Introduction to Time Series Analysis:

Laura Mayoral[1]

[1]Institute for Economic Analysis and Barcelona School of Economics

DG-ECFIN, Nov 2021

Handout 4

Appendix: Models for univariate non-stationary processes

Asymptotic theory with units roots

- Stationary AR(1):

$$X_t = \phi X_{t-1} + \varepsilon_t, \qquad (1)$$

where ε_t is a martingale difference sequence with $E\left(\varepsilon_t^2\right) < \infty$.

- The OLS estimator of ϕ is given by:

$$\hat{\phi} = \frac{\sum X_t X_{t-1}}{\sum X_{t-1}^2} = \phi + \frac{\sum X_{t-1}\varepsilon_t}{\sum X_{t-1}^2}$$

- By the CLT for dependent processes, It is simple to obtain that

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

- But:

- In order to obtain this distribution the assumption of $|\phi| < 1$ is crucial. In fact, if $\phi \to 1$, $T^{1/2}\left(\hat{\phi} - \phi\right) \xrightarrow{p} 0$! (Why??)

- To find out the asymptotic distribution of $\hat{\phi}$ one needs to use a new asymptotic theory!

- This theory is non-standard because it is not based on standard results (LLN and CLT) and asymptotic distributions are in general non-standard as well (i.e., they are not normal, t, Chi square or F).

Basic elements

- The asymptotic theory of unit root processes can be complicated
- Loosely speaking, its basic elements are
  - Asymptotic distributions are a (functional) of Brownian motions (rather than normal distributions).
  - The Functional central limit theorem (is used in place of the CLT.)
  - the Continuous mapping theorem

Non-standard asymptotic theory: Preliminary concepts
Brownian motion

- Consider a random walk process

$$X_t = X_{t-1} + \varepsilon_t, \tag{2}$$

where $\varepsilon_t \sim iid\ N(0,1)$ and $X_0 = 0$.

- Under the former assumptions, notice that 1) $X_t \sim N(0,t)$; 2) $X_s - X_t \sim N(0,(s-t))$ and 3) $X_s - X_t$ is independent of the change $(X_q - X_r)$ for any dates $t < s < r < q$.

- To see this, notice that

$$X_t = \sum_{i=1}^{t} \varepsilon_i.$$

- It follows that $X_t \sim N(0, t)$. Likewise, the change in the value of X between dates t and s, $t > s$

$$X_s - X_t = \varepsilon_{t+1} + \varepsilon_{t+2} + ... + \varepsilon_s \sim N(0, (s - t)).$$

- Furthermore, it is easy to check that $X_s - X_t$ is independent of the change $(X_q - X_r)$ for any dates $t < s < r < q$.

Let's now consider the change between two consecutive values $X_t - X_{t-1} = \varepsilon_t$ and assume that the change ε_t can be written as the sum of N individual *iid* processes with variance $\frac{1}{N}$, each happening at intervals of length $1/N$ between $t-1$ and t :

$$X_t - X_{t-1} = \varepsilon_t = e_{1t} + e_{2t} + ... + e_{Nt}$$

where $e_{it} \sim iid\ N(0, 1/N)$. Then, the process X_t is not only defined at integer values of t, but also at non-integer values $X_{t-i/N}$, that is

$$X_t - X_{t-i/N} = \sum_{j=i+1}^{N} e_{jt},\ i = 1, ...N.$$

The limit as $N \to \infty$ of X is a continuous-time process known as **Brownian motion** and the value of this process at t is denoted as $W(t)$.

Standard Brownian motion: definition

Let $W(.)$ be a continuous-time stochastic processes, associating each date $t \in [0,1]$ with the scalar $W(t)$ such that
a) $\mathrm{W}(0) = 0$,
b) For any dates $0 \leq t_1 \leq t_2 \leq ... \leq t_k \leq 1$, the changes $W(t_2) - W(t_1)$, $W(t_3) - W(t_2)$, ..., $W(t_k) - W(t_{k-1})$, are independent Gaussian variables with
$W(s) - W(t) \sim N(0, (s-t))$. In particular,
$W(t) = W(t)$-$W(0) \sim N(0, t)$
c) For any given realization, $W(t)$ is continuous in t with probability 1.
W is a standard Brownian Motion.

Remark: A realization of a random variable X, x, is a scalar. A realization of the random function $W(.)$, $W(t)$, is a random variable!

The functional Central limit theorem (FCLT).

- While the CLT establishes the convergence for random variables, the FCLT establishes conditions for convergence of **random functions**.

- Let ε_t be an iid(0,σ^2) sequence. Then, by the CLT we know that $\sqrt{T}\bar{\varepsilon}_T \xrightarrow{d} N\left(0, \sigma^2\right)$, where $\bar{\varepsilon}_T = T^{-1}\sum_{t=1}^{T} \varepsilon_t$ is the sample mean of ε_t.

Consider now an estimator of the sample mean that only considers the *rth* fraction of the observations, $r \in [0,1]$, that is

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

where $[Tr]$ denotes the integer part of Tr. Then, for any given realization, $X_T(r)$ is a step function in r:

$$X_T(r) = \begin{cases} 0 & 0 \le r < 1/T \\ \varepsilon_1/T & 1/T \le r < 2/T \\ (\varepsilon_1 + \varepsilon_2)/T & 2/T \le r < 3/T \\ \dots & \\ (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_T)/T & r = 1 \end{cases}$$

Now, notice that

$$\begin{aligned} \sqrt{T} X_T(r) &= T^{-1/2} \sum_{t=1}^{[Tr]} (\varepsilon_t) \\ &= \sqrt{[Tr]}\, T^{-1/2} \left(\sqrt{[Tr]} \right)^{-1} \sum_{t=1}^{[Tr]} (\varepsilon_t) \end{aligned}$$

and $\left(\sqrt{[Tr]} \right)^{-1/2} \sum_{t=1}^{[Tr]} (\varepsilon_t) \xrightarrow{d} N(0, \sigma^2)$ by the CLT and $\sqrt{[Tr]}/\sqrt{T} \to \sqrt{r}$. Then

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

- On the other hand, it is trivial to check that

$$\sqrt{T} \left(X_T(r_2) - X_T(r_1) \right) / \sigma \xrightarrow{d} N(0, r_2 - r_1),$$

and $X_T(r_2) - X_T(r_1)$ is independent of $(X_T(r_4) - X_T(r_3))$ if $r_1 < r_2 < r_3 < r_4$.

The functional central limit theorem

- The Functional limit theorem establishes that

$$\sqrt{T} X_T(.)/\sigma \xrightarrow{d} W(.) \tag{3}$$

Continuous mapping theorem (CMT)

- We saw that if $\{X_t\}$ is a collection of random variables, $X_t \overset{d}{\to} X$ and $g : \mathbb{R} \to \mathbb{R}$ is a continuous function, then $\mathrm{g}(X_t) \overset{d}{\to} g(X)$.

- A similar result holds for sequences of random functions.
- The CMT states that if $S_T(.) \overset{d}{\to} S(.)$ and g is a continuous functional, then $g(S_T(.)) \overset{d}{\to} g(S(.))$.

For instance, from (**??**) and the CMT it follows that

$$\sqrt{T} X_T(.) \overset{d}{\to} \sigma W(.).$$

Asymptotic theory and unit roots

- Consider now the random walk process $y_t = y_{t-1} + \varepsilon_t$ where $\{\varepsilon_t\}$ is an iid$\left(0, \sigma^2\right)$ sequence. Assuming that $y_0 = 0$, then $y_t = \sum_{t=1}^{T} \varepsilon_t$. Then, one can construct the stochastic function $X_T\left(r\right)$ as follows:

$$X_T\left(r\right) = \begin{cases} 0 & 0 \leq r < 1/T \\ y_1/T = \varepsilon_1/T & 1/T \leq r < 2/T \\ y_2/T = \left(\varepsilon_1 + \varepsilon_2\right)/T & 2/T \leq r < 3/T \\ ... & \\ y_T/T & r = 1 \end{cases}$$

$\mathrm{X}_t\left(r\right)$ is a step function whose values are given by y_i/T.

Now, consider the integral of $X_t(r)$ in [0,1]. It is clear that is is equal to the sum of the areas of each of the rectangles defined by y_i/T. The first rectangle has width 1/T and height equal to y_1/T, then its area is $\frac{y_1}{T^2}$. Doing the same for the remaining rectangles it follows that,

$$\int_0^1 X_T(r)\,dr = \frac{y_1}{T^2} + \frac{y_2}{T^2} + ... + \frac{y_T}{T^2} = T^{-2}\sum_{t=1}^{T} y_t.$$

Since $\sqrt{T}X_T(.) \xrightarrow{d} \sigma W(.)$, by the CMP $\int_0^1 \sqrt{T}X_T(r)\,dr \xrightarrow{d} \sigma \int_0^1 W(r)\,dr$, and therefore

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

A similar argument can be used to derive the asymptotic distribution of $\sum_{t=1}^{T} y_t^2$. Define the random function $S_T(.) = \left(\sqrt{T} X_T(r)\right)^2$

$$S_T(r) = \begin{cases} 0 & 0 \leq r < 1/T \\ y_1^2/T & 1/T \leq r < 2/T \\ y_2^2/T & 2/T \leq r < 3/T \\ \dots & \\ y_T^2/T & r = 1 \end{cases}.$$

It follows that $\int_0^1 S_T(r)\,dr = T^{-2} \sum_{t=1}^{T} y_t^2$. Therefore, since $\int_0^1 S_T(r)\,dr = \int_0^1 \left(\sqrt{T} X_T(r)\right)^2$, by the CMP

$$T^{-2} \sum_{t=1}^{T} y_t^2 \xrightarrow{d} \sigma^2 \int_0^1 W^2(r)\,dr.$$

- Consider now the asymptotic properties of the OLS estimator $\hat{\phi}$ of ϕ in (**??**) where the true process is a random walk.

Then,

$$T\left(\hat{\phi}-\phi\right)=\frac{T^{-1}\sum X_{t-1}\varepsilon_t}{T^{-2}\sum X_{t-1}^2}$$

The limit of the denominator is $\sigma^2 \int_0^1 W^2(r)\,dr$. As for the numerator, notice that $X_t^2 = X_{t-1}^2 + \varepsilon_t^2 + 2X_{t-1}\varepsilon_t$, and therefore

$$
\begin{aligned}
T^{-1}\sum X_{t-1}\varepsilon_t &= \frac{1}{2}\left(T^{-1}\sum(X_t^2 - X_{t-1}^2) - T^{-1}\sum \varepsilon_t^2\right) \\
&= \frac{1}{2}\left(T^{-1}X_T^2 - T^{-1}\sum \varepsilon_t^2\right) \\
&= \frac{1}{2}\left(T^{-1}X_T^2 - T^{-1}\sum_{t=1}^{T} \varepsilon_t^2\right) = \\
& \quad \frac{1}{2}\left(S_T(r) - T^{-1}\sum_{t=1}^{T} \varepsilon_t^2\right) \\
& \xrightarrow{d} \frac{1}{2}\sigma^2\left(W^2(1) - 1\right).
\end{aligned}
$$

Then

$$T\left(\hat{\phi} - \phi\right) \xrightarrow{d} \frac{\left(W^2(1) - 1\right)}{2\int_0^1 W^2(r)\,dr}.$$

Summarizing

- One cannot use standard asymptotic theory to obtain the distribution of $\hat{\phi}$ when the underlying process is a random walk,

- Using the FCLT and the CMP one can show that the OLS estimator converges to a non-standard distribution (functional of Brownian motions

- The rate of convergence of $\hat{\phi}$ to its limit is much faster than in the stationary case: T versus $\sqrt{T}$.

- For this reason, $\hat{\phi}$ is said to be a super-consistent estimator of ϕ.

Unit root tests

- Unit root tests are tests designed to determine whether a process contains a unit root (or more).

- One of the hypothesis, thus, is that the process contains (one or several) unit roots.

- The other hypothesis is a different model that is also plausible for the data at hand.

- Many possibilities for the alternative hypothesis, dependent on the characteristics of the process: stationarity, trend-stationarity, breaking trends, long memory....

- For instance, if the data looks trended the usual alternative hypothesis to the unit root+drift one is that the trend is given by a deterministic function (typically, a linear trend).

Unit root tests, II

- Very very large literature!! See summary: Xiao and Phillips (1999) .
- Pioneer work: the Dickey-Fuller test.
- Very simple idea: it is based on the $t-test$ associated to the coefficient of X_{t-1} in a regression of X_t on X_{t-1} and, possibly (although not always) lags of ΔX_t and some deterministic components.

- Difficulties:

- As we know, the asymptotic distribution of the relevant statistics is not standard (new tables of critical values are needed)

- Whether the “true” model and/or the regression model contain deterministic components has an impact on the asymptotic distribution!

- Thus, different tables of critical values should be used depending on these deterministic components.

The Dickey- Fuller test

- D-F test Goal: test for a unit root in X_t.

- Approach: it considers an autoregressive model (that nests the unit root model) and tests whether $\phi = 1$

$$X_t = \phi X_{t-1} + u_t \tag{4}$$

- Use a t-test to test for the significance of $\hat{\phi}$.

- But: distributions are not standard and one has to be careful with deterministic components!

Simplest case: DF test with uncorrelated disturbances

- Consider the simplest case: $u_t = \varepsilon_t$ is *iid*.

- We need to distinguish 4 cases, depending on whether the true model (TM) and the regression model (RM) contain determinist components

- Case 1. The true model (TM) is a random walk without drift $(\alpha = 0)$ and the regression model (RM) is an AR(1) process without constant

$$\begin{aligned} TM &: X_t = X_{t-1} + \varepsilon_t,\ X_0 = 0 \\ RM &: X_t = \phi X_{t-1} + \varepsilon_t \end{aligned} \tag{5}$$

where ε_t is an iid$\left(0, \sigma^2\right)$ sequence and X_0 are some initial conditions.

DF test with uncorrelated disturbances, III

- The hypotheses to be tested are

$$\begin{aligned} H_0 &: \phi = 1, \\ H_1 &: \phi < 1. \end{aligned}$$

or alternatively, subtracting X_{t-1} in both sides of (**??**), the regression model results

$$RM : \Delta X_t = \varphi X_{t-1} + \varepsilon_t, \tag{6}$$

where $\varphi = \phi - 1$, which gives the null of unit root versus the alternative hypothesis of stationarity

$$\begin{aligned} H_0 &: \varphi = 0, \\ H_1 &: \varphi < 0. \end{aligned}$$

DF test with uncorrelated disturbances, IV

- A $t-test$ then can be used for testing $\phi = 1$ in (**??**) or $\varphi = 0$ in **??** . The $t-tests$ are given by

$$t_T = \frac{\hat{\phi}-1}{\hat{\sigma}_{\hat{\phi}}} \text{ or } t_T = \frac{\hat{\varphi}}{\hat{\sigma}_{\hat{\varphi}}}$$

- Decision rule: reject H_0 if absolute value of t is larger than the critical value.

- But, what critical value???

- If X_t contains a unit root, t_T does not converge to a Normal distribution. It converges to the so-called 'Dickey-Fuller' distribution.

We now show how this distribution can be obtained. The t-tests are given by

$$t_T = \frac{T^{-1}\sum_{t=2}^{T} X_{t-1}\varepsilon_t}{\left(T^{-2}\sum_{t=2}^{T} X_{t-1}^2\right)^{1/2} s_T},$$

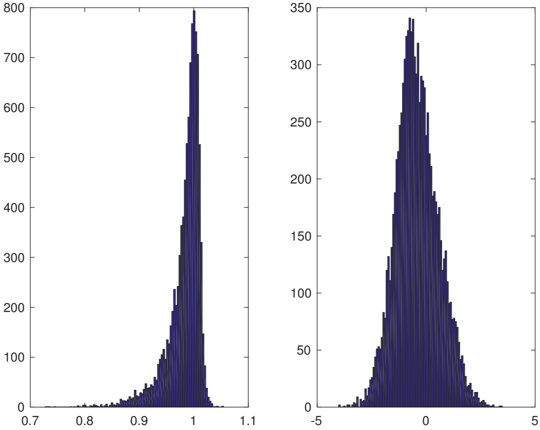
where $s_T^2 = (T-1)^{-1}\sum\left(X_t - \hat{\phi}X_{t-1}\right)^2$. Using the results in the previous section and the fact that $s_T^2 \xrightarrow{p} \sigma_\varepsilon^2$, it is easy to check that

$$t_T \xrightarrow{d} \frac{1/2\left(W^2(1) - 1\right)}{\left(\int_0^1 W^2(r)\,dr\right)^{1/2}}.$$

This distribution is non-standard and therefore, it has to be tabulated. Tables of critical values can be found in the Appendix of most time series books. For instance

$$P\left(t_T < -1{,}95\right) = 0{,}05.$$

Histograms of $\hat{\phi}$ and the t-statistic



Case 2.

- The true model (TM) is a random walk without drift and the regression model (RM) is also an AR(1) with a constant

$$\begin{aligned} TM &: X_t = X_{t-1} + \varepsilon_t, \; X_0 = x_0 \\ RM &: X_t = \alpha + \phi X_{t-1} + \varepsilon_t \end{aligned}$$

- The null and the alternative hypotheses can be postulated in the same way as above.

- However, the introduction of a constant in the model changes the asymptotic distribution of the $t-statistic$ associated to $\hat{\phi}$.
- In addition, introducing a constant in the RM makes the test invariant to the (unknown) value of the initial condition, x_0 (that can be equal to zero, as in Case 1, or not).

- It can be checked that now

$$t_T \xrightarrow{d} \frac{1/2\left(W^2(1)-1\right)-W(1)\int_0^1 W(r)\,dr}{\left(\int_0^1 W^2(r)\,dr-\left(\int_0^1 W(r)\,dr\right)^2\right)^{1/2}}.$$

- Then, one should use a different table of critical values as in Case 1.

- It would also be possible to use an F-test for the joint null hypothesis of $\alpha_0 = 0$ and $\phi = 1$.

The F-test is given by

$$F = \frac{(RSS_R - RSS_U)/r}{RSS_U/(T-k)} \tag{7}$$

where r is the number of restrictions to be tested (=2 in this case), RSS_R is the residual sum of squares of the restricted model, and RSS_U is the residual sum of squares of the unrestricted model. Critical values are also tabulated (see Hamilton, case 2).

Case 3.

- The true model (TM) is a random walk with drift and the regression model (RM) is an AR(1) process with constant

$$
\begin{aligned}
TM &: X_t = \alpha_0 + X_{t-1} + \varepsilon_t, \\
RM_1 &: X_t = \alpha + \phi X_{t-1} + \varepsilon_t,^{1}
\end{aligned}
$$

- The fact that X_t contains a drift changes dramatically the asymptotic distributions. It can be shown that

$$
\begin{pmatrix} T^{1/2}(\hat{\alpha} - \alpha_0) \\ T^{1/2}(\hat{\phi} - 1) \end{pmatrix} \xrightarrow{d} N\left(\begin{matrix} 0 \\ 0 \end{matrix}, \sigma^2 Q^{-1} \right)
$$

where

$$
Q = \begin{pmatrix} 1 & \alpha_0/2 \\ \alpha_0/2 & \alpha^2/3 \end{pmatrix}.
$$

[1] Or, alternatively

$$
RM_2 : \Delta X_t = \beta + \varphi X_{t-1} + \varepsilon_t.
$$

- Hence, in this case the asymptotic distributions are Gaussian.

- In fact the asymptotic distributions of $\hat{\alpha}$ and $\hat{\phi}$ are the same as those obtained in the regression model $Y_t = \alpha + \phi t + \varepsilon_t$.
- This is because the deterministic component of $X_t, \alpha_0 t$, (recall that if α_0 is different from zero, then $X_t = \alpha_0 t + \sum \varepsilon_t$), dominates the stochastic one, $\sum \varepsilon_t$.
- Finally, notice that the asymptotic distribution depends on two 'nuisance' parameters, α and σ. Furthermore, if $\alpha = 0$, the distribution is not the one above but the one described in 'case 2'. The asymptotic distribution of the t-test is $\mathrm{N}(0, 1)$.

Case 4.

- The true model (TM) is a random walk with drift and the regression model (RM) is an AR(1) process with constant and trend

$$\begin{aligned} TM &: X_t = \alpha + X_{t-1} + \varepsilon_t, \\ RM_1 &: X_t = \beta_0 + \beta_1 t + \phi X_{t-1} + \varepsilon_t. \end{aligned}$$

- In this case, the distribution of the $t-test$ of $\phi = 1$ is non-standard (functionals of BM, as in cases 1 and 2. See Case 4, Hamilton) and does not depend on nuisance parameters $\left(\sigma^2 \text{ or } \alpha\right)$.
- F-tests for the joint null of $\beta_1 = 0$ and $\phi = 1$ can also be applied and the corresponding critical values are tabulated.

Summary: Which is the correct RM to use?

- The Dickey-Fuller (DF) test is designed for testing the null of $\phi = 1$ or $\varphi = 0$ in three different regression models:

$$
\begin{aligned}
i)\ X_t &= \phi X_{t-1} + \varepsilon_t \text{ or } \Delta X_t = \varphi X_{t-1} + \varepsilon_t, \\
ii)\ X_t &= \alpha_0 + \phi X_{t-1} + \varepsilon_t \text{ or } \Delta X_t = \beta_0 + \varphi X_{t-1} + \varepsilon_t, \\
iii)\ X_t &= \alpha_0 + \beta t + \phi X_{t-1} + \varepsilon_t \text{ or } \Delta X_t = \beta_0 + \beta_1 t + \varphi X_{t-1} + \varepsilon_t.
\end{aligned}
$$

- What RM should be used?
  - One should a regression model that is plausible under both H_0 and H_1. That is, if the data looks trended, then iii) would offer a plausible specification under both hypothesis.
  - By default, include a constant and a trend in your RM. This means that you know that you are in Case 4, independently on the (unknown!!!) values of the deterministic components of the TM.