a 5% significance level, the critical value is -1.95.

Compare with the case $|\rho| < 1$:

$$\widehat{\rho} \to_p \rho,$$

$$T\widehat{\sigma}_{\widehat{\rho}}^2 = \frac{s^2}{\frac{1}{T}\sum_{t=2}^{T} Y_{t-1}^2} \to_p \frac{\sigma^2}{\frac{\sigma^2}{1-\phi^2}} = 1 - \phi^2$$

so

$$t = \frac{\sqrt{T}\,(\widehat{\rho} - \rho)}{\sqrt{T}\,\widehat{\sigma}_{\widehat{\rho}}}, t \to_d N(0, 1).$$

Then testing $H_0 : \{\rho = \phi\}$ vs. $H_A : \{\rho < \phi\}$ (when $|\phi| < 1$) with a t statistic, with a 5% significance level, the critical value is -1.65.

Which unit root test?

Recall the model

$$Y_t = \rho Y_{t-1} + \varepsilon_t, \ \ \varepsilon_t \ i.i.d. \ (0, \sigma^2) \text{ when } t > 0$$

$$Y_t = 0 \text{ when } t \leq 0$$

and $\rho = 1$ or $|\rho| < 1$;

let $\widehat{\rho}$ be the OLS estimate of ρ:

since $\widehat{\rho} \to_p \rho$, we can use the $T(\widehat{\rho} - 1)$ or the t statistic to test for a unit root testing $H_0 : \{\rho = 1\}$ vs $H_A : \{|\rho| < 1\}$.

However, when $|\rho| < 1$, so far we only considered processes Y_t that have $E(Y_t) = 0$. How about processes that are mean reverting and yet the mean to which they revert is not zero? Processes of this kind would be generated by

$$Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t \text{ with } \alpha \neq 0, |\rho| < 1$$

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

If this is the true model and we omit α, estimating $\widehat{\rho} = \frac{\sum_{t=2}^{T} Y_{t-1} Y_t}{\sum_{t=2}^{T} Y_{t-1}^2}$ instead, then $\widehat{\rho}$ is no longer a consistent estimate of ρ: however, $\widehat{\rho}$ converges in probability to a number smaller than one, so we can still rely on the $T(\widehat{\rho} - 1)$ or the t statistics to effectively test for a unit root.

"Case 1"

Estimate ρ via OLS in

$$Y_t = \rho Y_{t-1} + \varepsilon_t$$

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

When $\rho = 1$,

$$T(\hat{\rho} - 1) \to_d \frac{\frac{1}{2}\left(W(1)^2 - 1\right)}{\int_0^1 W(r)^2 dr}, \; t \to_d \frac{\frac{1}{2}\left(W(1)^2 - 1\right)}{\sqrt{\int_0^1 W(r)^2 dr}}$$

✠ Test:

Test $H_0 : \{\rho = 1\}$ vs. $H_A : \{|\rho| < 1\}$ with a t statistic (critical value is -1.95 at 5% significance level) (can also use the $T(\hat{\rho} - 1)$ statistic, the 5% critical value is -8.1).

"Case 2"

Estimate α, ρ via OLS in

$$Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t$$

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

Here $\widehat{\rho}$ is a consistent estimate of ρ regardless of α and ρ.

When $\rho = 1$, in order to have Y_t as a random walk (i.e., no linear trend) we also need $\alpha = 0$: we take it into account when computing the limit distribution of $T(\widehat{\rho} - 1)$ and of the t statistic $\frac{(\widehat{\rho}-1)}{\widehat{\sigma}_{\widehat{\rho}}}$.

When $\alpha = 0$, $\rho = 1$:

$$T(\widehat{\rho} - 1) \to_d \frac{\frac{1}{2}\left(W(1)^2 - 1\right) - W(1)\int_0^1 W(r)dr}{\int_0^1 W(r)^2 dr - \left(\int_0^1 W(r)dr\right)^2}$$

$$t \to_d \frac{\frac{1}{2}\left(W(1)^2 - 1\right) - W(1)\int_0^1 W(r)dr}{\sqrt{\int_0^1 W(r)^2 dr - \left(\int_0^1 W(r)dr\right)^2}}$$

★the limit distributions of $T(\widehat{\rho}-1)$ and of t when $\alpha = 0$ are not normal; they are also even more asymmetric than in Case 1

★the limit distribution of $\sqrt{T}\,\widehat{\alpha}$ when $\alpha = 0$ is not normal

✠ Test:

Test $H_0 : \{\rho = 1\}$ vs. $H_A : \{|\rho| < 1\}$ with a t statistic (critical value is -2.86 at 5% significance level) (can use the $T(\widehat{\rho}-1)$ statistic, the 5% critical value would be -14.1) (the limit distributions of the t and of the $T(\widehat{\rho}-1)$ statistics are computed under the assumption $\alpha = 0$).

Joint test, $H_0 : \left\{\alpha = 0, \rho = 1\right\}$ vs $H_A : \left\{\alpha \neq 0 \text{ \&/or } \rho \neq 1\right\}$ (the F test statistic associated to this test does not converge to $1/2\ \chi^2_2$: the 5% critical value is 4.59, as opposed to 2.99).

Which test then?

If Y_t does not have a unit root and $E(Y_t) \neq 0$, in Case 1 we overestimate ρ (in a probabilistic sense) a bit: the test will still be useful to detect a unit root, but it may have less power than a test in which a consistent estimate of ρ is used.

On the other hand, if If Y_t does not have a unit root and $E(Y_t) = 0$, then the two estimates of ρ (using Case 2 or Case 1) have the same limit distribution: however, the critical value for case 2 is smaller (-2.86 instead of -1.95), so in a finite sample there will be a higher proportion of Type 2 errors when using Case 2.

Finally, also notice that the t test has "one-sided" alternative, as opposed to the "two-sided" alternatives in the joint test in Case 2: one-sided alternative use more information (in this case, the knowledge that ρ is not bigger than 1) and this pays off because it gives more power.

The choice between the Case 1 and the Case 2 model then depends on how confident we can be of $\alpha = 0$ if $|\rho| < 1$: if we have no reasons to expect $\alpha = 0$ if $|\rho| < 1$, Case 2 should be preferred.

What if there is a linear trend?

If $\alpha \neq 0$ in $Y_t = \alpha + Y_{t-1} + \varepsilon_t$ $(t > 0)$, by repeated substitution

$$Y_t = \alpha t + \sum_{j=1}^{t} \varepsilon_j,$$

so the process has a linear trend, together with the random walk $\sum_{j=1}^{t} \varepsilon_j$.

"Case 3"

estimate α, ρ in

$$Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t$$

assuming ε_t $i.i.d.(0,\sigma^2)$.

When $\alpha \neq 0$, $\rho = 1$

$$T^{3/2}(\widehat{\rho} - 1) \to_d N\left(0, \frac{12}{\alpha^2}\sigma^2\right), \quad t \to_d N(0,1).$$

★even faster rate of convergence, and limit normality

✠ Test:

Test $H_0 : \{\rho = 1\}$ vs. $H_A : \{|\rho| < 1\}$ with a $T^{3/2}(\widehat{\rho} - 1)$ or a t statistic (the limit distributions of the $T^{3/2}(\widehat{\rho} - 1)$ and of the t statistics are computed under the assumption $\alpha \neq 0$)

"Case 4"

estimate α, ρ, δ in

$$Y_t = \alpha + \rho Y_{t-1} + \delta t + \varepsilon_t$$

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

When $\rho = 1, \delta = 0$:

★ the $T(\widehat{\rho} - 1)$ and the t statistics to test $H_0 : \{\rho = 1\}$ vs $H_A : \{|\rho| < 1\}$ do not converge to a $N(0,1)$.

✠ Test:

Test $H_0 : \{\rho = 1\}$ vs. $H_A : \{|\rho| < 1\}$ with a t statistic (critical value is -3.41 at 5% significance level) (can also use the $T(\widehat{\rho} - 1)$ statistic, the 5% critical value is -21.8) (the limit distributions of the t and of the $T(\widehat{\rho} - 1)$ statistics are computed under the assumption $\delta = 0$).

Joint test, $H_0 : \left\{\rho = 1, \delta = 0\right\}$ vs $H_A : \left\{\rho \neq 1 \text{ \&/or } \delta \neq 0\right\}$ (the F test statistic associated to this test does not converge to 1/2 χ^2_2: the 5% critical value is 6.25, as opposed to 2.99) .

Summarising

Case 4 seems to be the natural model when the data may have a linear trend.

Augmented Dickey Fuller test (ADF)

Allow for a more general dynamic structure:

$$Y_t = Y_{t-1} + u_t, \text{ when } t > 0$$

$$Y_t = 0 \text{ when } t \leq 0$$

what if u_t is (stationary) $AR(p-1)$ ($E(u_t) = 0$), instead of an independent process?

Let

$$u_t = \sum_{j=1}^{p-1} \zeta_j u_{t-j} + \varepsilon_t, \text{ where } \varepsilon_t \text{ is i.i.d.}(0, \sigma^2)$$

notice that u_t is observable, because

$$u_t = \Delta Y_t$$

so

$$Y_t = Y_{t-1} + u_t = Y_{t-1} + \sum_{j=1}^{p-1} \zeta_j u_{t-j} + \varepsilon_t$$

$$= Y_{t-1} + \sum_{j=1}^{p-1} \zeta_j \Delta Y_{t-j} + \varepsilon_t$$

Case 1

Estimate (via OLS) ρ, ζ_1, .., ζ_{p-1}, in the model

$$Y_t = \rho Y_{t-1} + \sum_{j=1}^{p-1} \zeta_j \Delta Y_{t-j} + \varepsilon_t$$

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

When $\rho = 1$:

★ the t statistic to test $H_0 : \{\rho = 1\}$ vs $H_A : \{|\rho| < 1\}$ behaves asymptotically as in Case 1 of the basic D-F test (i.e. the limit properties of $\widehat{\rho}$ are not affected by the knowledge, or lack of, of ζ_1, .., ζ_{p-1})

★ the limit properties of $\widehat{\zeta}_1$, .., $\widehat{\zeta}_{p-1}$ are not affected by the knowledge, or lack of, of ρ, so the limit properties of $\widehat{\zeta}_1$, .., $\widehat{\zeta}_{p-1}$ are the same ones as those of the OLS estimates in the (stationary) AR($p-1$) model

$$\Delta Y_t = \sum_{j=1}^{p-1} \zeta_j \Delta Y_{t-j} + \varepsilon_t.$$

Case 2

Estimate (via OLS) α, ρ, ζ_1, .., ζ_{p-1}, in the model

$$Y_t = \alpha + \rho Y_{t-1} + \sum_{j=1}^{p-1} \zeta_j \Delta Y_{t-j} + \varepsilon_t$$

(ε_t $i.i.d.(0, \sigma^2)$).

When $\alpha = 0$, $\rho = 1$:

★ the t statistic to test $H_0 : \{\rho = 1\}$ vs $H_A : \{|\rho| < 1\}$ and the F statistic to jointly test $H_0 : \left\{\alpha = 0, \rho = 1\right\}$ vs $H_A : \left\{\alpha \neq 0 \text{ \&/or } \rho \neq 1\right\}$ behave asymptotically as in Case 2 of the basic D-F test (i.e. the limit properties of $\widehat{\alpha}$ and $\widehat{\rho}$ are not affected by the knowledge, or lack of, of ζ_1, .., ζ_{p-1})

★ the limit properties of $\widehat{\zeta}_1$, .., $\widehat{\zeta}_{p-1}$ are not affected by the knowledge, or lack of, of α or of ρ, so the limit properties of $\widehat{\zeta}_1$, .., $\widehat{\zeta}_{p-1}$ are the same ones as those of the OLS estimates in the (stationary) AR($p - 1$) model

$$\Delta Y_t = \sum_{j=1}^{p-1} \zeta_j \Delta Y_{t-j} + \varepsilon_t.$$

Case 3

Estimate (via OLS) α, ρ, $\zeta_1, .., \zeta_{p-1}$, in the model

$$Y_t = \alpha + \rho Y_{t-1} + \sum_{j=1}^{p-1} \zeta_j \Delta Y_{t-j} + \varepsilon_t$$

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

When $\alpha \neq 0$, $\rho = 1$:

★ the t statistic to test $H_0 : \{\rho = 1\}$ vs $H_A : \{|\rho| < 1\}$ behaves asymptotically as in Case 3 of the basic D-F test (i.e. the limit properties of $\widehat{\alpha}$ and $\widehat{\rho}$ are not affected by the knowledge, or lack of, of $\zeta_1, .., \zeta_{p-1}$)

★ the limit properties of $\widehat{\zeta}_1, .., \widehat{\zeta}_{p-1}$ are not affected by the knowledge, or lack of, of α and of ρ, so the limit properties of $\widehat{\zeta}_1, .., \widehat{\zeta}_{p-1}$ are the same ones as those of the OLS estimates in the (stationary) AR($p-1$) model

$$\Delta Y_t = \alpha + \sum_{j=1}^{p-1} \zeta_j \Delta Y_{t-j} + \varepsilon_t.$$

Case 4

Estimate (via OLS) α, ρ, $\zeta_1, .., \zeta_{p-1}$, in the model

$$Y_t = \alpha + \rho Y_{t-1} + \delta t + \sum_{j=1}^{p-1} \zeta_j \Delta Y_{t-j} + \varepsilon_t$$

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

When $\delta = 0$, $\rho = 1$:

★ the t statistic to test $H_0 : \{\rho = 1\}$ vs $H_A : \{|\rho| < 1\}$ and the F statistic to jointly test $H_0 : \left\{\rho = 1, \delta = 0\right\}$ vs $H_A : \left\{\rho \neq 1 \text{ \&/or } \delta \neq 0\right\}$ behave asymptotically as in Case 4 of the basic D-F test (the limit properties of $\widehat{\alpha}$, of $\widehat{\rho}$ and of $\widehat{\delta}$ are not affected by the knowledge, or lack of, of $\zeta_1, .., \zeta_{p-1}$).

★ the limit properties of $\widehat{\zeta}_1, .., \widehat{\zeta}_{p-1}$ are not affected by the knowledge, or lack of, of α, of ρ and of δ, so the limit properties of $\widehat{\zeta}_1, .., \widehat{\zeta}_{p-1}$ are the same ones as those of the OLS estimates in the (stationary) AR($p - 1$) model

$$\Delta Y_t = \alpha + \sum_{j=1}^{p-1} \zeta_j \Delta Y_{t-j} + \varepsilon_t.$$

Summarising:

★once that the lags $\Delta Y_{t-1},...,\Delta Y_{t-p+1}$ have been added to the model, we can just test if $\rho = 1$ using the t or the F statistic, and refer to the "basic" (ie, with no lags) case for the limit distributions.

This is a very useful result, because it means that we do not have to adjust the limit distributions to the structure of u_t: the adjustment is made automatically by the t or by the F statistic.

★The result that the limit properties of $\widehat{\zeta}_1, .., \widehat{\zeta}_{p-1}$ are the same ones as those of the estimates in the (stationary) AR($p-1$) and therefore do not depend on ρ is very useful as well, because we can use it to determine the order $p-1$ of the AR($p-1$) structure when indeed $p-1$ is unkown.

★If we don't know $p-1$, we can select the order of the AR model for u_t using an information criterion; otherwise, we may select a tentative order, say, *pmax* (obviously, *pmax*> *p*), and test if $\widehat{\zeta}_p, \ldots, \widehat{\zeta}_{pmax-1}$ are not statistically significant.

The hypotesis of an AR($p-1$) model for u_t is rather general, because it corresponds to an AR(p) model for Y_t (at least, when no linear trends are present). We can see it by looking, for example, at the Case 1 representation

$$Y_t = \rho Y_{t-1} + \sum_{j=1}^{p-1} \zeta_j \Delta Y_{t-j} + \varepsilon_t$$

$$Y_t - \rho Y_{t-1} - \sum_{j=1}^{p-1} \zeta_j \Delta Y_{t-j} = \varepsilon_t$$

Using the lag operator, replacing
Y_{t-1} by LY_t, Δ by $(1-L)$ and Y_{t-j} by $L^j Y_t$,

$$Y_t - \rho Y_{t-1} - \sum_{j=1}^{p-1} \zeta_j \Delta Y_{t-j}$$

$$= \left(1 - \rho L - \sum_{j=1}^{p-1} \zeta_j (1-L) L^j \right) Y_t$$

and

$$
\begin{aligned}
& 1-\rho L-\sum_{j=1}^{p-1} \zeta_j (1-L) L^j \\
&= 1-\rho L-(1-L)\sum_{j=1}^{p-1} \zeta_j L^j \\
&= 1-\rho L-(1-L)\zeta_1 L-(1-L)\zeta_2 L^2-\ldots \\
&\quad -(1-L)\zeta_{p-1} L^{p-1} \\
&= 1-\rho L-\zeta_1 L+\zeta_1 L^2-\zeta_2 L^2+\zeta_2 L^3-\ldots \\
&\quad -\zeta_{p-1} L^{p-1}+\zeta_{p-1} L^p \\
&= 1+(-\rho-\zeta_1)L+(\zeta_1-\zeta_2)L^2+\ldots \\
&\quad +(\zeta_{p-2}-\zeta_{p-1})L^{p-1}+\zeta_{p-1} L^p \\
&= 1-(\rho+\zeta_1)L-(\zeta_2-\zeta_1)L^2-\ldots \\
&\quad -(\zeta_{p-1}-\zeta_{p-2})L^{p-1}-(-\zeta_{p-1})L^p
\end{aligned}
$$

so

$$
\begin{aligned}
\phi_1 &= \rho+\zeta_1 \\
\phi_2 &= \zeta_2-\zeta_1 \\
&\ldots \\
\phi_{p-1} &= \zeta_{p-1}-\zeta_{p-2} \\
\phi_p &= -\zeta_{p-1}
\end{aligned}
$$

We can also notice that the ϕ_j are such that

$$
\begin{aligned}
& \phi_1 + \phi_2 + \ldots + \phi_{p-1} + \phi_p \\
= \ & \rho + \zeta_1 + \zeta_2 - \zeta_1 + \ldots + \zeta_{p-1} - \zeta_{p-2} - \zeta_{p-1} \\
= \ & \rho
\end{aligned}
$$

so

when $\rho = 1$,

$\phi_1 + \phi_2 + \ldots + \phi_{p-1} + \phi_p = 1.$

An alternative regression for DF/ADF

Consider again, for example, the regression model for Case 2:

$$Y_t = \alpha + \rho Y_{t-1} + \sum_{j=1}^{p-1} \zeta_j \Delta Y_{t-j} + \varepsilon_t$$

($\varepsilon_t \; i.i.d. (0, \sigma^2)$). Subctracting Y_{t-1} by both sides, we get

$$\Delta Y_t = \alpha + (\rho - 1) Y_{t-1} + \sum_{j=1}^{p-1} \zeta_j \Delta Y_{t-j} + \varepsilon_t$$

This model is equivalent to the previous one, but instead of testing $H_0\{\rho = 1\}$ we then test $H_0\{\rho - 1 = 0\}$.

The test is equivalent to the previous one (so, it also uses the same limit distribution).

Of course, it is also possible to adapt the other cases (Case 1 to Case 4) to test $H_0\{\rho - 1 = 0\}$ instead.

Phillips and Perron test (PP)

Allow for a more general dynamic structure:

$$Y_t = Y_{t-1} + u_t, \text{ when } t > 0$$

$$Y_t = 0 \text{ when } t \leq 0$$

what if u_t is (stationary and invertible) $ARMA(p,q)$ (with $E(u_t) = 0$), instead of an independent process?

Case 1

Let $\widehat{\rho} = \frac{\sum_{t=2}^{T} Y_t Y_{t-1}}{\sum_{t=2}^{T} Y_{t-1}^2}$,

$$T(\widehat{\rho} - 1) \rightarrow_d \frac{\frac{1}{2}\left(W(1)^2 - 1\right)}{\int_0^1 W(r)^2 dr} + \nu$$

where ν is a shift term.

This can be consistently estimated: call that estimate $\widehat{\nu}$, we can test for a unit root using

$$T(\widehat{\rho} - 1) - \widehat{\nu} \rightarrow_d \frac{\frac{1}{2}\left(W(1)^2 - 1\right)}{\int_0^1 W(r)^2 dr}$$

★Case 2, Case 3 and Case 4 work in the same way (the shift term v may be different).

★The same considerations for the choice Case 1 vs Case 2, and Case 3 vs Case 4 apply.

★$\widehat{\rho}$ is still "superconsistent" (compare with $|\rho| < 1$: $\widehat{\rho}$ would in general inconsistent, in this case)

★ the PP test works in a more general set up than the ADF

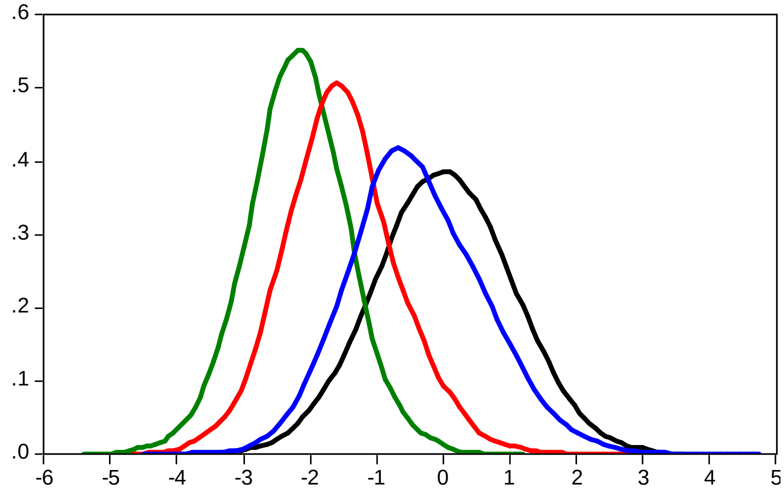
★ the ADF has more power than the PP if p is known; otherwise, the performance of the two tests are not much different.

Appendix

- The distributions of the Dickey and Fuller *t* statistics

- Which Case in the unit root test?

The distributions of the Dickey and Fuller *t* statistics



Note: Generated using 5000 repetitions and $T = 1000$.

Note: Black, $N(0,1)$; Blue, Case 1; Red, Case 2, Green Case 4.

Which Case in the unit root test?

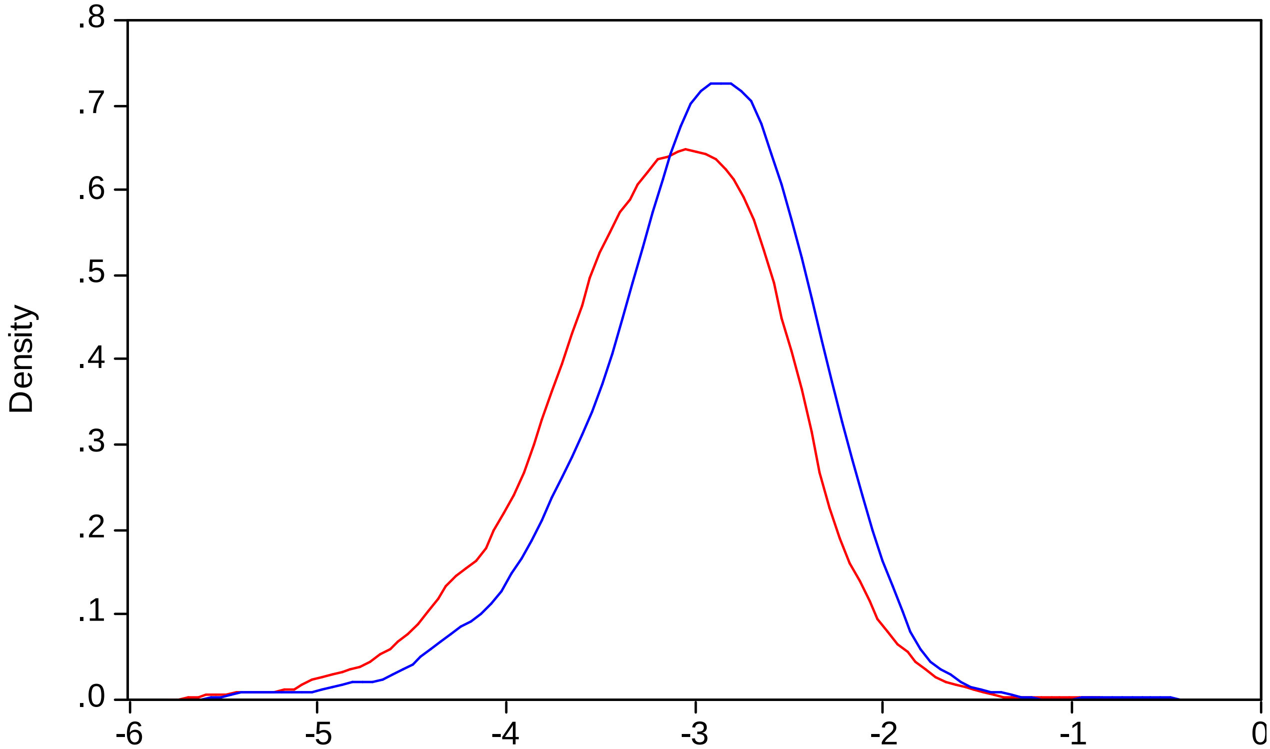
Case 1 and Case 2 both have the same null hypothesis,

$$Y_t = Y_{t-1} + \varepsilon_t, \text{ i.e., } \rho = 1.$$

If indeed $\rho = 1$, then both tests will NOT Reject the null hypothesis with probability 95% (as we set the size to 5%). So, we can only choose between the two tests if we look at what happens when in fact the null hypothesis is not correct and $|\rho| < 1$.

Two alternatives are possible: "$c = 0$", i.e, $Y_t = \rho Y_{t-1} + \varepsilon_t$, and "$c \neq 0$", i.e. $Y_t = c + \rho Y_{t-1} + \varepsilon_t$.

✠ $c = 0$. In this example, we used $T = 100$ and $\rho = 0.85$:

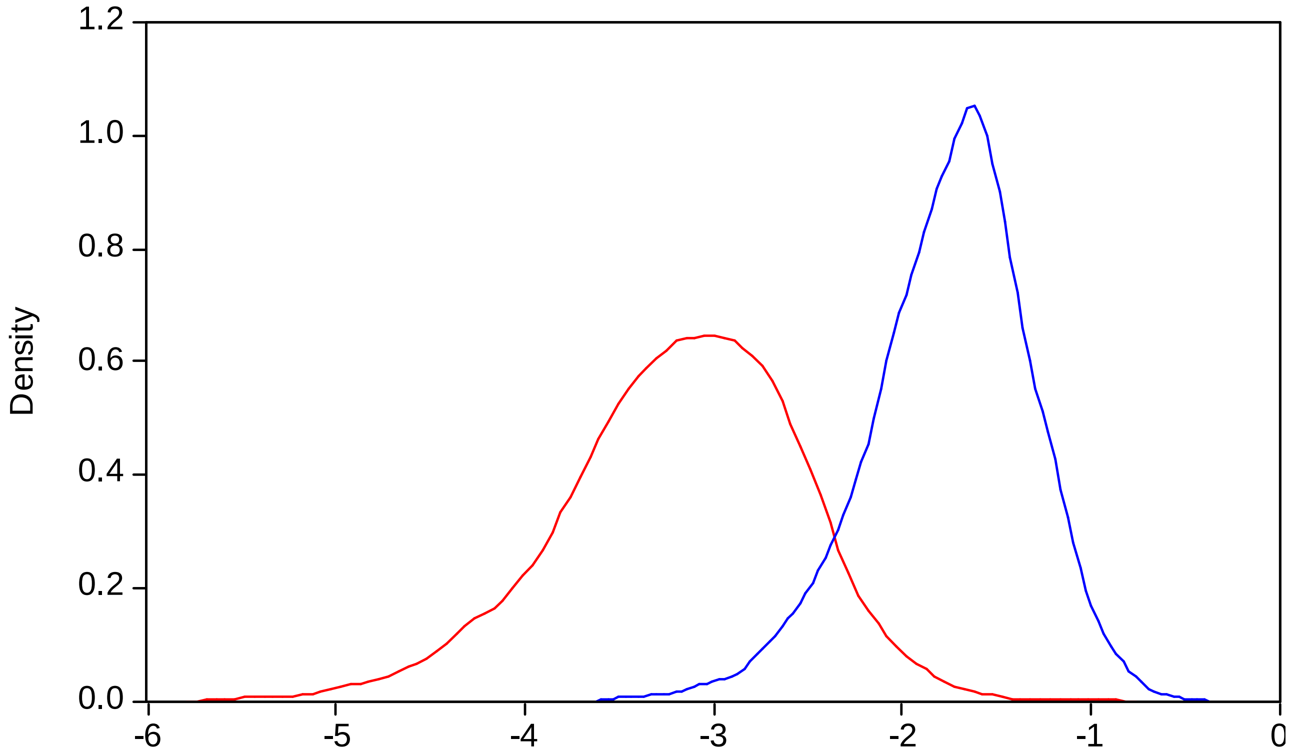


The BLUE distribution is the distribution of the standardized t statistic if Case 1 is estimated, and the RED if Case 2 is estimated (note that the theoretical limit distribution of $\widehat{\rho}$ is the same, the apparent difference in the distribution of t is only due to the sample variability).

The critical value for Case 1 is -1.95, and in our example, 97.9% was below it (i.e., in 97.9% of the samples we correctly concluded that $|\rho| < 1$);

The critical value for Case 2 is -2.86, and in our example, 67.7% was below it (i.e., in 67.7% of the samples we correctly concluded that $|\rho| < 1$).

✠ $c \neq 0$. The distribution of the estimate of ρ and of the standardized t under Case 2 are unaffected. Under case 1, however, ρ is no longer consistently estimated. Here we kept $T = 100$ and $\rho = 0.85$ but set $c = 2.5$:



The BLUE distribution is for the standardized t statistic if Case 1 is estimated, and the RED if Case 2 is estimated (note that the theoretical limit distributions of t are no longer same; the RED distribution is the same as in the case with $c = 0$).

Case 1: in our example in 29.3% of the samples we correctly concluded $|\rho| < 1$;

Case 2: in our example in 67.7% of the samples we correctly concluded $|\rho| < 1$.