Time Series Analysis:

Introduction to time series and forecasting

Handout 4: Univariate nonstationary processes: processes with unit roots.

Laura Mayoral

IAE and Barcelona GSE

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!

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

- We will focus on process with non-constant means.

Non-constant means

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

- In many occasions, they are trended.

- The approach we'll follow:

- Find transformations that remove the trend component so that the resulting process is statiotionary.

- Apply to the transformed process the techniques for stationary process.

Types of trend: Deterministic vs stochastic trends

- Two ways of capturing the trend component of a process:

- Deterministic trends: the trend is a non-random function of time.

  - Typically, it will be a polynomial in t:

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

- Stochastic trends: The trend is a random variable.

  - unit roots

Models with trends: Trend stationary and Unit root processes.

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 process

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

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.

- TS models were very popular in the 80's. But nowadays most people believe that stochastic trends (unit roots) are more appropriate to model economic time series.

- Still, when fitting a model to the data most times one starts by testing for unit roots.

- Due to its importance in applied work, in the following we will mostly focus on models with stochastic trends (unit roots).

Unit root processes: examples

- Simplest case: random walks and random walks 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,

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

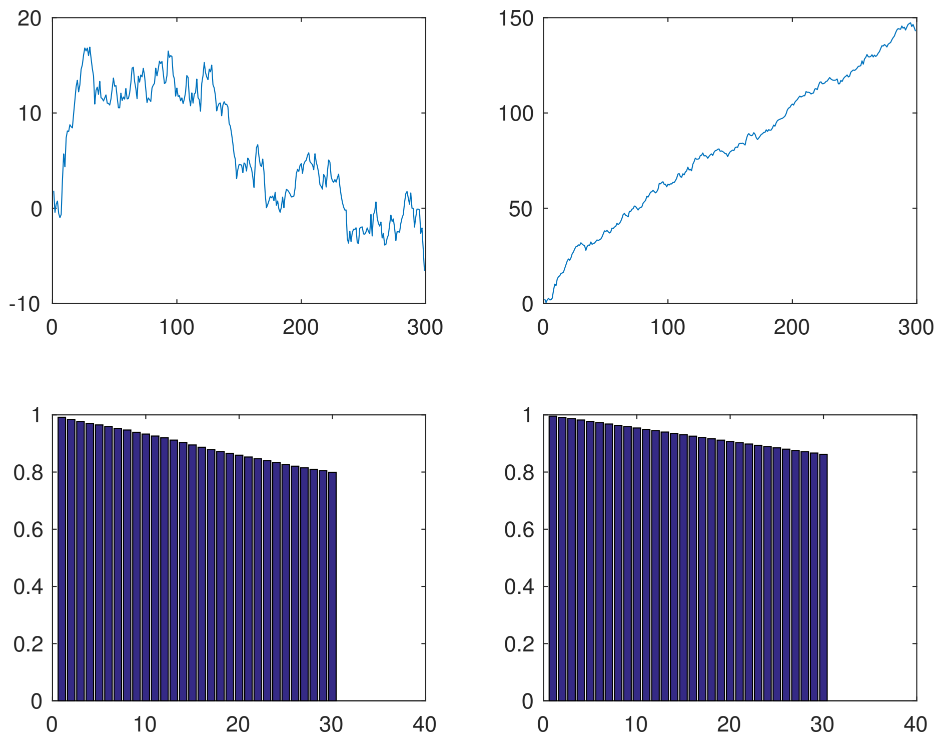
and $E(X_t) = X_0$ and $var(X_t) = t\sigma^2$.

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

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

and by back-substitution

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



- 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 specifically, 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

- More generally, d will be a positive integer number.

- 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 2 and 3.

- 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.

Deterministic components in integrated processes

- The term β is a deterministic component and plays different roles for different values of d.

- If $d = 0$, β represents a constant term such that the mean of $\{X_t\}$ is given by $\mu = \beta/(1 - \phi_1 - ... - \phi_p)$.

- If $d = 1$, β is the rate of growth

- if $d = 2$, β is the rate at which the growth rate increases.

Units roots, logarithms and rates of growth

- If the variable is in logs, a unit root in that variable implies that its rate of growth of is stationary.

$$(1-L)\log X_t = u_t,$$

- 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

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

Some issues arising when dealing with unit root processes

- To simplify consider the simplest case: you fit an AR(1) model to a process that in reality is a random walk (i.e., the AR parameter is $\phi = 1$. This are some of the problems you might encounter.

1. The autoregressive coefficient is biased towards zero.

- If 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.)

- 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!)

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).

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

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).

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!!

- This handout provides a brief explanation of why the above facts occur

- Introduces tests for detecting the existence of unit roots.

Asymptotic theory with units roots

- Stationary AR(1):

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

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 is used in place of the LLN

Non-standard asymptotic theory: Preliminary concepts

Brownian motion

- Consider a random walk process

$$X_t = X_{t-1} + \varepsilon_t, \tag{5}$$

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 $e_{it} \sim iid\ N\left(0, 1/N\right)$. 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 $W\left(t\right)$.

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) 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:

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

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

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

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

The functional central limit theorem

- The Functional limit theorem establishes that

$$\sqrt{T} X_T\left(.\right)/\sigma \xrightarrow{d} W\left(.\right) \tag{6}$$

Continuous mapping theorem (CMT)

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

- A similar result holds for sequences of random functions.

- The CMT states that if $S_T\left(.\right) \stackrel{d}{\to} S\left(.\right)$ and g is a continuous functional, then $g\left(S_T\left(.\right)\right) \stackrel{d}{\to} g\left(S\left(.\right)\right)$.

For instance, from (6) and the CMT it follows that

$$\sqrt{T} X_T\left(.\right) \stackrel{d}{\to} \sigma W\left(.\right).$$

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:

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

$X_t\left(r\right)$ 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 $S_T(.) = \left(\sqrt{T} X_T(r)\right)^2$

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

- 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 (4) 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 $\phi = 1$

$$X_t = \phi X_{t-1} + u_t \tag{7}$$

- 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

$$
\begin{aligned}
TM &: \quad X_t = X_{t-1} + \varepsilon_t, \;\; X_0 = 0 \\
RM &: \quad X_t = \phi X_{t-1} + \varepsilon_t
\end{aligned} \tag{8}
$$

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

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

or alternatively, subtracting X_{t-1} in both sides of (8), the regression model results

$$RM : \Delta X_t = \varphi X_{t-1} + \varepsilon_t, \tag{9}$$

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

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

DF test with uncorrelated disturbances, IV

- A $t-test$ then can be used for testing $\phi = 1$ in (8) or $\varphi = 0$ in 9 . 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.

- 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

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

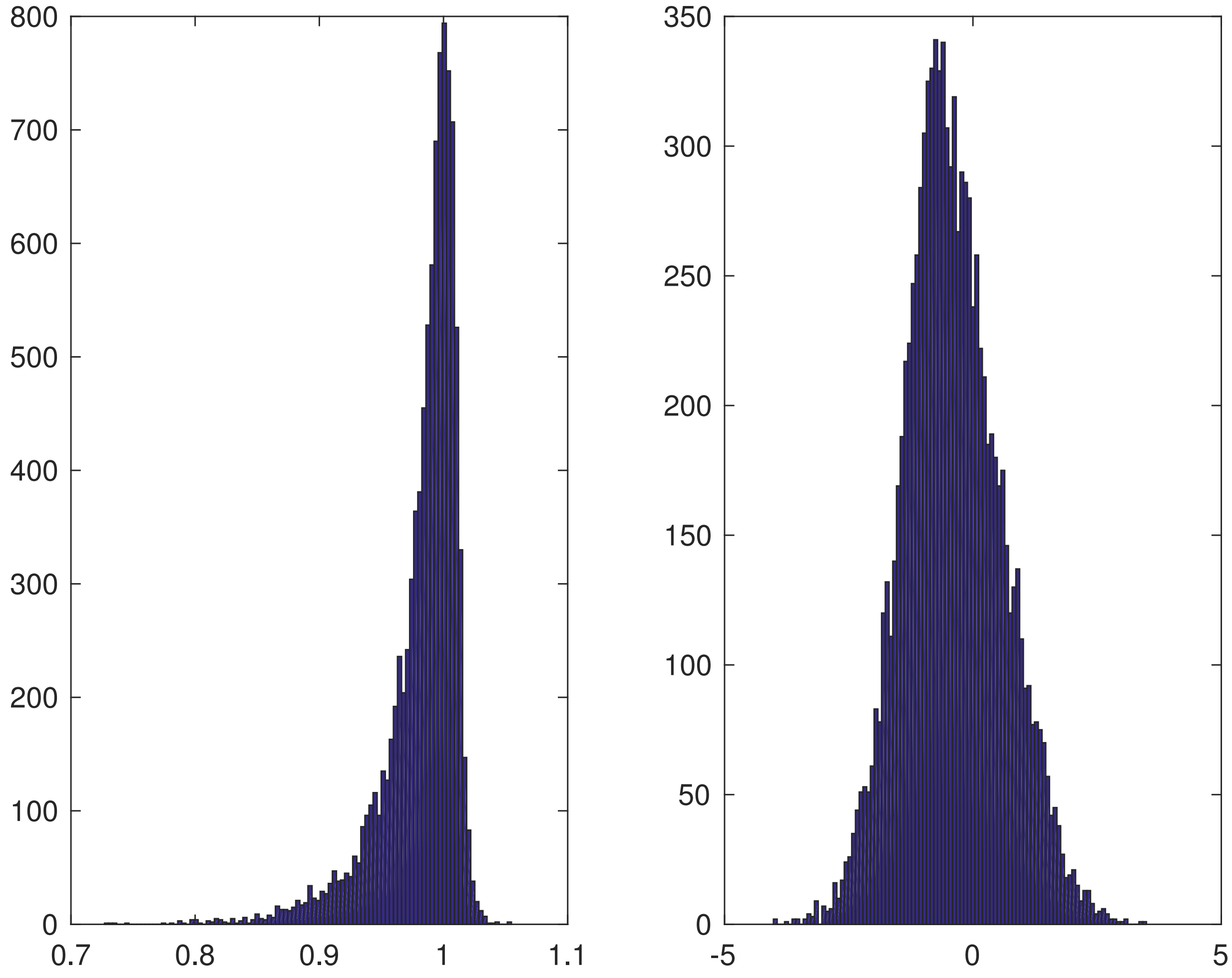
where $s_T^2 = (T-1)^{-1}\sum\left(X_t - \hat{\phi}X_{t-1}\right)^2$. Using the results in the previous section and the fact that $s_T^2 \xrightarrow{p} \sigma_\varepsilon^2$, it is easy to check that

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

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

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

- 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

$$F = \frac{\left(RSS_R - RSS_U\right)/r}{RSS_U/\left(T-k\right)} \tag{10}$$

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

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

■ 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

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

- 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 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

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

- 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.

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 :

$$X_t = \alpha + \beta t + \phi X_{t-1} + u_t. \tag{11}$$

- 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,

$$\left(\gamma_0/\lambda^2\right)^{1/2} t_T - \{1/2\left(\lambda^2-\gamma_0\right)/\lambda\} \times \{T\,\hat{\sigma}_{\hat{\phi}}/s_T\} \xrightarrow{d} \Lambda \tag{12}$$

where Λ is the Case 4 distribution for the case where $u_t = \varepsilon_t$ is an iid sequence, $\gamma_0 = Var(u_t)$, $\lambda^2 = \sigma^2\psi\left(1\right)^2 = \gamma_0 + 2\sum_{i=1}^{\infty}\gamma_i$, 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 (11) by OLS

2. Compute the $t-test$ associated to the hypothesis $\phi = 1$

3. Estimate the other elements in equation (12). 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 $\hat{u}_t = X_t - \hat{\alpha} - \hat{\phi} X_{t-1}$.

5. 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.

6. Construct the corrected t-statistic and use the critical values corresponding to Case 4.

Similar ideas can be used to generalize case 2 in the previous section.

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:

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

where $\phi^*(z) = \sum_{j=1}^{p-1}\phi_j^* z^j$ and $\phi_j^* = \sum_{i=j+1}^{p}\phi_i$. This is the so-called Beveridge-Nelson decomposition of $\phi(z)$.

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

$$\begin{aligned}
(1-L)\,X_t &= (\delta_1 - 1)X_{t-1} + \frac{\varepsilon_t}{\phi\,(L)} \\
\phi\,(L)\,\Delta X_t &= \phi\,(L)\,(\delta_1 - 1)X_{t-1} + \varepsilon_t \\
\Delta X_t &= \underbrace{\phi\,(1)\,(\delta_1 - 1)}_{=\alpha}X_{t-1} + \\
&\quad \phi^*\,(L)\,\delta_1 \Delta X_{t-1} + (\phi\,(L) - 1)\Delta X_t + \varepsilon_t \qquad (13) \\
\Delta X_t &= \alpha X_{t-1} + \sum_{k=1}^{p-1} \varphi_k \Delta X_{t-k} + \varepsilon_t. \qquad (14)
\end{aligned}$$

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

$$H_0 : \alpha = 0.$$

vs

$$H_1 : \alpha < 1$$

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 (14) 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 $\left(y_0^\alpha,\ y_t^\alpha\right) = \left(y_0,\ \left(1 - \bar{\alpha} L\right) y_t\right)$, 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 properties.

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.

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:

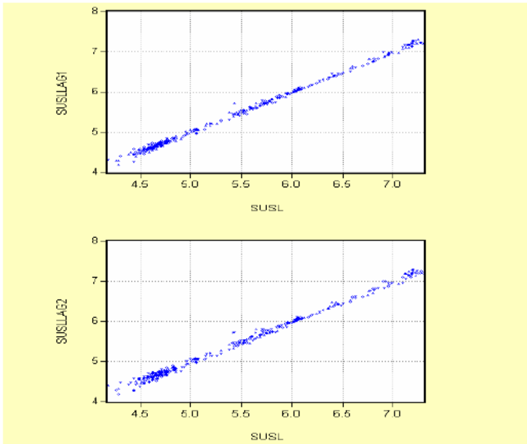
Unit root and stationary processes differ in their implications at infinite time horizons, but for any given finite number of observations, there is a representation from either class of models that could account for all the observed features of the data.

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

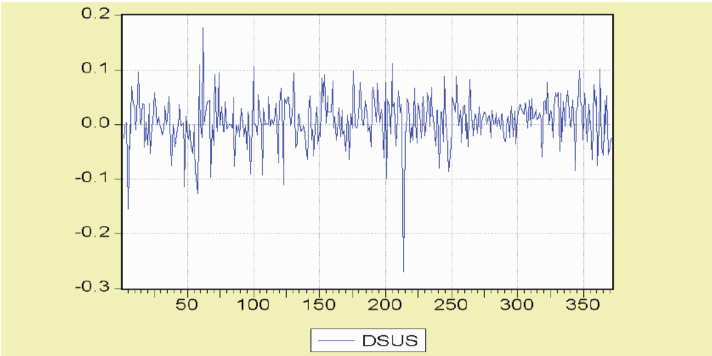
- **The goal of unit root tests is to find a parsimonious representation that gives a reasonable approximation to the true process, as opposed to determining whether or not the true process is literally** $I(1)$.

Some examples

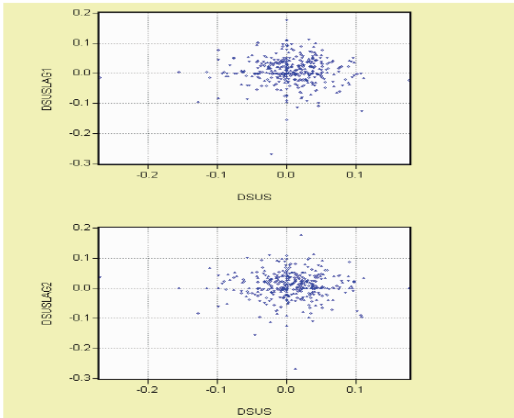
Asset price levels (logs) US market



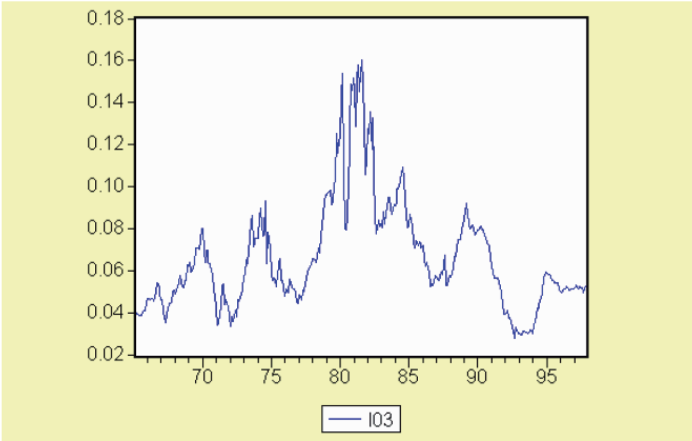
US stock market returns 1980-2000



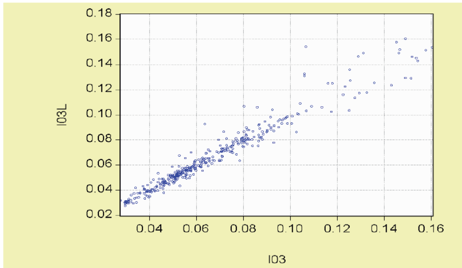
Looking at information in past prices: returns and past returns



Interest rate USA

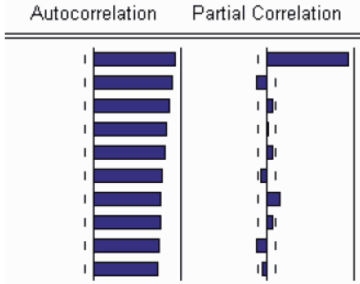


Scatter plot: unveiling information of past interest rate level



ACF and PAC for interest rates

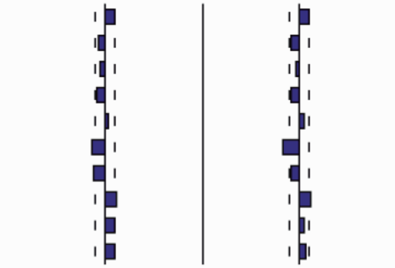
Date: 09/26/01 Time: 14:01
Sample: 1965:03 1997:12
Included observations: 394



Lag	AC	PAC	Q-Stat	Prob
1	0.971	0.971	374.09	0.000
2	0.935	-0.123	722.20	0.000
3	0.904	0.066	1048.2	0.000
4	0.875	0.009	1354.5	0.000
5	0.851	0.072	1645.0	0.000
6	0.825	-0.061	1918.9	0.000
7	0.808	0.152	2181.9	0.000
8	0.797	0.078	2438.6	0.000
9	0.780	-0.121	2685.0	0.000
10	0.757	-0.057	2918.1	0.000

AC and PAC of interest rate changes

Date: 09/27/01 Time: 14:54
Sample: 1965:03 1997:12
Included observations: 393



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.113	0.113	5.0741	0.024
		2	-0.077	-0.091	7.4383	0.024
		3	-0.046	-0.027	8.2977	0.040
		4	-0.090	-0.090	11.495	0.022
		5	0.034	0.050	11.944	0.036
		6	-0.142	-0.173	20.000	0.003
		7	-0.128	-0.092	26.552	0.000
		8	0.120	0.120	32.365	0.000
		9	0.099	0.054	36.362	0.000
		10	0.099	0.067	40.317	0.000

Regression for the DF test

Dependent Variable: I03
Method: Least Squares
Date: 10/16/01 Time: 14:02
Sample(adjusted): 1965:04 1997:12 [redacted]
Included observations: 393 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001939	0.000843	2.300072	0.0220
I03(-1)	0.971338	0.011798	82.33083	0.0000

R-squared	0.945462	Mean dependent var	0.066449
Adjusted R-squared	0.945323	S.D. dependent var	0.026379
S.E. of regression	0.006168	Akaike info criterion	-7.333761
Sum squared resid	0.014876	Schwarz criterion	-7.313538
Log likelihood	1443.084	F-statistic	6778.366
Durbin-Watson stat	1.749519	Prob(F-statistic)	0.000000

ADF test for interest rates

ADF Test Statistic	-2.513452	1% Critical Value*	-3.4491
		5% Critical Value	-2.8692
		10% Critical Value	-2.5708

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(I03)
Method: Least Squares
Date: 09/27/01 Time: 13:19
Sample(adjusted): 1965:06 1997:12
Included observations: 391 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
I03(-1)	-0.029996	0.011934	-2.513452	0.0124
D(I03(-1))	0.136205	0.050465	2.699030	0.0073
D(I03(-2))	-0.074735	0.050636	-1.475938	0.1408
C	0.002030	0.000852	2.381844	0.0177

R-squared	0.036752	Mean dependent var	3.54E-05
Adjusted R-squared	0.029285	S.D. dependent var	0.006222
S.E. of regression	0.006130	Akaike info criterion	-7.340916
Sum squared resid	0.014544	Schwarz criterion	-7.300316
Log likelihood	1439.149	F-statistic	4.921887
Durbin-Watson stat	2.001991	Prob(F-statistic)	0.002288