Chapter 5 Univariate time-series analysis

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. The simplest 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 unconditional moments.

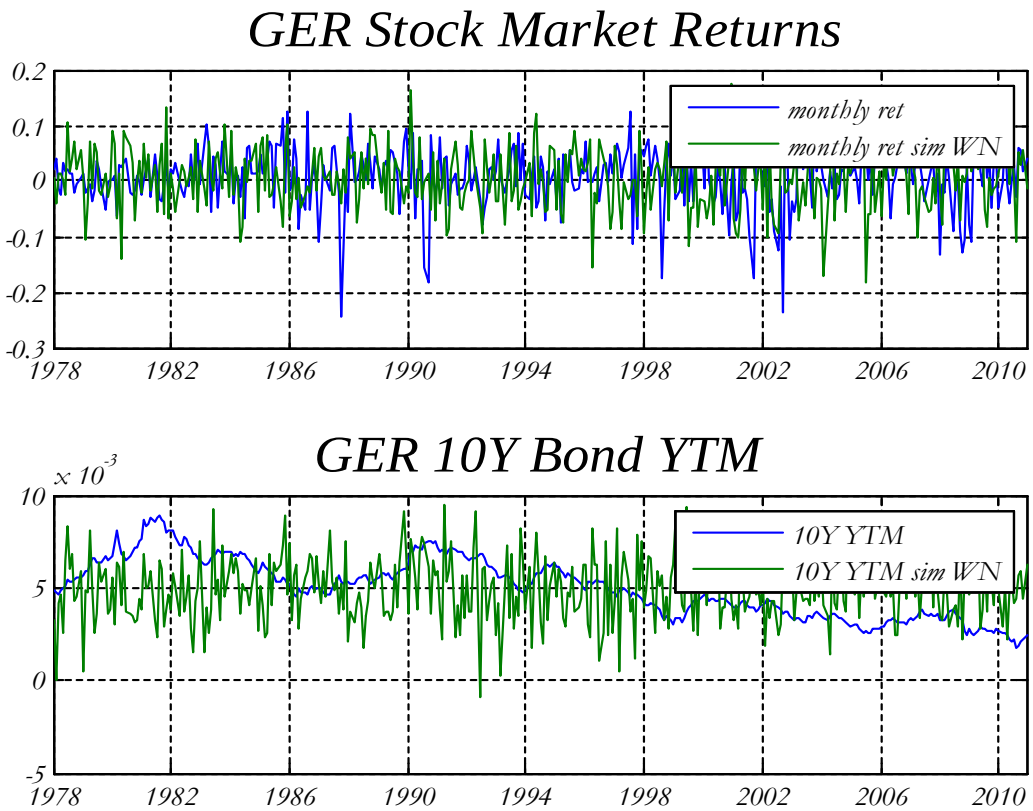
Better models

The model:

$$\begin{aligned} x_t &= \alpha + \epsilon_t,\ \epsilon_t \sim n.i.d.\left(0, \sigma_\epsilon^2\right), \\ \hat{\alpha} &= \frac{1}{T}\sum_{t=1}^{T} x_t,\ \hat{\sigma}_\epsilon^2 = \sum_{t=1}^{T}\frac{1}{T}\left(x_t - \hat{\alpha}\right)^2 \end{aligned}$$

Reflect the traditional approach to portfolio allocation, but it does not reflect the data. At high frequency the variance is not constant and predictable, at low frequency returns are persistent and predictable. To construct more realistic models, we concentrate on univariate models first to consider then multivariate models.

Better models



While the CER gives a plausible representation for the 1-month returns, the behaviour over time of the YTM of the 10-Year bond does not resemble at all that of the simulated data.

ARMA modelling

A more general and more flexible class of models emerges when combinations of ϵ_t are used to model x_t. We concentrate on a class of models created by taking linear combinations of the white noise, the autoregressive moving average (ARMA) models:

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

Analysing time-series models

To illustrate empirically all 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.

Conditional and Unconditional Densities

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.

Conditional Densities

Consider the AR(1) model. The moments of the density of x_t conditional upon x_{t-1} are immediately obtained from the relevant process:

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

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}:

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

Unconditional Densities

Unconditional moments are derived by substituting recursively from 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.

Insert Clicker 1 here

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,

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

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:

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

Insert Clicker 2 here

Stationarity

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:

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

General ARMA processes

The Wold decomposition theorem warrants that any stationary stochastic process can be expressed as the sum of a deterministic and a stochastic moving-average component:

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

Represent the polynomial $b(L)$ as the ratio of two polynomials of lower order:

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

This is an ARMA process. Stationary requires that the roots of $c\left(L\right)$ lie outside the unit circle. Invertibility of the MA component require that the roots of $a\left(L\right)$ lie outside the unit circle.

General ARMA processes

Consider the simplest case, the ARMA(1,1) process:

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

The above equation is equivalent to:

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

Which shows that the ratio of two finite lag polynomials allows us to model an infinite lag polynomial.

General ARMA processes

We then have,

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

Hence,

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

General ARMA processes

For example, suppose $c(L)\, x_t = a(L)\, \epsilon_t$ and you want to find $x_t = d(L)\, \epsilon_t$. Parameters in $d(L)$ are most easily found by writing $c(L)\, d(L) = a(L)$ and by matching terms in $L^{j\cdot}$ For an illustration suppose $a(L) = 1 + a_1 L$, $c(L) = 1 + c_1 L$. Multiplying out $d(L)$ we have

$$(1 + c_1 L)\left(1 + d_1 L + d_2 L^2 + ...d_n L^n\right) = 1 + a_1 L$$

Matching powers of L,

$$\begin{aligned} d_1 &= a_1 - c_1 \\ c_1 d_1 + d_2 &= 0 \\ c_1 d_2 + d_3 &= 0 \\ c_1 d_{n-1} + d_n &= 0 \end{aligned}$$

$$x_t = \epsilon_t + (a_1 - c_1)\, \epsilon_{t-1} - c_1 (a_1 - c_1)\, \epsilon_{t-2} + ... (-c_1)^{n-1} (a_1 - c_1)\, \epsilon_{t-n}$$

Insert Clicker 3 here

Persistence and the linear model

Persistence of time-series destroys one of the crucial properties for implementing valid estimation and inference in the linear model.
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

$$E(\boldsymbol{\epsilon} \mid \mathbf{X}) = \mathbf{0}. \tag{3}$$

Hypothesis (3) implies that

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

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

$$y_t = a_0 + a_1 y_{t-1} + \epsilon_t.$$

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 (3) 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 (say 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.
- the Monte-Carlo simulation is a sort of controlled experiment. To overcome the potential dependence of the set of random numbers drawn on the sequence of simulated white-noise residuals, the DGP is replicated many times.

How serious is the problem?

We report the averages across replications in the following figure .

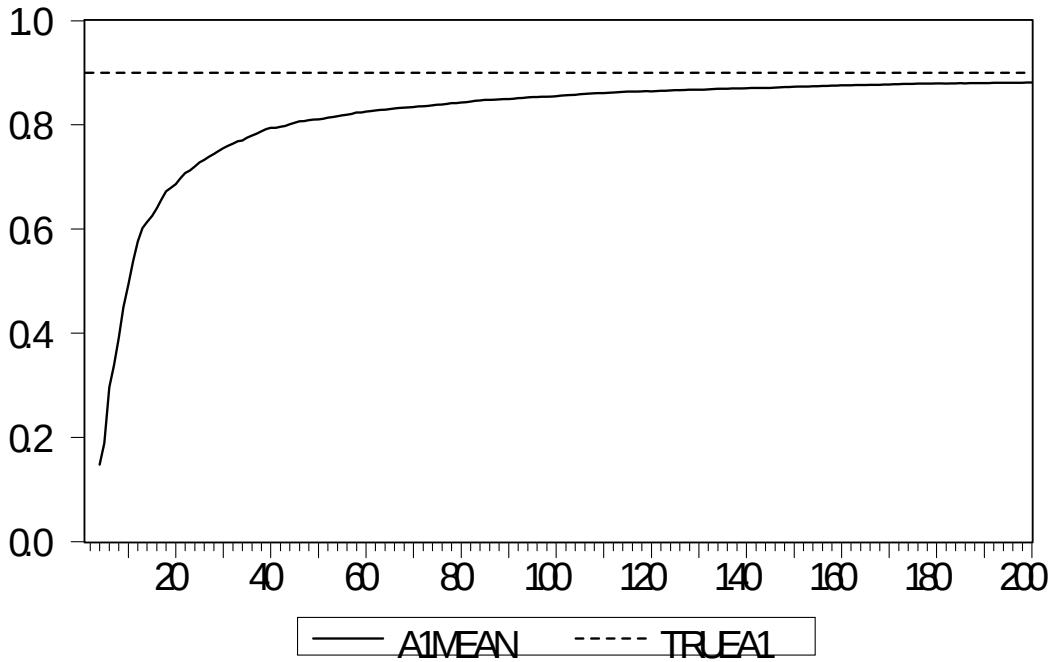


Figure: Small sample bias

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

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.
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.

Convergence

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.
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.

Central limit theorem

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

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

  ω converges in distribution to a standard normal.
- 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, $p\lim g\left(X_T\right) = g\left(a\right)$.

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:

$$X_T + Y_T \text{ converges in distribution to } (a + Y);$$
$$Y_T / a_T \text{ converges in distribution to } (Y/a);$$
$$Y_T \cdot a_T \text{ converges in distribution to } (Y \cdot a).$$

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

Mann-Wald

- 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 $E(\mathbf{z}_t \epsilon_t) = 0$, then

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

Insert Clicker 4 here

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.

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

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 ε_{t+1} 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.

MLE of an AR process

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 $\hat{}$, 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).

MLE of an AR process

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:

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

MLE of an AR process

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

Insert Clicker 5 here

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

$$X_t \mid \mathbf{X}_{t-1} \sim N\left(\lambda x_{t-1}, \sigma^2\right), \tag{4}$$

$$D\left(\mathbf{X}_T^1 \mid \lambda, \sigma^2\right) = D\left(X_1 \mid \lambda, \sigma^2\right) \prod_{t=2}^{T} D\left(X_t \mid \mathbf{X}_{t-1}, \lambda, \sigma^2\right). \tag{5}$$

From (5), the likelihood function clearly involves $T-1$ conditional densities and one unconditional density. The conditional densities are given by (4), 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:

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

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.

Putting ARMA models at work

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

- PRE WHITENING: make sure that the time series is stationary..
- MODEL SELECTION: **Information criteria** are a useful tool to this end. The Akaike's information criteria (**AIC**) and the .Schwarz Bayesian Criterion (**SBC)** are the most commonly used criteria:

$$\begin{aligned} AIC &= -2\log(L) + 2(p+q) \\ SBC &= -2\log(L) + \log(n)(p+q) \end{aligned}$$

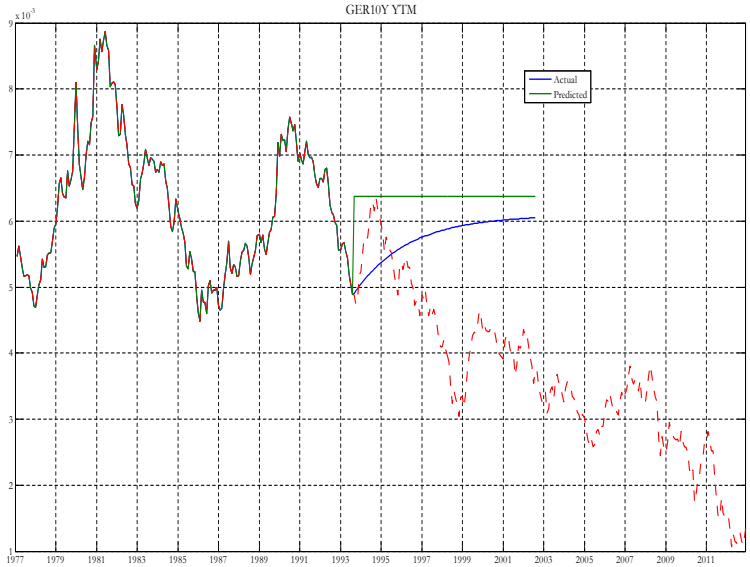
- MODEL CHECKING: residual tests. Make sure that residuals are not autocorrelated and check whether their distribution is normal, also ex-post evaluation technique based on RMSE and MAE are implemented (Diebold-Mariano, Giacomini-White).
- FORECASTING, the selected model is typically simulated forward after estimation of the estimation of parameters to produce forecasts for the variable of interests at the relevant horizon

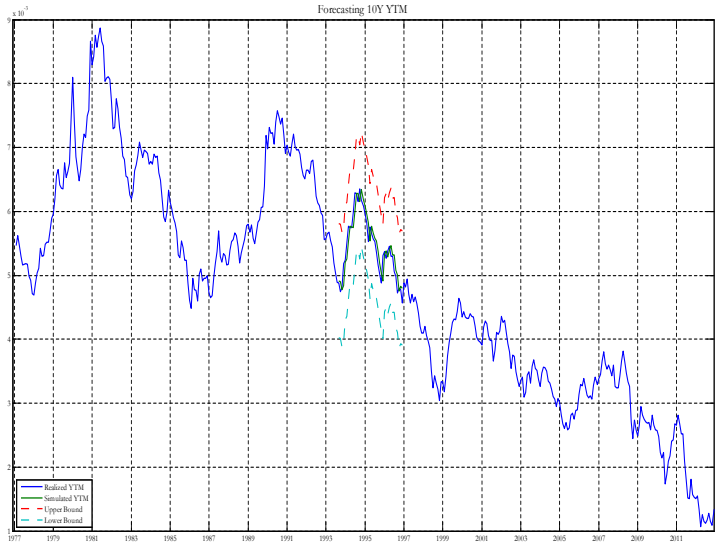
An Illustration

To illustrate how ARMA model can be put at work consider the case of forecasting the YTM of 10-year German bonds from 1994:1 onward, given the availability of data over the period 1978-1993. Estimation can be performed in MATLAB by using the appropriate specification in the GARCH procedure. The ML estimation delivers the following results:

$$y_t^{10} = \underset{(0.0004137)}{0.00014225} + \underset{(0.061035)}{0.9764}\, y_{t-1}^{10} + \underset{(3.2616e-08)^{0.5}}{(0.00000007)^{0.5}}\hat{u}_t$$

An Illustration





Insert Clicker 6 here

Trends

Forecasting using an ARMA models exploits two features of the data: mean-reversion and persistence.
Unfortunately many financial time series do not feature mean reversion as they behave like non-stationary time series.
Non-stationarity of time-series is a possible manifestation of a trend.
Consider, for example, the random walk process with a drift:Consider, for example, the random walk process with a drift:

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

Recursive substitution yields

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

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

Integrated Series

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 $\mathrm{I}(d)$. 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).

Deterministic vs Stochastic Trends

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

$$z_t = \alpha + \beta t + \epsilon_t.$$

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 (**??**) 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.

Insert Clicker 7 here

Univariate decompositions of time-series

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:

$$D(L) = C(L) - C(1).$$

Univariate decompositions of time-series

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

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

where $C^*(L)$ is a polynomial of order $q - 1$.
By equating (**??**) to (**??**), we have:

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

Univariate decompositions of time-series

$$\Delta x_t = \mu + C^*(L)\,\Delta\epsilon_t + C(1)\,\epsilon_t.$$

By integrating , we 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:

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

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:

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

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:

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

and the Beveridge and Nelson decomposition yields:

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

BN 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:

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

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 (**??**) 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

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;
- 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.
- Note that in general the variance of innovations might have an economic interpretation and theory might suggest different patterns of correlations between innovations in the permanent and transitory components that do differ from a perfectly negative correlation. 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.

Nelson decomposition

–

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.
the Hodrick−Prescott filter is derived by minimizing the following expression:

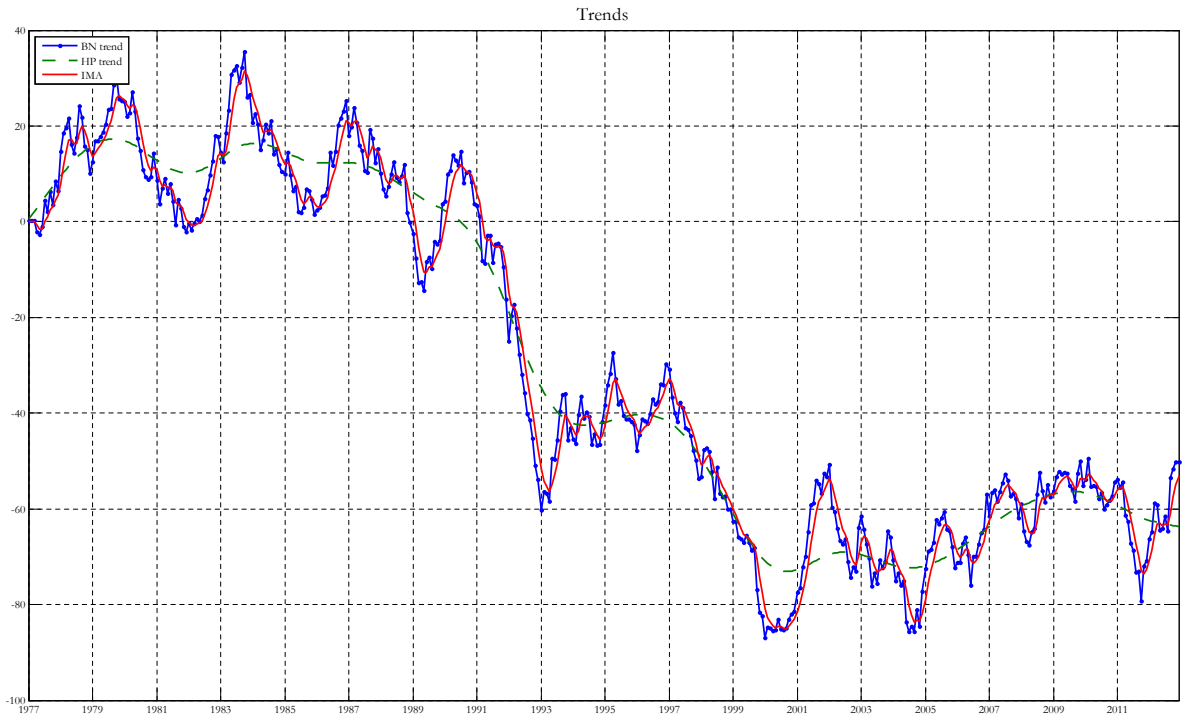
$$\sum_{t=1}^{T} (x_t - TR_t)^2 + \lambda \sum_{t=2}^{T-1} \left[(TR_{t+1} - TR_t)^2 - (TR_t - TR_{t-1})^2 \right].$$

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.

Nelson decomposition

–

Figure 8 reports the Beveridge−Nelson trend and the Hodrick−Prescott trend (with $\lambda = 14400$) for the data generated in the previous section.



Insert Clicker 8 here

Asset Allocation with a simple TVER model: the SOP method

Total stock market returns in local currency can be expressed as follows:

$$\begin{aligned} 1 + H^s_{t+1} &\equiv \frac{P_{t+1} + D_{t+1}}{P_t} = \frac{P_{t+1} + D_{t+1}}{P_t} \\ &= \frac{P_{t+1}}{P_t} + \frac{D_{t+1}}{P_t} = \frac{P_{t+1}}{P_t} + \frac{D_{t+1}}{P_{t+1}} \frac{P_{t+1}}{P_t} \\ &= \left(1 + \frac{D_{t+1}}{P_{t+1}}\right) \frac{P_{t+1}}{P_t} \end{aligned}$$

taking logs:

$$h^s_{t+1} \simeq (p_{t+1} - p_t) + \frac{D_{t+1}}{P_{t+1}}$$

Asset Allocation with a simple TVER model

In case of investment in a foreign currency we have a third term capturing exchange rate fluctuations.

$$h_{t+1}^{s,l} \simeq (p_{t+1} - p_t) + \frac{D_{t+1}}{P_{t+1}} + (e_{t+1} - e_t)$$

Asset allocation with the simplest TVER model can be applied by specifying univariate time series model to predict each part of the total return.
The investor uses monthly data available over the period 1978-2003 to find the tangency portfolio for an investment over the period 2004-2007. The risky assets available for portfolio allocation are German, US and UK shares and the German 10-Year government bond. The risk free asset is the German short-term rate.

Asset Allocation with a simple TVER model

German stock market returns:

$$
\begin{aligned}
r^{GER}_{t,t+36} &= \left(p^{GER}_{t+36} - p^{GER}_{t}\right) + \sum_{i=j}^{36} \frac{D^{GER}_{t+j}}{P^{GER}_{t+j}} \\
\left(p^{GER}_{t+36} - p^{GER}_{t}\right) &= E_t\left(p^{GER}_{t+36} - p^{GER}_{t}\right) + u_{1,t+36} \\
E_t\left(p^{GER}_{t+36} - p^{GER}_{t}\right) &= \beta^{GER}_0 + \beta^{GER}_1 \left(p^{GER}_{t} - \frac{1}{36}\sum_{j=1}^{36} p^{GER}_{t-j}\right) \\
\sum_{j=1}^{36} \frac{D^{GER}_{t+j}}{P^{GER}_{t+j}} &= E_t\left(\sum_{i=1}^{36} \frac{D^{GER}_{t+j}}{P^{GER}_{t+j}}\right) + u_{2,t+36} \\
E_t\left(\sum_{j=1}^{36} \frac{D^{GER}_{t+j}}{P^{GER}_{t+j}}\right) &= \sum_{i=0}^{35} \frac{D^{GER}_{t-j}}{P^{GER}_{t-j}}
\end{aligned}
$$

Asset Allocation with a simple TVER model

US and UK stock market returns :

$$
\begin{aligned}
r^i_{t,t+36} &= \left(p^i_{t+36} - p^i_t\right) + \sum_{j=1}^{36} \frac{D^i_{t+j}}{P^i_{t+j}} + \left(e^i_{t+36} - e_t\right) \\
\left(p^i_{t+36} - p^i_t\right) &= E_t\left(p^i_{t+36} - p^i_t\right) + u^i_{1,t+36} \\
E_t\left(p^i_{t+36} - p^i_t\right) &= \beta^i_0 + \beta^i_1\left(p^i_t - \frac{1}{36}\frac{\beta^i_2}{\beta^i_1}\sum_{j=1}^{36} p^i_{t-j}\right) \\
\sum_{i=1}^{36} \frac{D^i_{t+i}}{P^i_{t+i}} &= \sum_{j=0}^{35} \frac{D^i_{t-j}}{P^i_{t-j}} + u^i_{2,t+36} \\
e^i_{t+36} &= e^i_t + u^i_{3,t+36}, \quad i = UK, US
\end{aligned}
$$

The simple CER model is maintained for the German bond returns: