Univariate time-series analysis

Carlo Favero and Celso Brunetti

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

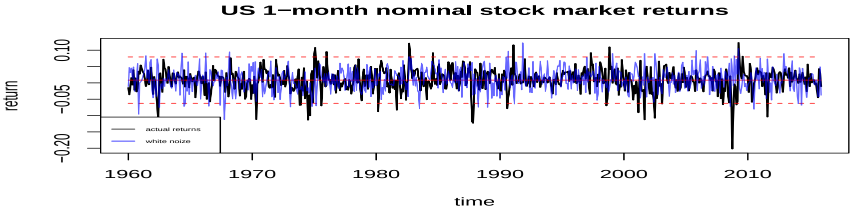
The model:

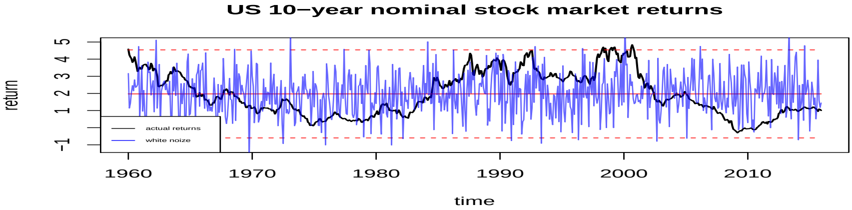
$$x_t = \alpha + \epsilon_t, \;\; \epsilon_t \sim n.i.d. \left(0, \sigma_\epsilon^2\right),$$

$$\hat{\alpha} = \frac{1}{T}\sum_{i=t}^{T} x_t, \;\; \hat{\sigma}_\epsilon^2 = \sum_{i=t}^{T}\frac{1}{T}\left(x_t - \hat{\alpha}\right)^2$$

Reflect the traditional approach to portfolio allocation, but it does not reflect the data. At high frequency (daily, intra-day) 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





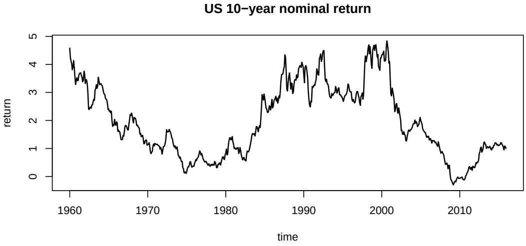
While the CER gives a plausible representation for the 1-month returns, the behaviour over time of the YTM of the 10-Year returns does not resemble at all that of the simulated data.

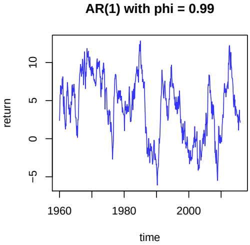
ARMA modelling

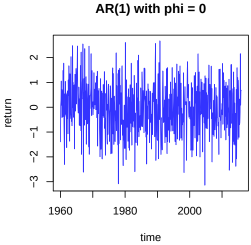
A more general and more flexible class of models emerges when combinations of ϵ_t are used to model x_t. We concentrate on a class of models created by taking linear combinations of the white noise, the autoregressive moving average (ARMA) models:

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

An Illustration







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 its expression 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.

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

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

Conditional Densities: An Example

Consider the following MA model

$$
\begin{aligned}
x_t &= 0.8 + \epsilon_t + 0.4\epsilon_{t-1} \\
\epsilon_t &\sim n.i.d.\left(0, \sigma_\epsilon^2\right)
\end{aligned}
$$

State which of the following is correct:
(a) $E\left(x_t \mid x_{t-1}\right) = 0.8 + \epsilon_t + 0.4\epsilon_{t-1}$
(b) $E\left(x_t \mid x_{t-1}\right) = 0.8 + 0.4\epsilon_{t-1}$
(c) $Var\left(x_t \mid x_{t-1}\right) = 1.16\sigma_\epsilon^2$
(d) $Var\left(x_t \mid x_{t-1}\right) = \sigma_\epsilon^2$

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,

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

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:

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

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:

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

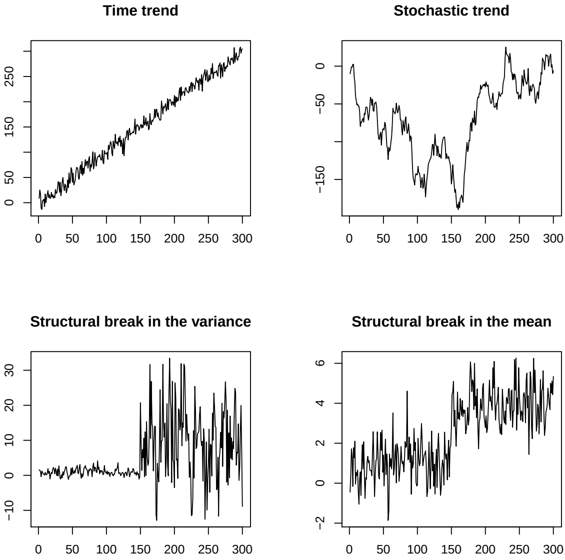
Stationarity

consider the following model

$$\begin{aligned} x_{t+1} &= 0.05 + x_t + \epsilon_{t+1} \\ \epsilon_{t+1} &\sim n.i.d. \left(0, \sigma_\epsilon^2\right) \end{aligned}$$

State which of the following is correct:
(a) $E\left(x_{t+5} \mid x_t\right) = 0.25 + x_t$
(b) $E\left(x_t\right) = 0.05$
(c) $Var\left(x_{t+5} \mid x_t\right) = 5\sigma_\epsilon^2$
(d) $Var\left(x_{t+5} \mid x_t\right) = \sigma_\epsilon^2$
(e) $E\left(x_t\right) = 0.05t + x_0$

Stationarity



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:

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

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,

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

Hence,

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

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 $L^{j.}$ For an illustration suppose $a(L) = 1 + a_1 L$, $c(L) = 1 + c_1 L$. Multiplying out $d(L)$ we have

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

Matching powers of L,

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

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

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

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

Hypothesis (3) implies that

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

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 $E\left(\epsilon_t \mid y_{t-1}\right) = 0$, $E\left(\epsilon_{t-1} \mid y_{t-1}\right) \neq 0$ and (3) breaks down.

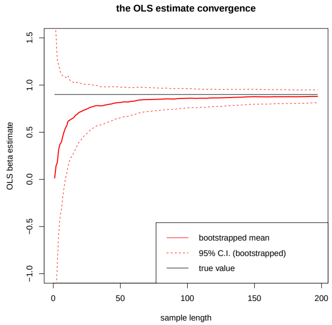
How serious is the problem?

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

How serious is the problem?

We report the averages across replications in the following figure .



We note that the estimate of a_1 is heavily biased in small samples, but the bias decreases as the sample gets larger. One can show analytically that the average of the OLS estimate of a_1 is $a_1 \left(1 - \frac{2}{T}\right)$.

Implications

When using time-series models it is of crucial importance

- to specify models for stationary series
- have available large samples of observations .

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.

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 parameters to produce forecasts for the variable of interests at the relevant horizon.

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:

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

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 $\left(\sum_{i=0}^{t-1} \epsilon_{t-i}\right)$ 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 (4) 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.

Testing for Stationarity

There are several statistical tests that can help assessing whether a time series is stationary. We analyze only two of them:

- Dickey-Fuller and Augmented Dickey-Fuller (ADF)
- ADF-GLS (GLS: Generalized Least Squared)

The concept of stationarity strongly depends on the sample: different sub-samples of the same time series may have different characteristics.

DF Test

Consider a simple AR(1) model

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

a unit root is present if $\rho = 1$ and the model would be non-stationary.
We can rewrite the model as follows:

$$(1 - L)y_t = \Delta y_t = (\rho - 1)y_{t-1} + \epsilon_t = \delta y_{t-1} + \epsilon_t$$

Testing for a unit root implies testing the null:
$H_0 : \delta = 0$ which is equivalent to $H_0 : \rho = 1$
The alternative is: $H_1 : |\rho| < 1$
The test is run over the residual terms rather than the actual data, it is not possible to use the t-distribution to provide critical values. Critical values are provided by Dickey-Fuller (Dickey and Fuller, 1979). Note: The null hypothesis implies a unit root, if a series is stationary the null should be rejected.

ADF Test

There are three main versions of the test:

$$\Delta y_t = \delta y_{t-1} + \epsilon_t$$

$$\Delta y_t = \alpha + \delta y_{t-1} + \epsilon_t \ \text{ constant}$$

$$\Delta y_t = \alpha + \beta t + \delta y_{t-1} + \epsilon_t \ \text{ constant and trend}$$

The Augmneted Dickey-Fuller is:

$$\Delta y_t = \alpha + \beta t + \delta y_{t-1} + \gamma_1 \Delta y_{t-1} + \gamma_2 \Delta y_{t-2} + ... + \gamma_p \Delta y_{t-p} + \epsilon_t$$

ADF-GLS Test

ADF-GLS is very similar to the ADF but it filters the series (GLS)

$$\bar{y}_t = y_t - (\frac{c}{T})y_{t-1}$$

In other words, $\rho = 1 - \frac{c}{T}$
where T is the number of observations.
This testing procedure dominates other procedures in terms of power.
It demeans and de-trends the series locally to perform a more efficient estimation of δ.
Standard t-dstribution does not apply (similar to ADF). (See Elliott, Rothenberg and Stock, 1992).

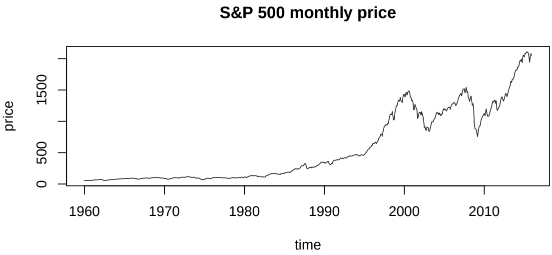
Time Series Analysis: Step 1

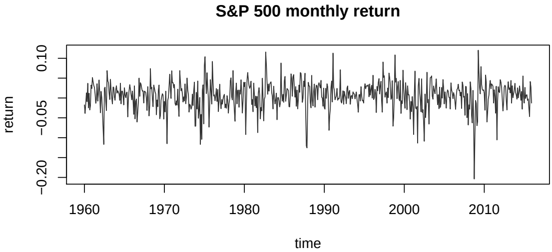
Plot the data!

Always look at the data as a first step. A visual analysis may provide valuable information on:

- Stationarity (trends, regime switching, etc.)
- Special events (outliers) – are they real or errors?
- Linkages to economic and finance theory/intuition – the S&P500 declines in recessions while VIX increases
- Missing data (real or errors?)

Time Series Analysis: Step 1





Time Series Analysis: Step 2

Test for stationarity.

- If null is rejected: no need to transform the series
- If fail to reject the null: first difference – or log-difference (e.g. stock returns) or growth rates (e.g. GDP), etc.
  - Re-start from Step 1 for the transformed series
  - Re-run tests for stationarity for the transformed series

Time Series Analysis: Step 2

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max
-240.165   -3.621    0.665    6.607  111.323

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
z.lag.1    0.003293   0.001321   2.492    0.013 *
z.diff.lag 0.185867   0.038269   4.857 1.49e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 26.92 on 668 degrees of freedom
Multiple R-squared:  0.04939,	Adjusted R-squared:  0.04654
F-statistic: 17.35 on 2 and 668 DF,  p-value: 4.497e-08


Value of test-statistic is: 2.4917

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
      Min        1Q    Median        3Q       Max
-0.189253 -0.013584  0.006696  0.025579  0.133906

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.76414    0.04681 -16.324   <2e-16 ***
z.diff.lag  0.04002    0.03865   1.035    0.301
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03461 on 668 degrees of freedom
Multiple R-squared:  0.3688,	Adjusted R-squared:  0.3669
F-statistic: 195.1 on 2 and 668 DF,  p-value: < 2.2e-16


Value of test-statistic is: -16.3244

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

Time Series Analysis: Step 2 (cont.)

```
###############################################
# Elliot, Rothenberg and Stock Unit Root Test #
###############################################

Test of type DF-GLS
detrending of series with intercept


Call:
lm(formula = dfgls.form, data = data.dfgls)

Residuals:
     Min       1Q   Median       3Q      Max
-227.599   -3.263    0.945    7.114  114.184

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
yd.lag        0.002961   0.001495   1.980   0.0481 *
yd.diff.lag1  0.192627   0.039030   4.935 1.01e-06 ***
yd.diff.lag2 -0.050672   0.039682  -1.277   0.2021
yd.diff.lag3  0.089182   0.039972   2.231   0.0260 *
yd.diff.lag4  0.043378   0.039922   1.087   0.2776
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 26.88 on 662 degrees of freedom
Multiple R-squared:  0.06035,   Adjusted R-squared:  0.05325
F-statistic: 8.503 on 5 and 662 DF,  p-value: 8.164e-08


Value of test-statistic is: 1.9804

Critical values of DF-GLS are:
                 1pct  5pct 10pct
critical values -2.57 -1.94 -1.62
```

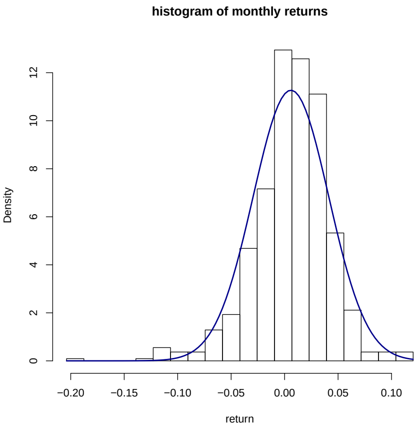
```
###############################################
# Elliot, Rothenberg and Stock Unit Root Test #
###############################################

Test of type DF-GLS
detrending of series with intercept


Call:
lm(formula = dfgls.form, data = data.dfgls)

Residuals:
      Min        1Q    Median        3Q       Max
-0.174022 -0.012319  0.007701  0.026038  0.161466

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
yd.lag       -0.29226    0.04767  -6.131 1.50e-09 ***
yd.diff.lag1 -0.36259    0.05217  -6.951 8.73e-12 ***
yd.diff.lag2 -0.36829    0.04994  -7.374 4.95e-13 ***
yd.diff.lag3 -0.23818    0.04522  -5.267 1.88e-07 ***
yd.diff.lag4 -0.15981    0.03847  -4.154 3.70e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03565 on 662 degrees of freedom
Multiple R-squared:  0.3354,    Adjusted R-squared:  0.3304
F-statistic: 66.83 on 5 and 662 DF,  p-value: < 2.2e-16


Value of test-statistic is: -6.1307

Critical values of DF-GLS are:
                 1pct  5pct 10pct
critical values -2.57 -1.94 -1.62
```

Time Series Analysis: Step 3

Summary statistics (conditional and unconditional)

- Compute moments: mean, median, standard deviation, skewness and kurtosis
- Test for Normality and plot the distribution
- Compute and plot autocorrelation function
  - The number of lags in the autocorrelation depends on data frequency
  - The autocorrelation function is very informative on how to specify the ARMA model

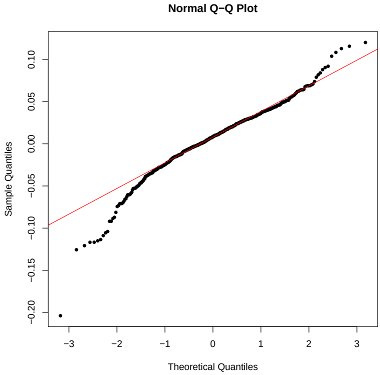
Time Series Analysis: Step 3



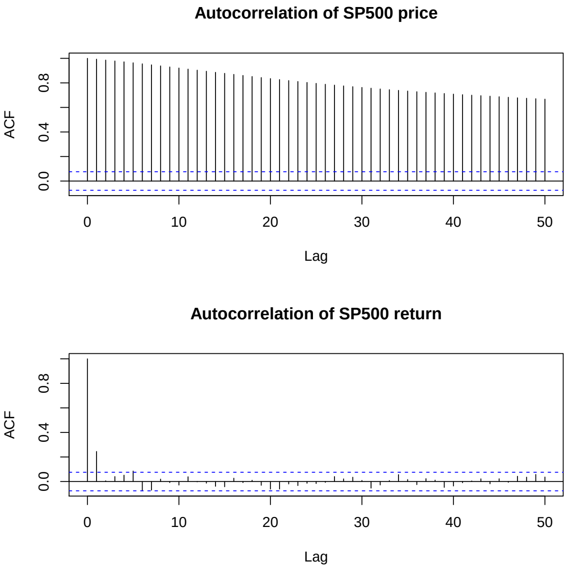
SP500	
Min.	-0.2039
1st Qu.	-0.0126

statistic	p.value	method	data.name
0.958598618005397	$8.01600113266016e-13$	Shapiro-Wilk normality test	SW

statistic	parameter	p.value	method	data.name
56.2513510566111	10	$1.84192959773455e-08$	Box-Ljung test	LB



Time Series Analysis: Step 3



Time Series Analysis: Step 4

ARMA modeling

```
> auto.arima(regdata$ret_m_w, d = 0, max.q = 0, seasonal = F, ic = "bic")
Series: regdata$ret_m_w
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1    mean
      0.2448  0.0059
s.e.  0.0374  0.0018

sigma^2 estimated as 0.001179:  log likelihood=1313.06
AIC=-2620.11   AICc=-2620.08   BIC=-2606.58
```

Time Series Analysis: Step 5

Diagnostics (conditional and unconditional)

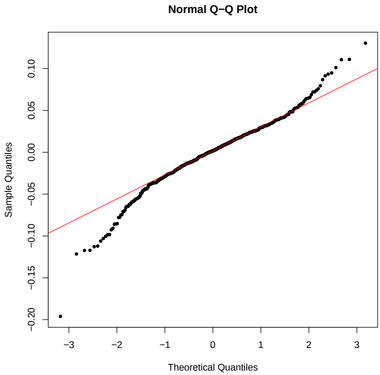
- Are the residuals normally distributed?
  - Test for Normality and plot the distribution

- Did we capture all the persistence?
  - Plot the autocorrelation function of the residuals

Time Series Analysis: Step 5

statistic	p.value	method	data.name
0.964405292256005	$1.06207374907505e-11$	Shapiro-Wilk normality test	ARresiduals

statistic	parameter	p.value	method	data.name
21.6936001982079	10	0.0167439163223797	Box-Ljung test	ARresiduals



Time Series Analysis: Step 5