An Introduction to univariate financial time series analysis

Carlo Favero
Dept of Finance, Bocconi University

1 Introduction: what is a time-series?

Time-series is a sequence

$$\{x_1, x_2, ..., x_T\} \text{ or } \{x_t\}, \ t = 1, ..., T,$$

where t is an index denoting the period in time in which x occurs. We shall treat x_t as a random variable; hence, a time-series is a sequence of random variables ordered in time. Such a sequence is known as a stochastic process. The probability structure of a sequence of random variables is determined by the joint distribution of a stochastic process. A possible probability model for such a joint distribution is:

$$x_t = \alpha + \epsilon_t, \ \epsilon_t \sim n.i.d. \left(0, \sigma_\epsilon^2\right),$$

i.e., x_t is normally independently distributed over time with constant variance and mean equal to α. In other words, x_t is the sum of a constant and a *white-noise* process. If a white-noise process were a proper model for financial time-series, forecasting would not be very interesting as the best forecast for the moments of the relevant time series would be their uncoditional moments. However this is certainly not the case for all financial time series. Consider the dataset STOCKINT.XLS which contains, in Excel format, retrieved from Datastream, quarterly time-series data for stock index and valuation ratio and consumer price index for US, Germany and the UK, over the sample period 1973:1-2010:4.

TOTMKUS_PI: US Datastream Stock Market Price Index;
TOTMKUS_RI: US Datastream Stock Market Total Return Index;
TOTMKUS_DY: US Datastream Stock Dividend Yield;
TOTMKUS_PE: US Datastream Stock Price Earning;
TOTMKUK_PI: UK Datastream Stock Market Price Index;

TOTMKUK_RI: UK Datastream Stock Market Total Return Index;
TOTMKUK_DY: UK Datastream Stock Dividend Yield;
TOTMKUK_PE: UK Datastream Stock Price Earning;
TOTMKBD_PI: GER Datastream Stock Market Price Index;
TOTMKBD_RI: GER Datastream Stock Market Total Return Index;
TOTMKBD_DY:GER Datastream Stock Dividend Yield;
TOTMKBD_PE: GER Datastream Stock Price Earning;
USDOLLR: US $ TO UK £ (WMR) - EXCHANGE RATE
USCPI7500F:US CONSUMER PRICE INDEX;
UKCPI7500F:US CONSUMER PRICE INDEX;
BDCPI7500F:US CONSUMER PRICE INDEX;
BDBRYLD: GERMANY BENCHMARK BOND 10 YR (DS) - RED. YIELD
BDINTER3: BD FIBOR - 3MONTH (MTH.AVG.)

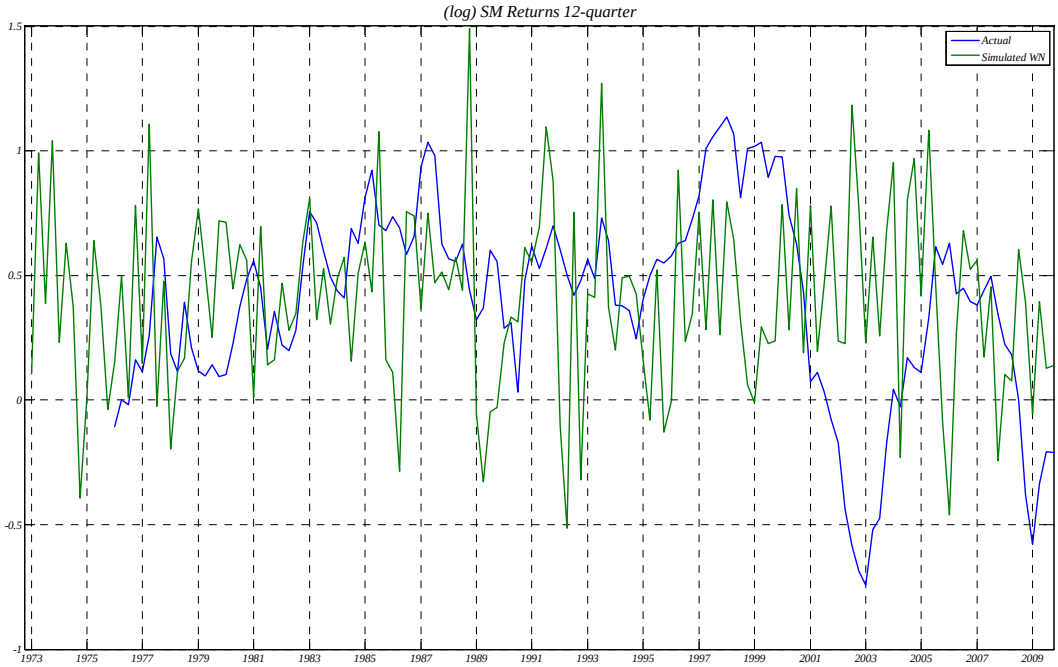
To assess the behaviour of financial time-series against the benchmark of a white-noise process plus a constant, we run first the MATLAB programmes **datatran.m** to load the data from the EXCEL file STOCKINT.XLS and perform the necessary data transformation:

We then run the following programme to generate artificial time series and compare them with the actual ones:

```
%to be run after datatran_int.m
%defining the parameters
alpha_1q=mean(us_ret_1r(2:end));
alpha_12q=mean(us_ret_12r(13:end));
beta_1q=var(us_ret_1r(2:end));
beta_12q=var(us_ret_12r(13:end));
alpha_ldp=mean(us_ldp(4:end));
beta_ldp=var(us_ldp(4:end));
%simulate artificial series
Num=size(us_ret_1r);
Wn_1q=alpha_1q+sqrt(beta_1q)*normrnd(0,1,Num,1);
Wn_12q=alpha_12q+sqrt(beta_12q)*normrnd(0,1,Num,1);
Wn_ldp=alpha_ldp+sqrt(beta_ldp)*normrnd(0,1,Num,1);
%plot the result
figure
h1=plot(t',us_ret_2r,t',Wn_1q,'-','LineWidth',2);
title(' (log) SM Returns 1-quarter','fontname','times','fontangle','italic','fontsize',18);
 set(gca,'fontname','times','fontangle','italic','fontsize',12,'gridlinestyle',':');
 set(gca,'xtick',[1:8:rows(t')]);
 set(gca,'xlim',[0 rows(t')]);
set(gca,'xticklabel','1973|1975|1977|1979|1981|1983|1985|1987|1989|1991|1993|1995|1997|1999|2001|
```

grid;
set(gcf,'color','w');
h1=legend('Actual','Simulated WN',1);

The first routine imports the data, implements a number of transformation to generate financial returns and ratios, the second routine generates artificial series defined as a constant (8.03) plus a normal random variable with zero mean and a given standard deviation, moments are chosen to calibrate those of quarterly and bi-annual returns. The graph shows clearly that the simple model described above does not describe the actual behaviour of all the time-series we are interested in.



To construct more realistic models, we concentrate on univariate models first to consider then multivariate models. In univariate models one should use combinations of ϵ_t. We concentrate on a class of models created by taking linear combinations of the white noise, the

autoregressive moving average (ARMA) models:

$$\begin{aligned}
AR(1): &\quad x_t = \rho x_{t-1} + \epsilon_t, \\
MA(1): &\quad x_t = \epsilon_t + \theta \epsilon_{t-1}, \\
AR(p): &\quad x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + ... + \rho_p x_{t-p} + \epsilon_t, \\
MA(q): &\quad x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}, \\
ARMA(p,q): &\quad x_t = \rho_1 x_{t-1} + ... + \rho_p x_{t-p} + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}.
\end{aligned}$$

Note that each of these models can be easily put to action to generate the equivalent time-series by modifying appropriately and running the following program in MATLAB, which generates an AR(1) series:

```
%Generate artificial AR series
%defining the parameters
rho=0.5;
%simulate AR series
AR_1q=NaN(size(us_p));
AR_12q=NaN(size(us_p));
AR_1q(2,1)=us_ret_1r(2,1);
ind=rows(us_p);
for i=3:ind
AR_1q(i,:)=rho*AR_1q(i-1)+alpha_1q*(1-rho)+normrnd(0,1,1)*sqrt((1-
rho^2)*beta_1q);
end
AR_12q(13,1)=us_ret_12r(13,1);
for i=14:ind
AR_12q(i,:)=0.5*AR_12q(i-1)+alpha_12q*(1-rho)+normrnd(0,1,1)*sqrt((1-
rho^2)*beta_12q);
end
```

The program generates a sample of observations from an AR(1) model with $\rho = 0.5$. The series is first initialized for the first observations, the command `series` then generates the series for the specified process, each observation being 0.5 times the previous observation plus a random disturbance drawn from a serially independent standard normal distribution.

2 Analysing time-series: fundamentals

To illustrate empirically all the fundamentals we consider a specific member of the ARMA family, the AR model with drift,

$$\begin{aligned}
x_t &= \rho_0 + \rho_1 x_{t-1} + \epsilon_t, \\
\epsilon_t &\sim n.i.d.\left(0, \sigma_\epsilon^2\right).
\end{aligned} \tag{1}$$

Given that each realization of our stochastic process is a random variable, the first relevant fundamental is the density of each observation. In particular, we distinguish between conditional and unconditional densities. Having introduced these two concepts, we define and discuss stationarity, generalize the form of our specific member to the whole family of ARMA models, and conclude this section with a discussion of deterministic and stochastic trends and de-trending methods. Note that at this introductory stage we concentrate almost exclusively on univariate models, for the sake of exposition. After the completion of our introductory tour, we shall concentrate on multivariate models.

2.1 Conditional and unconditional densities

We distinguish between conditional and unconditional densities of a time-series. The unconditional density is obtained under the hypothesis that no observation on the time-series is available, while conditional densities are based on the observation of some realization of random variables. In the case of time-series, we derive unconditional density by putting ourselves at the moment preceding the observation of any realization of the time-series. At that moment the information set contains only the knowledge of the process generating the observations. As observations become available, we can compute conditional densities. As distributions are summarized by their moments, let us illustrate the difference between conditional and unconditional densities by looking at our AR(1) model.

The moments of the density of x_t conditional upon x_{t-1} are immediately obtained from (1):

$$\begin{aligned} E\left(x_t \mid x_{t-1}\right) &= \rho_0 + \rho_1 x_{t-1}, \\ Var\left(x_t \mid x_{t-1}\right) &= \sigma_\epsilon^2, \\ Cov\left[\left(x_t \mid x_{t-1}\right), \left(x_{t-j} \mid x_{t-j-1}\right)\right] &= 0 \text{ for each } j. \end{aligned}$$

To derive the moments of the density of x_t conditional upon x_{t-2}, we need to substitute x_{t-2} from (1) for x_{t-1}:

$$\begin{aligned} E\left(x_t \mid x_{t-2}\right) &= \rho_0 + \rho_0\rho_1 + \rho_1^2 x_{t-2}, \\ Var\left(x_t \mid x_{t-2}\right) &= \sigma_\epsilon^2\left(1+\rho_1^2\right), \\ Cov\left[\left(x_t \mid x_{t-2}\right), \left(x_{t-j} \mid x_{t-j-2}\right)\right] &= \rho_1\sigma_\epsilon^2, \quad \text{for } j = 1, \\ Cov\left[\left(x_t \mid x_{t-2}\right), \left(x_{t-j} \mid x_{t-j-2}\right)\right] &= 0, \quad \text{for } j > 1. \end{aligned}$$

Finally, unconditional moments are derived by substituting recursively from (1) to express x_t as a function of information available at

time t_0, the moment before we start observing realizations of our process.

$$\begin{aligned} E(x_t) &= \rho_0\left(1+\rho_1+\rho_1^2+...+\rho_1^{t-1}\right)+\rho_1^t x_0, \\ Var(x_t) &= \sigma_\epsilon^2\left(1+\rho_1^2+\rho_1^4+...+\rho_1^{2t-2}\right), \\ \gamma(j) &= Cov(x_t, x_{t-j}) = \rho_1^j Var(x_t), \\ \rho(j) &= \frac{Cov(x_t, x_{t-j})}{\sqrt{Var(x_t)\,Var(x_{t-1})}} = \frac{\rho_1^j Var(x_t)}{\sqrt{Var(x_t)\,Var(x_{t-1})}}. \end{aligned}$$

Note that $\gamma(j)$ and $\rho(j)$ are functions of j, known respectively as the autocovariance function and the autocorrelation function.

2.2 Stationarity

A stochastic process is strictly stationary if its joint density function does not depend on time. More formally, a stochastic process is stationary if, for each $j_1, j_2, ..., j_n$, the joint distribution,

$$f(x_t, x_{t+j_1}, x_{t+j_2}, x_{t+j_n}),$$

does not depend on t.

A stochastic process is covariance stationary if its two first unconditional moments do not depend on time, i.e. if the following relations are satisfied for each h, i, j:

$$\begin{aligned} E(x_t) &= E(x_{t+h}) = \mu, \\ E\left(x_t^2\right) &= E\left(x_{t+h}^2\right) = \mu_2, \\ E(x_{t+i}x_{t+j}) &= \mu_{ij}. \end{aligned}$$

In the case of our AR(1) process, the condition for stationarity is $|\rho_1| < 1$. When such a condition is satisfied, we have:

$$\begin{aligned} E(x_t) &= E(x_{t+h}) = \frac{\rho_0}{1-\rho_1}, \\ Var(x_t) &= Var(x_{t+h}) = \frac{\sigma_\epsilon^2}{1-\rho_1^2}, \\ Cov(x_t, x_{t-j}) &= \rho_1^j Var(x_t). \end{aligned}$$

On the other hand, when $|\rho_1| = 1$, the process is obviously non-stationary:

$$\begin{aligned} E(x_t) &= \rho_0 t + x_0, \\ Var(x_t) &= \sigma_\epsilon^2 t, \\ Cov(x_t, x_{t-j}) &= \sigma_\epsilon^2(t-j). \end{aligned}$$

2.3 ARMA processes

Before introducing the fundamentals of time-series, we have asserted that white-noise processes are too simplistic to describe economic time-series and one can obtain a closer fit by considering combinations of white noises. We have then introduced ARMA models and discussed the fundamentals to understand their properties, but we have not yet shown that ARMA models represent combinations of white-noise processes. We show this by considering a time-series as a polynomial distributed lag of a white-noise process:

$$\begin{aligned} x_t &= \epsilon_t + b_1\epsilon_{t-1} + b_2\epsilon_{t-2} + ... + b_n\epsilon_{t-n} \\ &= \left(1 + b_1 L + b_2 L^2 + ... + b_n L^n\right)\epsilon_t \\ &= b(L)\epsilon_t, \end{aligned}$$

where L is the lag operator. The Wold decomposition theorem, which states that any stationary stochastic process can be expressed as the sum of a deterministic and a stochastic moving-average component, warrants generality of our representation. However, in order to describe successfully a time-series, a very high order in the polynomial $b(L)$ is required. This feature can be problematic for estimation, given the usual limitations for sample sizes. This potential problem is resolved, if the polynomial $b(L)$ can be represented as the ratio of two polynomials of lower order:

$$\begin{aligned} x_t = b\left(L\right)\epsilon_t = \frac{a\left(L\right)}{c\left(L\right)}\epsilon_t, \\ c\left(L\right)x_t = a\left(L\right)\epsilon_t. \end{aligned} \tag{2}$$

Equation (2) is an ARMA process. The process is stationary when the roots of $c\left(L\right)$ lie outside the unit circle. The MA component is invertible when the roots of $a\left(L\right)$ lie outside the unit circle. Invertibility of the MA component allows it to be represented as an autoregressive process.

To illustrate how the autocovariance and the autocorrelation functions of an ARMA model are derived, we consider the simplest case, the ARMA(1,1) process:

$$\begin{aligned} x_t = c_1 x_{t-1} + \epsilon_t + a_1\epsilon_{t-1}, \\ \left(1 - c_1 L\right)x_t = \left(1 + a_1 L\right)\epsilon_t. \end{aligned} \tag{3}$$

Equation (3) is equivalent to:

$$\begin{aligned} x_t &= \frac{1+a_1L}{1-c_1L}\epsilon_t \\ &= (1+a_1L)\left(1+c_1L+(c_1L)^2+...\right)\epsilon_t \\ &= \left[1+(a_1+c_1)L+c_1(a_1+c_1)L^2+c_1^2(a_1+c_1)L^3+...\right]\epsilon_t. \end{aligned}$$

Then,

$$\begin{aligned} Var(x_t) &= \left[1+(a_1+c_1)^2+c_1^2(a_1+c_1)^2+...\right]\sigma_\epsilon^2 \\ &= \left[1+\frac{(a_1+c_1)^2}{1-c_1^2}\right]\sigma_\epsilon^2, \\ Cov(x_t,x_{t-1}) &= \left[(a_1+c_1)+c_1(a_1+c_1)+c_1^2(a_1+c_1)+...\right]\sigma_\epsilon^2 \\ &= \left[(a_1+c_1)+\frac{c_1(a_1+c_1)^2}{1-c_1^2}\right]\sigma_\epsilon^2. \end{aligned}$$

Hence,

$$\begin{aligned} \rho(1) &= \frac{Cov(x_t,x_{t-1})}{Var(x_t)} \\ &= \frac{(1+a_1c_1)(a_1+c_1)}{1+c_1^2+2a_1c_1}. \end{aligned}$$

Successive values for $\rho(j)$ are obtained from the recurrent relation $\rho(j) = c_1\rho(j-1)$ for $j \geq 2$.

3 Persistence: Monte-Carlo experiment

Persistence of time-series destroys one of the crucial properties for implementing valid estimation and inference in the linear model. We have already seen that in the context of the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

The following property is required to implement valid estimation and inference

$$E(\boldsymbol{\epsilon} \mid \mathbf{X}) = \mathbf{0}. \tag{4}$$

Hypothesis (4) implies that

$$E(\boldsymbol{\epsilon}_i \mid \mathbf{x}_1, ...\mathbf{x}_i, ..., \mathbf{x}_n) = 0, \quad (i = 1, ..., n).$$

Think of the simplest time-series model for a generic variable y:

$$y_t = a_0 + a_1y_{t-1} + \epsilon_t.$$

Clearly, if $a_1 \neq 0$, then, although it is still true that $E(\epsilon_t \mid y_{t-1}) = 0$, $E(\epsilon_{t-1} \mid y_{t-1}) \neq 0$ and (4) breaks down.

How serious is the problem? To assess intuitively the consequences of persistence, we construct a small Monte-Carlo simulation on the short sample properties of the OLS estimator of the parameters in an AR(1) process.

A Monte-Carlo simulation is based on the generation of a sample from a known data generating process (DGP). First we generate a set of random numbers from a given distribution (here a normally independent white-noise disturbance) for a sample size of interest (200 observations) and then construct the process of interest (in our case, an AR(1) process). When a sample of observations on the process of interest is available, then we can estimate the relevant parameters and compare their fitted values with the known true value. For this reason the Monte-Carlo simulation is a sort of controlled experiment. The only potential problem with this procedure is that the set of random numbers drawn is just one possible outcome and the estimates are dependent on the sequence of simulated white-noise residuals. To overcome this problem in a Monte-Carlo study, the DGP is replicated many times. For each replication we obtain a set of estimates, and compute averages across replications of the estimated parameters, to assess these averages against the known true values.

We report the averages across replications in Figure 1.

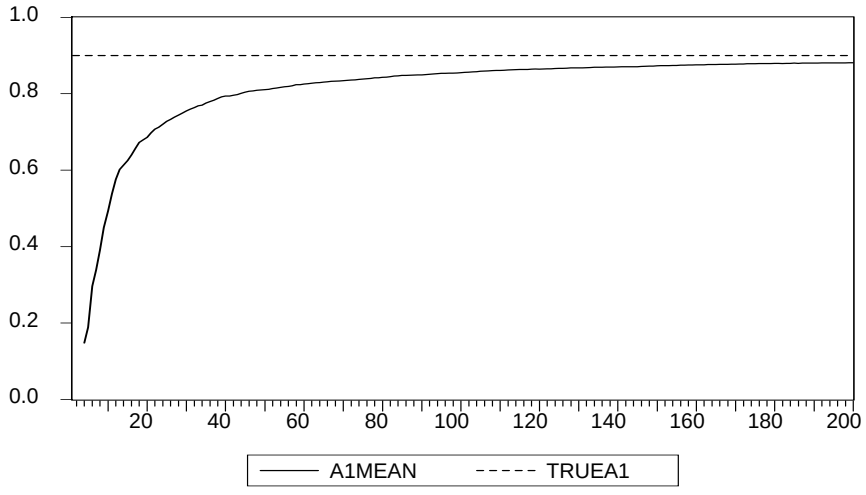


Figure 1: Small sample bias

From the figure we note that the estimate of a_1 is heavily biased

in small samples, but the bias decreases as the sample gets larger, and disappears eventually. One can show analytically that the average of the OLS estimate of a_1 is $a_1\left(1-\frac{2}{T}\right)$. This is an interesting result, which can be generalized. For stationary time-series, the correlation, which destroys the orthogonality between residuals and regressors in the linear regression model, tends to disappear as the distance between observations increases. Therefore, as we shall show in the next section, the finite sample results can be extended to time-series by considering large samples. Such an aim is obtained by introducing asymptotic theory.

4 Traditional solution: asymptotic theory

Stationary time-series feature time-independent distributions, as a consequence, the effect of any specific innovation disappears as time elapses. We show in this section that the intuition of the simple Monte-Carlo simulation can be extended and asymptotic theory can be used to perform valid estimation and inference when modelling *stationary* time-series.

4.1 Basic elements of asymptotic theory

In this section we introduce the elements of asymptotic theory necessary to illustrate how the results in estimation and inference for the linear model applied to cross-sectional data in Chapter 1 can be extended to time-series models.[1]

Consider a sequence $\{X_T\}$ of random variables with the associated sequence of distribution functions $\{F_T\} = F_1, ..., F_T$, we give the following definitions of convergence for X_T.

4.1.1 Convergence in distribution.

Given a random variable X with distribution function F, X_T converges in distribution to X if the following equality is satisfied:

$$\lim_{T\to\infty} P\left\{X_T < x_0\right\} = P\left\{X < x_0\right\},$$

for all x_0, where the function $F(x)$ is continuous.

4.1.2 Convergence in probability.

Given a random variable X with distribution function F, X_T converges in probability to X if, for each $\epsilon > 0$, the following relation holds:

$$\lim_{T\to\infty} P\left\{|X_T - X| < \epsilon\right\} = 1.$$

Note that convergence in probability implies convergence in distribution.

[1] For a formal treatment of these topics, see White (1984).

4.1.3 Central limit theorem (formulation of Lindeberg−Levy).

Given a sequence $\{X_T\}$ of identically and independently distributed random variables with mean μ and finite variance σ^2, define

$$\bar{X} = \frac{1}{T}\sum_{i=1}^{T} X_i,$$
$$\omega = \sqrt{T}\frac{\left(\bar{X} - \mu\right)}{\sigma}.$$

ω converges in distribution to a standard normal.

4.1.4 Slutsky's Theorem.

For any random variable X_T, such that $p\ \lim\ X_T = a$, where a is a constant, given a function $g(\cdot)$ continuous at a, $p\ \lim\ g\left(X_T\right) = g\left(a\right)$.

4.1.5 Cramer's Theorem.

Given two random variables X_T and Y_T, such that Y_T converges in distribution to Y and X_T converges in probability to a constant a, the following relationships hold:

$X_T + Y_T$ converges in distribution to $(a + Y)$;

Y_T/a_T converges in distribution to (Y/a);

$Y_T \cdot a_T$ converges in distribution to $(Y \cdot a)$.

Note that all theorems introduced so far extend to vectors of random variables.

4.1.6 Mann−Wald Theorem.

Consider a vector $\mathbf{z}_t$ $(k \times 1)$ of random variables which satisfies the following property:

$$p\ \lim\ T^{-1}\sum_{t=1}^{T}\mathbf{z}_t\mathbf{z}_t' = \mathbf{Q},$$

where $\mathbf{Q}$ is a positive definite matrix. Consider also a sequence ϵ_t of random variables, identically and independently distributed with zero mean and finite variance, σ^2, for which finite moments of each order are defined. If $E\left(\mathbf{z}_t\epsilon_t\right) = 0$, then

$$p\ \lim\ T^{-1}\sum_{t=1}^{T}\mathbf{z}_t\epsilon_t = \mathbf{0}, \sqrt{\frac{1}{T}}\sum_{t=1}^{T}\mathbf{z}_t\epsilon_t \xrightarrow{d} N\left(0, \sigma^2\mathbf{Q}\right).$$

4.2 Application to models for stationary time-series

Consider the following time-series model:

$$y_t = \gamma y_{t-1} + \epsilon_t,$$

where y_t is a stationary variable and $|\gamma| < 1$. As already shown, $E(y_t \epsilon_{t-i}) \neq 0$ and the OLS estimator of γ is biased.

By applying the Mann−Wald result, we can derive the asymptotic distribution of the OLS estimator of γ, $\widehat{\gamma}$:

$$\widehat{\gamma} \xrightarrow{d} N\left[\gamma, \sigma^2 \mathbf{Q}^{-1}\right],$$

and all the finite sample results available for cross-section can be extended to stationary time-series just by considering large-sample theory.

5 Estimation of ARMA models. The Maximum Likelihood Method

The Maximum Likelihood Method

- The likelihood function is the joint probability distribution of the data, treated as a function of the unknown coefficients
- The maximum likelihood estimator (MLE) consists of value of the coefficients that maximize the likelihood function
- The MLE selects the value of parameters to maximize the probability of drawing the data that have been effectively observed

5.1 MLE of an MA process

MLE of an MA process

Consider an MA process for a return r_{t+1} :

$$r_{t+1} = \theta_0 + \varepsilon_{t+1} + \theta_1 \varepsilon_t$$

The time series of the residuals can be computed as

$$\begin{aligned} \varepsilon_{t+1} &= r_{t+1} - \theta_0 - \theta_1 \varepsilon_t \\ \varepsilon_0 &= 0 \end{aligned}$$

If ε_{t+1} is normally distributed, than we have

$$f\left(\varepsilon_{t+1}\right)=\frac{1}{\left(2\pi\sigma_{\varepsilon}^{2}\right)^{1/2}}\exp\left(-\frac{\varepsilon_{t+1}^{2}}{2\sigma_{\varepsilon}^{2}}\right)$$

MLE of an MA process

If the $\varepsilon_{t+1}s$ are independent over time the likelihood function can be written as follows

$$\begin{aligned}f\left(\varepsilon_{1},\varepsilon_{2},...\varepsilon_{t+1}\right) &= \prod_{i=1}^{T} f\left(\varepsilon_{i}\right)\\ &= \prod_{i=1}^{T}\frac{1}{\left(2\pi\sigma_{\varepsilon}^{2}\right)^{1/2}}\exp\left(-\frac{\varepsilon_{i}^{2}}{2\sigma_{\varepsilon}^{2}}\right)\end{aligned}$$

The MLE chooses $\theta_0, \theta_1, \sigma_{\varepsilon}^{2}$ to maximize the probability that the estimated model has generated the observed data-set. The optimum is not always found analically, iterative search is the standard method.

5.2 MLE of an AR process

Before considering solutions to the problems generated by non-stationarity it is instructive to illustrate such problems from a different perspective. Consider a vector $\mathbf{x}_t$ containing observations on time-series variables at time t. A sample of T time-series observations on all the variables is represented as:

$$\mathbf{X}_{T}^{1}=\begin{bmatrix}\mathbf{x}_{1}\\ .\\ .\\ .\\ \mathbf{x}_{T}\end{bmatrix}.$$

In general, estimation is performed by considering the joint sample density function, known also as the likelihood function, which can be expressed as $D\left(\mathbf{X}_{T}^{1}\mid\mathbf{X}_{0},\boldsymbol{\theta}\right)$. The likelihood function is defined on the parameter space Θ, given the observation of the observed sample $\mathbf{X}_{T}^{1}$ and of a set of initial conditions $\mathbf{X}_{0}$. One can interpret such initial conditions as the pre-sample observations on the relevant variables (which are usually unavailable). In case of independent observations the likelihood function can be written as the product of the density functions for each observation. However, this is not the relevant case for time-series, as time-series observations are in general sequentially correlated. In the case of time-series, the sample density is constructed using the concept of sequential conditioning. The likelihood function, conditioned with

respect to initial conditions, can always be written as the product of a marginal density and a conditional density:

$$D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = D\left(\mathbf{x}_1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) D\left(\mathbf{X}_T^2 \mid \mathbf{X}_1, \boldsymbol{\theta}\right).$$

Obviously,

$$D\left(\mathbf{X}_T^2 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = D\left(\mathbf{x}_2 \mid \mathbf{X}_1, \boldsymbol{\theta}\right) D\left(\mathbf{X}_T^3 \mid \mathbf{X}_2, \boldsymbol{\theta}\right),$$

and, by recursive substitution:

$$D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = \prod_{t=1}^{T} D\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right).$$

Having obtained $D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right)$, we can in theory derive $D\left(\mathbf{X}_T^1, \boldsymbol{\theta}\right)$ by integrating with respect to X_0 the density conditional on pre-sample observations. In practice this could be intractable analytically, as $D\left(X_0\right)$ is not known. The hypothesis of stationarity becomes crucial at this stage, as stationarity restricts the memory of time-series and limits the effects of pre-sample observations to the first observations in the sample. This is why, in the case of stationary processes, one can simply ignore initial conditions. Clearly, the larger the sample, the better, as the weight of lost information becomes smaller. Moreover, note that even by omitting initial conditions, we have:

$$D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = D\left(\mathbf{x}_1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) \prod_{t=2}^{T} D\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right).$$

Therefore, the likelihood function is separated in the product on $T-1$ conditional distributions and one unconditional distribution. In the case of non-stationarity, the unconditional distribution is undefined. On the other hand, in the case of stationarity, the DGP is completely described by the conditional density function $D\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right)$.

5.2.1 The first-order autoregressive process.

To give more empirical content to our case, let us consider again the case of the univariate first-order autoregressive process,

$$x_t \mid \mathbf{X}_{t-1} \sim N\left(\lambda x_{t-1}, \sigma^2\right), \tag{5}$$

$$D\left(\mathbf{X}_T^1 \mid \lambda, \sigma^2\right) = D\left(x_1 \mid \lambda, \sigma^2\right) \prod_{t=2}^{T} D\left(x_t \mid \mathbf{X}_{t-1}, \lambda, \sigma^2\right). \tag{6}$$

From (6), the likelihood function clearly involves $T-1$ conditional densities and one unconditional density. The conditional densities are

given by (5), the unconditional density can be derived only in the case of stationarity:

$$\begin{aligned} x_t &= \lambda x_{t-1} + u_t, \\ u_t &\sim N.I.D\left(0, \sigma^2\right). \end{aligned}$$

We can obtain by recursive substitution:

$$x_t = u_t + \lambda u_{t-1} + ... + \lambda^{n-1} u_1 + \lambda^n x_0.$$

Only if $|\lambda| < 1$, the effect of the initial condition disappears and we can write the unconditional density of x_t as:

$$D\left(x_t \mid \lambda, \sigma^2\right) = N\left(0, \frac{\sigma^2}{1-\lambda^2}\right).$$

Under stationarity we can derive the exact likelihood function:

$$\begin{aligned} D\left(\mathbf{X}_T^1 \mid \lambda, \sigma^2\right) = (2\pi)^{-\frac{T}{2}} \sigma^{-T} \left(1-\lambda^2\right)^{\frac{1}{2}} \\ \exp\left[-\frac{1}{2\sigma^2}\left(\left(1-\lambda^2\right) x_1^2 + \sum_{t=2}^{T} \left(x_t - \lambda x_{t-1}\right)^2\right)\right], \end{aligned} \tag{7}$$

and estimates of the parameters of interest are derived by maximizing this function. Note that $\widehat{\lambda}$ cannot be derived analytically, using the exact likelihood function; but it requires conditioning the likelihood and operating a grid search. Note also that, in large samples, using the approximate likelihood function by dropping the first observation works only under the hypothesis of stationarity. When the first observation is dropped and the approximate likelihood function is considered, one can show that the Maximum Likelihood (ML) estimate of λ coincides with the OLS estimate.

6 Putting ARMA models at work

There are three main steps in the Box-Jenkins approach:

- PRE WHITENING: make sure that the time series is stationary, in other word make sure that the model at hand is ARMA and not ARIMA. This is acheived in univariate time-series via differencing, as we shall see there are alternative, and more interesting, ways of achieving stationarity in multivariate time-series analysis.

- MODEL SELECTION: the aim of model selection is to look for the best ARMA specification, this is not an easy task as different

specification perform very closely. **Information criteria** are a useful tool to this end. They are model selection criteria based on the log-likelihood function and the can be used to select p and q in an ARMA model. The Akaike's information criteria (**AIC**) and the .Schwarz Bayesian Criterion (**SBC)** are the most commonly used criteria. they are defined as follows:

$$AIC = -2\log(L) + 2(p+q)$$
$$SBC = -2\log(L) + \log(n)(p+q)$$

where p and q are the length of the AR and MA polynomials and n is the number of observations. The model are ranked from the best to the worst in such a way that best model is the one that minimizes the chosen criterion. Note that the criterion is minimized when the likelihood function penalized for the lack of parsimony (i.e. over-parameterization) is maximized

- MODEL CHECKING: residual tests. Make sure that residuals are not autocorrelated and check whether their distribution is normal, also ex-post evaluation technique based on RMSE and MAE are implemented (Diebold, Giacomini-White).

Add here an application to illustrate how the procedure can be implemented.

7 Non Stationarity

7.1 Deterministic and stochastic trends

Figure **??** at the beginning of this chapter shows that macroeconomic time-series, besides being persistent, often feature upward trends. Non-stationarity of time-series is a possible manifestation of a trend. Consider, for example, the random walk process with a drift:

$$x_t = a_0 + x_{t-1} + \epsilon_t,$$
$$\epsilon_t \sim n.i.d.\left(0, \sigma_\epsilon^2\right).$$

Recursive substitution yields

$$x_t = x_0 + a_0 t + \sum_{i=0}^{t-1} \epsilon_{t-i}, \tag{8}$$

which shows that the non-stationary series contains both a deterministic ($a_0 t$) and a stochastic $\left(\sum_{i=0}^{t-1} \epsilon_{t-i}\right)$ trend.

An easy way to make a non-stationary series stationary is differencing:

$$\Delta x_t = x_t - x_{t-1} = (1 - L)\, x_t = a_0 + \epsilon_t.$$

In general, if a time-series needs to be differenced d times to become stationary, then it is integrated of order d or I(d). Our random walk is I(1). When the d-th difference of a time-series x, $\Delta^d x_t$, can be represented by an ARMA(p, q) model, we say that x_t is an integrated moving-average process of order p, d, q and denote it as ARIMA(p, d, q).

Compare the behaviour of an integrated process with that of a trend stationary process. Trend stationary processes feature only a deterministic trend:

$$z_t = \alpha + \beta t + \epsilon_t. \tag{9}$$

The z_t process is non-stationary, but the non-stationarity is removed simply by regressing z_t on the deterministic trend. Unlike this, for integrated processes like (8) the removal of the deterministic trend does not deliver a stationary time-series. Deterministic trends have no memory while integrated variables have an infinite one. Both integrated variable and deterministic trend exhibit systematic variations, but in the latter case the variation is predictable, whereas in the other one it is not. This point is easily seen in Figure 2, where we report three series for a sample of 200 observations. The series are generated in MATLAB by running the following program:

```
smpl 1 1
genr ST1=0
genr ST2=0
smpl 2 200
series ST1=0.1+ST1(-1)+nrnd
series ST2=0.1+ST2(-1)+nrnd
series DT=0.1*@trend+nrnd
```

We have a deterministic trend (DT) generated by simulating equation (9) with $\alpha = 0, \beta = 0.1$, and a white noise independently distributed as a standard normal (`nrnd`), and two integrated series (ST1 and ST2), which are random walks with a drift of 0.1. The only difference between ST1 and ST2 is in the realizations from the error terms, which are different drawings from the same serially independent standard normal distribution.

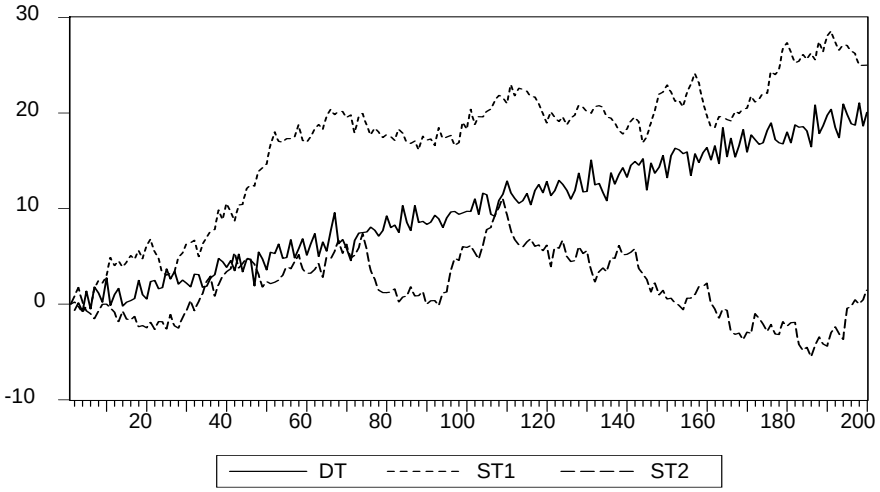


Figure 2: Deterministic (DT) and stochastic (ST1 and ST2) trends

7.2 Univariate decompositions of time-series

The general solution proposed to the problem introduced in the previous section is the search for a stationary representation of non-stationary time-series. This has been done both in univariate and multivariate frameworks. We discuss briefly methodologies used in a univariate framework, to move swiftly to decompositions in a multivariate framework, which are at the heart of our discussion of modern macroeconometrics.

Beveridge and Nelson (1981) provide an elegant way of decomposing a non-stationary time-series into a permanent and a temporary (cyclical) component by applying ARIMA methods. For any non-stationary time-series x_t integrated of the first order, the Wold decomposition theorem could be applied to its first difference, to deliver the following representation:

$$\begin{aligned} \Delta x_t &= \mu + C(L)\,\epsilon_t, \\ \epsilon_t &\sim n.i.d.\left(0, \sigma_\epsilon^2\right), \end{aligned}$$

where $C(L)$ is a polynomial of order q in the lag operator. Consider now the polynomial $D(L)$, defined as:

$$D(L) = C(L) - C(1). \tag{10}$$

Given that $C(1)$ is a constant, also $D(L)$ will be of order q. Clearly,

$$D(1) = 0,$$

therefore, 1 is a root of $D(L)$, and

$$D(L) = C^*(L)(1-L), \tag{11}$$

where $C^*(L)$ is a polynomial of order $q-1$.

By equating (10) to (11), we have:

$$C(L) = C^*(L)(1-L) + C(1),$$

and

$$\Delta x_t = \mu + C^*(L)\Delta\epsilon_t + C(1)\epsilon_t. \tag{12}$$

By integrating (12), we finally have:

$$\begin{aligned} x_t &= C^*(L)\epsilon_t + \mu t + C(1) z_t \\ &= C_t + TR_t, \end{aligned}$$

where z_t is a process for which $\Delta z_t = \epsilon_t$. C_t is the cyclical component and TR_t is the trend component made of a deterministic and a stochastic trend. Note that the trend component can be represented as:

$$TR_t = TR_{t-1} + \mu + C(1)\epsilon_t.$$

7.2.1 Beveridge–Nelson decomposition of an IMA(1,1) process

Consider the process:

$$\Delta x_t = \epsilon_t + \theta\epsilon_{t-1}, \quad 0 < \theta < 1.$$

In this case:

$$\begin{aligned} C(L) &= 1 + \theta L, \\ C(1) &= 1 + \theta, \\ C^*(L) &= \frac{C(L) - C(1)}{1-L} = -\theta. \end{aligned}$$

The Beveridge and Nelson decomposition gives the following result:

$$\begin{aligned} x_t &= C_t + TR_t \\ &= -\theta\epsilon_t + (1+\theta) z_t. \end{aligned}$$

7.2.2 Beveridge–Nelson decomposition of an ARIMA(1,1) process

Consider the process:

$$\Delta x_t = \rho\Delta x_{t-1} + \epsilon_t + \theta\epsilon_{t-1}.$$

Here:

$$C(L) = \frac{1+\theta L}{1-\rho L},$$
$$C(1) = \frac{1+\theta}{1-\rho},$$
$$C^*(L) = \frac{C(L) - C(1)}{1-L} = -\frac{\theta+\rho}{(1-\rho)(1-\rho L)},$$

and the Beveridge and Nelson decomposition yields:

$$x_t = C_t + TR_t = -\frac{\theta+\rho}{(1-\rho)(1-\rho L)}\epsilon_t + \frac{1+\theta}{1-\rho}z_t.$$

7.2.3 Deriving the Beveridge–Nelson decomposition in practice

The practical derivation of a Beveridge and Nelson decomposition for any ARIMA process is easily derived by applying a methodology suggested by Cuddington and Winters (1987). For any I(1) process, the stochastic trend can be represented as:

$$TR_t = TR_{t-1} + \mu + C(1)\epsilon_t. \tag{13}$$

The decomposition can then be applied in the following steps:

1. identify the appropriate ARIMA model and estimate ϵ_t and all the parameters in μ and $C(1)$;
2. given an initial value for TR_0, use (13) to generate the permanent component of the time-series;
3. generate the cyclical component as the difference between the observed value in each period and the permanent component.

The above procedure gives the permanent component up to a constant. If the precision of this procedure is unsatisfactory, one can use further conditions to identify the decomposition more precisely. For example, one can impose the condition that the sample mean of the cyclical component is zero, to pin down the constant in the permanent component.

To illustrate how the procedure works in practice, we have simulated an ARIMA(1,1,1) model in E-Views for a sample of 200 observations, by running the following program:

```
smpl 1 2
```

```
genr x=0
smpl 1 200
genr u=nrnd
smpl 3 200
series x=x(-1)+0.6*x(-1)-0.6*x(-2)+u+0.5*u(-1)
```

From the previous section we know the exact Beveridge and Nelson decomposition of our x_t:

$$\begin{aligned} x_t &= C_t + TR_t \\ &= -\frac{1.1}{(1-0.6)(1-0.6L)}\epsilon_t + \frac{1.5}{0.4}z_t, \\ TR_t &= TR_{t-1} + \frac{1.5}{0.4}\epsilon_t. \end{aligned}$$

We can, therefore, generate the permanent component of X and the transitory component as follows:

```
smpl 1 2
genr p=0
smpl 3 200
series TR= TR(-1)+(1.5/0.4)*u
genr CYCLE=X-TR
```

Figure 3 reports the series X, TR and CYCLE.

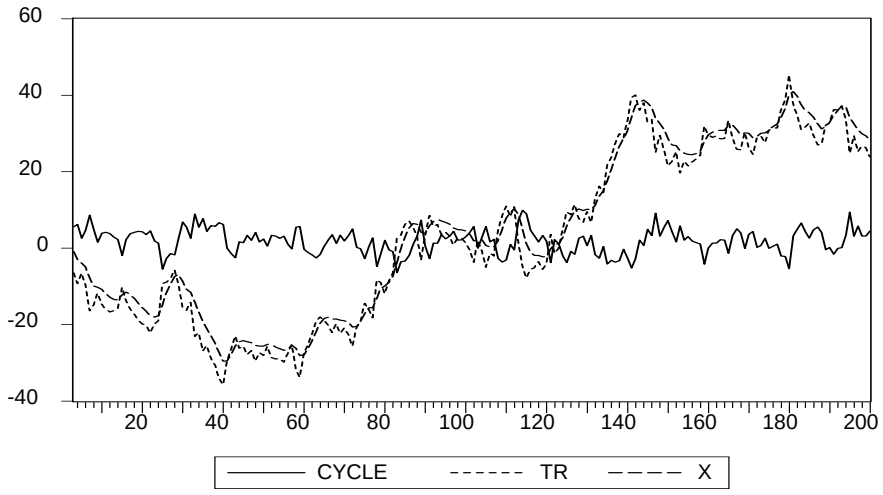


Figure 3: A Beveridge−Nelson decomposition of an ARIMA(1,1,1) process

This is exactly the procedure that we follow in practice, except that we estimate parameters rather than impute them from a known DGP.

7.2.4 Assessing the Beveridge−Nelson decomposition

The properties of the permanent and temporary components of an integrated time-series delivered by the Beveridge−Nelson decomposition are worth some comments. The innovations in the permanent and the transitory components are perfectly negatively correlated; moreover, the trend component is more volatile than the actual time-series as the negative correlation between the permanent and the transitory components acts to smooth the original time-series. These results are easily seen for the simplest case we have already discussed. For example, in the case of the IMA(1,1) process, the correlation between the innovations in the permanent and transitory components is minus one and the variance of the innovation in the trend component is $(1.5/0.4)^2 \sigma_\epsilon^2 > \sigma_\epsilon^2$. Note that in general the variance of innovations might have an economic interpretation and economic theory might suggest different patterns of correlations between innovations from a perfectly negative correlation. As we shall see in one of the next chapters, an interesting pattern is the absence of correlation between the innovations in the cycle and the trend components of an integrated time-series. In general, different restrictions on the correlation between the trend and the cycle components lead to the identification of different stochastic trends for integrated time-series. As a consequence, the Beveridge−Nelson decomposition is not unique. In general, neither are all univariate decompositions. To see this point more explicitly we can compare the Beveridge−Nelson trend with the trend extracted using an alternative technique which has been recently very successful in time-series analysis: the Hodrick−Prescott filter.

Hodrick and Prescott proposed their method to analyse postwar US business cycles in a working paper circulated in the early 1980s and published in 1997. The Hodrick−Prescott (HP) filter computes the permanent component TR_t of a series x_t by minimizing the variance of x_t around TR_t, subject to a penalty that constrains the second difference of TR_t. That is, the Hodrick−Prescott filter is derived by minimizing the following expression:

$$\sum_{t=1}^{T} (x_t - TR_t)^2 + \lambda \sum_{t=2}^{T-1} \left[(TR_{t+1} - TR_t)^2 - (TR_t - TR_{t-1})^2 \right].$$

The penalty parameter λ controls the smoothness of the series, by controlling the ratio of the variance of the cyclical component and the variance of the series. The larger the λ, the smoother the TR_t approaches a linear trend. In practical applications λ is set to 100 for annual data, 1600 for quarterly data and 14400 for monthly data.

In Figure 4 we report the Beveridge−Nelson trend and the Hodrick−Prescott trend (with $\lambda = 100$) for the data generated in the previous section.

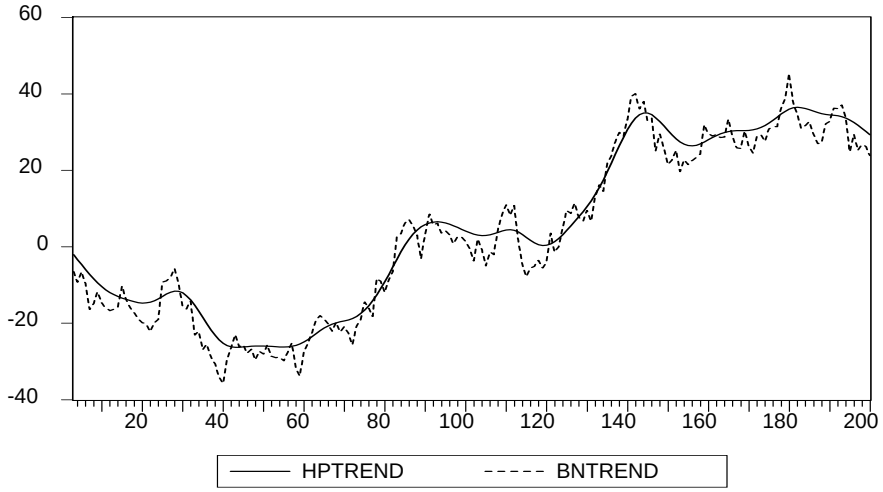


Figure 4: Trend components: Hodrick−Prescott versus Beveridge−Nelson

Note that the Beveridge−Nelson trend is more volatile than the Hodrick−Prescott trend. It is possible to increase the volatility of the Hodrick−Prescott trend by reducing the parameter λ; however, the Hodrick−Prescott filter reaches at most the volatility of the actual time-series, which, as we already know, is smaller than the volatility of the Beveridge−Nelson trend.

The Hodrick−Prescott filter has the advantage of removing the same trend from all time-series; this might be desirable as some theoretical models, as, for example, real business cycle models, which indicate that macroeconomic variables share the same stochastic trend. However, Harvey and Jaeger (1993) showed that the use of such a filter can lead to the identification of spurious cyclical behaviour. The authors predicate a different approach to modelling time-series, known as structural time-series modelling, which we do not consider in our analysis, as it is less closely related to theoretical macroeconomic models, but certainly merits some attention (Harvey and Koopman 1996; Maravall 1995).

The comparison between the Hodrick−Prescott and the Beveridge−Nelson trend reinforces the argument of non-uniqueness of univariate decomposition made before. Moreover, we are left with the problem of how to use the filtered series in applied macroeconometrics and how to relate them to theoretical models. The empirical counterparts of theoretical macroeconomic models are multivariate time-series. Theoretical models often predict that different time-series share the same stochastic trend.

The natural question at this point is if the problem of non-stationarity in time-series can be resolved by considering multivariate models. In this context, stationarity is obtained by considering combinations of non-stationary time-series sharing the same stochastic trend. If possible, it would justify the identification of trends by relating them to macroeconomic theory.

8 A Solved Exercise on univariate time-series models

Question 1

1. Consider the following three AR(1) models . Which one might be a reasonable model for daily stock market prices?
   (a) $y_t = y_{t-1} + e_t$
   (b) $y_t = 0.8y_{t-1} + u_t$
   (c) $y_t = 0.2y_{t-1} + v_t$
2. How would you derive the autocorrelation function of these models?
3. Derive the MA representation for the generic AR(1) model $y_t = \alpha y_{t-1} + e_t$.

Question 2

1. Which are the main steps of the Box-Jenkins approach to ARMA modeling?
2. Write the unconditional moments (mean and variance) of the generic AR(1) process: $y_t = \alpha_0 + \alpha_1 y_{t-1} + e_t$
3. Is it possible to estimate a MA(1) model using OLS?

Question 3 (See also Exercise 11 Ch. 5 Brooks)
You observe the following correlograms for four time series. Suggest the most appropriate ARMA specification for each series.(Hint: All charts refer either to an AR or a MA model)

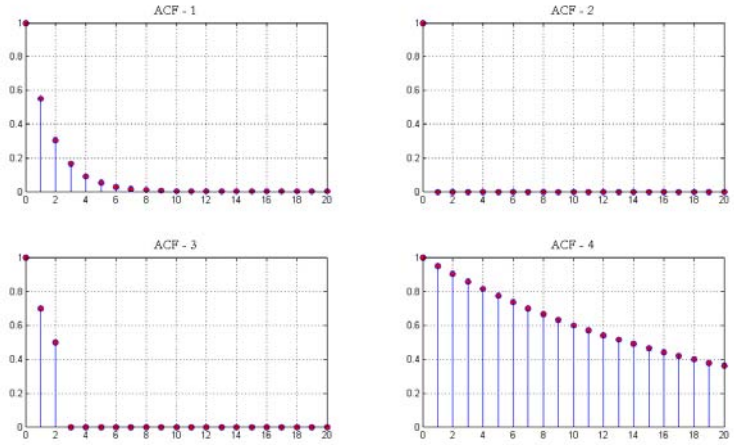


Figure 5:

8.1 Solutions

Question 1

1. The correct answer is (a). In fact, the process generating daily prices is usually assumed to be a Random Walk.

2. The autocovariance function for a generic AR(1) up to lag k can be easily derived as (there is no intercept since we consider de-meaned time series):

$$cov(y_t, y_{t-k}) = E\left[y_t y_{t-k}\right] = E\left[\sum_{i=0}^{\infty}\alpha^i e_{t-i}\sum_{i=0}^{\infty}\alpha^i e_{t-k-i}\right] = \alpha^k\sum_{i=0}^{\infty}\alpha^{2i}e_i^2 = \alpha^k\frac{\sigma_e^2}{1-\alpha^2} = \alpha^k\sigma_y^2$$

Where σ_y^2 is the unconditional variance of y and σ_e^2 is the variance of residuals (see also exercise 2).The autocorrelation function is:

$$corr(y_t, y_{t-k}) = \frac{cov[y_t, y_{t-k}]}{\sigma^2} = \frac{\alpha^k\sigma_y^2}{\sigma^2} = \alpha^k$$

3. The equation for a generic AR(1) model is (there is no intercept since we consider de-meaned time series):

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

By an easy substitution:

$$y_t = \alpha(\alpha y_{t-2} + e_{t-1}) + e_t$$

Thus:

$$y_t = \alpha^2 y_{t-2} + \alpha e_{t-1} + e_t$$

If you iterate the substitution you get to:

$$y_t = \alpha^n y_{t-n} + \sum_{i=0:n-1} \alpha^i e_{t-i}$$

For a stationary process (see also exercise 5) $\alpha < 1$. Thus for $n \longrightarrow \infty$:

$$y_t = \sum_{i=0}^{\infty} \alpha^i e_{t-i}$$

Which is the infinite MA representation.

Question 2

1. There are three main steps in the Box-Jenkins approach:

- PRE WHITENING: make sure that the time series is stationary. If it is not, transform it in order to make it stationary (usually taking first differences).
- MODEL SELECTION: look for the best ARMA specification; **information criteria** are a useful tool in this part.
- MODEL CHECKING: residual tests. Make sure that residuals are not autocorrelated and check whether their distribution is normal.

2.Consider the AR(1) stationary process:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + e_t$$

With residuals volatility equal to σ.

The unconditional mean coincides with the long run expected value of the time series, which is:

$$E_t\left[y_{t+n}\right] = \alpha_0 \sum_{i=i}^{n} \alpha^{i-1} + \alpha_1^i y_{t-i}$$

For $n \longrightarrow \infty$, since $\alpha_1 < 1$:

$$E[y] = \frac{\alpha_0}{1-\alpha_1}$$

The unconditional variance coincides with the long run expected value of the time series, which is:

$$V_t[y_{t+n}] = \sum_{i=i}^{n} \alpha^{2(i-1)} \sigma_e^2$$

For $n \longrightarrow \infty$, since $\alpha_1 < 1$:

$$\sigma_y^2 = \frac{\sigma_e^2}{1-\alpha_1^2}$$

3. The answer is no. The least square problem for equation:

$$Y = \beta X + e$$

Is the solution to the linear system:

$$X'X\hat{\beta} = XY$$

This solution is given by the vector of coefficients $\hat{\beta}$ which minimizes the sum of squared errors. The AR(1) model:

$$y_{t+1} = \alpha_0 + \alpha_1 y_t + e_{t+1}$$

Is approximately compatible with the OLS assumptions,as we can write

$$y_{t+1} \backsim f(y_t|\sigma_e^2, \alpha_0, \alpha_1)$$
$$y_{t:t+n} \backsim \prod_{i=0:n-1} f(y_{t+i}|\sigma_e^2, \alpha_0, \alpha_1)$$

and discard the distribution of the first observation in large samples.

It can be shown that the problem of fitting this distribution with the most appropriate values for $\sigma_e^2, \alpha_0, \alpha_1$ (which is the Maximum Likelihood Estimation problem) coincides with the OLS problem.

However, an MA(1) model entails some form of dependency; as a matter of fact:

$$y_{t+1} \backsim f(e_t|\sigma^2) = f(y_t|y_{t-1}, \sigma_e^2, \alpha_0, \alpha_1)$$

And then

$$y_{t:t+n} \backsim f(y_{t+n}|y_{t+n-1,}\sigma_e^2, \alpha_0, \alpha_1) f(y_{t+n-1}|y_{t+n-1,}\sigma_e^2, \alpha_0, \alpha_1) ... f(y_{t+1}|y_{t,}\sigma_e^2, \alpha_0, \alpha_1)$$

This more complicated form cannot be solved as an OLS problem and requires Maximum Likelihood Estimation.

Question 3

In graphical representations of ACF, we can say in general that AR process show autocorrelations that decline smoothly along with lags, while in MA models autocorrelations different from zero are observed only in correspondence of lagged terms significantly different from zero.

- Time Series 1: this is clearly an AR process, and probably an AR(1) process with a root positive but rather small in absolute value
- Time Series 2: well....there is no autocorrelation in this process; we are observing the ACF of a **white noise**
- Time Series 3: only the first two lags show an autocorrelation different from zero, then the ACF falls to zero. This is a MA(2) model.
- Time Series 4: Thus is again an AR process, but this time we are considering a very persistent process. For instance, it might be an AR(1) with a root very close to one.

Forecasting a stationary process: an application

9 A MATLAB Exercise on Univariate Time-Series Models

Preparation Steps

Step 1 **(Data loading)** Import in MATLAB all the data contained in the file STOCKINT2011.XLS and described in the lecture notes.

Step 2 **(Data transformation)** Let the stock market indices be denoted as P_{1t} for the **US**, P_{2t} for **UK**, P_{3t} for the **Germany** and similarly for the dividend yields as DP_{1t}, DP_{2t} and DP_{3t}. For each country

create 1-quarter, 4-quarter, 8-quarter, 12-quarter annualized stock market total returns (hint: log return series as Δp_{it} = ln P_{it} - ln P_{it-1}). Construct the dividend series for each country as D_{it} = P_{it} * (DP_{it}/100), dividend growth series as Δd_{it} = lnD_{it} - lnD_{it-1} and log dividend yield series as $d_{it} - p_{it} = \ln(DP_{it}/100)$. Plot these series and carry out a cross-country comparison for each investment horizon. Then consider the following three models over the sample period from 1976Q1 to 2010Q4.

Basic Forecasting Models in Financial Econometrics (Ex. 1, Ex. 1, and Ex. 3: estimation and with-in-sample prediction)

Ex 1. **(AR(*1*) Model for log Return)**

Consider the following AR(1) model of log returns for each of the countries:

$$\Delta p_{it} = \phi_{0i} + \phi_{1i}\Delta p_{it-1} + \varepsilon_{p,it} \tag{14}$$

Estimate the parameter vector $\gamma_i = (\mu_i, \phi_i)\prime$ for $i = 1, 2, 3$ via OLS in MATLAB. Compute the corresponding t-statistics and $R.^2$ Plot your results. Are stock returns predictable from past returns (i.e. is ϕ_{1i} statistically significant)?

Ex. 2. **(MA(*1*) Model for Dividend Growth)**

Consider the following MA(1) dividend growth model for each of the countries:

$$\Delta d_{it} = \eta_{0i} + \eta_{1i}\varepsilon_{d,it-1} + \varepsilon_{d,it} \tag{15}$$

According to your results is dividend growth predictable from past dividend growth innovations?

Ex. 3. **(ARMA(*1,1*) Model for Dividend Growth)**

Consider the following ARMA(1,1) dividend growth model for each of the countries:

$$\Delta d_{it} = \mu_{0i} + \mu_{1i}\Delta d_{it-1} + \mu_{2i}\varepsilon_{d,it-1} + \varepsilon_{d,it} \tag{16}$$

Repeat the analysis outlined above and comment your results.

More Exercises

Ex. 4. **(Asset Allocation Strategies)**

Consider US data and four alternative portfolio allocation strategies in the period 1997Q1:2009Q4:

A. safe strategy;

B. buy and hold strategy;

C. econometric strategy (i.e. optimal ex ante);

D. ex-post strategy (i.e. optimal ex post) .

Forecast your returns at the end of each year. Which strategy would you choose to invest your money at 1997Q1?

Ex. 5. **(AR(*1*) Model for log Return)**

Consider the following AR(1) model of log returns for each of the countries the over the sample 1974:1 1999:4:

$$dp_{t+1}^i = \beta_0^i + \beta_1^i dp_t^i + \varepsilon_{t+1}^i \tag{17}$$

5.1 **(Model Estimate)** Estimate the parameter vector $\gamma_i = (\beta_0^i, \beta_1^i)\prime$ for $i = 1, 2, 3$ via OLS in MATLAB. Compute the corresponding t-statistics and R^2.

5.2 **(Out-of-Sample Forecasting)** Evaluate the forecasting performance of this *AR(1)* model when forecasting one-step ahead (next quarter dividends) over the sample 2000:1-2009:4.

- **5.2.1** First, you should produce 40 forecasts for each time point, as well as their associated 95 per cent confidence interval.

  5.2.2 Second, evaluate the forecasts comparing to the realized values, where you can do this via a graphical analysis.

  5.2.3 In the end, you have to rank the forecasting performance of this model for different countries.

Ex. 6. **(Predictability of Return from Dividend Yields)**

Define the k-period cumulative return from period $t+1$ through period $t+k$, as follows:

$$\mathbf{r}_{t,t+k} = \sum_{i=1}^{k} \mathbf{r}_{t+i}$$

Run a the following predictive regressions for UK, US and Germany

$$\mathbf{r}^i_{t,t+k} = \alpha^i_0 + \alpha^i_1 dp^i_t + \varepsilon^i_{t+k} \tag{18}$$

Where dp^i_t is the log dividend/price ratio for the country i. Make comparisons across countries (i.e. given the horizon) and within country (i.e. given the country). Are stock returns predictable from dividend yields?

10 References

Amisano, G., and Giannini, C. (1997). *Topics in Structural VAR Econometrics*. 2nd edn, New York: Springer Verlag.

Banerjee, A., Dolado, J. J., Hendry, D. F. and Smith, G. W. (1986). 'Exploring Equilibrium Relationships in Econometrics through Static Models: some Monte Carlo Evidence'. *Oxford Bulletin of Economics and Statistics*, August.

Banerjee, A., Dolado, J. J., Galbraith, J. W. and Hendry, D. F. (1993). *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data*. Oxford University Press.

Banerjee, A., and Hendry, D. F. (1992). 'Testing Integration and Cointegration'. *Oxford Bulletin of Economics and Statistics*, 54, Special Issue.

Beveridge, S., and Nelson, C. (1981). 'A New Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to the Measurement of Business Cycle'. *Journal of Monetary Economics*, 151-174.

Bhargava, A. (1986). 'On the Theory of Testing for Unit Roots in Observed Time Series'. *Review of Economic Studies*, 53: 369-384.

Cuddington, J. T., and Winters, L. A. (1987). 'The Beveridge-Nelson decomposition of economic time-series. A quick computational method'. *Journal of Monetary Economics*, 19: 125-127.

DeJong, D. N., and Whiteman, C. H. (1991). 'Reconsidering Trends and random walks in macroeconomic time-series'. *Journal of Monetary Economics*, 28: 221-254.

Dickey, D. A., and Fuller, W. A. (1981). 'Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root'. *Econometrica*, 49:1057-1072.

Doornik, J., and Hendry, D. F. (1994). *PcFiml 8.0. Interactive Econometric Modelling of Dynamic Systems*. London: International Thomson Publishing.

Enders, W. (1995). *Applied Econometric Time Series*. New York: Wiley Series in Probability and Mathematical Statistics.

Engle, R. F., and Granger, C. W. J. (1987). 'Cointegration and Error Correction: Representation, Estimation and Testing'. *Econometrica*, 55: 251-276.

Engle, R. F., Granger, C. W. J., and Hallman, J. J. (1987). 'Combining Short and Long Run Forecasts: an Application of Seasonal Cointegration to Monthly Electricity Sales Forecasting'. Unpublished paper, University of California San Diego.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*. New York: J. Wiley and Sons.

Giannini, C. (1992). 'Topics in Structural VAR Econometrics'. *Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag.

Giannini, C., Lanzarotti, S., and Seghelini, M. (1994). 'A Traditional Interpretation of Macroeconomi Fluctuations: the Case of Italy'. *European Journal of Political Economy*.

Gonzalo, J. (1994). 'Five Alternative Methods of Estimating Long-run Equilibrium Relationships'. *Journal of Econometrics*, 60: 203-233.

Granger, C. W. J. (1986). 'Developments in the Study of Cointegrated Economic Variables'. *Oxford Bulletin of Economic and Statistics*, 48: 213-228.

Granger, C. W. J., and Newbold, P. (1974). 'Spurious Regressions in Econometrics'. *Journal of Econometrics*, 2: 111-120.

Hamilton, J. (1994). *Time-Series Analysis*. Princeton: Princeton University Press.

Hansen, H., and Juselius, K. (1995). *CATS in RATS. Cointegration Analysis of Time Series*. Evanston: Estima.

Harvey, A. C., and Jaeger, A. (1993). 'Detrending, Stylized Facts and the Business Cycles'. *Journal of Applied Econometrics*.

Harvey, A. C., and Koopman, S. J.(1996) 'Multivariate Structural Time Series Models' in C. Heij et al. (eds.), *Systematic Dynamics in Economic and Financial Models.* New York: John Wiley.

Hatanaka, M. (1987). *Time-Series-based-econometrics*. Oxford: Oxford University Press.

Hendry, D. F. (1987) 'Econometric Modelling with Cointegrated Variables'. *Oxford Bulletin of Economic and Statistics*, 48: 201-212.

Hendry, D. F. (1995). *Dynamic Econometrics.* Oxford: Oxford University Press.

Hendry, D. F., and Ericsson, N. (1991). 'Modeling the demand for narrow money in the United Kingdom and the United States'. *European Economic Review*, 35: 833-886.

Hodrick, R. J., and Prescott, E. C. (1997). 'Postwar U.S. Business Cycles: An Empirical Investigation'. *Journal of Money, Credit, and Banking*, 29: 116.

Horvath, M. T. K., and Watson, M. W. (1995). 'Testing for cointegration when some of the cointegrating vectors are known'. *Econometric Theory*, 11: 984-1014.

Johansen, S. (1988). 'Statistical Analysis of Cointegration Vectors'. *Journal of Economics Dynamics and Control*, March.

Johansen, S. (1989). 'Likelihood-based Inference on Cointegration: Theory and Applications'. Mimeo, University of Copenhagen.

Johansen, S. (1992). *Identifying Restrictions of Linear Equations.* University of Copenaghen, Institute of Mathematical Statistics.

Johansen, S. (1994). 'The role of the Constant and Linear Terms in Cointegration Analysis of Nonstationary Variables'. *Econometric Reviews*, 13: 205-229.

Johansen, S. (1995). *Likelihood Based Inference on Cointegration in the Vector Autoregressive Model.* Oxford: Oxford University Press.

Johansen, S., and Bartlett, A. (1999). 'Correction Factor for Tests on the Cointegrating Relationships'. Available at

http://www.iue.it/Personal/Johansen/Welcome.html.

Johansen, S., and Juselius, K. (1990). 'Maximum Likelihood Estimation and Inference on Cointegration with Applications to the Demand for Money'. *Oxford Bulletin of Economics and Statistics*, May.

Johansen, S., and Nielsen, B. G. (1993). 'Asymptotics for the cointegration rank test tests in the presence of intervantion dummies'. *Manual for the simulation program DisCo.* Manual and program are available at the URL http://www.nuff.ox.ac.uk/users/nielsen/disco.html.

King, R. G., Plosser, C. I., Stock, J. H., and Watson, M. W. (1991). 'Stochastic trends and economic fluctuations'. *American Economic Review*, 81, 4: 820-840.

Kiwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). 'Testing the Null Hypothesis of Stationarity against the alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?'. *Journal of Econometrics*, 54: 159-178.

Leeper, E. M., Sims, C. A., and Zha, T. (1996). 'What does monetary policy do?'. *Brookings Paper on Economic Activity*, 2: 1-79.

Maddala, G. S., and Kim, I. M. (1998). *Unit roots, Cointegration and Structural Change*. Cambridge: Cambridge University Press.

Maravall, A. (1995) 'Unobserved Components in Economic Time Series' in M. H. Pesaran and M. Wickens (eds.), *Handbook of Applied Econometrics. Macroeconomics*, Oxford: Basil Blackwell.

Mellander, E., Vredin, A., and Warne, A. (1992). 'Stochastic trends and economic fluctuations in a small open economy'. *Journal of Applied Econometrics*, 7: 369-394.

Mosconi, R. (1998). *Malcolm. The theory and practice of cointegration analysis in RATS*. Venezia: Cafoscarina.

Nelson, C. R., and Plosser, C. I. (1982). 'Trends versus Random Walks in economic time-series. Some evidence and implications'. *Journal of Monetary Economics*, 10: 139-162.

Osterwald-Lenum, M. (1992). 'A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics'. *Oxford Bulletin of Economics and Statistics*, 54: 461-472.

Perron, P. (1987). 'The Great Crash, the Oil Price Shock and the Unit Root Hypothesis'. Research Paper, Department of Economics, University of Montreal, Quebec, Canada.

Phillips, P. C. B. (1991). 'Optimal Inference in Cointegrated Systems'. *Econometrica*, 59: 283-306.

Phillips, P. C. B., and Perron, P. (1988). 'Testing for a unit root in time-series regression'. *Biometrika*, 75: 335-346.

Reimers, H. E. (1992). 'Comparison of test for multivariate cointegration'. *Statistical Papers*, 33: 335-359.

Rudebusch, G. D. (1993). 'The Uncertain Unit Root in Real GNP'. *The American Economic Review*, 83: 264-272.

Saikkonen, P. (1993). 'Estimation of cointegration Vectors with Linear Restrictions'. *Econometric Theory*, 9: 19-35.

Salmon, M. (1982). 'Error Correction Mechanisms'. *The Economic Journal*, 92: 615-629.

Sargan, D. (1964). 'Wages and Prices in the United Kingdom: A Study in Econometric Methodology'. In P. E. Hart, G. Mills and J. K. Whitaker (eds.), *Econometric Analysis of National Economic Planning*. London: Butterworth, 25-63.

Sargan, J. D., and Barghava, A. (1983). 'Testing Residuals from Least Squares Regression for Being Generated by Gaussian Random Walk'. *Econometrica*, 51: 153-174.

Sims, C. (1980). 'Macroeconomics and Reality'. *Econometrica*, 48:1-49.

Sims, C., Stock, J. and Watson, M. (1990). 'Inference in Linear Time Series Models with Some Unit Roots'. *Econometrica*, 58, 1: 113-144.

Spanos, A. (1986). *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.

Stock, J. H. (1987). 'Asymptotic Properties of the Least Squares Estimator of Cointegrating Vectors'. *Econometrica*, 55: 1035-1056.

Stock, J. H., and West, K. D. (1987). *Integrated Regressors and the Test of the Permanent Income Hypothesis*. Princeton University, Woodrow Wilson School D.P. 127.

White, H. (1984). *Asymptotic Theory for Econometricians*. London: Academic Press.

Zivot, E. and Andrews, D. W. K. (1992). 'Further Evidence on the Great Crash, the Oil-Price Shock and the Unit Root Hypothesis'. *Journal of Business Economics and Statistics*, 10: 251-270.