Introduction to Time Series Analysis:

Laura Mayoral[1]

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

DG-ECFIN, Nov 2021

Handout 4

Models for univariate non-stationary processes

Introduction

- Most economic and business time series are nonstationary and, therefore, the type of models that we have studied cannot (directly) be used.
- Nonstationary can occur in many ways: non constant means, non-constant variances, seasonal patterns, etc.
- For nonstationary process a *Wold representation*-type theorem does not exist
- Modelling is much more complicated, as we have an infinite number of models to choose from!

Anna Karenina effect

"Happy families are all alike; every unhappy family is unhappy in its own way". (Anna Karenina, Tolstoy).

- Models for stationary processes are all alike, and the same is NOT true for nonstationary ones!

- This handout introduces several approaches for modelling non-stationary time series.

- We will focus primarily on processes whose mean (and variance) change over time.

Models for processes with non constant means

- Most economic series do not have a constant mean and therefore are not stationary.

- The approach we'll follow:

  - STEP 1: Look for models that are 1) simple and 2) capable of capturing the main features of the data (trends, high persistence, non-constant means, etc)

  - STEP 2: Apply a transformation that removes the non-stationarity component so that the resulting process is stationary.

  - STEP 3: Apply the techniques for stationary processes to the transformed process.

Main question of this handout:

- The main question we will try to answer in this handout is how to choose a (non-stationary) model for our data?

- To do that we will

- Consider (a small set) of candidate models that are able to reproduce some of the characteristics observed in the data

- Use statistical tests to decide on the "right" model

- More specifically: unit root tests.

Models for processes non-constant means: deterministic vs stochastic trends

- Two popular candidates:

- Trend-stationary models: sum of a deterministic function of time (for instance a linear trend) and a stationary component Typically, it will be a polynomial in t:

$$\tau(t) = \beta_0 + \beta_1 t + ... + \beta_s t^s,$$

- Integrated processes: The trend is itself a random variable. (Models with unit roots)

- But there are other models:

- For instance, models with structural breaks
- See Hansen (2001) for a survey on those models

Trend stationary model:

Trend stationary model:

- It is the sum of a deterministic trend and a stationary process.

$$X_t = \tau(t) + \psi(L)\varepsilon_t.$$

where $\psi(L)\varepsilon_t$ is a stationary process.

- In most applications $\tau(t) = \beta_0 + \beta_1 t$, is simply a polynomial in t of degree 1.

- This process is often called **trend-stationary** because if one subtracts the non-random trend from X_t the result is a stationary process.

Unit root processes

$$X_t = X_{t-1} + \beta + \psi(L)\varepsilon_t \tag{1}$$

where $\psi(1) \neq 0$.

- X_t can also be written as $(1-L)X_t = \beta + \psi(L)\varepsilon_t$.

- It is said to be a unit root process because $L = 1$ is a root of the autoregressive polynomial.
[Some notation: $(1-L) = \Delta$.]

- The transformed process $(1-L)X_t = \Delta X_t = X_t - X_{t-1}$ is stationary and describes the changes (or the growth rate if X_t is in logs) of the series X_t.

Unit root processes: examples

- Simplest case: random walk and random walk with drift.

- **Random walk**. If $\psi(L) = 1$ and $\beta = 0$ in (1), then $\{X_t\}$ is a random walk sequence,

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

- Assuming that $X_t = 0$ for all $t < 0$ and that X_0 is a fixed finite initial condition then, by backward substitution,

$$X_t = X_0 + \sum_{i=1}^{t} \varepsilon_i.$$

and $E(X_t) = X_0$ for $t \geq 0$ and $var(X_t) = t\sigma^2$.

- **Random walk with drift.** To introduce an upward or downward trend component it is only needed to include a constant in (2) . The *random walk with drift* model is defined as

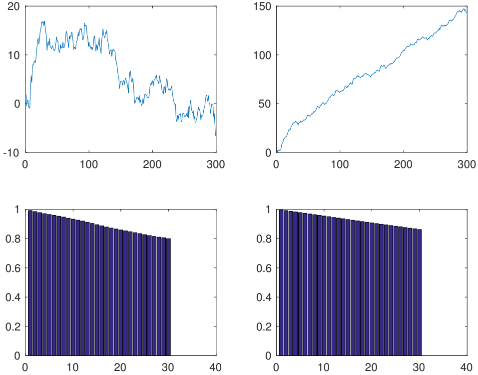
$$X_t = \beta + X_{t-1} + \varepsilon_t \tag{3}$$

and by back-substitution

$$X_t = \beta + (\beta + X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \ldots = X_0 + \beta t + \sum_{i=1}^{t} \varepsilon_i .$$

A simulation of a random walk

- The graphs above represent a random walk and a random walk with drift (top graphs) and their corresponding sample autocorrelations (botton graphs), computed with simulated data.



Drift or no drift, that is the question...

- Notice that the behavior of the ACF and the PACF for the random walk with or without drift is fairly similar.

- To decide whether to include or not a constant in the model we need to look at the plot of the original data.

  - If the data looks trended include a constant.

Beyond the random walk: ARIMA models.

- ARIMA: Autoregressive Integrated Moving Average

- Consider the process X_t

$$X_t = X_{t-1} + u_t$$

where u_t is a stationary process; therefore $u_t = \mu + \psi(L)\varepsilon_t$

- If u_t admits an ARMA(p,q) representation, then X_t is ARIMA(p,1,q).

ARIMA(p,d,q), II

- More generally, X_t admits the following representation:

$$\phi_p(L)(1-L)^d X_t = \beta^* + \theta_q(L)\varepsilon_t,$$

where

- $d = 1$ in this case,
- $\phi_p(L) = 1 - \phi_1 L - ... - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + ... + \theta_q L^q$

are the AR and MA polynomials, respectively.

- $\beta^* = \mu\phi(1)$.

ARIMA(p,d,q), III

- The parameter d represents the number of times X_t must be differenced to achieve a stationary transformation.

- Typically, $d \in \{0, 1, 2\}$. The case $d = 0$ corresponds to the ARMA case, studied in Handouts 1 and 2.

- X_t is said to be an integrated process of order d, or $I(d)$ in short.

- $I(0)$ processes are stationary while $I(1)$ and $I(2)$ are not.

- Models with unit roots have become very popular in applied research.

- Reasons:

  - Simple and yet, they seem to fit well economic data

  - The transformed stationary process (i.e., the resulting process after differences are taken) is easy to interpret.

Units roots, logarithms and rates of growth

- If the variable is in logs, its first difference is the rate of growth:
- To see this notice that

$$\begin{aligned}(1-L)\log X_t &= \log(X_t/X_{t-1}) \\ &= \log(1+(X_t-X_{t-1})/X_{t-1})\end{aligned}$$

and if the change is small, using the approximation $\log(1+x)\sim x$ if x is close to zero, then

$$(1-L)\log X_t \approx (X_t-X_{t-1})/X_{t-1}.$$

Success of I(d) processes

1. Simplicity: by differencing d times one obtains a stationary process.

2. The required transformation is (usually) easy to interpret:

    - In the I(1) case, if the variables are in logs, ΔX_t is the rate of growth of X_t

3. Empirically plausible.
    - Nelson and Plosser (1982): most US macroeconomic series are better represented as I(d) rather than as TS models.

    - Nice features: persistence of shocks, unbounded confidence intervals for long-horizon forecasts, etc

4. Technically feasible: development of the asymptotic theory for integrated processes.

Main question: how to choose the "right" model for your potentially non-stationary data?

- We want to fit a model to y_t, then we need to answer these questions
  - Is it stationary or not?
  - If stationary, fit an ARMA, AR, etc.
  - If non-stationary, choose a non-stationary model
- How? Unit root tests

- If the data is non-stationary, is it the same to choose one model or another?
- No!
- The different models have different statistical properties, they might provide different answers to the same questions
- Thus, you might end up adopting different conclusions
- Therefore, a lot of emphasis in the literature on the importance of choosing a "good" model.

Some implications/problems associated to unit root processes

- If X_t is a unit root process,

  1. AR OLS estimates are biased towards zero $\longrightarrow$ bias can be large for the usual size of macroeconomic data $\longrightarrow$ bad forecasts if the unit root is not imposed.
  2. Standard inference does not hold $\longrightarrow$ Wrong inference if one uses the traditional critical values.
  3. Spurious regressions. Let X_t and Y_t be unit root processes. If one regresses one on another, estimates are not consistent in general (important exception: X_t and Y_t are cointegrated)

- Huge literature on detecting unit roots!

Problems if unit roots go undetected, I

1. Bad forecasts: The autoregressive coefficient is biased towards zero.

- To simplify consider the simplest case: Y_t follows a random walk and you fit and AR(1) model to this data: the AR coefficient is biased downwards (i.e., you will tend to find values of $\hat{\phi}$ that are smaller than 1, we'll see why!).

- The bias may be important if some sizes are small or moderate. Since the sample size of most macroecomic series is typically short, this is a problem in practise.

- One implication: forecasts based on AR(1) models may perform quite badly in comparison to forecasts based on random walks (despite the fact that the AR(1) nests the random walk model!)

Problems if unit roots go undetected, II

2. Standard inference doesn't hold

- Remember that the validity of LLNs and CLTs relied on stationarity+ergodicity. Processes with unit roots are not stationary so they do not hold.

- The asymptotic distribution of the autoregressive parameter is non-normal. It is a **functional of Brownian motions**.

- These distributions are called non-standard since they are not any of the standard distributions (i.e., Normal, χ^2, t or F distributions).

- Hence, if one uses standard critical values, the corresponding inference will be wrong.

Problems if unit roots go undetected, III

3.The problem of spurious regressions.

- Two independent unit root processes may look related even if they are independent.

- That is, if X_t and Y_t are independent unit root variables, the estimate of the coefficient β in the regression

$$Y_t = \alpha + \beta X + a_t$$

does not tend to zero (in fact, $\hat{\beta}$ converges to a random variable).

- Hence, one can obtain 'spurious' relationships between variables (see Granger and Newbold, 1974 and Phillips, 1986).

- See here http://mayoral.iae-■ csic.org/timeseries_insead/examplespurious.pdf for some examples of non-sense regressions

Summarizing...

- If a process contains a unit root but it is not taken into account we might have
  - Estimation problems
  - Inference problems
  - Non-sense relationships between variables.
- Thus, detecting unit roots is very important!!

3. Unit root Tests

Unit root tests

- Unit root tests are tests designed to determine whether a process contains a unit root (or more) or not.

- 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....

Unit root tests, II

- Very very large literature!! See early summary: Xiao and Phillips (1999) .

- Pioneer work: the Dickey-Fuller test, then many other other tests

- Dickey-Fuller test: based on a very simple idea: Compute 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.

The Dickey- Fuller test

- D-F test Goal: test for a unit root in X_t (versus different types of alternative hypothesis).

- Approach (simplest case): consider an autoregressive model (that nests the unit root model) where u_t is a white noise process.

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

- $H_0 : \phi = 1$: non-stationary (unit root); $H_1 : \phi < 1$, stationary.

- Then, use a t-test to test $\phi = 1$ vs. $\phi < 1$

But, some difficulties arise:

1. The distribution is “non-standard”: different critical values
2. The distribution changes depending on the deterministic components included in the model
3. We need to account for serial correlation in the residuals (otherwise the distribution will also be different!)

First problem: what are the critical values one should use?

- Can we employ normal critical values, as usual??
- Answer: NO!
- The distribution of the t-statistic under the null hypothesis of $\phi = 1$ is NOT normal
- As we know, when ϕ is 1 (unit root), the properties of the OLS estimator change completely!

- Why? LLN and CLT don't work anymore.

- These results need to be replaced by very different tools (Functional central limit theorem and continous mapping theorem)

- The asymptotic distribution of $\hat{\phi}$ (the OLS estimator of ϕ looks VERY different from a normal distribution.)

Asymptotic distribution of $\hat{\phi}$ if ϕ=1

- For the simple case outlined above (no deterministic components, u_t is white noise

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

where W(.) is a Brownian motion (=continuous time stochastic process, see the Appendix for its definition).

(See the appendix of this handout for details on the derivation of this distribution)

- This distribution is very different from the stationary one

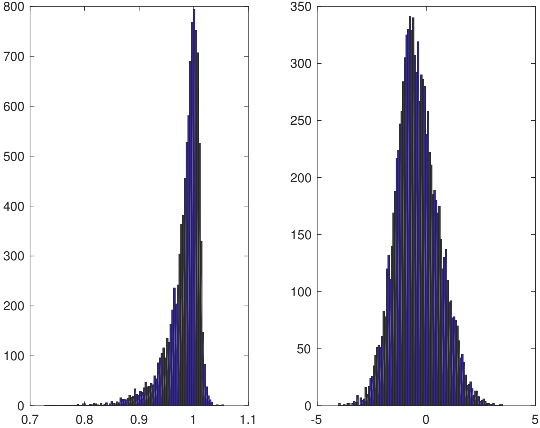
- Recall that if $\phi < 1$, then

$$T^{1/2}\left(\hat{\phi}_{ols} - \phi\right) \xrightarrow{d} N\left(0, \frac{\sigma^2}{var(X_t)}\right) = N\left(0, \left(1 - \phi^2\right)\right).$$

What is different now?

- Many things!
- If $\phi = 1$, $\hat{\phi}$ is still consistent (converges to the true value 1)
- $\hat{\phi}$ converges at a rate T (rather than $T^{1/2}$, the usual rate): $\hat{\phi}$ is super consistent
- The distribution is NOT normal, it's not even symmetric!
- it's skewed to the left

A simulation: histograms of $\hat{\phi}$ and the t-statistic if $\phi = 1$



- Small sample bias of $\hat{\phi}$:

- We mentioned that if $\phi = 1$ then $\hat{\phi}$ is very often well below that value.

- you see why this is the case in that graph!

Dickey-fuller test, II

- The contribution of Dickey and Fuller: develop the distribution of $\hat{\phi}$ and its t-statistic if $\phi = 1$
- To do that, they had to use a totally different asymptotic theory (non-standard):
- The tools employed to get there are different: functional central limit theorem and continuous mapping theorem.
- The limit process is (a functional) of a Brownian motion

Second complication: inclusion of deterministic components

- The model above doesn't include deterministic components.

- But we should always include some (variables can look trended, can have non-zero means, etc).

- Now, denote by

- "regression model": the model you actually run and

- "true model" the model that generates the data (unobserved, under the null hypothesis, a random walk –with or without drift).

- Problem: the asymptotic distribution changes depending on
  - the deterministic components in the true model (unobserved!)
  - and in the regression model (observed).

An example

- True model: $X_t = \mu + \phi X_{t-1} + u_t$, u_t white noise.
- Consider estimation of these regressions models in the stationary case ($\phi < 1$):

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

$$(2) X_t = \alpha + \phi X_{t-1} + \varepsilon_t$$

$$(3) X_t = \alpha + \beta t + \phi X_{t-1} + \varepsilon_t$$

- Notice that in the stationary case, the distribution of the t-statistic associated to $\phi = \phi_0$ in any of these three models is the same (N(0,1))
- The distribution of the t-statistic is different if: μ is zero or not (unobserved!) and models (1), (2) and (3) are run!

Example, II

- If $\phi = 1$, then the distribution of the t-statistic is different in the regression models (1), (2) and (3)

- And something even worse:

- The distribution also changes depending on whether the "true model" has a drift or not, that is

  - $X_t = X_{t-1} + u_t$: random walk (no trend)
  - $X_t = \mu + X_{t-1} + u_t$: random walk with drift (trended process)

Then, what are the critical values we should use in the D-F test?

- It follows that the combination of 1) different true models (unit root with or without drift) and different regression models generate different "cases".

- See Hamilton for a full description.

- We next explain how to deal with this problem in the simplest case: u_t is a white noise

Simplest case: DF test with uncorrelated disturbances

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

- To avoid complications ALWAYS run regression models with $(\alpha + \beta t)$

- By doing this, the distribution of the t-test is invariant to the true value of μ (can be zero or different from zero).

- Case 4 (Hamilton's notation). The true model (TM) is a random walk with or without drift and the regression model (RM) is an AR(1) and contains a constant and a trend

$$
\begin{aligned}
TM &: X_t = \mu + X_{t-1} + \varepsilon_t & (5) \\
RM &: X_t = \alpha + \beta t + \phi X_{t-1} + \varepsilon_t & (6)
\end{aligned}
$$

where ε_t is an $\text{iid}\left(0, \sigma^2\right)$ sequence

DF test with uncorrelated disturbances, III

- The hypotheses to be tested are

$$H_0 \quad : \quad \phi = 1,$$
$$H_1 \quad : \quad \phi < 1.$$

or alternatively, subtracting X_{t-1} in both sides of (6), the regression model results

$$RM : \Delta X_t = \alpha + \beta t + \varphi X_{t-1} + \varepsilon_t, \tag{7}$$

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

$$H_0 \quad : \quad \varphi = 0,$$
$$H_1 \quad : \quad \varphi < 0.$$

DF test with uncorrelated disturbances, IV

- A $t - test$ then can be used for testing $\phi = 1$ in (6) or $\varphi = 0$ in 7 . The $t - test$s 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.

- Big advantage of including a trend in the regression model: the distribution of the t-statistic is invariant to the value of μ. (case 4 distribution)

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?
  - 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.

And a third complication: Unit root tests with correlated u_t.

- In the previous section we have assumed that the innovation $u_t = \varepsilon_t$ was an *iid* sequence. This framework is very narrow since most real processes do not fall in this category.

- If u_t is a general stationary process, the distributions described above do not longer hold.

- There exist two main approaches that are able to solve this problem: the Phillips-Perron approach and the Augmented DF test.

First approach: Phillips-Perron correction.

- Consider the random walk process X_t, $X_t = X_{t-1} + u_t$, where u_t is a stationary process that admits a Wold representation $u_t = \psi(L)\,\varepsilon_t$ and ε_t is an *iid* sequence.

- Assume that the following regression model is employed to test for a unit root in X_t :

$$X_t = \alpha + \beta t + \phi X_{t-1} + u_t. \qquad (8)$$

- If X_t is stationary, the OLS estimate of ϕ is inconsistent if u_t is autocorrelated (why?)

- However, if X_t contains a unit root $(\phi = 1)$, it can be shown that $\hat{\phi}$ is still super-consistent (converges to 1 in probability at a rate T).

- Phillips-Perron idea: Use the t-statistic in the AR(1) regression, as before.

- However, he fact that u_t is autocorrelated changes the distribution of the $t-test$, and therefore the critical values in Case 4 tables cannot be directly employed here.

- Phillips and Perron showed that a function of the $t-test$ does converge to the distribution described in the 'Case 4' section above. More specifically,

$$\left(\gamma_0/\lambda^2\right)^{1/2} t_T - \left\{1/2\left(\lambda^2-\gamma_0\right)/\lambda\right\} \times \left\{T\hat{\sigma}_{\hat{\phi}}/s_T\right\} \xrightarrow{d} \Lambda \quad (9)$$

where Λ is the Case 4 distribution for the case where $u_t = \varepsilon_t$ is an iid sequence, $\gamma_0 = Var(u_t)$, $\lambda^2 = \sigma^2\psi(1)^2 = \gamma_0 + 2\sum_{i=1}^{\infty}\gamma_i$, where γ_i is the *ith* autocovariance of u_t, $\hat{\sigma}_{\hat{\phi}}$ is the standard error of $\hat{\phi}$ and s_T^2 is an estimator of the variance of ε_t.

Steps of the Phillips-Perron approach

1. Estimate the RM (8) by OLS
2. Compute the $t - test$ associated to the hypothesis $\phi = 1$
3. Estimate the other elements in equation (9). This basically entails to estimate γ_0 and λ.
4. The former can be estimated simply as

$$\hat{\gamma}_0 = T^{-1} \sum_{t=1}^{T} \hat{u}_t^2.$$

where $\hat{u}_t = X_t - \hat{\alpha} - \hat{\phi} X_{t-1}$.

- As for $\lambda^2 = \sigma^2 \psi(1)^2$ (also called **the long-run variance**) there exists many estimators that can be employed. A popular one is the Newey-West estimator

$$\hat{\lambda}^2 = \hat{\gamma}_0 + 2 \sum_{j=1}^{q} (1 - j/(q+1)) \hat{\gamma}_j$$

where $\gamma_j = T^{-1} \sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j}$.

- Phillips (1987) established the consistency of $\hat{\lambda}^2$ provided that q,the lag truncation parameter, goes to infinity as the sample size T grows and provided that q grows sufficiently slowly relative to T. More specifically, T, $q \to \infty$ and $q/T^{1/4} \to 0$. **Remark:** This is an asymptotic result and does not tell us exactly how q should be chosen in small samples.
- Construct the corrected t-statistic and use the critical values corresponding to Case 4.

Second approach: the Augmented Dickey-Fuller test

- Consider the case where u_t is an AR(p) process, that is,

$$X_t = \delta_1 X_{t-1} + u_t$$

$$u_t = \varepsilon_t / \phi(L)$$

where $\phi(L) = (1 - \delta_2 L)...(1 - \delta_p L) = (\phi_0 + \phi_1 L + ...\phi_p L^p)$ with $\phi_0 = 1$.

- Notice that the polynomial $\phi(z)$ can be written in the following way:

$$
\begin{aligned}
\phi(z) &= \sum_{i=0}^{p}\phi_i - \sum_{i=1}^{p}\phi_i + \left(\sum_{i=1}^{p}\phi_i - \sum_{i=2}^{p}\phi_i\right) z \\
&\quad + \left(\sum_{i=2}^{p}\phi_i - \sum_{i=3}^{p}\phi_i\right) z^2 + \\
&\quad ... + \left(\sum_{i=p-1}^{p}\phi_i - \phi_p\right) z^{p-1} + (\phi_p) z^p \\
&= \phi(1) - (1-z)\sum_{i=1}^{p}\phi_i - (1-z)\sum_{i=2}^{p}\phi_i z \\
&\quad - (1-z)\sum_{i=3}^{p}\phi_i z^2 - ... - (1-z)\phi_p z^{p-1} \\
&= \phi(1) - (1-z)\phi^*(z)
\end{aligned}
$$

- This implies that $X_t = \delta_1 X_{t-1} + \frac{\varepsilon_t}{\phi(L)}$ can be written as

$$
\begin{aligned}
(1-L)\,X_t &= (\delta_1 - 1)X_{t-1} + \frac{\varepsilon_t}{\phi(L)} \\
\phi(L)\,\Delta X_t &= \phi(L)\,(\delta_1 - 1)X_{t-1} + \varepsilon_t \\
\Delta X_t &= \underbrace{\phi(1)\,(\delta_1 - 1)}_{=\alpha}X_{t-1} + \\
&\quad \phi^*(L)\,\delta_1 \Delta X_{t-1} + (\phi(L) - 1)\Delta X_t + \varepsilon_t \qquad (10) \\
\Delta X_t &= \alpha X_{t-1} + \sum_{k=1}^{p-1} \varphi_k \Delta X_{t-k} + \varepsilon_t. \qquad (11)
\end{aligned}
$$

- Augmented DF test:
- Hypotheses:

$$H_0 : \phi = 0.$$

vs

$$H_1 : \phi < 1$$

- Model to be run:

$$\Delta X_t = \beta_0 + \beta_1 t + \phi X_{t-1} + \sum_{k=1}^{p-1} \varphi_k \Delta X_{t-k} + \varepsilon_t$$

- Notice that under the null hypothesis, α is equal to zero. Thus, the null hypothesis can be formulated as

Summarizing

- The Agmented Dickey-Fuller test consists of regressing ΔX_t on X_{t-1} and p lags of ΔX_{t-1} (and possibly, some deterministic components, as discussed before.

- Dickey and Fuller showed that if the true DGP is such that $\alpha = 1$ and $u_t = \varepsilon_t / \phi(L)$, and regression model (11) is considered, then the test based on the t-test of $\alpha = 0$ has the same asymptotic distribution as in the non-autocorrelated case.

- When deterministic components are considered, the same arguments can be applied and the fundamental result still applies.

How to choose p?

- In practice, the order p is unknown and perhaps, it is infinite (for instance, if u_t contains an MA component).

- Using the results in Berk (1974), Said and Dickey (1984) showed that we still can fit a finite-order AR(p) model to approximate the AR(∞) one and use the DF critical values corresponding to the uncorrelated case.

- In practice, information criteria can be used to select the order of the polynomial of lags of ΔX_t.

Other approaches to test for unit roots.

- The ADF and the PP unit root tests are known to suffer severe finite-sample **power** and **size** problems:

- Power: ADF and PP tests have low power if the true DGP (=data generating process) is a stationary process with a large autoregressive root. (See, e.g., DeJong et al., *J. of Econometrics*, 1992).

- Size: Both the ADF and the PP tests are known to have severe size distortions (they over-reject the null) when the series has a large negative moving average root.

For instance Schwert, JBES, 1989 showed that if the true DGP has a unit root with a MA component where $\theta = -0{,}8$, then size=100 %!. (Important macroeconomic process that tend to present this type of behavior are, for instance, inflation or the unemployment rate).

Other approaches to test for unit roots, II

- A variety of alternative procedures have been proposed to resolve these problems.

- To improve power, generalized least squares detrending has become very popular.

- To improve size: the use new information criteria to determine the number of lags that should be introduced in the regression to control for short-run correlation has improved size considerably.

Generalized least squares detrending, Elliot et al. (1996)

- Elliot et al. (ERS, Econometrica, 1996) proposed a modification to the DF approach that increases power substantially.

- They showed that if there is not deterministic components in the model, then the asympotic power of the DF test is close to the asymptotic optimal power bound.

- However, whenever deterministic components are present (constant terms and/or trends), then power can be improved significantly by modifying the method employed to estimate the coefficients of the deterministic components.

Generalized least squares detrending

- The approach is as follows. Consider the DGP

$$\begin{aligned} y_t &= \psi' z_t + u_t, \\ u_t &= \alpha u_{t-1} + \varpi_t, \end{aligned}$$

where z_t is a set of deterministic components and ϖ_t is a stationary process.

Elliot et al. proposed local to unity GLS detrending of the data.

- This amounts to define $(y_0^\alpha,\ y_t^\alpha) = (y_0,\ (1-\bar{\alpha}L)\, y_t)$, for some chosen $\bar{\alpha} = 1 + \bar{c}/T$. [Elliot et al. determined by simulation which is the optimal value for $\bar{c}$]

- The GLS detrended series is defined as

$$\tilde{y}_t = y_t - \hat{\psi}' z_t$$

where $\hat{\psi}$ minimizes $S(\bar{\alpha}, \psi) = \left(y^{\bar{\alpha}} - \psi' z_t^{\bar{\alpha}}\right)' \Sigma^{-1} \left(y^{\bar{\alpha}} - \psi' z_t^{\bar{\alpha}}\right)$, where Σ is the variance-covariance matrix of ϖ_t.

- Then, ERS recommended using the DF-GLS statistic as the $t - statistic$ for testing $\beta_0 = 0$ in the regresion

$$\Delta \tilde{y}_t = \beta_0 \tilde{y}_{t-1} + \sum_{i=1}^{k} \beta_i \Delta \tilde{y}_{t-i} + e_{tk}.$$

- The power gains of the DF-GLS with respect to the standard DF test are impressive.

- However, tests exhibit strong size distortions when ϖ_t contains an MA(1) with a negative coefficient.

Correcting size distortions: Ng and Perron's method.

- Ng and Perron (2001) proposed a new method to select k, the number of lags to be included in the ADF.
- Notice that if ϖ_t contains an MA component, then we might need a lot of lags (in theory, an infinite number) to correct for the autocorrelation in ϖ_t.
- Ng and Perron argued that standard information criteria (AIC, BIC...) tend to select a k that is too low if ϖ_t has a negative MA component. As a consequence of this, size can be very poor in these situations.
- They introduced a modification to standard IC (the Modified AIC and the Modified BIC, –MAIC and MBIC–) that is able to select a larger k in these situations.
- Finally, they also propose a new set of tests (the MZ-GLS tests+MAIC, MBIC) that have very good power and size

Test of stationarity (versus non-stationarity)

- All the unit-root test studied up to now are designed for testing the null hypothesis of a unit root.

- Then, unless evidence against the unit root is found the null would not be rejected.

- This favors the non-rejection of the unit root.
- Kwiatkowski et al. (1992) proposed to reverse the hypothesis. That is, to test the null of stationarity versus the alternative of unit root.

Some tips for running unit root tests

- DF test are outdated.
- USE GLS-DF tests instead.
- Always include a constant and a trend in the model
- You can use the MAIC and MBIC to improve the size of the test
- You can also run tests of stationarity (KPSS test), where the null is I(0).

Unit roots with STATA

See this STATA code in the website of the course: unit$_r$*oots.do*

`http://mayoral.iae-csic.org/timeseries2021/unit_roots.do`

On the observational equivalence of Unit root and covariance-stationary processes

- In 1982 Nelson and Plosser showed that many macroeconomic series are better described by unit roots than by deterministic time trends. This changed the way macroeconomist modelled this type of data.

- Since then, large literature on detecting unit root tests in the data has been produced.

- However, several authors have argued that the question of whether a process contains a unit root is inherently answerable on the basis of a finite sample.

- The argument goes as follows:

- For any unit root process there exists a stationary processes that will be impossible to distinguish from the unit root representation for any given sample size T (¡∞). Why? One can consider a stationary AR processes with one of the roots arbitrarily close to 1 in such a way that it displays very similar characteristics.

- The converse proposition is also true: for any stationary process and a given sample size T, there exists a unit root process that will be impossible to distinguish from the stationary representation.
- For instance, a unit root process with an MA component with a root very close to 1 (such that the unit root and the MA root almost cancel out) would be indistinguisible from a stationary process for any finite sample size.

- Conclusion:

Unit root and stationary processes differ in their implications at infinite time horizons, but for any given finite number of observations, there is a representation from either class of models that could account for all the observed features of the data.

- Thus, if both the unit root and the stationary representations are possible, what is the meaning of unit root tests??

- **The goal of unit root tests is to find a parsimonious representation that gives a reasonable approximation to the true process, as opposed to determining whether or not the true process is literally** $I(1)$.