Econometrics II

Notes

William Radaic Peron

EESP-FGV

November 20, 2020

Contents

Chapter 1

Introduction

1.1 Motivation

This course will be dedicated to *time series analysis.* Informally, a *time series* is any type of data collected over time – or, more formally, it is the realization of a stochastic process indexed in time. We usually denote the time series as follows:

$$y_1, ..., y_T; \quad \{y_t\}_{i=1}^{T}; \quad \{y_t\}_t$$

Time series analysis is useful for a number of different applications:

- **Forecasting.**
  - Uni and multivariate models
  - **ARIMA** models: mean and confidence interval forecasting
  - **ARCH** models: variance forecasting – especially useful in finance for volatility and risk
- **Dynamics.** Evaluate the impact of one variable in another over time.
  - Multivariate models including VAR, ECM
  - Contemporaneous lagged structural relations

It is important to address a first and simple question. **Why time series are different from other data?** The answer is also simple but incredibly relevant: *time series observations are not serially independent!*

$$Y_t \not\perp\!\!\!\perp Y_{t-j}$$

In fact, they don't even have to be identically distributed:

$$F_{Y_t} \neq F_{Y_{t-j}}$$

This means that the essential *iid* hypothesis for traditional Econometrics *does not hold.* This means that we'll have to make some adjustments to our methods. That is the task of time series analysis.

1.2 Statistics with dependence

Let's begin with a proper definition of a time series.

1.2.1 Definition of a time series

Suppose that we have a probability space $(\Omega, S, \mathbb{P})$. Ω is the sample space; S is the set of all events; $\mathbb{P}$ is a measure of probability $\mathbb{P} : S \to [0, 1]$. From this, we define a random variable $Y : \Omega \to \mathbb{R}$. A realization of this r.v. is denoted by $y = Y(\omega)$ with fixed ω.

From this, we can define multiple random variables in the same sample space, indexed by integers:

$$Y = \{..., Y_{t-2}, Y_{t-1}, Y_t, ...\}$$

This is equivalent to writing:

$$Y : \Omega x \mathbb{Z} \to \mathbb{R}$$

We now arrive at our formal definition of a time series: $\{Y_t, t \in \mathbb{Z}\}$ is a time-indexed stochastic process.

- $Y(\cdot, t) : \Omega \to \mathbb{R}$ is a r.v. for fixed t.
- $Y(\omega, \cdot) : \mathbb{Z} \to \mathbb{R}$ is a *sequence of real numbers* for a fixed ω. In other words, this represents the *observed time series.*
- For fixed t, ω, $Y(\omega, t) \in \mathbb{R}$.

1.2.2 Unconditional expectation

An important concept to make clear here is *unconditional expectation.* With fixed t,

$$\mathbb{E}(Y_t) = \int_{-\infty}^{\infty} x f_{Y_t}(x) dx$$

Note the Y_t subscript on the probability density function f_{Y_t}. This means that $\mathbb{E}(Y_t)$ is not calculated with the values assumed by Y_{t-1}, Y_{t+1}. This raises an important problem: *how would you be able to estimate* $\mathbb{E}(Y_t)$*?* Note that we only observe $Y_t = y_t$, i.e., one realization of the r.v.

1.2.3 Statistical dependence

For any random variables X, Y, we can define multiple measures of dependence:

- **Linear:** $Cov(X, Y) \equiv \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
- **Quadratic:** $Cov(X^2, Y^2)$
- **General:** $Cov(f(X), g(Y))$. This is a measure of covariance between two general functional forms of X and Y.

With this general definition, we arrive at an equivalent definition for independent random variables:

- $F_{X,Y}(x, y) = F_X(x) * F_Y(y)$, i.e., joint pdf is equal to the product of the marginal pdfs.
- $Cov(f(X), g(Y)) = 0$ for every pair of bounded functions f, g.

From this, we now define the *autocovariance and autocorrelation functions.*

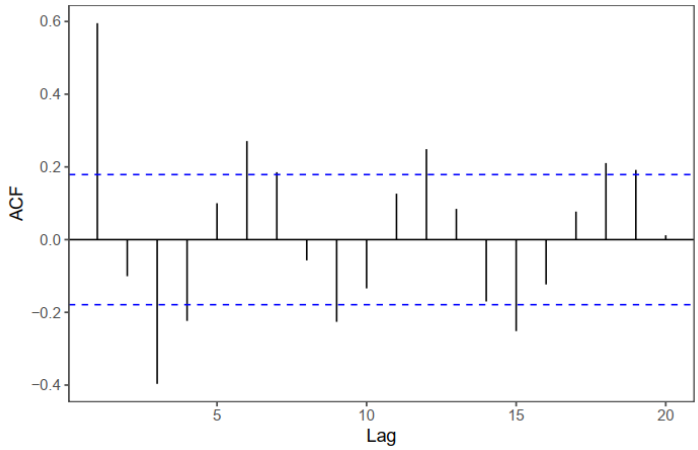
Definition 1.2.1. $\gamma_{j,t} := Cov(Y_t, Y_{t-j})$ *is the* ***autocovariance function*** *for a given time series* $\{Y_t, t \in \mathbb{Z}\}$.

Definition 1.2.2. $\rho_{j,t} := \dfrac{\gamma_{j,t}}{\sqrt{\gamma_{0,t}\gamma_{0,t-j}}}$ *is the* ***autocorrelation function*** *for a given time series* $\{Y_t, t \in \mathbb{Z}\}$.

Note that, if *iid* holds:

$$\gamma_{j,t} = \begin{cases} 0 & j \neq 0, \forall t \\ Var(Y) & otherwise \end{cases}$$

This is an example of an autocorrelation function.



1.3 Asymptotic theory with dependence

Some form of asymptotic theory is needed to enable *any kind of statistical analysis.* Namely, we need to have some form of Law of Large Numbers (LLN) and Central Limit Theorem (CLT) that are analogous to the *iid* environment. This will be achieved in our setting with some conditions called *stationarity* and *ergodicity.*

1.3.1 Stationarity

Definition 1.3.1. *A process* $\{Y_t, t \in \mathbb{Z}\}$ *is* ***strictly stationary*** *if, for all finite set of indexes* $\{t_1, ..., t_r\}$ *and for all* $m \in \mathbb{Z}$, $F(y_{t_1}, ..., y_{t_r}) = F(y_{t_1+m}, ..., y_{t_r+m})$ *holds, where* $F(y_{t_1}, ..., y_{t_r})$ *is the joint cdf of* $(Y_{t_1}, ..., Y_{t_r})$.

More informally, a given process is called *strictly stationary* if its statistical properties depend only on the *relative position* between observations, and not its *absolute position.*

We'll usually adopt a weaker definition of stationarity for our models. Henceforth, we will refer to stationarity in this sense.

Definition 1.3.2. *A process* $\{Y_t, t \in \mathbb{Z}\}$ *is* ***stationary*** *(or weakly stationary) if there exists* $\mu \in \mathbb{R}$ *and* $\{\gamma_j\}_{j \in \mathbb{N}}$ *such that:*

- $\mathbb{E}(Y_t) = \mu, \quad \forall t$
- $\mathbb{E}[(Y_t - \mu)(Y_{t-j} - \mu)] = \gamma_j, \quad \forall (t, j) \in \mathbb{N}^2$

Note that, from the second condition in the definition, we have $\mathbb{E}(Y_t - \mu)^2 = \gamma_0 \in \mathbb{R}, \forall t \in \mathbb{N}$. In other words, *the unconditional variance of the time series is constant.*

Some important remarks on stationarity:

- Stationarity does not imply strict stationarity
- Stricy stationarity does not imply stationarity
- Every strictly stationary process with finite variance is stationary
- Every iid process is strictly stationary
- Every strictly stationary process is identically distributed
- A stationary process is not necessarily identically distributed

1.3.2 Ergodicity

Stationarity is not enough to guarantee that we have even a Law of Large Numbers. To see why that is the case, consider the following example:

$$Y_t = X + \varepsilon_t, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2), \quad X \sim \mathcal{N}(0, 1), \quad X \perp\!\!\!\perp \varepsilon_t$$

Is this process stationary? No, because the sample *time average* $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$ does not converge to the population *ensemble average* $\mathbb{E}(Y_t) = \mu$.

We need some condition that guarantees that the dependence structure of the time series decays as the observation get further from each other. That is the intuition behind *ergodicity.*

Definition 1.3.3. *A strictly stationary process* $\{Y_t, t \in \mathbb{Z}\}$ *is called* ***ergodic*** *if*

$$\lim_{J\to\infty} \frac{1}{J} \sum_{j=1}^{J} Cov[f(X_1), g(X_j)] = 0,$$

for all pairs of bounded functions f, g.

This is a kind of mean asymptotic independence, in which the asymptotic independence would be defined by $Cov[f(X_1), g(X_J)] \to 0$ as $J \to \infty$.

Now, we can define a Law of Large Numbers – also called the *Ergodic Theorem.*

Theorem 1.3.1. *Given an ergodic stochastic process* $\{Y_t, t \in \mathbb{Z}\}$ *such that* $\mathbb{E}|Y_1| < \infty$,

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} Y_t = \mathbb{E}(Y_1) \quad \textit{almost sure}$$

This theorem is the generalization of the strong LLN. However, it presupposes *strict stationarity,* which is a very strong assumption most of the time. Fortunately, this theorem gave rise to other definitions that arrive at our objective, namely, a LLN for the first two moments.

Definition 1.3.4. *A stationary process* $\{Y_t, t \in \mathbb{Z}\}$ *is said to be* ***ergodic for the mean*** *if*

$$\frac{1}{T} \sum_{t=1}^{T} Y_t \to_p \mathbb{E}(Y_t), \quad T \to \infty$$

Definition 1.3.5. *A stationary process* $\{Y_t, t \in \mathbb{Z}\}$ *is said to be* ***ergodic for the second moment*** *if, for every* j,

$$\frac{1}{T-j}\sum_{t=j+1}^{T} Y_t Y_{t-j} \to_p \mathbb{E}(Y_t Y_{t-j}), \quad T \to \infty$$

Proposition 1.3.2. $\sum_{j=0}^{\infty} |\gamma_j| < \infty$ *is a suficient condition for ergodicity for the mean.*

Proof. Let $Z_t := Y_t - \mu$ and $\bar{Z}_t := \frac{1}{T}\sum_{t=1}^{T} Z_t$, where $\{Y_t, t \in \mathbb{Z}\}$ is a stationary process. We will show that $\bar{Z}_t$ converges to 0 in mean square.

$$\begin{aligned}
\mathbb{E}\left(\bar{Z}_T^2\right) &= \mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\right)\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\right)\right] = \frac{1}{T^2}\mathbb{E}\left(\sum_{s=1}^{T}\sum_{t=1}^{T} Z_s Z_t\right) \\
&= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}\left(Z_s Z_t\right) = \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_{s-t} = \frac{1}{T}\sum_{j=-T+1}^{T-1}\frac{T-|j|}{T}\gamma_j \\
&\leq \frac{1}{T}\sum_{j=-T+1}^{T-1}\frac{T-|j|}{T}|\gamma_j| \leq \frac{1}{T}\sum_{j=-T+1}^{T-1}|\gamma_j| \to 0
\end{aligned}$$

□

1.3.3 A Central Limit Theorem for time series

The conditions that guarantee the existence of a CLT for stationary and ergodic processes are much more envolving than in the *iid* environment. However, we have a relatively simple result that will be useful to us in time series analysis. It will now be presented without proof.

Theorem 1.3.3. *Let* $\{Y_t, t \in \mathbb{Z}\}$ *be a* ***linear*** *stationary process, i.e., that can be written in the form* $Y_t = \mu + \sum_{j=0}^{\infty}\psi_j \varepsilon_{t-j}$, *where* $\varepsilon \sim_{iid} (0, \sigma^2)$ *and* $\sum_{j=-\infty}^{\infty}|\psi_j| < \infty$. *Then,*

$$\sqrt{T}(\bar{Y}_t - \mu) \to_d \mathcal{N}(0, \omega^2),$$

where $\omega^2 := \sum_{j=-\infty}^{\infty}\gamma_j < \infty$

Chapter 2

ARMA Models

ARMA is a class of models that we'll employ frequently in time series analysis. Let's begin with some definitions.

2.1 White noise

We call *white noise* stationary time series with mean zero that do not have serial correlation.

Definition 2.1.1. *$\{Y_t, t \in \mathbb{Z}\}$ is **white noise**, denoted by $Y_t \sim wn(0, \sigma^2)$, if*

$$\mathbb{E}(Y_t) = 0; \quad \mathbb{E}(Y_t, Y_{t-j}) = \begin{cases} \sigma^2 & j = 0 \\ 0 & j \neq 0 \end{cases}$$

This is the most simple time series – except for the *iid* case, where independence also holds. It will be the building block for a number of processes that we will study.

2.2 Moving Average processes

Let's begin with the simplest form of MA processes: MA(1).

Definition 2.2.1. *A stationary process $\{Y_t, t \in \mathbb{Z}\}$ is called **MA(1)**, or a **moving average of order 1**, if it follows the following form:*

$$Y_t = c + \varepsilon_t + \theta\varepsilon_{t-1}, \quad \varepsilon_t \sim wn(0, \sigma^2),$$

where c, θ are constant.

2.2.1 Moments of an MA(1) model

The expected value of an MA(1) is:

$$\mu \equiv \mathbb{E}(Y_t) = \mathbb{E}(c + \varepsilon_t + \theta\varepsilon_{t-1}) = c$$

With this result, we can rewrite the model as:

$$(Y_t - \mu) = \varepsilon_t + \theta\varepsilon_{t-1}$$

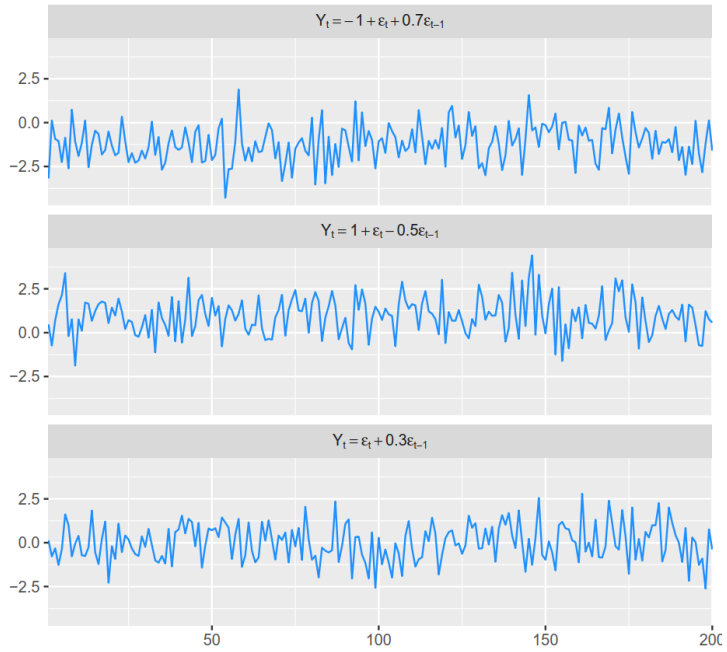
Multiplying both sides by $(Y_{t-j} - \mu)$ yields:

$$\begin{aligned}(Y_t - \mu)(Y_{t-j} - \mu) &= (\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-j} + \theta\varepsilon_{t-j-1}) \\ &= \varepsilon_t\varepsilon_{t-j} + \theta\varepsilon_t\varepsilon_{t-j-1} + \theta\varepsilon_{t-1}\varepsilon_{t-j} + \theta^2\varepsilon_{t-1}\varepsilon_{t-j-1}\end{aligned}$$

Applying the expected value operator to both sides, we have the autocovariances of the model.

$$\gamma_j \equiv \mathbb{E}\left[(Y_t - \mu)(Y_{t-j} - \mu)\right] = \begin{cases} (1+\theta^2)\,\sigma^2 & j = 0 \\ \theta\sigma^2 & j = \pm 1 \\ 0 & |j| > 1 \end{cases}$$

2.2.2 Some examples of MA(1) processes



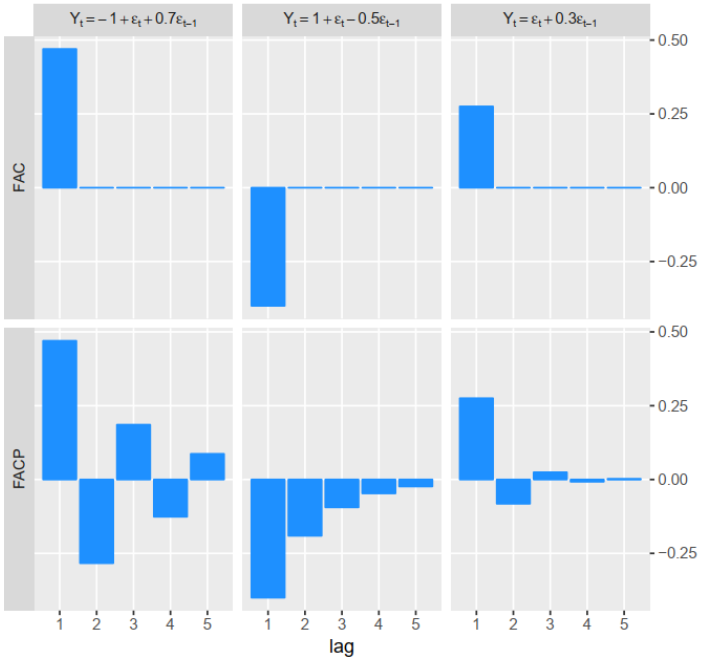
2.2.3 Deriving the Autocorrelation function of the MA(1) process

While deriving the moments of the MA(1), it became clear that the process is stationary and ergodic to the mean. Note that the time average $\bar{y}_t$ converges to $\mathbb{E}(Y_t)$, the *ensemble average*, the absolute sum of all covariances is clearly finite ($\gamma_j = 0, \forall j > 1$) and the dependence structure depends only on the relative positions of the observations.

Let's use the results of the autocovariances to construct the ACF:

$$\rho_j \equiv \frac{\gamma_j}{\gamma_0} = \begin{cases} 1 & j = 0 \\ \frac{\theta}{1+\theta^2} & j = \pm 1 \\ 0 & |j| > 1 \end{cases}$$

Note that the ACF of an MA(1) process is *truncated* in zero for lags greater than 1.



2.3 Generalizing the MA model

We can now generalize the MA(1) model for a moving average of order q.

Definition 2.3.1. *A stationary process* $\{Y_t, t \in \mathbb{Z}\}$ *is called* ***MA(q)****, or a* ***moving average of order q****, if it follows the following form:*

$$Y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}, \quad \varepsilon \sim wn(0, \sigma^2),$$

where $c, \theta_1, ..., \theta_q \in \mathbb{R}, q \in \mathbb{Z}^+$.

2.3.1 Moments of an MA(q) process

The expected value of an MA(q) is:

$$\mu \equiv \mathbb{E}(Y_t) = \mathbb{E}(c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}) = c$$

Again, using the first result, we can rewrite the model as:

$$(Y_t - \mu) = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

Multiplying both sides by $(Y_{t-j} - \mu)$, we have

$$\begin{aligned}(Y_t - \mu)(Y_{t-j} - \mu) &= \left(\sum_{k=0}^{q} \theta_k \varepsilon_{t-k}\right)\left(\sum_{k=0}^{q} \theta_k \varepsilon_{t-j-k}\right) \\ &= \sum_{k=0}^{q}\sum_{\ell=0}^{q} \theta_k \theta_\ell \varepsilon_{t-k}\varepsilon_{t-j-\ell},\end{aligned}$$

where $\theta_0 = 1$. Applying the expectation operator, we have:

$$\gamma_j = \begin{cases} (\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \cdots + \theta_q \theta_{q-j})\,\sigma^2 & |j| = 0, 1, \ldots, q \\ 0 & |j| > q \end{cases}$$

2.3.2 Deriving the Autocorrelation function of the MA(q) process

Again, we can clearly see that the MA(q) model is *stationary* and *ergodic.* Note that the time average $\bar{y_t}$ converges to $\mathbb{E}(Y_t)$, the *ensemble average*, the absolute sum of all covariances is clearly finite ($\gamma_j = 0, \forall j > q$) and the dependence structure depends only on the relative positions of the observations.

The autocorrelation function is given by:

$$\rho_j \equiv \frac{\gamma_j}{\gamma_0} = \begin{cases} 1 & j = 0 \\ \frac{\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \cdots + \theta_q\theta_{q-j}}{1+\theta_1^2+\cdots+\theta_q^2} & |j| = 1, 2, \ldots, q \\ 0 & |j| > q \end{cases}$$

Now, the ACF is truncated in zero for lags *greater than* q.

2.4 The MA(∞) model

Consider a special case of a MA(q) model where $q \to \infty$. This yields a moving average of infinite order, MA(∞).

Definition 2.4.1. *A stationary process* $\{Y_t, t \in \mathbb{Z}\}$ *is called* ***MA(∞)****, or a* ***moving average of infinite order****, if it follows the following form:*

$$Y_t = c + \sum_{i=0}^{\infty} \theta_i \varepsilon_{t-i}, \quad \sim wn(0, \sigma^2),$$

where $c, \theta_1, ..., \theta_q \in \mathbb{R}, \theta_0 = 1$.

We also assume that $\sum_{i=0}^{\infty} |\theta_i| < \infty$. This guarantees that the process is *ergodic.*[1] With this assumption, we can obtain the moments of the MA(∞) simply by taking the limit of the finite case MA(q) – because it enables us to exchange the order between the sum and the expectation operator.

This means that $\mu = c$, as in the previous cases, and:

$$\gamma_j = \left(\sum_{i=0}^{\infty} \theta_{j+i}\theta_i \right) \sigma^2$$

2.5 The Wold Decomposition

This result motivates all ARMA models. It can be defined informally as "any stationary process has a MA(∞) representation".

Theorem 2.5.1. ***Wold Representation Theorem.*** *Any process* $\{Y_t, t \in \mathbb{Z}\}$ *purely nondeterministic can be written as*

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

where $\varepsilon_t = Y_t - \pi(Y_t|1, Y_{t-1}, Y_{t-2}, ...)$*, i.e.,* ε_t *is the error of the linear projection of* Y_t *in* $(1, Y_{t-1}, Y_{t-2}, ...)$.

[1] Details in Hamilton (1994), Appendix, 3.A.

2.6 Autoregressive models

Again, we'll begin with its simplest form, AR(1).

Definition 2.6.1. *A stationary process $\{Y_t, t \in \mathbb{Z}\}$ is called **AR(1)**, or an **autoregressive process of order 1**, if it follows the following form:*

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim wn(0, \sigma^2),$$

where c, ϕ are constant.

2.6.1 Moments of an AR(1) process

With AR($\cdot$) models, we will work in the opposite direction when it comes to stationarity. We'll first *assume* that is holds, and then provide reasoning for why the assumption is valid.

With the assumption of stationarity, we can take expectations and variances on both sides:

$$\mu = c + \phi\mu \iff \mu = \mathbb{E}(Y_t) = \frac{c}{1-\phi}$$

$$\gamma_0 = \phi^2\gamma_0 + \sigma^2 \iff \gamma_0 = Var(Y_t) = \frac{\sigma^2}{1-\phi^2}$$

Using the first result, we can rewrite the model as:

$$(Y_t - \mu) = \phi(Y_{t-1} - \mu) + \varepsilon_t$$

Multiplying both sides by $(Y_{t-j} - \mu)$ and taking expectations yields:

$$\gamma_j = \phi\gamma_{t-j}, \quad j = 1, 2, ...$$

2.6.2 Some examples of AR(1) series

2.6.3 Autocorrelation function of an AR(1) process

Given that $\gamma_j = \phi\gamma_{t-j}$, it is easy to see that the autocovariance is given by:

$$\gamma_j = \phi^{|j|}\gamma_0, \quad j \in \mathbb{Z}$$

Therefore, $\rho_j = \dfrac{\gamma_j}{\gamma_0} = \phi^{|j|}$.

2.6.4 Partial Autocorrelation Function

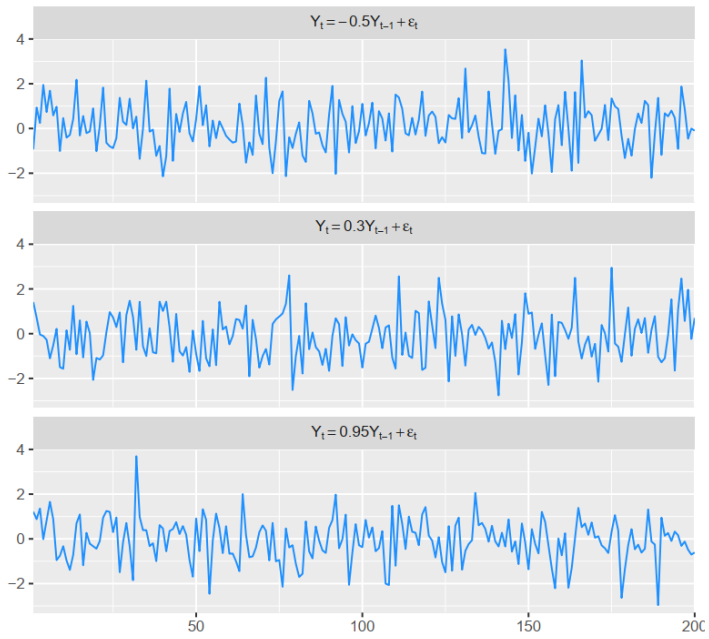
Note that Y_t, Y_{t-2} *are correlated.* Can we isolate the correlation between Y_t, Y_{t-2} from the effects of Y_{t-1}?

$$Cor(Y_t, Y_{t-2}|Y_{t-1}) = Cor(c + \phi Y_{t-1} + \varepsilon_t, Y_{t-2}|Y_{t-1}) = 0$$

This is the intuition behind the *partial autocorrelation function* (PACF).

Definition 2.6.2. *The **partial autocorrelation function** of a stationary process $\{Y_t, t \in \mathbb{Z}\}$ is given by:*

$$\alpha_j = \begin{cases} \text{Cor}\,(Y_t, Y_{t-1}) =: \rho_1 & j = 1 \\ \text{Cor}\,(Y_t, Y_{t-j} \mid Y_{t-1}, \ldots, Y_{t-j+1}) & j \geq 2 \end{cases}$$



To estimate the ACF of a given time series, we need to use its sample equivalent and a version of the Law of Large Numbers, presented in the previous section, because we're only looking for correlations – i.e., population moments. To estimate the PACF, that is not enough. We're now looking for *partial correlation.*

It so happens that OLS gives us the *ceteris paribus* effects. Note that a general form for β is given by: $\beta = \frac{Cov(X,Y)}{Var(X)}$. Therefore, we can estimate using OLS the following models for every j:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + ... + \alpha_j Y_{t-j} + u_t$$

The last coefficient of *each regression*, $\hat{\alpha}_j$, is a consistent estimator for α_j. It is important to highlight, here, that a new model shall be estimated *for each* j, as it guarantees that the coefficient α_j will be conditional on all t *prior* to j.

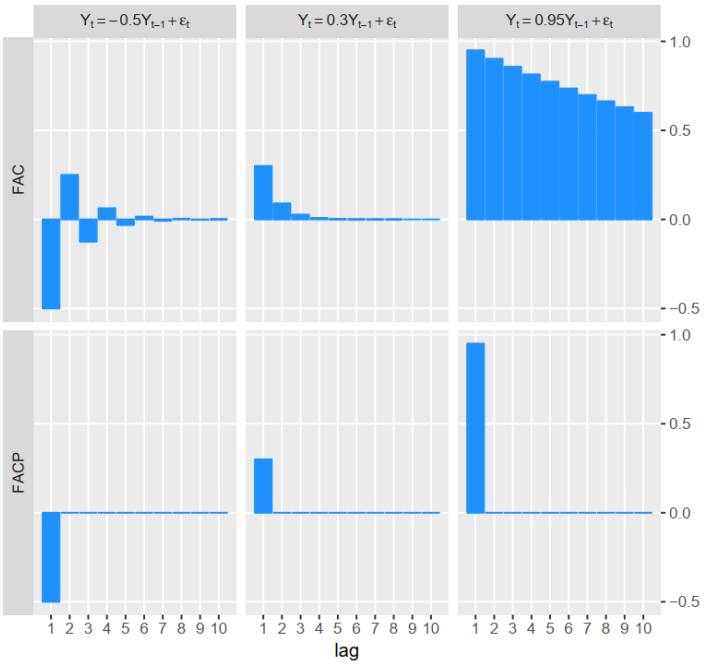
The following plots showcase ACFs and PACFs for AR(1) processes.

2.6.5 Conditions for stationarity

When is an AR(1) process stationary? Note that:

$$\begin{aligned} Y_t &= c + \phi Y_{t-1} + \varepsilon_t \\ &= c + \phi\left(c + \phi Y_{t-2} + \varepsilon_{t-1}\right) + \varepsilon \\ &= c + \phi\left(c + \phi\left(c + \phi Y_{t-3} + \varepsilon_{t-2}\right) + \varepsilon_{t-1}\right) + \varepsilon_t \\ &\dots \\ &= c\sum_{j=0}^{k-1}\phi^j + \phi^k Y_{t-k} + \sum_{j=0}^{k-1}\phi^j \varepsilon_{t-j} \end{aligned}$$

Assuming that $|\phi| < 1$ and taking the limit $k \to \infty$, we have:



$$Y_t = \frac{c}{1-\phi} + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

The first term follows from the sum of an infinite geometric sequence. This means that *an AR(1) process can be written as a MA(∞)* with $\sum_{j=0}^{\infty} |\theta_j| < \infty$. Note that this is equivalent to saying that the Wold Representation Theorem holds, with $\mu = \frac{c}{1-\phi}$, $\psi_j = \phi^j$. This guarantees that the AR(1) process is *stationary* and *ergodic*.

2.7 Generalizing the AR model

Definition 2.7.1. *A stationary process $\{Y_t, t \in \mathbb{Z}\}$ is called **AR(p)**, or an **autoregressive process of order p**, if it follows the following form:*

$$Y_t = c + \phi Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim wn(0, \sigma^2),$$

where $c, \phi_1, ..., \phi_p$ are constant.

2.7.1 Moments of an AR(p) process

Assuming stationarity, we can apply again expectations on both sides:

$$\mu = c + \phi_1 \mu + \cdots + \phi_p \mu \iff \mu = \frac{c}{1 - \phi_1 - \cdots - \phi_p}$$

Using this result, we can rewrite the model as:

$$(Y_t - \mu) = \phi_1 (Y_{t-1} - \mu) + \cdots + \phi_p (Y_{t-p} - \mu) + \varepsilon_t$$

Multiplying both sides by $(Y_{t-j} - \mu)$ and taking expectations, we have:

$$\gamma_j = \begin{cases} \phi_1\gamma_{j-1} + \cdots + \phi_p\gamma_{j-p} & j = 1, 2, \ldots \\ \phi_1\gamma_1 + \cdots + \phi_p\gamma_p + \sigma^2 & j = 0 \end{cases}$$

Note that the last term in γ_0 is implied by $\mathbb{E}(\varepsilon_t)(Y_t - \mu) = \sigma^2$.

2.7.2 ACF of an AR(p) process

Dividing the previous result by γ_0 yields:

$$\rho_j = \phi_1\rho_{j-1} + ... + \phi_p\rho_{j-p}$$

Evaluating at $j = 1, 2, ..., p-1$ and using $p_i = p_{-i}$, we have the following system of difference equations (aka. Yule-Walker Equations):

$$\begin{cases} \rho_1 = \phi_1 + \phi_2\rho_1 + \cdots + \phi_p\rho_{p-1} \\ \rho_2 = \phi_1\rho_1 + \phi_2 + \cdots + \phi_p\rho_{p-2} \\ \vdots \\ \rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \cdots + \phi_p \end{cases}$$

To solve this, we need to find $\rho_1, \rho_2, ..., \rho_j$ as functions of $\phi_1, \phi_2, ..., \phi_j$. The first equation above implies that further correlations from lag j will decay exponentially[2]. This means that the ACF pattern of an AR(p) looks like the one from the simple AR(1) model.

2.8 The Lag Operator

Definition 2.8.1. *Given a process* $\{Y_t, t \in \mathbb{Z}\}$, *the* ***lag operator*** *is defined by:*

$$\begin{aligned} &LY_t := Y_{t-1} \\ &L^2Y(t) := L(LY_t) = L(Y_{t-1}) = Y_{t-2} \\ &\vdots \\ &L^jY(t) := L(L(L\ldots LY_t) = Y_{t-j} \end{aligned}$$

The lag operator is also commutative with multiplication and distributive with regards to addition:

$$L(cY_t) = c(LY_t)$$
$$L(Y_t + X_t) = LY_t + LX_t$$

2.8.1 The lag operator as a polynomial

Note that we can use the lag operator as a *polynomial.* We can now rewrite an AR(p) with zero mean as:

$$(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)Y_t = \varepsilon_t$$

Note that the term multiplying Y_t *is a polynomial in* L. We denote this by:

$$\Phi_p(L)Y_t = \varepsilon_t$$

[2] Review this.

Analogously, we can rewrite an MA(q) process as:

$$\begin{aligned} Y_t &= \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q} \\ &= \left(1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q\right)\varepsilon_t \\ &\equiv \Theta_q(L)\varepsilon_t \end{aligned}$$

We would also like to define an operator $(1-\phi L)^{-1}$ such that:

$$(1-\phi L)^{-1}(1-\phi L) = 1$$

$(1-\phi L)^{-1}$ is well defined when $|\phi| < 1$ and the following condition holds[3]:

$$(1-\phi L)^{-1} := 1 + \phi L + \phi^2 L^2 + \phi^3 L^3 + ...$$

From this, we can rewrite the AR(1) as a MA(∞) by multiplying the AR by $(1-\phi L)^{-1}$ on both sides:

$$Y_t = (1-\phi L)^{-1}\varepsilon_t$$

The $(1-\phi L)^{-1}$ operator will be very useful to translate models between AR and MA representations, aside from highlighting the conditions of stationarity for the process.

2.8.2 Stationarity and the lag operator

We can factor out the polynomial of an AR(p) process as:

$$1 - \phi_1 L - ... - \phi_p L^p = (1-\lambda_1 L)...(1-\lambda_p L),$$

where $\lambda_j = \frac{1}{a_j} \forall j = 1, ..., p$ and $a_1, ..., a_p$ are the p roots of a polynomial of p-th degree. This means that we can rewrite the AR(p) process as:

$$(1-\lambda_1 L)...(1-\lambda_p L)Y_t = \varepsilon_t$$

If $|\lambda_p| < 1$ (or, equivalently, $|a_j > 1)$ $\forall j = 1, ..., p$, then *the inverse polynomial exists* and we can write the AR(p) process as a MA(∞) – which *we know to be stationary*:

$$\begin{aligned} Y_t &= (1-\lambda_1 L)^{-1} \dots (1-\lambda_p L)^{-1} \varepsilon_t \\ &=: \left(1 + \psi_1 L + \psi_2 L^2 + \dots\right)\varepsilon_t \\ &=: \Psi_\infty(L)\varepsilon_t \end{aligned}$$

2.9 Finally, the ARMA(p,q) process

An ARMA(p,q) model is created by combining an AR(p) with a MA(q).

Definition 2.9.1. *A stationary process* $\{Y_t, t \in \mathbb{Z}\}$ *is called* ***ARMA(p,q)****, or an* ***autoregressive-moving average process of order (p,q)****, if it follows the following form:*

$$Y_t = c + \phi Y_{t-1} + ... + \phi_p Y_{t-p} + \theta_1\varepsilon_{t-1} + ... + \theta_q\varepsilon_{t-q} + \varepsilon_t, \quad \varepsilon_t \sim wn(0, \sigma^2),$$

where $c, \theta_1, ..., \theta_p, \phi_1, ..., \phi_q$ *are constant,* $p, q \in \mathbb{Z}^+$.

[3] Hamilton (1994), p. 27-29.

Using the lag operator yields an alternate form for the ARMA(p,q) process:

$$\left(1 - \phi_1 L - \cdots - \phi_p L^p\right) Y_t = c + \left(1 + \theta_1 L + \cdots + \theta_q L^q\right) \varepsilon_t$$
$$\Phi_p(L) Y_t = c + \Theta_q(L) \varepsilon_t$$

2.9.1 Stationarity and invertibility of an ARMA(p,q) process

Stationarity depends only on the AR part of the process, because all MA(·) are stationary. It is sufficient to verifiy that *the roots of the polynomial* $\Phi_p(L)$ *are out of the unit circle:*

$$\Phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$$

Invertibility depends only on the MA part of the process, because it needs to be able to be rewritten as a linear combination of its past values plus the contemporaneous error term ε_t:

$$Y_t = \alpha + \sum_{s=1}^{\infty} \pi Y_{t-s} + \varepsilon_t$$

for some α and $\{\pi_j\}$.

Consider, for example, the case of $MA(1)$ with $\mu = 0$

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

which can be rewritten as

$$\varepsilon_t = y_t - \theta \varepsilon_{t-1}$$

Repeated substitution of this relation for the lagged ε_{t-s} terms yields

$$\begin{aligned} \varepsilon_t &= y_t - \theta \left(y_{t-1} - \theta \varepsilon_{t-2}\right) \\ &= y_t - \theta y_{t-1} + \theta^2 \varepsilon_{t-2} \\ &\cdots \\ &= y_t - \theta y_{t-1} + \ldots + (-\theta)^p y_{t-p} + (-\theta)^{p+1} \varepsilon_{t-p+1} \end{aligned}$$

If $|\theta| < 1$, then the last term in this expression tends to zero in mean-square as $p \to \infty$, so that it make sense to write

$$\varepsilon_t = y_t + \sum_{s=1}^{\infty} (-\theta)^s y_{t-s}$$

Or

$$y_t = \varepsilon_t + \sum_{s=1}^{\infty} (-\theta)^s y_{t-s}$$

so $|\theta| < 1$ is the sufficient condition for a $MA(1)$ process to be invertible. (Powell, Conditions for Stationarity and Invertibility, UC Berkeley.)

In other words, because AR(·) models *with roots of the polynomial outside of the unit circle* are invertible, being able to write the MA(q) part of the process as an AR(∞) with the root condition is sufficient to guarantee invertibility.

2.9.2 Moments of an ARMA(p,q) process

If the process is stationary, $\Phi_p^{-1}(L)$ exists and we can rewrite ARMA (p, q) as MA(∞)

$$Y_t = \mu + \Psi_\infty(L)\varepsilon_t$$

where

$$\mu \equiv \frac{c}{\Phi(1)}; \quad \Phi(1) = 1 - \sum_{j=1}^{p} \phi_j; \quad \Psi_\infty(L) \equiv \Phi_p(L)^{-1}\Theta_q(L)$$

From the results derived for MA(q) we have for $q = \infty$

$$\mathbb{E}(Y_t) = \mu$$
$$\gamma_j = \left(\sum_{i=0}^{\infty} \psi_{j+i}\psi_j\right)\sigma^2$$

where $\psi_0 = 1$

2.9.3 ACF of an ARMA(p,q) process

It is usually easy to identify an AR(p) or MA(q) visually by inspecting its ACF and PACF, because AR's PACF is truncated on p, MA's ACF is truncated on q. For ARMA(p,q) models it is more complicated: both functions are not truncated! Note, however, that in that case, the ACF decays geometrically after lag q and the PACF decays geometrically after lag p.

Model	**ACF**	**PACF**
AR(p)	Decays	Truncated after lag p
MA(q)	Truncated after lag q	Decays
ARMA(p,q)	Decays after lag q	Decays after lag p

2.10 Testing for time dependence

We've seen that a sufficient condition for ergodicity is convergence of the absolute sum of all covariances. This presents a problem: how can we *estimate* these covariances?

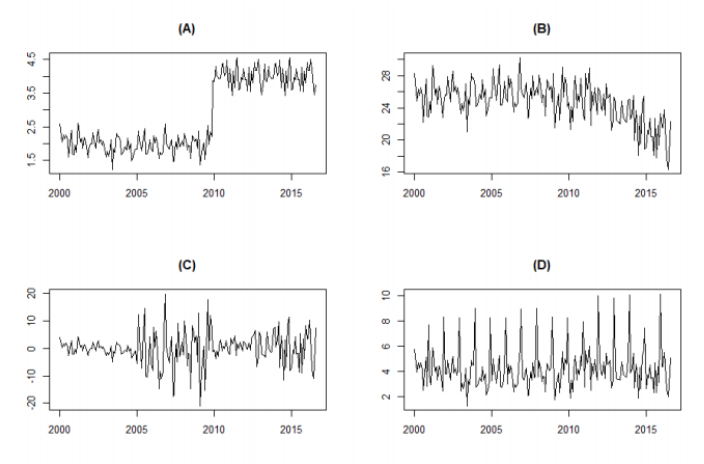
Let Z_t be our series to be tested. Denote the autocovariance of order j as $\gamma_j := Cov(Z_t, Z_{t-j})$. We can try to estimate these parameters with its sample equivalents:

$$\bar{z_t} := \frac{1}{T}\sum_{t=1}^{T} z_t$$

$$\hat{\gamma}_j := \frac{1}{T-j-1}\sum_{t=j+1}^{T}(z_t - \bar{z_t})(z_{t-j} - \bar{z_t})$$

But this is not as simple as it seems. We know that $\hat{\gamma}_j$ converges almost sure to γ_j *if the process is ergodic.* If it *isn't,* the information from $\hat{\gamma}_j$ may not be reliable – after all, we won't have a Law of Large Numbers!

Our solution to this problem won't be very rigorous here. We'll plot $\{\hat{\gamma}_j, j \in \mathbb{N}\}$ and check if it looks stationary. If the series passes this intuitive test, we can assume that $\{\hat{\gamma}_j, j \in \mathbb{N}\}$ will be informative about $\{\gamma_j, j \in \mathbb{N}\}$.

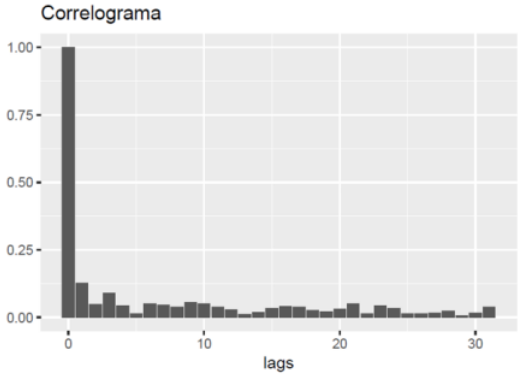


The visual inspection should focus on two main factors: (i) constant mean over time; (ii) constant variance over time. Here are some examples of series to be inspected:

After plotting the time series and assuring that it is well behaved, we can plot its *correlogram:* $\{\hat{\rho}_j := \frac{\hat{\gamma}_j}{\hat{\gamma}_0}, j \in \mathbb{N}\}$. If its sum looks convergent, we will assume that the process is *stationary* and *ergodic* – which will enable us to use sample equivalents as representations of population parameters.

2.10.1 Hypothesis testing

Consider this correlogram:



This series appears to not be correlated with its past. How can we test this?

$$H_0 = \rho_j = 0, \forall j \neq 0$$

This implies, in theory, that we would need to test infinite correlations. In practice, we limit the range to an arbitrary J. Let $\hat{\rho} := (\hat{\rho}_1, \hat{\rho}_2, ..., \hat{\rho}_J)^T, \rho = (\rho_1, \rho_2, ..., \rho_J)^T$. Under the null, $\rho = 0$, and as $T \to \infty$:

$$\sqrt{T}\hat{\rho} \to \mathcal{N}(0, I_J)$$

The intuition here is that, under H_0, $\hat{\rho}$ is a sequence of *iid* variables with mean zero and variance-covariance matrix I_J – which makes the CLT valid.

Given this result, we can now create a statistic that does not depend on the multivariate normal distribution. We will square and sum the expression to arrive at a Chi-squared distribution. This enables us to test the hypothesis with a Wald statistic. Under the null:

$$W_T = T\hat{\rho}^T\hat{\rho} = T\sum_{j=1}^{J}\hat{\rho}_j{}^2 \to \chi^2_J$$

2.10.2 Testing autocorrelations or regressions?

Note that inferring about the autocorrelations is intimately related to inferring in a regression of a time series on its past values. This can be understood by remembering the linear projection interpretation of OLS. Ordinary Least Squares estimation *always* reports the parameters of the linear projection of Y in X, no matter how the model is specified!

Consider the following model:

$$\begin{aligned} Y_t &= \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_J Y_{t_J} + \varepsilon_t \\ &= \alpha + X_t\beta + \varepsilon_t \end{aligned}$$

where $\beta := (\beta_1, \ldots, \beta_J)^T$ and $X_t = (Y_{t-1}, \ldots, Y_{t-j})$. If we define the coefficients of the model above as the parameters of the linear projection of Y_t on the unit vector and X_t, $\alpha = \mu_Y - \mu_X\beta$ where $\mu_Y = \mathrm{E}(Y_t)$ and $\mu_X = \mathrm{E}(X_t)$

Using this result, we have:

$$Y_t - \mu_Y = (X_t - \mu_X)\beta + \varepsilon_t$$

This means that β can be written as:

$$\beta = \mathbb{E}\left[(X_t - \mu_X)^T(X_t - \mu_X)\right]^{-1}\mathbb{E}\left[(X_t - \mu_X)^T(Y_t - \mu_Y)\right] = \Gamma^{-1}\gamma$$

where the matrix Γ^{-1} is symmetric with diagonal elements all equal to $\gamma_1, \gamma_2, \ldots, \gamma_{J-1}$, due to the assumed stationarity. Note that $\mathbb{E}(Y_t - \mu_Y)^2 = \mathbb{E}(Y_{t-j} - \mu_{Y-j})^2 = \gamma_0 \forall j$.

Thus, $\beta = \overrightarrow{0} \iff \gamma = \overrightarrow{0}$, because Γ is a positive definite matrix. This means that *testing* $\beta = 0$ *is equivalent to testing* $\gamma = 0$.

It is important to highlight that this analysis is based upon the inference of $\gamma_j = \mathbb{E}(z_t - \bar{z}_T)(z_{t-j} - \bar{z}_T)$. If we were interested in *other types of relations between* Z_t *and its past, the analysis would have to be adapted* – for example, Z_t^2. It would be necessary to check again for stationarity and ergodicity.

Chapter 3

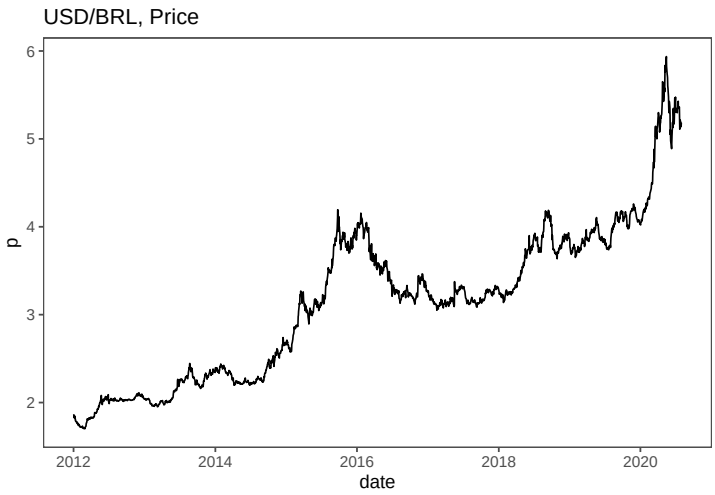
Problem 1: Modelling exchange rates

Loading the database and creating dummy variables:

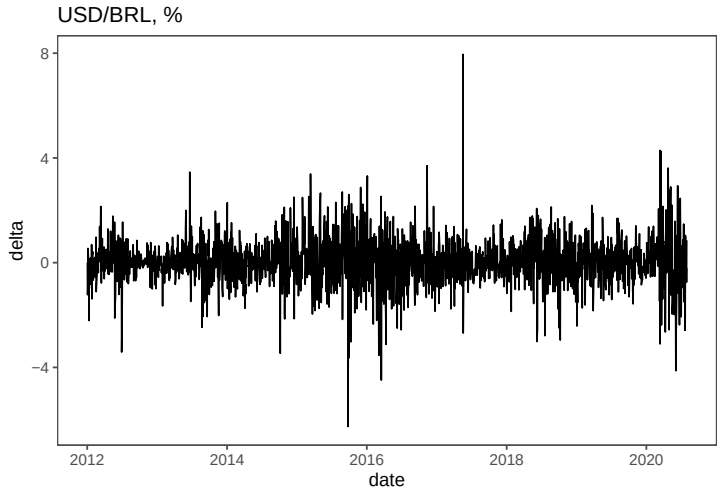
```
df <- read_excel("RS_USD.xlsx")


names(df)[names(df) == "R$/US$"] <- "p"

names(df)[names(df) == "Variação (em %)"] <- "delta"

names(df)[names(df) == "Data"] <- "date"

sign <- as.numeric(df$delta > 0)

count <- c(1:2153)

df <- data.frame(count, df, sign)
```

Before constructing our models, we need to check (intuitively) if the series at hand is *stationary* and *ergodic*. For this, we're going to plot the time series, its autocorrelations and partial autocorrelations.

```
pplot <- ggplot(data = df, aes(x = date, y = p)) + geom_line() +
    ggtitle("USD/BRL, Price") + theme_few()
pplot
```

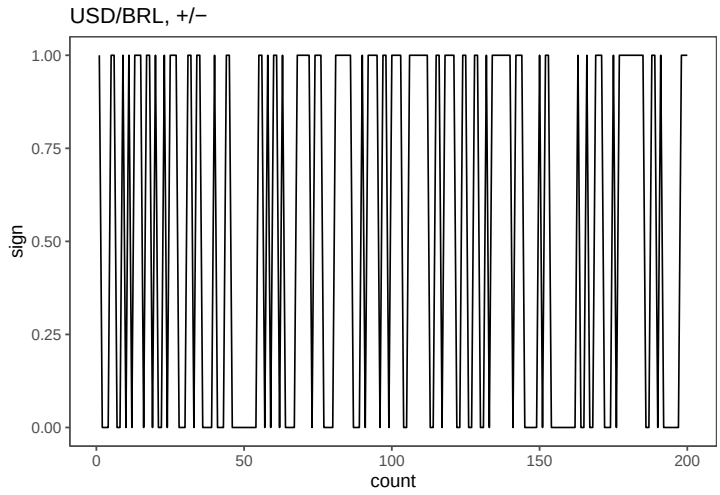


```r
deltaplot <- ggplot(data = df, aes(x = date, y = delta)) + geom_line() +
    ggtitle("USD/BRL, %") + theme_few()
deltaplot
```

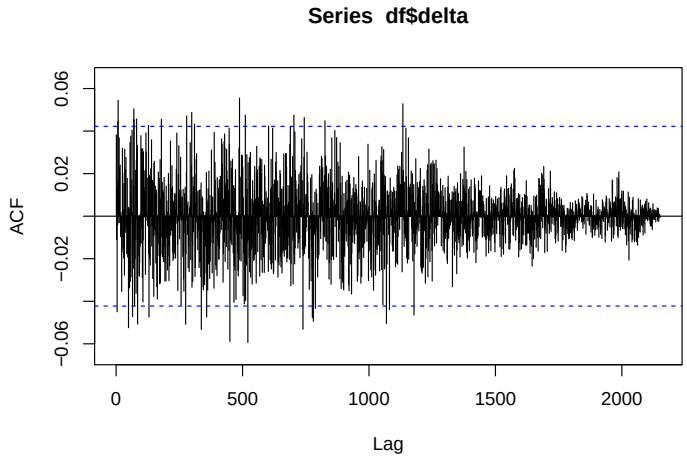



```r
dummyplot <- ggplot(data = df, aes(x = count, y = sign)) + geom_line() +
    ggtitle("USD/BRL, +/-") + xlim(1, 200) + theme_few()
dummyplot
```

```
## Warning: Removed 1953 row(s) containing missing values (geom_path).
```




```r
# For delta

acf_delta <- Acf(df$delta, lag.max = 5000)
```




```
acf_test_values <- acf_delta$acf/sd(acf_delta$acf)

head(data.frame(acf_test_values))
```

```
##   acf_test_values
## 1      37.9547672
## 2       1.4506537
## 3      -0.4173129
## 4       0.2125873
## 5      -1.7053782
## 6       0.5358210
```

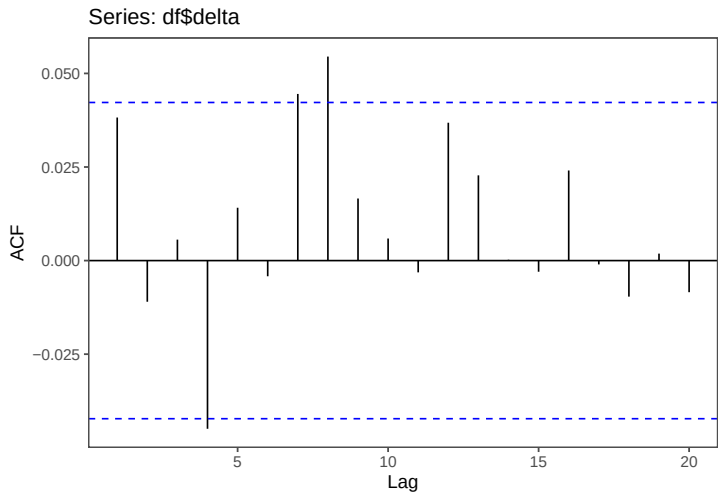
```
facst <- ggAcf(df$delta, type = "correlation", lag.max = 20,
    plot = T) + theme_few()
faclt <- ggAcf(df$delta, type = "correlation", lag.max = 5000,
    plot = T) + theme_few()

facpst <- ggPacf(df$delta, type = "correlation", lag.max = 100,
    plot = T) + theme_few()
```

```
## Warning: Ignoring unknown parameters: type
```

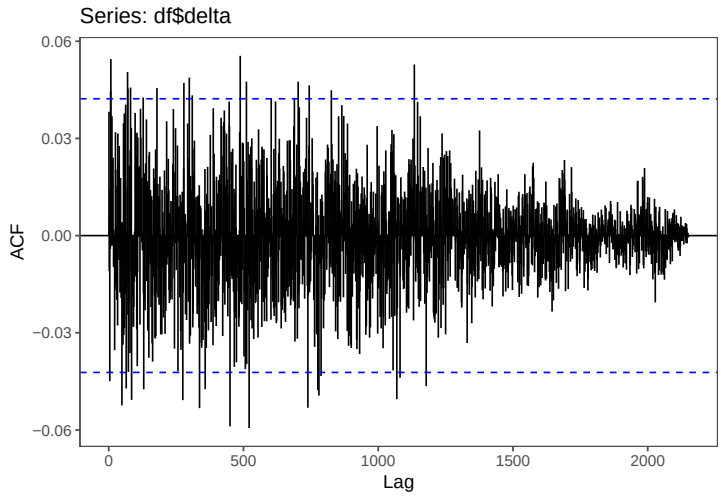
```
facplt <- ggPacf(df$delta, type = "correlation", lag.max = 5000,
    plot = T) + theme_few()
```

```
## Warning: Ignoring unknown parameters: type
```

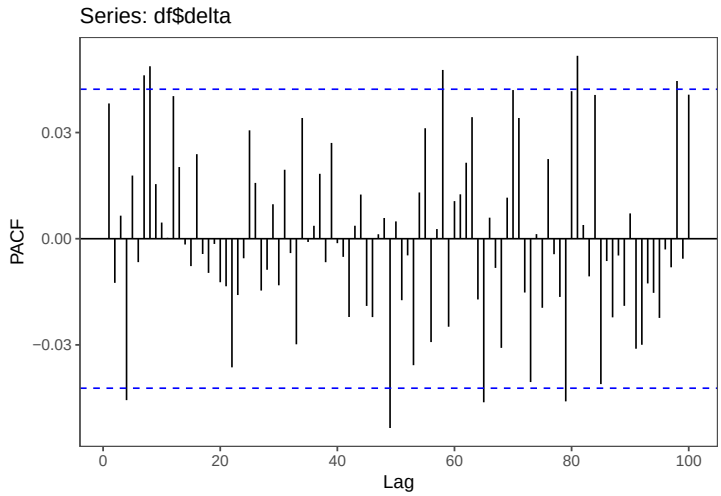
```
facst
```



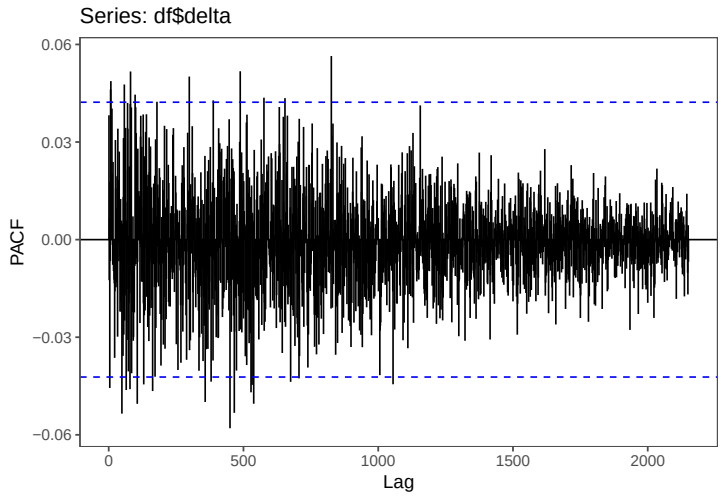

```
faclt
```



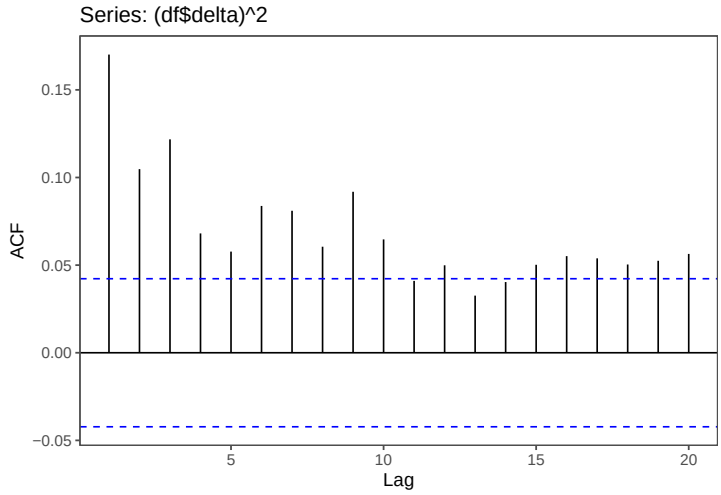

```
facpst
```



```
facplt
```




```
facst2 <- ggAcf((df$delta)^2, type = "correlation", lag.max = 20,
    plot = T) + theme_few()
facst2
```



Let's now create our first ARMA models (equivalent to ARIMA with 2nd argument = 0). We'll begin with the first hypothesis: $\mathbb{P}(+) = \mathbb{P}(-)$. Modelling this with an AR(1), we have:

$$Sign_{t+1} = \alpha + \beta Sign_t + \varepsilon, \qquad \varepsilon \sim wn(0, \sigma^2)$$

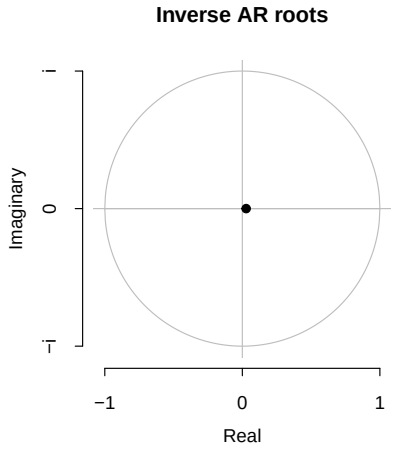
In R, we'll use the package *forecast* to construct this model:

```
AR1sign <- Arima(df$sign, order = c(1, 0, 0))
summary(AR1sign)
```

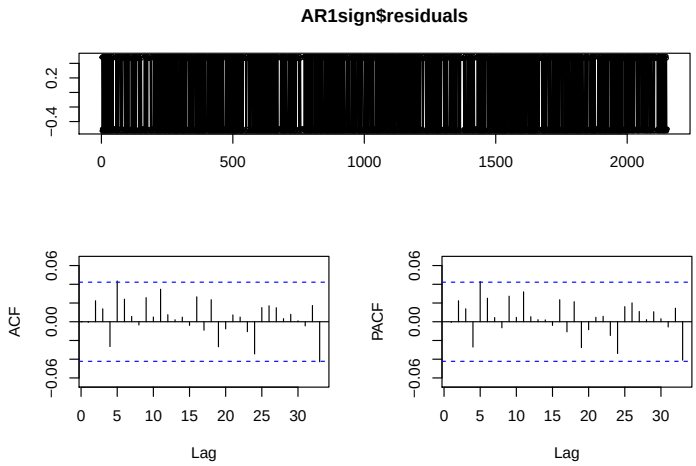
```
## Series: df$sign
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##          ar1    mean
##       0.0278  0.5165
```

```
## s.e.  0.0215  0.0111
##
## sigma^2 estimated as 0.2498:  log likelihood=-1560.63
## AIC=3127.26   AICc=3127.27   BIC=3144.28
##
## Training set error measures:
##                       ME      RMSE       MAE  MPE MAPE     MASE          ACF1
## Training set 2.157119e-05 0.4995356 0.4990724 -Inf  Inf 1.027755 -0.0006313563
```

```
plot(AR1sign)
```




```
tsdisplay(AR1sign$residuals)
```



With the results of the summary, we can now apply a hypothesis test for our first question.[1]

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

[1] Testing β is equivalent to testing γ.

$\frac{a\hat{r}_1 - ar_1}{s.e.(ar_1)}$:

```
AR1sign$coef[1]/sqrt(AR1sign$var.coef[1, 1])
```

```
##      ar1
## 1.287942
```

The second hypothesis in the problem refers to the delta of the variation:

$$\mathbb{E}(\Delta|+) \neq \mathbb{E}(\Delta|-).$$

$$\Delta_{t+1} = \alpha + \beta Sign_t + \varepsilon, \qquad \varepsilon \sim wn(0, \sigma^2).$$

```
lmsignt <- lm(delta ~ lag(df$sign, k = 1), data = df)
summary(lmsignt)
```

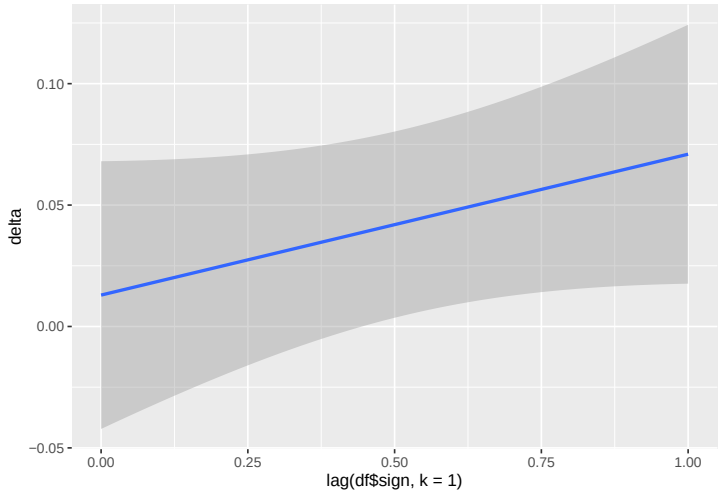
```
##
## Call:
## lm(formula = delta ~ lag(df$sign, k = 1), data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3285 -0.4706 -0.0060  0.4655  7.9442
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           0.01293    0.02811   0.460    0.646
## lag(df$sign, k = 1)   0.05802    0.03911   1.484    0.138
##
## Residual standard error: 0.9066 on 2150 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.001023,   Adjusted R-squared:  0.000558
## F-statistic: 2.201 on 1 and 2150 DF,  p-value: 0.1381
```

```
ggplot(df, aes(x = lag(df$sign, k = 1), y = delta)) + geom_smooth(method = "lm")
```

```
## Warning: Use of `df$sign` is discouraged. Use `sign` instead.
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```



$$\Delta_{t+1} = \alpha + \beta_1 \Delta_t + \beta_2 Sign_t + \varepsilon, \qquad \varepsilon \sim wn(0, \sigma^2)$$

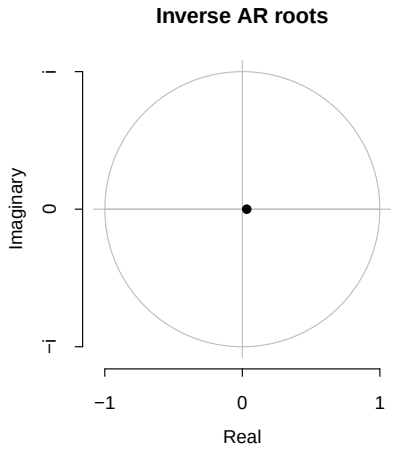
```
AR1delta <- Arima(df$delta, order = c(1, 0, 0), xreg = lag(df$sign,
    k = 1))
summary(AR1delta)
```

```
## Series: df$delta
## Regression with ARIMA(1,0,0) errors
##
## Coefficients:
##          ar1  intercept    xreg
##       0.0321     0.0343  0.0166
## s.e.  0.0306     0.0351  0.0556
##
## sigma^2 estimated as 0.8219:  log likelihood=-2840.96
## AIC=5689.93   AICc=5689.94   BIC=5712.62
##
## Training set error measures:
##                        ME      RMSE      MAE MPE MAPE      MASE         ACF1
## Training set 1.147232e-05 0.9059341 0.643417 NaN  Inf 0.7252668 0.0003998294
```

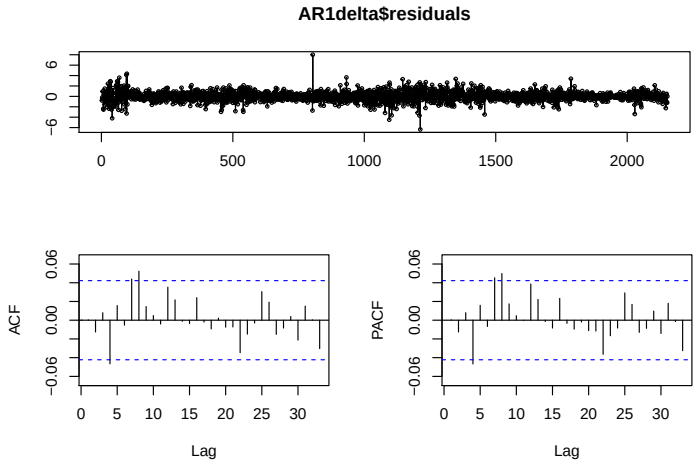
```
AR1delta$coef[1]/sqrt(AR1delta$var.coef[1, 1])
```

```
##     ar1
## 1.04747
```

```
plot(AR1delta)
```




```
tsdisplay(AR1delta$residuals)
```



The last hypothesis in the problem refers to the variance:

$$\mathbb{E}(\Delta_{t+1}^2|\Delta_t).$$

$$\Delta_{t+1}^2 = \alpha + \beta\Delta_t^2 + \varepsilon, \qquad \varepsilon \sim wn(0, \sigma^2)$$

```
AR1var <- Arima((df$delta)^2, order = c(1, 0, 0))
summary(AR1var)
```

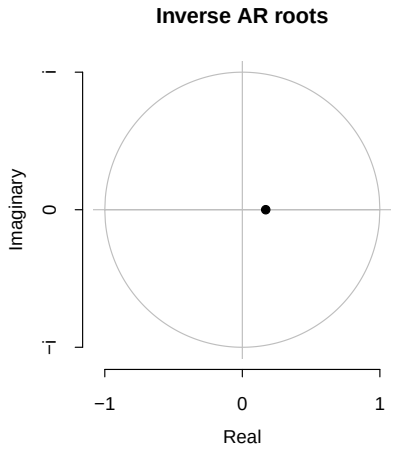
```
## Series: (df$delta)^2
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##          ar1    mean
##       0.1701  0.8238
## s.e.  0.0212  0.0582
##
## sigma^2 estimated as 5.026:  log likelihood=-4792.09
```

```
## AIC=9590.17   AICc=9590.18   BIC=9607.2
##
## Training set error measures:
##                         ME     RMSE       MAE  MPE MAPE      MASE         ACF1
## Training set -6.299133e-05 2.240784 0.9197335 -Inf  Inf 0.8595655 -0.01320252
```

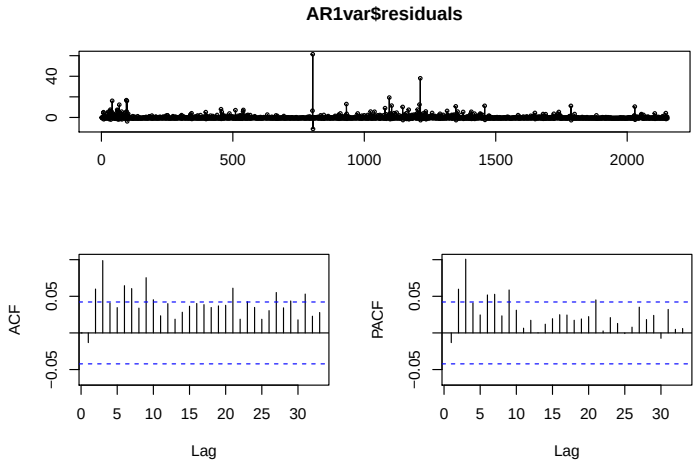
```
AR1var$coef[1]/sqrt(AR1var$var.coef[1, 1])
```

```
##      ar1
## 8.011092
```

```
plot(AR1var)
```




```
tsdisplay(AR1var$residuals)
```



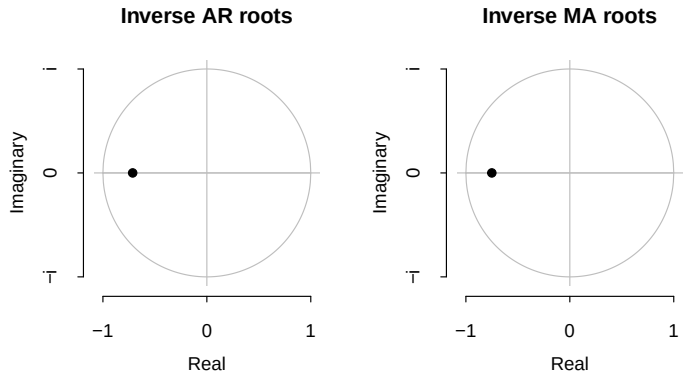
Now, let's run *auto.arima*.

```
aadelta <- auto.arima(df$delta, stepwise = F)
summary(aadelta)
```

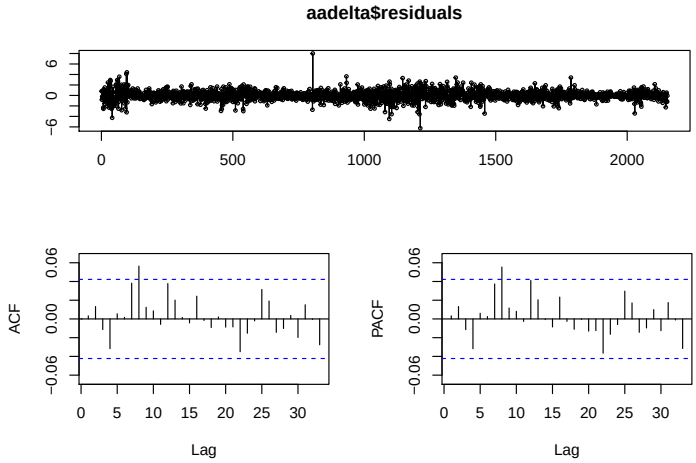
```
## Series: df$delta
## ARIMA(1,0,1) with non-zero mean
##
```

```
## Coefficients:
##           ar1     ma1    mean
##       -0.7138  0.7506  0.0433
## s.e.   0.1486  0.1399  0.0199
##
## sigma^2 estimated as 0.8208:  log likelihood=-2840.87
## AIC=5689.74   AICc=5689.76   BIC=5712.44
##
## Training set error measures:
##                       ME      RMSE       MAE MPE MAPE    MASE        ACF1
## Training set 2.610156e-05 0.9053381 0.6434199 NaN  Inf 0.72527 0.003320553
```

```
plot(aadelta)
```




```
tsdisplay(aadelta$residuals)
```



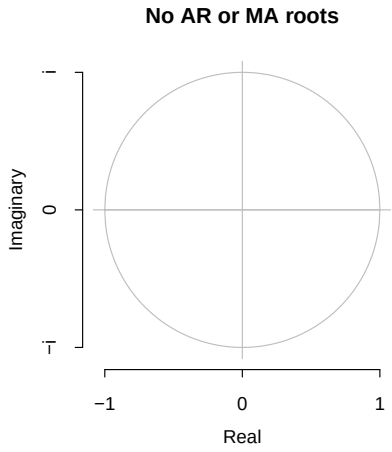

```
aasign <- auto.arima(df$sign, stepwise = F)
summary(aasign)
```

```
## Series: df$sign
## ARIMA(0,0,0) with non-zero mean
```

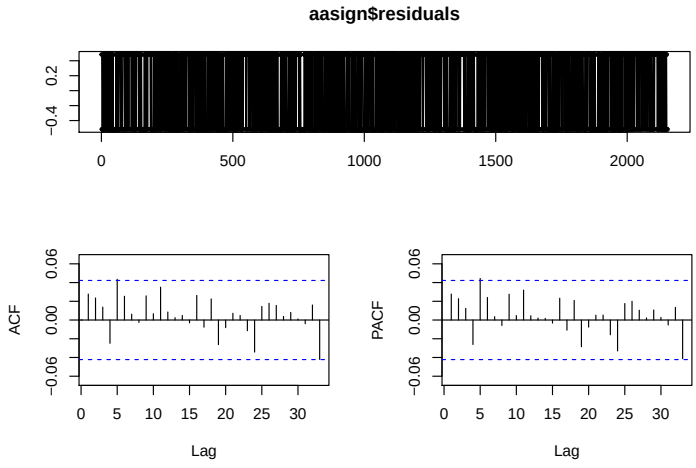
```
##
## Coefficients:
##         mean
##       0.5165
## s.e.  0.0108
##
## sigma^2 estimated as 0.2498:  log likelihood=-1561.46
## AIC=3126.91   AICc=3126.92   BIC=3138.26
##
## Training set error measures:
##                       ME      RMSE       MAE  MPE MAPE     MASE       ACF1
## Training set -2.382602e-13 0.4997281 0.4994563 -Inf  Inf 1.028545 0.02773919
```

```
plot(aasign)
```

```
## Warning in plot.Arima(aasign): No roots to plot
```



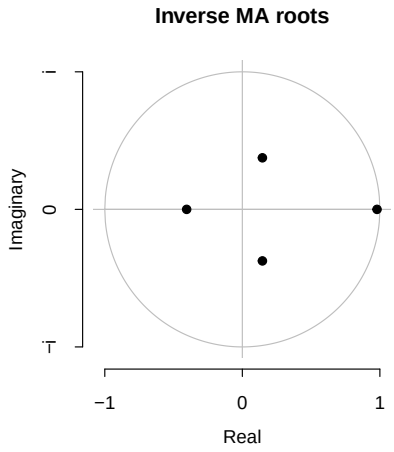

```
tsdisplay(aasign$residuals)
```



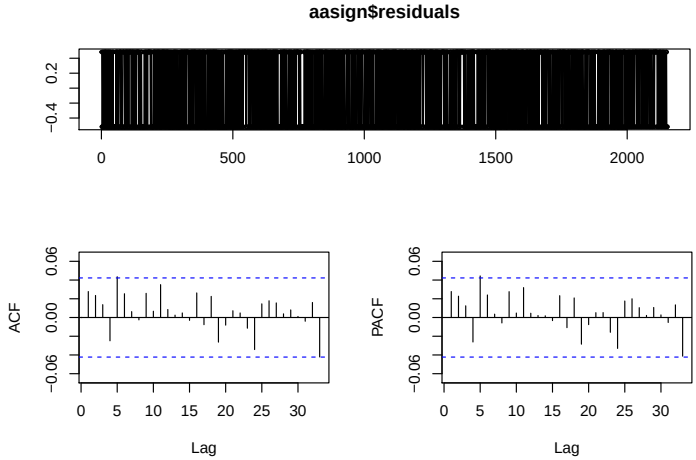

```
aavar <- auto.arima((df$delta)^2, stepwise = F)
summary(aavar)
```

```
## Series: (df$delta)^2
## ARIMA(0,1,4)
##
## Coefficients:
##           ma1      ma2     ma3      ma4
##       -0.8662  -0.0671  0.0228  -0.0641
## s.e.   0.0215   0.0284  0.0280   0.0214
##
## sigma^2 estimated as 4.892:  log likelihood=-4761.22
## AIC=9532.45   AICc=9532.48   BIC=9560.82
##
## Training set error measures:
##                       ME     RMSE       MAE  MPE MAPE      MASE         ACF1
## Training set -0.02170361 2.209213 0.8904046 -Inf  Inf 0.8321553 0.0001189823
```

```
plot(aavar)
```



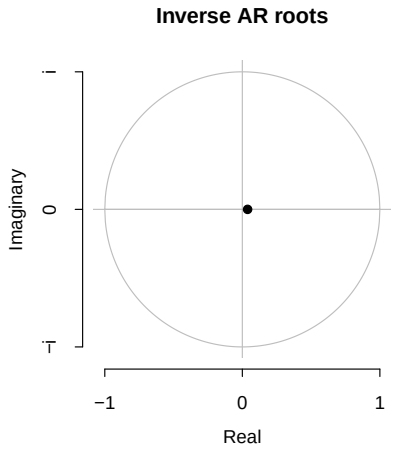

```
tsdisplay(aasign$residuals)
```



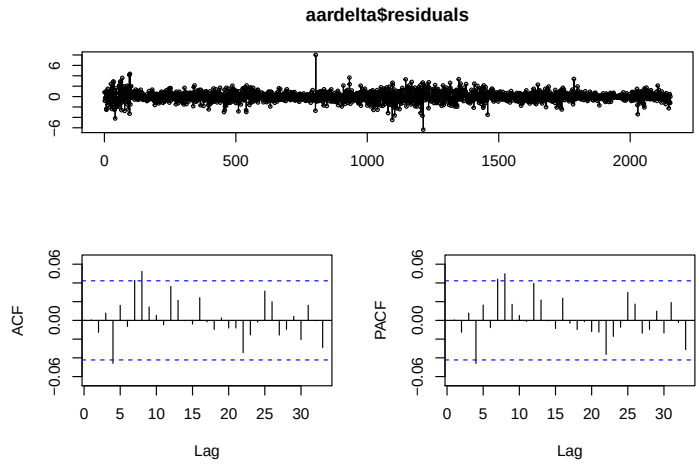

```
aardelta <- auto.arima(df$delta, max.q = 0, stepwise = F)
summary(aardelta)
```

```
## Series: df$delta
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##          ar1    mean
##       0.0382  0.0433
## s.e.  0.0215  0.0203
##
## sigma^2 estimated as 0.8215:  log likelihood=-2842.26
## AIC=5690.52   AICc=5690.53   BIC=5707.54
##
## Training set error measures:
##                        ME      RMSE       MAE MPE MAPE      MASE         ACF1
## Training set -1.464203e-05 0.9059233 0.6434299 NaN  Inf 0.7252813 0.0004805619
```

```
plot(aardelta)
```




```
tsdisplay(aardelta$residuals)
```

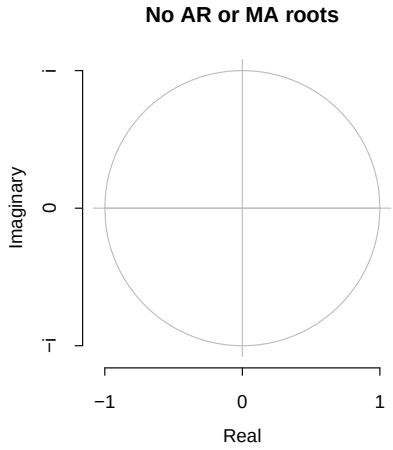



```
aarsign <- auto.arima(df$sign, max.q = 0, stepwise = F)
summary(aarsign)
```

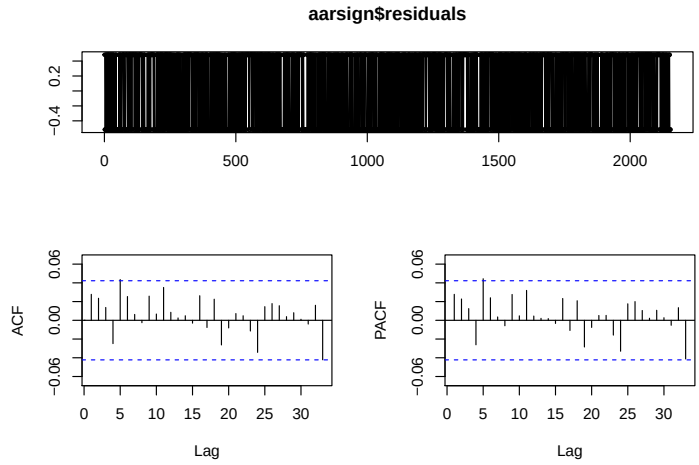
```
## Series: df$sign
## ARIMA(0,0,0) with non-zero mean
##
## Coefficients:
##         mean
##       0.5165
## s.e.  0.0108
##
## sigma^2 estimated as 0.2498:  log likelihood=-1561.46
## AIC=3126.91   AICc=3126.92   BIC=3138.26
##
## Training set error measures:
##                       ME      RMSE       MAE  MPE MAPE     MASE       ACF1
## Training set -2.382602e-13 0.4997281 0.4994563 -Inf  Inf 1.028545 0.02773919
```

```
plot(aarsign)
```

```
## Warning in plot.Arima(aarsign): No roots to plot
```



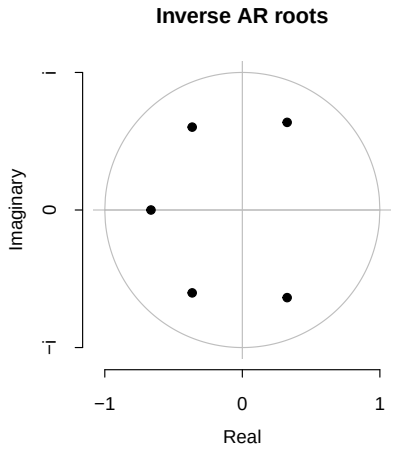

```
tsdisplay(aarsign$residuals)
```



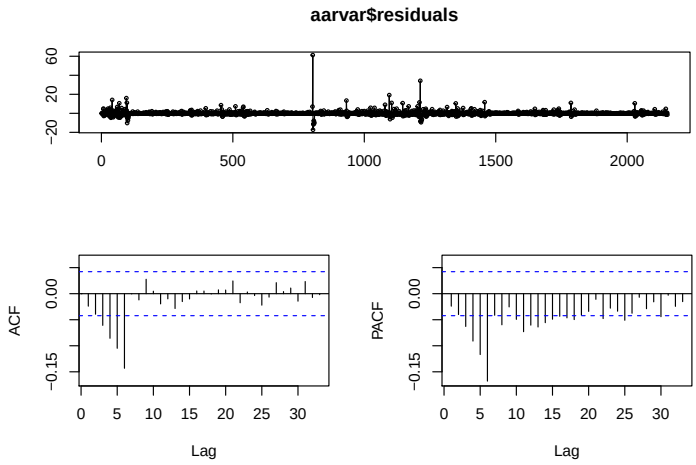
```r
aarvar <- auto.arima((df$delta)^2, max.q = 0, stepwise = F)
summary(aarvar)
```

```
## Series: (df$delta)^2
## ARIMA(5,1,0)
##
## Coefficients:
##           ar1      ar2      ar3      ar4      ar5
##       -0.7420  -0.5845  -0.4042  -0.2873  -0.1687
## s.e.   0.0213   0.0259   0.0274   0.0258   0.0212
##
## sigma^2 estimated as 5.465:  log likelihood=-4878.95
## AIC=9769.89   AICc=9769.93   BIC=9803.94
##
## Training set error measures:
##                        ME     RMSE       MAE  MPE MAPE      MASE       ACF1
## Training set -3.234587e-05 2.334504 0.9170278 -Inf  Inf 0.8570369 -0.0239684
```

```r
plot(aarvar)
```




```r
tsdisplay(aarvar$residuals)
```



$$\Delta_{t+1} = c + \beta\Delta_t + \varepsilon$$

```
AR1_2 <- Arima(df$delta, order = c(1, 0, 0))

summary(AR1_2)
```

```
## Series: df$delta
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##          ar1    mean
##       0.0382  0.0433
## s.e.  0.0215  0.0203
##
## sigma^2 estimated as 0.8215:  log likelihood=-2842.26
## AIC=5690.52   AICc=5690.53   BIC=5707.54
##
## Training set error measures:
##                        ME      RMSE       MAE MPE MAPE      MASE         ACF1
## Training set -1.464203e-05 0.9059233 0.6434299 NaN  Inf 0.7252813 0.0004805619
```

```
confint(AR1_2, level = 0.95)
```

```
##                  2.5 %     97.5 %
## ar1       -0.003988796 0.08042284
## intercept  0.003511958 0.08308440
```

Chapter 4

Problem 2: Estimating ARMA models

Consider an ARMA($\cdot$) model of the form:

$$\begin{aligned} Y_t =& C + \phi_1 Y_{t-1} + \phi Y_{t-2} + \cdots + \phi_P Y_{t-p} + \varepsilon_t \\ &+ \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \end{aligned}$$

with white noise ε_+:

$$\begin{aligned} \mathbb{E}(\varepsilon_t) &= 0 \\ E(\varepsilon_t \varepsilon_\tau) &= \begin{cases} \sigma^2, & \forall t = \tau \\ 0, & c \cdot c \end{cases} \end{aligned}$$

We'll use MLE, with $\Theta = (c, \phi_1, ..., \phi_p, \theta_1, ..., \theta_q)'$ a vector of parameters. This means that we need to calculate the pdf: $f_{Y_T, Y_{T-1}, ...}(y_T, y_{T-1}, ...; \Theta)$. We usually assume a normal distribution for ε.

4.1 MLE for Gaussian AR(1)

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t, \quad (\Theta = (c, \theta, \sigma^2)')$$

$$\mathbb{E}(Y_1) = \mu = \frac{c}{1-\phi}$$

$$\mathbb{E}(Y_1 - \mu)^2 = \frac{\sigma^2}{1-\phi^2}$$

The likelihood function is given by:

$$f_{Y_1}(y_1; \Theta) * \Pi_{t=2}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \Theta)$$

$$\mathcal{L}(\Theta) = log\, f_{Y_1}(y_1; \Theta) + \sum_{t=2}^{T} log\, f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \Theta)$$

4.1.1 Conditional MLE

$$\Pi_{t=2}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\Theta)$$

For the gaussian case, the minimization problem is equivalent to minimizing:

$$\sum_{t=2}^{T}(y_t - c - \phi y_{t-1})^2$$

This means that the *conditional MLE is equivalent to OLS!*

4.2 Likelihood funciton for AR(p)

$$Y_t = c + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim wn(0, \sigma^2), \quad \Theta = (c, \phi_1, ..., \phi_p, \sigma^2)$$

$$\mu = \frac{c}{1 - \phi_1 - ... - \phi_p}$$

$$\sigma^2 V_p \text{is the var-cov matrix}$$

$$f_{Y_p, Y_{p-1}, ...}(y_p, y_{p-1}, ...; \Theta) * \Pi_{t=p+1}^{T} f_{Y_t|Y_{t-1},...,Y_{t-p}}(y_t|y_{t-1}, ..., y_{t-p}; \Theta)$$

The first term represents the distribution of the p first observations. The second term is the *prediction error decomposition.*

4.2.1 Conditional MLE

Again, we can apply traditional OLS estimation in the conditional MLE case, because it minimizes:

$$\sum_{t=p+1}^{T} (Y_t - c - \phi_1 Y_{t-1} - ... - \phi_p Y_{t-p})^2$$

Thus, the variance estimator is also the one that we are used to:

$$\hat{\sigma}^2 = \frac{1}{T-p} \sum_{t=p+1}^{T} \left(y_t - \hat{c} - \hat{\phi}_1 y_{t-1} - \cdots - \hat{\phi}_p y_{t-p}\right)^2$$

4.3 MLE for Gaussian MA(·)

Let's now begin with the Conditional MLE.

$$Y_t = \mu + \theta \varepsilon_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim wn(0, \sigma^2), \quad \Theta = (\mu, \theta, \sigma^2)$$

$$\mathcal{L}(\Theta) = log\, f_{Y_t, Y_{t-1}, ..., Y_1|\varepsilon_0 = 0}(y_T, y_{T-1}, ..., y_1|\varepsilon_0 = 0; \Theta)$$

$$\mathcal{L}(\Theta) = -\frac{T}{2} log\,(2\pi) - \frac{T}{2} log\,(\sigma^2) - \sum_{t=1}^{T} \frac{\varepsilon_t^2}{2\sigma^2}$$

If $|\theta| < 1$, the effects of imposing $\varepsilon_0 = 0$ diminish over time and the conditional MLE is a good approximation. Otherwise, it is not – the effects build up over time. This is why an essential condition for an ARMA(p,q) model is *invertibility.*

The same idea can be generalized for an MA(q) model. If we fix the first q values of ε as 0: $\varepsilon_0 = \varepsilon_{-1} = ... = \varepsilon_{-q+1} = 0$, we can iterate on the values of the innovation terms:

$$\varepsilon_t = y_t - \mu - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}.$$

4.4 Conditional MLE estimation for ARMA(p,q) models

Uniting both of the results shown above, we can apply conditional MLE estimation for an ARMA(p,q) model:

$$Y_t = c + \phi Y_{t-1} + ... + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad \varepsilon_t \sim wn(0, \sigma^2)$$

This is possible by taking fixed initial values for $y_0 = (y_0, ..., y_{-p+1})', \varepsilon_0 = (\varepsilon_0, ..., \varepsilon_{-q+1})'$ as given. Then, we apply the log-likelihood function:

$$\mathcal{L}(\Theta) = -\frac{T}{2} log\,(2\pi) - \frac{T}{2} log\,(\sigma^2) - \sum_{t=1}^{T} \frac{\varepsilon_t^2}{2\sigma^2}$$

Chapter 5

Problem 3: Identification of ARMA models

In this problem, we'll be tackling the issue of *identification* of an ARMA model. Namely, we will employ the *Box-Jenkins* model selection strategy, based upon the concept of *parsimony*.

The principle of *parsimony* is inspired on the trade-off between *fit*, i.e., R^2, and *degrees of freedom*. "Box and Jenkins argue that parsimonious models produce better forecasts than overparametrized models". (p. 76)

The Box-Jenkins strategy is divided in three main stages:

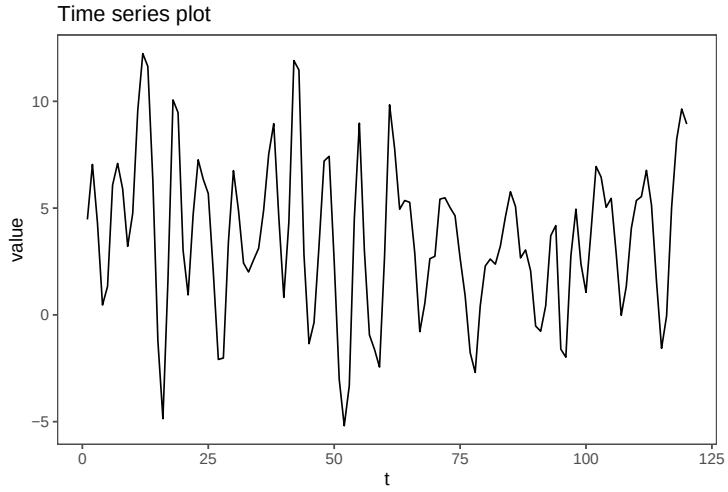
- Identification;
- Estimation;
- Diagnostic checking.

These estimations depend upon two essential conditions (discussed in earlier problems and lectures): *stationarity* and *invertibility*. Stationarity, as we have discussed earlier, is necessary to effectively *employ econometric methods* and to infer characteristics of a population through a given sample. Enders also points out that t-statistics and Q-statistics are based upon the assumption that the data are stationary (p. 77). This implies a condition on the AR process of an ARMA model (roots of characteristic polynomial outside of unity circle).

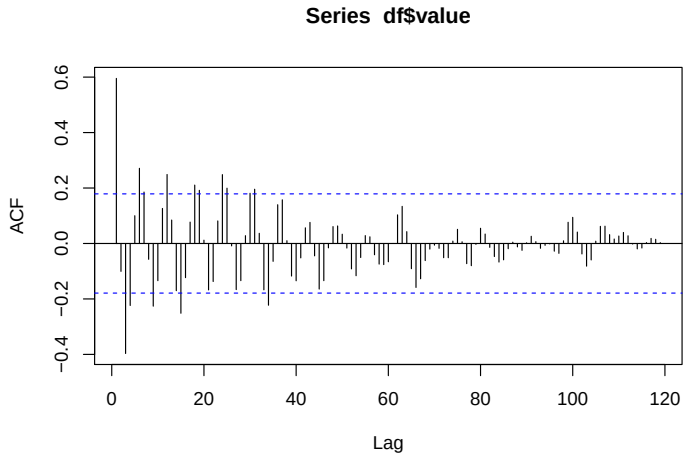
Furthermore, the model shall be *invertible* – i.e., if it can be represented by a finite or convergent AR model. This implies a condition on the MA process – i.e., if it can be written as an AR(∞).

We're going to check these conditions intuitively by plotting the ACFs and PACFs of the time series:

```
df <- data.frame(df)

pplot <- ggplot(data = df, aes(x = t, y = value)) + geom_line() +
    ggtitle("Time series plot") + theme_few()
pplot
```




```
acf_ts <- Acf(df$value, lag.max = 5000)
```




```
acf_test_values <- acf_ts$acf/sd(acf_ts$acf)

head(data.frame(acf_test_values))
```

```
##   acf_test_values
## 1       6.5814152
## 2       3.9180772
## 3      -0.6619326
## 4      -2.6109255
## 5      -1.4713722
## 6       0.6589976
```

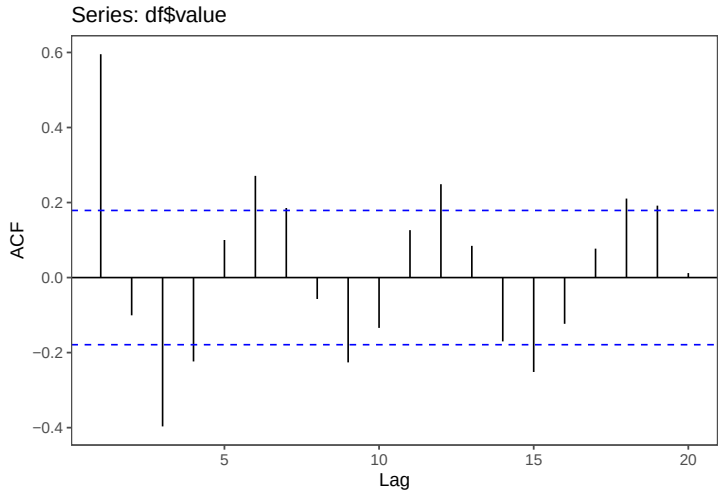
```
facst <- ggAcf(df$value, type = "correlation", lag.max = 20,
plot = T) + theme_few()
faclt <- ggAcf(df$value, type = "correlation", lag.max = 5000,
plot = T) + theme_few()

facpst <- ggPacf(df$value, type = "correlation", lag.max = 100,
plot = T) + theme_few()
```

```
## Warning: Ignoring unknown parameters: type
```

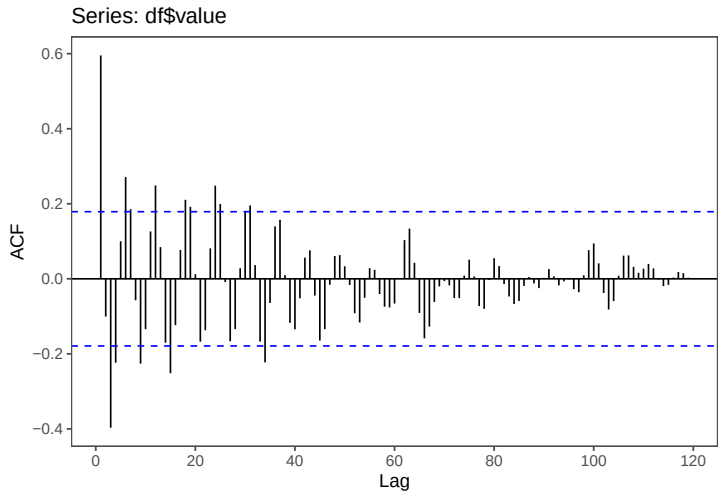
```
facplt <- ggPacf(df$value, type = "correlation", lag.max = 5000,
plot = T) + theme_few()
```

```
## Warning: Ignoring unknown parameters: type
```

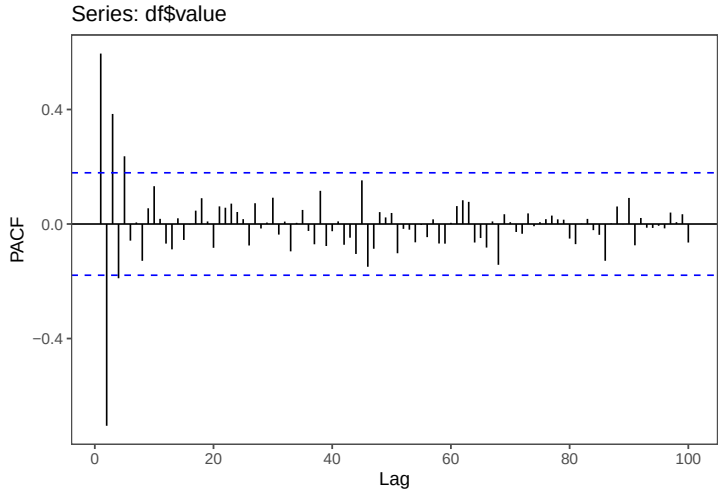
```
facst
```



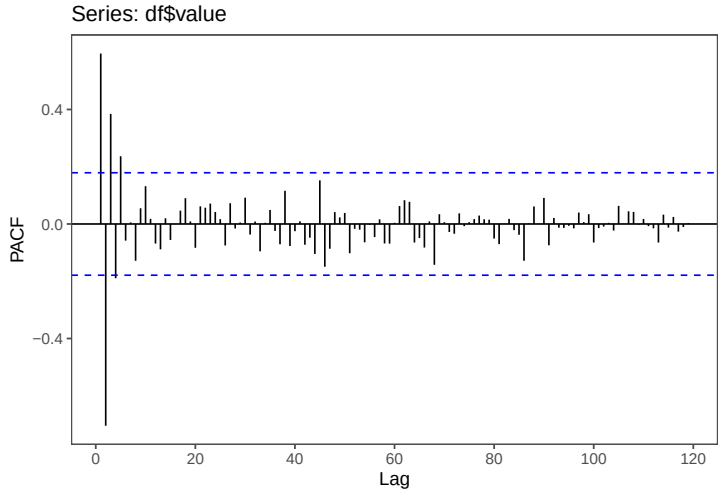

```
faclt
```




```
facpst
```




```
facplt
```



Aside from usual methods, we'll employ the following criteria:

- Akaike Information Criterion (AIC).

$$AIC = T * ln(SSR) + 2n$$

- Schwartz Bayesian Criterion (SBC).

$$SBC = T * ln(SSR) + n * ln(T)$$

n denotes the number of parameters estimated (an useful metric given the importance of the degrees of freedom). T denotes the number of *usable* observations. Note that, when comparing different models, it is important to *fix* T to ensure that the AIC and SBC values are comparable and are capturing only variations in the actual model and not the effect of changing T.

The objective with these criteria is to *minimize* their values. "As the fit of the model improves, the AIC and SBC will approach -∞." (p. 70) AIC and SBC have different advantages and drawbacks: while the former is biased toward overparametrization and more powerful in small

samples, SBC is consistent and has superior large sample properties. If both metrics point to the same model, we should be fairly confident that it is, indeed, the correct specification.

It is also important to apply hypothesis tests to the estimates of the population parameters μ, σ^2 and $\rho_s - \bar{y}, \hat{sigma}^2, r_s$, respectively. Worthy of note here is r_s, which presents the following distributions under the null that y_t is stationary with $\varepsilon_t \sim \mathcal{N}$:

$$Var(r_s) = T^{-1} \qquad for\, s = 1$$

$$Var(r_s) = T^{-1}(1 + 2\sum_{j=1}^{s-1} r_j^2) \qquad for\, s > 1$$

The Q-statistic is also introduced by Enders in this chapter. It is used to test whether a group of autocorrelations is significantly different from zero.

$$Q = T\sum_{k=1}^{s} r_k^2$$

Under the null of $r_k = 0 \forall k$, Q is asymptotically χ^2 with s degrees of freedom. "Certainly, a white-noise process (in which all autocorrelations should be zero) would have a Q value of zero". (p. 68)

An alternative form for Q is presented by Ljung and Box (1978):

$$Q = T(T+2)\sum_{k=1}^{s} \frac{r_k^2}{(T-k)}$$

Furthermore, it is also important to check whether the residuals of the model are actually *white noise*. This can be done via the Q-statistic, which *should not result in the rejection of the null*. If that is not the case, the model specified is not the best one available, as there's still a relevant underlying variable (y or ε).

Let's now perform the *estimation stage*. This shall be done via the function *auto.arima* from the package *forecast*.

```
aa_model <- auto.arima(df$value, num.cores = 24, max.d = 0, max.D = 0,
stepwise = F)

summary(aa_model)
```

```
## Series: df$value
## ARIMA(2,0,1) with non-zero mean
##
## Coefficients:
##          ar1      ar2    ma1    mean
##       0.7524  -0.5545  0.797  3.5305
## s.e.  0.0818   0.0813  0.064  0.3485
##
## sigma^2 estimated as 3.002:  log likelihood=-235.87
## AIC=481.75   AICc=482.27   BIC=495.68
##
## Training set error measures:
```

```
##                       ME     RMSE      MAE      MPE    MAPE      MASE
## Training set 0.01363546 1.703559 1.426984 5.237664 82.3373 0.5612557
##                      ACF1
## Training set 0.004670695
```

```
print("t-values: ")
```

```
## [1] "t-values: "
```

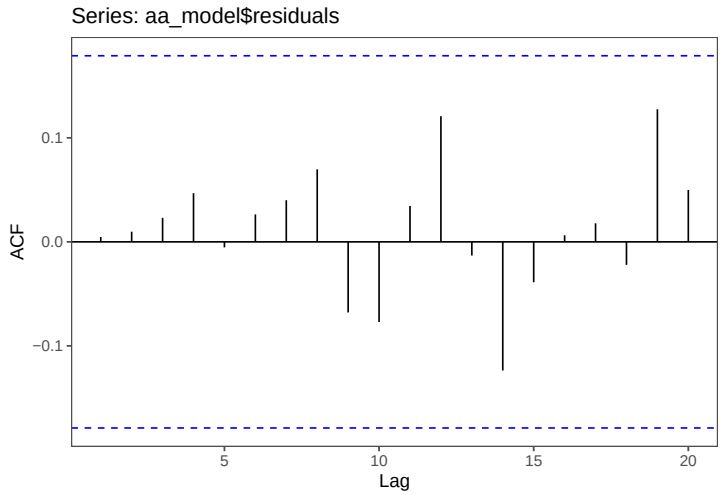
```
aa_t <- matrix(NA, nrow = 4)

for (i in c(1:4)) {

    aa_t[i] <- aa_model$coef[i]/sqrt(aa_model$var.coef[i, i])

}

aa_t <- data.frame(aa_t)

aa_t
```

```
##        aa_t
## 1  9.203615
## 2 -6.822352
## 3 12.444488
## 4 10.129782
```

```
aa_q <- Box.test(aa_model$residuals, lag = aa_model$arma[1] +
    aa_model$arma[2])
aa_q
```

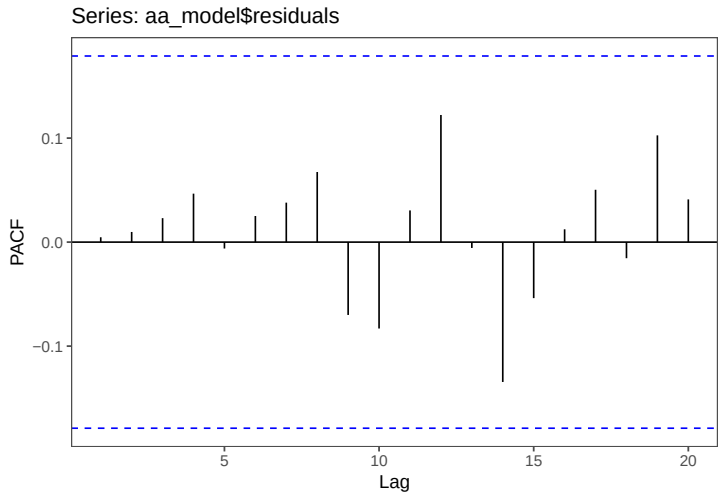
```
##
##  Box-Pierce test
##
## data:  aa_model$residuals
## X-squared = 0.078351, df = 3, p-value = 0.9943
```

```
ggAcf(aa_model$residuals, type = "correlation", lag.max = 20,
plot = T) + theme_few()
```




```
ggPacf(aa_model$residuals, type = "correlation", lag.max = 20,
plot = T) + theme_few()
```

```
## Warning: Ignoring unknown parameters: type
```



The results of *auto.arima* imply that the best model is an ARMA(2,1):

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim wn(0, \sigma^2)$$

Furthermore, the Q-statistic *(Box.test)* seems to indicate that ε_t is truly white noise.

Let's now run some different models and compare them against the results of *auto.arima*. We'll begin with some overspecified model. First, an ARMA(2,2):

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim wn(0, \sigma^2)$$

```
arma22 <- Arima(df$value, order = c(2, 0, 2))

summary(arma22)
```

```
## Series: df$value
## ARIMA(2,0,2) with non-zero mean
##
## Coefficients:
##          ar1      ar2     ma1     ma2    mean
##       0.7417  -0.5502  0.8113  0.0149  3.5311
## s.e.  0.1442   0.0950  0.1692  0.1631  0.3514
##
## sigma^2 estimated as 3.028:  log likelihood=-235.87
## AIC=483.74   AICc=484.48   BIC=500.46
##
## Training set error measures:
##                       ME     RMSE      MAE      MPE     MAPE      MASE
## Training set 0.01360342 1.703508 1.426237 4.791703 81.74486 0.5609619
##                      ACF1
## Training set 0.001161589
```

```
arma22_t <- matrix(NA, nrow = 5)

for (i in c(1:5)) {
```

```
    arma22_t[i] <- arma22$coef[i]/sqrt(arma22$var.coef[i, i])

}

arma22_t <- data.frame(arma22_t)

arma22_t
```

```
##      arma22_t
## 1  5.14206433
## 2 -5.78853038
## 3  4.79577309
## 4  0.09134106
## 5 10.04968297
```

```
arma22_q <- Box.test(arma22$residuals, lag = arma22$arma[1] +
    arma22$arma[2])
arma22_q
```

```
##
##  Box-Pierce test
##
## data:  arma22$residuals
## X-squared = 0.26458, df = 4, p-value = 0.992
```

The t-value of $ma2$ is not able to reject the null hypothesis. Furthermore, the Q-statistic *(Box.test)* seems to indicate that ε_t is truly white noise.

Now, an ARMA(3,1):

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \Phi_3 y_{t-3} + \theta_1 \varepsilon_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim wn(0, \sigma^2)$$

```
arma31 <- Arima(df$value, order = c(3, 0, 1))

summary(arma31)
```

```
## Series: df$value
## ARIMA(3,0,1) with non-zero mean
##
## Coefficients:
##          ar1     ar2     ar3     ma1    mean
##       0.7610  -0.565  0.0112  0.7923  3.5312
## s.e.  0.1215   0.137  0.1174  0.0825  0.3516
##
## sigma^2 estimated as 3.028:  log likelihood=-235.87
## AIC=483.74   AICc=484.48   BIC=500.46
##
## Training set error measures:
##                       ME     RMSE      MAE      MPE     MAPE      MASE
## Training set 0.01359734 1.703504 1.426202 4.779284 81.71699 0.5609482
##                      ACF1
## Training set 0.0008885573
```

```
arma31_t <- matrix(NA, nrow = 5)

for (i in c(1:5)) {

    arma31_t[i] <- arma31$coef[i]/sqrt(arma31$var.coef[i, i])

}

arma31_t <- data.frame(arma31_t)

arma31_t
```

```
##      arma31_t
## 1  6.26539086
## 2 -4.12346904
## 3  0.09501299
## 4  9.59788435
## 5 10.04172094
```

```
arma31_q <- Box.test(arma31$residuals, lag = arma31$arma[1] +
    arma31$arma[2])
arma31_q
```

```
##
##  Box-Pierce test
##
## data:  arma31$residuals
## X-squared = 0.25911, df = 4, p-value = 0.9923
```

The t-value of $ar3$ is not able to reject the null hypothesis. Furthermore, the Q-statistic *(Box.test)* seems to indicate that ε_t is truly white noise.

Now, let's try some *underspecified models*. Beginning with an ARMA(2,0):

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim wn(0, \sigma^2)$$

```
arma20 <- Arima(df$value, order = c(2, 0, 0))

summary(arma20)
```

```
## Series: df$value
## ARIMA(2,0,0) with non-zero mean
##
## Coefficients:
##          ar1      ar2    mean
##       1.0226  -0.7153  3.5194
## s.e.  0.0634   0.0635  0.2694
##
## sigma^2 estimated as 4.24:  log likelihood=-256.37
## AIC=520.74   AICc=521.09   BIC=531.89
##
## Training set error measures:
##                       ME     RMSE      MAE      MPE     MAPE      MASE      ACF1
## Training set 0.01106086 2.033314 1.630805 73.16585 128.4039 0.6414218 0.2915253
```

```
arma20_t <- matrix(NA, nrow = 3)

for (i in c(1:3)) {

    arma20_t[i] <- arma20$coef[i]/sqrt(arma20$var.coef[i, i])

}

arma20_t <- data.frame(arma20_t)

arma20_t
```

```
##    arma20_t
## 1  16.12226
## 2 -11.26848
## 3  13.06561
```

```
arma20_q <- Box.test(arma20$residuals, lag = arma20$arma[1] +
    arma20$arma[2])
arma20_q
```

```
##
##  Box-Pierce test
##
## data:  arma20$residuals
## X-squared = 13.728, df = 2, p-value = 0.001045
```

The Q-statistic indicates that there is an ommited variable – namely, ε_{t-1} that we have just excluded from the model.

Now, an ARMA(1,1):

$$y_t = c + \Phi_1 y_{t-1} + +\theta_1 \varepsilon_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim wn(0, \sigma^2)$$

```
arma11 <- Arima(df$value, order = c(1, 0, 1))

summary(arma11)
```

```
## Series: df$value
## ARIMA(1,0,1) with non-zero mean
##
## Coefficients:
##          ar1     ma1    mean
##       0.4476  0.9244  3.6027
## s.e.  0.0831  0.0323  0.6234
##
## sigma^2 estimated as 4.026:  log likelihood=-253.74
## AIC=515.48   AICc=515.82   BIC=526.63
##
## Training set error measures:
##                       ME     RMSE      MAE      MPE     MAPE      MASE
## Training set 0.006283661 1.981184 1.621537 51.07775 142.3988 0.6377767
##                    ACF1
## Training set 0.2028683
```

```
arma11_t <- matrix(NA, nrow = 3)

for (i in c(1:3)) {

    arma11_t[i] <- arma11$coef[i]/sqrt(arma11$var.coef[i, i])

}

arma11_t <- data.frame(arma11_t)

arma11_t
```

```
##    arma11_t
## 1  5.387948
## 2 28.624391
## 3  5.779126
```

```
arma11_q <- Box.test(arma11$residuals, lag = arma11$arma[1] +
    arma11$arma[2])
arma11_q
```

```
##
##  Box-Pierce test
##
## data:  arma11$residuals
## X-squared = 12.668, df = 2, p-value = 0.001775
```

Again, the Q-statistic indicates that there is an ommited variable – namely, y_{t-1} that we have just excluded from the model.

Finally, let's compare the AIC and BIC values for all these models.

```
criteria <- matrix(NA, nrow = 5, ncol = 3)

aa_criteria <- data.frame("ARMA(2,1)*", aa_model$aic, aa_model$bic)

names(aa_criteria) <- c("Model", "AIC", "BIC")

arma22_criteria <- data.frame("ARMA(2,2)", arma22$aic, arma22$bic)

names(arma22_criteria) <- c("Model", "AIC", "BIC")

arma31_criteria <- data.frame("ARMA(3,1)", arma31$aic, arma31$bic)

names(arma31_criteria) <- c("Model", "AIC", "BIC")

arma20_criteria <- data.frame("ARMA(2,0)", arma20$aic, arma20$bic)

names(arma20_criteria) <- c("Model", "AIC", "BIC")

arma11_criteria <- data.frame("ARMA(1,1)", arma11$aic, arma11$bic)

names(arma11_criteria) <- c("Model", "AIC", "BIC")


criteria <- rbind.data.frame(aa_criteria, arma22_criteria, arma31_criteria,
    arma20_criteria, arma11_criteria)

criteria
```

```
##         Model      AIC      BIC
## 1 ARMA(2,1)* 481.7460 495.6834
## 2  ARMA(2,2) 483.7376 500.4625
```

```
## 3  ARMA(3,1) 483.7369 500.4619
## 4  ARMA(2,0) 520.7381 531.8880
## 5  ARMA(1,1) 515.4753 526.6253
```

As we can clearly see, the model chosen by *auto.arima* is the optimal choice according both to AIC and BIC.

Chapter 6

Forecasting

Suppose that you observe a time series up to period T, $Y_1, Y_2, ..., Y_T$, and would like to forecast its value in $T+1$, or, more generically, up to a time horizon $h \geq 1$: $Y_{T+1}, ..., Y_{T+h}$.

Let $g(Y_1, ..., Y_T)$ be a general predictor of Y_{T+1} built with information up to period T. We can measure its utility with its *mean squared error (MSE)*:

$$MSE[g(Y_1, ..., Y_T)] := \mathbb{E}[(Y_{T+1} - g(Y_1, ..., Y_T))^2]$$

We know that the conditional mean is the predictor that minimizes MSE:

$$\mathbb{E}\left(Y_{T+1} \mid Y_T, \ldots, Y_1\right) = \text{argmin}_g \, \mathbb{E}\left[\left(Y_{t+1} - g\left(Y_1, Y_2, \ldots, Y_T\right)\right)^2\right]$$

Unfortunately, we usually don't have the functional form of $\mathbb{E}\left(Y_{T+1} \mid Y_T, \ldots, Y_1\right)$, and postulate its best linear form:

$$\pi\left(Y_{T+1} \mid Y_T, \ldots, Y_1\right) = \alpha + \beta_0 Y_T + \beta_1 Y_{T-1} + \ldots \beta_{T-1} Y_1$$

Denote the best linear predictor with information up to T as $Y_{T+1|T} := \pi(Y_{T+1}|Y_t, ..., Y_1)$, and its estimated version as $\hat{Y}_{T+1|T} := \hat{\pi}(Y_{T+1}|Y_t, ..., Y_1) = \hat{\alpha} + \hat{\beta}_0 Y_T + \hat{\beta}_1 Y_{T-1} + \ldots \hat{\beta}_{T-1} Y_1$.

6.1 Forecasting with an AR(1) model

Remember from Definition 2.6.1 that an AR(1) model has the following form:

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t$$

6.1.1 Forecast

Given that $\mathbb{E}(\varepsilon_{t+1}|Y_T) = 0$ (as $\varepsilon \sim wn(0, \sigma^2)$), we have:

$$Y_{T+1|T} := \pi\left(Y_{T+1} \mid Y_T, \ldots, Y_1\right) = c + \phi Y_T$$

For $T+2$, we have:

$$\begin{aligned} Y_{T+2|T} &:= \pi\left(Y_{T+2} \mid Y_T, \ldots, Y_1\right) \\ &= \pi\left(\pi\left(Y_{T+2} \mid Y_{T+1}, Y_T, \ldots, Y_1\right) \mid Y_T, \ldots, Y_1\right) \\ &= \pi\left(c + \phi Y_{T+1} \mid Y_T, \ldots, Y_1\right) \\ &= c + \phi\left(c + \phi Y_T\right) = (1+\phi)c + \phi^2 Y_T \end{aligned}$$

The clear pattern here yields, more generally:

$$Y_{T+h|T} = \left(1 + \phi + \phi^2 + \cdots + \phi^{h-1}\right) c + \phi^h Y_T$$

6.1.2 Forecast error

It is also easy to verify that the *forecast error* for $h = 1$ is given by:

$$u_{T+1|T} := Y_{T+1} - Y_{T+1|T} = Y_{T+1} - c - \phi Y_T = \varepsilon_{T+1}$$

For $h = 2$, we have:

$$\begin{aligned} u_{T+2|T} &:= Y_{T+2} - Y_{T+2|T} \\ &= c + \phi Y_{T+1} + \varepsilon_{T+2} - \left(c + \phi Y_{T+1|T}\right) \\ &= \phi\left(Y_{T+1} - Y_{T+1|T}\right) + \varepsilon_{T+2} \\ &= \phi u_{T+1|T} + \varepsilon_{T+2} = \phi\varepsilon_{T+1} + \varepsilon_{T+2} \end{aligned}$$

More generally, for a given time horizon h:

$$u_{T+h|T} = \left(\varepsilon_{T+h} + \phi\varepsilon_{T+h-1} + \cdots + \phi^{h-1}\varepsilon_{T+1}\right)$$

6.1.3 Mean reversion

Note that, as h increases, the prediction *reverts to the unconditional mean:*

$$\lim_{h\to\infty} Y_{T+h|T} = c \lim_{h\to\infty} \sum_{i=0}^{h-1} \phi^i + Y_T \lim_{h\to\infty} \phi^h = \frac{c}{1-\phi} =: \mu$$

The variance of the forecast error is given by:

$$Var(u_{T+h|T}) = (1 + \phi^2 + \phi^4 + ... + \phi^{2(h-1)})\sigma^2$$

Taking the limit as $h \to \infty$, the variance of the forecast error *also approaches the unconditional variance of the process:*

$$\lim_{h\to\infty} \mathbb{V}\left(u_{T+h|T}\right) = \sigma^2 \lim_{h\to\infty} \sum_{i=0}^{h-1} \phi^{2i} = \frac{\sigma^2}{1-\phi^2} =: \gamma_0$$

6.2 Forecasting with an AR(p) model

6.2.1 Forecast

The same procedure can be applied to an AR(p). For $h = 1$:

$$Y_{T+1|T} = \pi\left(Y_{T+1} \mid Y_T, \ldots, Y_1\right) = c + \phi_1 Y_T + \cdots + \phi_p Y_{T-p+1}$$

For a given h, proceed recursively:

$$\begin{aligned}
Y_{T+2|T} &= c + \phi_1 Y_{T+1|T} + \phi_2 Y_T + \phi_3 Y_{T-1} + \cdots + \phi_p Y_{T+2-p} \\
Y_{T+3|T} &= c + \phi_1 Y_{T+2|T} + \phi_2 Y_{T+1|T} + \phi_3 Y_T + \cdots + \phi_p Y_{T+3-p} \\
Y_{T+4|T} &= c + \phi_1 Y_{T+3|T} + \phi_2 Y_{T+2|T} + \phi_3 Y_{T+1|T} + \cdots + \phi_p Y_{T+4-p} \\
&\vdots \\
Y_{T+h|T} &= c + \phi_1 Y_{T+h-1|T} + \phi_2 Y_{T+h-2|T} + \cdots + \phi_p Y_{T+h-p|T}
\end{aligned}$$

6.2.2 Forecast error

For $h = 1$, the forecast error is given by:

$$u_{T+1|T} := Y_{T+1} - Y_{T+1|T} = \varepsilon_{T+1}$$

As h increases:

$$\begin{aligned}
u_{T+2|T} &= \phi_1 u_{T+1|T} + \varepsilon_{T+2} \\
u_{T+3|T} &= \phi_1 u_{T+2|T} + \phi_2 u_{T+1|T} + \varepsilon_{T+3} \\
&\vdots \\
u_{T+h|T} &= \phi_1 u_{T+h-1|T} + \cdots + \phi_{h-1} u_{T+1|T} + \varepsilon_{T+h},
\end{aligned}$$

where $\phi_{h-1} = 0$ for $h - 1 > p$.

6.3 Forecasting with a MA(1) model

Remember from 2.2.1 that a MA(1) model is given by

$$Y_t = c + \theta \varepsilon_{t-1} + \varepsilon_t$$

$$Y_{T+1|T} := \pi(Y_{T+1}|Y_T, ..., Y_1)$$

We have seen that, if the MA(1) is invertible, we can write it as an AR(∞). This would mean, however, that the forecast errors would depend on *all past values* – notwithstanding the decreasing dependence, due to ergodicity. Given a fixed ε_0, we can reconstruct the entire error series:

$$\begin{aligned}
\varepsilon_1 &= Y_1 - c - \theta \varepsilon_0 \\
\varepsilon_2 &= Y_2 - c - \theta \varepsilon_1 \\
&\vdots \\
\varepsilon_T &= Y_T - c - \theta \varepsilon_{T-1}
\end{aligned}$$

If we do not know ε_0, we can approximate it by its (known!) mean, $\tilde{\varepsilon}_0 = 0$. If T is large enough and $|\theta| < 1$, this is a good approximation. That is the case because for large T, the influence of the assumption $\tilde{\varepsilon}_0 = 0$ is geometrically diminished over time, given $|\theta| < 1$.

6.3.1 Forecast

We are now able to construct the forecast using the estimated $\widetilde{\varepsilon}_T$.

$$Y_{T+1|T} := c + \theta\widetilde{\varepsilon}_T$$

Iteratively for larger horizons:

$$Y_{T+2|T} = c + \theta\widetilde{\varepsilon}_{T+1} = c + \theta\left(Y_{T+1|T} - c - \theta\widetilde{\varepsilon}_T\right) = c$$

Note that *the MA(1) process is not predictable for* $h > 1$. The forecast immediately mean reverts at $h > 1$.

6.3.2 Forecast error

The forecast error, assuming $\widetilde{\varepsilon}_T \approx \varepsilon_T$, is given by:

$$u_{T+1|T} := Y_{T+1} - Y_{T+1|T} = \varepsilon_{T+1}$$

This means that the variance of the forecast error is σ^2. For larger horizons, the variance converges to the unconditional variance:

$$u_{T+h|T} := Y_{T+h} - Y_{T+h|T} = \varepsilon_{T+h} + \theta\varepsilon_{T+h-1}, \forall h > 1$$

Thus, $Var(u_{T+h|T}) = (1+\theta)^2\sigma^2 = \gamma_0, \forall h > 1$.

6.4 Forecasting with a MA(q) model

We can proceed iteratively for the MA(q) model.

$$\begin{aligned}
Y_{T+1|T} &= c + \theta_1\widetilde{\varepsilon}_T + \cdots + \theta_q\widetilde{\varepsilon}_{T+1-q} \\
Y_{T+2|T} &= c + \theta_2\widetilde{\varepsilon}_T + \cdots + \theta_q\widetilde{\varepsilon}_{T+2-q} \\
Y_{T+3|T} &= c + \theta_3\widetilde{\varepsilon}_T + \cdots + \theta_q\widetilde{\varepsilon}_{T+3-q} \\
&\vdots \\
Y_{T+q|T} &= c + \theta_q\widetilde{\varepsilon}_T
\end{aligned}$$

Note that, after q periods, the prediction is the unconditional mean c.

6.5 Forecast with an ARMA(p,q) model

Let's combine the procedures presented for AR(p) and MA(q). For $h = 1$:

$$Y_{T+1|T} = c + \phi_1 Y_T + \cdots + \phi_p Y_{T-p+1} + \theta_1\widetilde{\varepsilon}_T + \cdots + \theta_q\widetilde{\varepsilon}_{T+1-q},$$

where $\{\widetilde{\varepsilon}_T, ..., \widetilde{\varepsilon}_{T+1-q}\}$ were obtained from the reconstruction process explained in the MA(q) case. However, here we need, aside from the first q innovation values, also the *last p observations before the first one*. That is the case because of the expression for $\widetilde{\varepsilon}_1$:

$$\widetilde{\varepsilon}_1 = Y_1 - c - \phi_1 Y_0 - \cdots - \phi_p Y_{p-1} - \theta_1\widetilde{\varepsilon}_0 - \theta_2\widetilde{\varepsilon}_{-1} - \cdots - \theta_q\widetilde{\varepsilon}_{1-q}$$

Alternatively, we can start reconstructing the residuals from $p + 1$.

6.5.1 Forecast

From this, we can obtain the forecasts for any horizon $h \geq 1$, given $h > p > q$:

$$\begin{aligned}
Y_{T+1|T} &= c + \phi_1 Y_T + \cdots + \phi_p Y_{T+1-p} + \theta_1 \widetilde{\varepsilon}_T + \cdots + \theta_q \widetilde{\varepsilon}_{T+1-q} \\
Y_{T+2|T} &= c + \phi_1 Y_{T+1|T} + \cdots + \phi_p Y_{T+2-p} + \theta_2 \widetilde{\varepsilon}_T + \cdots + \theta_q \widetilde{\varepsilon}_{T+2-q} \\
Y_{T+3|T} &= c + \phi_1 Y_{T+2|T} + \cdots + \phi_p Y_{T+3-p} + \theta_3 \widetilde{\varepsilon}_T + \cdots + \theta_q \widetilde{\varepsilon}_{T+3-q}
\end{aligned}$$

$$\vdots$$

$$\begin{aligned}
Y_{T+q|T} &= c + \phi_1 Y_{T+q-1|T} + \cdots + \phi_p Y_{T+q-p} + \theta_q \widetilde{\varepsilon}_T \\
Y_{T+q+1|T} &= c + \phi_1 Y_{T+q|T} + \cdots + \phi_p Y_{T+q+1-p} \\
&\vdots \\
Y_{T+h|T} &= c + \phi_1 Y_{T+h-1|T} + \cdots + \phi_p Y_{T+h-p|T}
\end{aligned}$$

6.6 Confidence intervals for forecasts

Given a forecast $\hat{Y}_{T+h|T}$, we'd like to have a *confidence interval* for it. In practice, forecasting involves two types of errors:

- **Estimation error:** $Y_{T+h|T} - \hat{Y}_{T+h|T}$. This is due to the estimation of the ARMA model, as we are always working with a sample.
- **Forecast error:** $u_{T+h|T} = Y_{T+h} - Y_{T+h|T}$. This is the portion of the error that would exist event if we knew the true population parameters.

From this, we have the *aggregate error:*

$$Y_{T+h} - \hat{Y}_{T+h|T} = (Y_{T+h} - Y_{T+h|T}) + (Y_{T+h|T} - \hat{Y}_{T+h|T})$$

6.6.1 Normally distributed errors

A first way to tackle the issue of confidence intervals for forecasts is to *assume normality*: $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.

In that case, we know that $Y_{T+h}, \hat{Y}_{T+h|T}$ will also be normally distributed, as the first term consists of a combination of past errors and past values of Y, and the second term is asymptotically normal given the properties of ARMA estimators. This means that the confidence interval is known to us, for $\alpha = 5\%$:

$$\left[\widehat{Y}_{T+h|T} \pm 1,96\sqrt{\mathbb{V}\left(u_{T+h|T}\right)}\right]$$

We clearly do not observe $Var(u_{T+h|T})$, but we know its functional form and, thus, can consistently estimate it. For an AR(1), for example:

$$\widehat{\mathbb{V}}\left(u_{T+h|T}\right) = \widehat{\sigma}^2 \sum_{i=0}^{h-1} \widehat{\phi}^{2i}$$

Note that, even if we assume that the innovation terms are normally distributed, we *still need an asymptotic argument* to guarantee that $\widehat{Y}_{T+h|T} \xrightarrow{p} Y_{T+h|T}$ and $\widehat{\mathbb{V}}\left(u_{T+h|T}\right) \xrightarrow{p} \mathbb{V}\left(u_{T+h|T}\right)$.

6.6.2 Bootstrapping

What if the errors *are not normally distributed?* This makes the issue a lot more complicated. Note that, even if we apply the CLT to argue that $Y_{T+h|T} - \hat{Y}_{T+h|T} \sim \mathcal{N}(\cdot)$, what makes us believe that $(Y_{T+h} - Y_{T+h|T})$ is also normally distributed? In the last subsection, this condition was valid because of the distribution of the errors. There's no asymptotical argument to be made.

In practice, we don't know the distribution of the error terms. In fact, they are *rarely normally distributed!* A procedure to construct confidence intervals in this setting is called *bootstrap.* For ARMA models, it consists in generating simulated samples from the actual sample and repeat the estimations for each simulated sample. This yields an arbitrarily large number of estimates *from the same data.* From this, we can use its *empirical distribution* to find the confidence interval.

Bootstrapping has a number of advantages over the traditional procedures for obtaining a confidence interval. In many cases, the convergence is faster than the usual $1/\sqrt{T}$, and it frequently performs better in small sample environments. It can also be applied in settings in which asymptotic theory is very intricate. In practice, we almost always use some sort of bootstrap method.

The essential idea behind bootstrapping is to assume that ε_t is *iid* with a cdf F. Suppose that you have T residuals, $\hat{\varepsilon}_t$. Its empirical distribution is simply:

$$\hat{F}_T(v) = \frac{\{no.\, residuals \leq v\}}{T}, \quad v \in \mathbb{R}$$

The Law of Large Numbers guarantees that, as $T \to \infty$,

$$\hat{F}_T(v) \to_{a.s.} F(v)$$

This means that the empirical distribution of the residuals becomes arbitrarily close to the real distribution as T grows.

Procedure

Bootstrapping for an ARMA(p,q) model involves the following steps:

1. - Estimate ARMA(p,q)

$$Y_t = c + \sum_{j=1}^{p} \phi_j Y_{t-j} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

   - Calculate the residuals of the regression:

$$\hat{\varepsilon}_t := Y_t - (\hat{c} + \sum_{j=1}^{p} \hat{\phi}_j Y_{t-j} + \sum_{j=1}^{q} \hat{\theta}_j \varepsilon_{t-j})$$

   - If the residuals do not have mean 0, create the centered residuals:

$$\tilde{\varepsilon}_t = \hat{\varepsilon}_t - \frac{1}{t} \sum_{t=1}^{T} \hat{\varepsilon}_t$$

2.
   - Select at random, with replacement, a sample with $T + m$ elements, $m >> 0$:

$$\{\varepsilon_1^*, ..., \varepsilon_{T+m}^*\}$$

   - Create a series $\{Y_t^*\}_{t=1}^{T+m}$:

$$Y_t^* = Y_t, 1 \leq t \leq max(p, q)$$

$$Y_t^* = \hat{c} + \sum_{j=1}^{p} \hat{\phi}_j Y_{t-j} + \sum_{j=1}^{q} \hat{\theta}_j \varepsilon_{t-j}^* + \varepsilon_t^*, max(p, q) < t \leq T + m$$

3.
   - Using the simulated sample $\{Y_t^*\}_{t=1}^{T+m}$, create a forecast for $h > 0$ periods using the estimated coefficients *obtained with the real sample.*
   - This yields a vector of dimension h containing the forecasts in the form:

$$(\hat{Y}_{T+1}^*, ..., \hat{Y}_{T+h}^*)$$

   - Repeat steps 2 and 3 for S times. Create a matrix with the results.
   - This yields a S x h matrix where each row is equal to the aforementioned vector.

Chapter 7

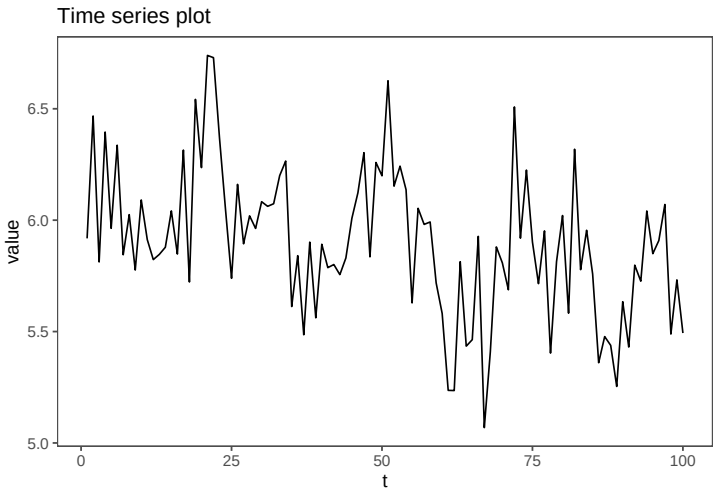
Problem 4: Cross-validation and bootstrap

In this problem, we'll be tackling the issue of *forecasting* of an ARMA model. The problem is split in two parts: (i) *cross-validation*; and (ii) *bootstrapping*.

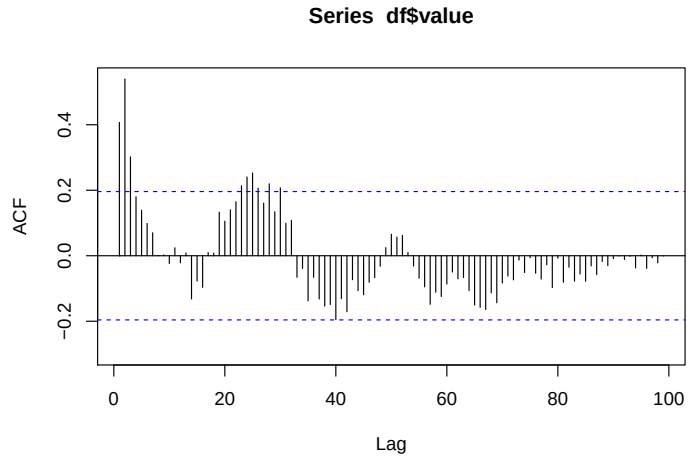
7.1 Identification and estimation

First, let's identify the best model for our time series.

```
df <- data.frame(df)

pplot <- ggplot(data = df, aes(x = t, y = value)) + geom_line() +
    ggtitle("Time series plot") + theme_few()
pplot
```



```
acf_ts <- Acf(df$value, lag.max = 5000)
```




```
acf_test_values <- acf_ts$acf/sd(acf_ts$acf)

head(data.frame(acf_test_values))
```

```
##   acf_test_values
## 1        6.176432
## 2        2.515951
## 3        3.335438
## 4        1.864909
## 5        1.112884
## 6        0.858639
```

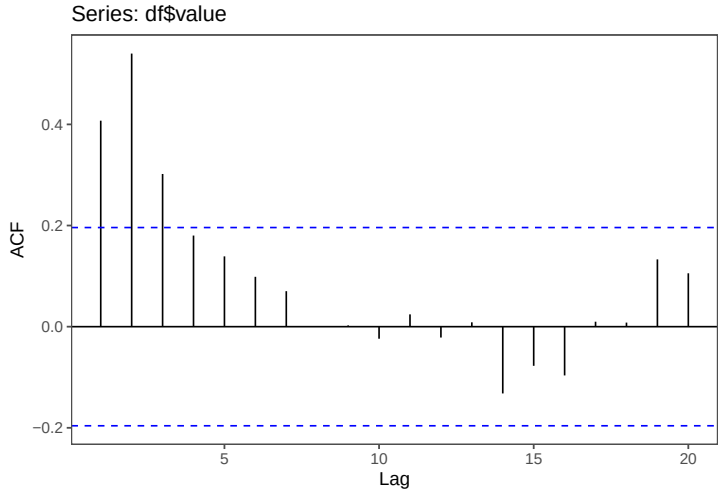
```
facst <- ggAcf(df$value, type = "correlation", lag.max = 20,
plot = T) + theme_few()
faclt <- ggAcf(df$value, type = "correlation", lag.max = 5000,
plot = T) + theme_few()

facpst <- ggPacf(df$value, type = "correlation", lag.max = 100,
plot = T) + theme_few()
```

```
## Warning: Ignoring unknown parameters: type
```

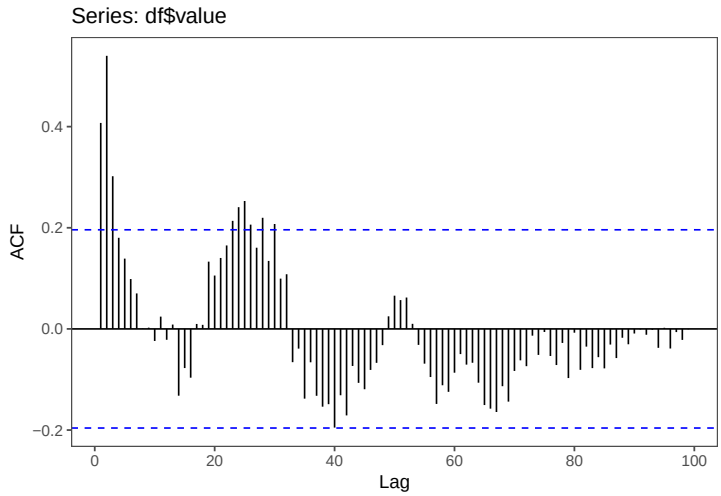
```
facplt <- ggPacf(df$value, type = "correlation", lag.max = 5000,
plot = T) + theme_few()
```

```
## Warning: Ignoring unknown parameters: type
```

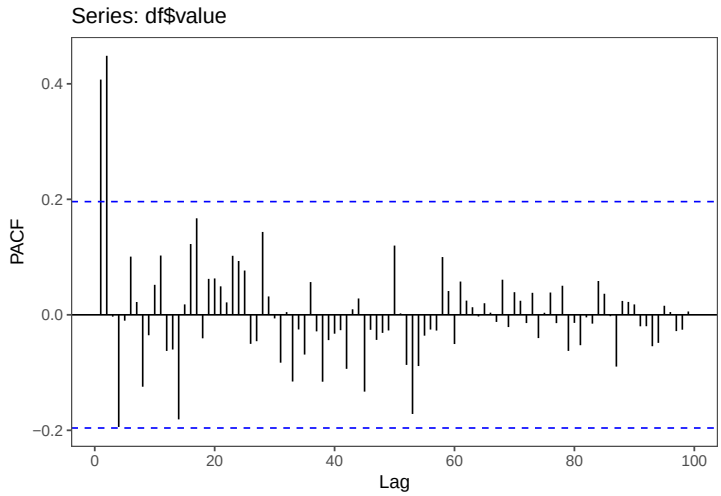
```
facst
```



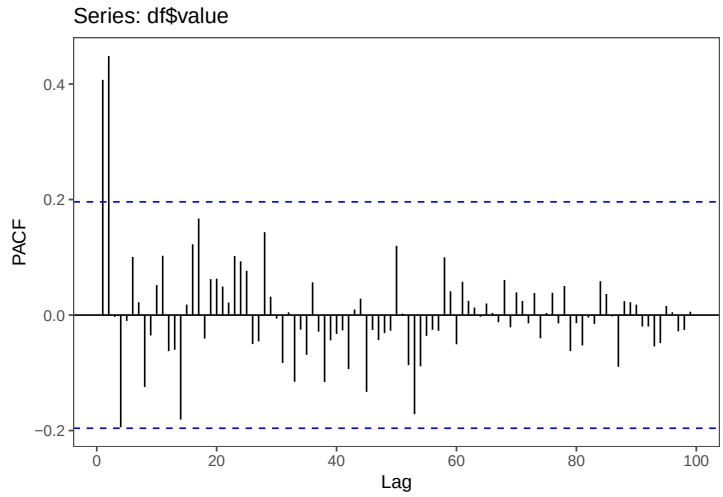

```
faclt
```




```
facpst
```



```
facplt
```



We'll now use the function *auto.arima* from the package *forecast* to identify and estimate the model.

```
aa_model <- auto.arima(df$value, num.cores = 24, max.d = 0, stepwise = F)

summary(aa_model)
```

```
## Series: df$value
## ARIMA(0,0,3) with non-zero mean
##
## Coefficients:
##          ma1     ma2     ma3    mean
##       0.1814  0.6647  0.4001  5.8982
## s.e.  0.0852  0.0750  0.0949  0.0562
##
## sigma^2 estimated as 0.0667:  log likelihood=-5.42
## AIC=20.85   AICc=21.49   BIC=33.88
##
## Training set error measures:
##                       ME      RMSE       MAE        MPE     MAPE      MASE
## Training set -0.002315954 0.2530428 0.2131067 -0.2268814 3.612855 0.7314965
##                     ACF1
## Training set 0.03868106
```

```
print("t-values: ")
```

```
## [1] "t-values: "
```

```
aa_t <- matrix(NA, nrow = aa_model$arma[1] + aa_model$arma[2])

for (i in c(1:4)) {

    aa_t[i] <- aa_model$coef[i]/sqrt(aa_model$var.coef[i, i])

}

aa_t <- data.frame(aa_t)
```

```
aa_t
```

```
##         aa_t
## 1   2.128691
## 2   8.861580
## 3   4.216481
## 4 105.004537
```

```
aa_q <- Box.test(aa_model$residuals, lag = aa_model$arma[1] +
    aa_model$arma[2])
aa_q
```

```
##
##  Box-Pierce test
##
## data:  aa_model$residuals
## X-squared = 0.35002, df = 3, p-value = 0.9504
```

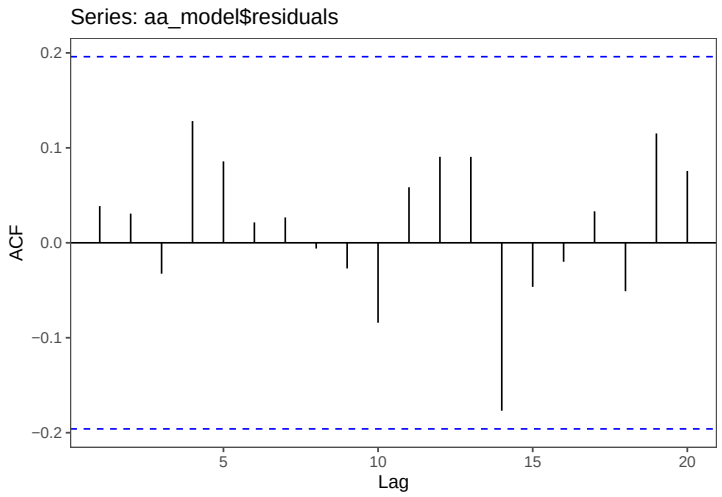
```
criteria <- matrix(NA, nrow = 1, ncol = 3)

aa_criteria <- data.frame("MA(3)*", aa_model$aic, aa_model$bic)

names(aa_criteria) <- c("Model", "AIC", "BIC")

aa_criteria
```

```
##    Model      AIC      BIC
## 1 MA(3)* 20.84963 33.87549
```

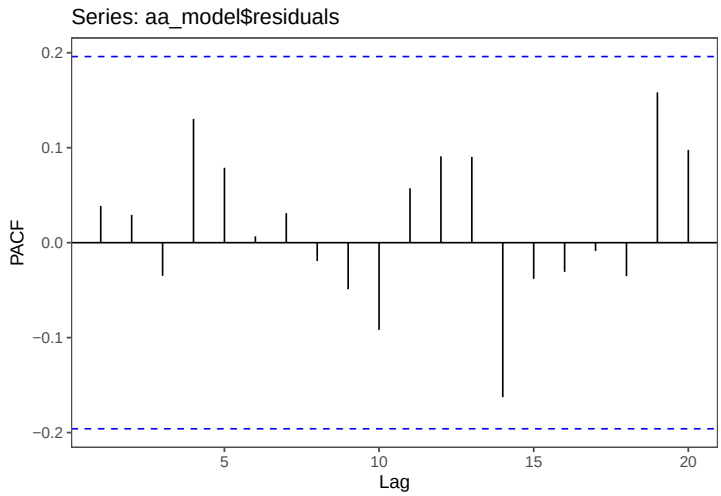
```
fac_e <- ggAcf(aa_model$residuals, type = "correlation", lag.max = 20,
plot = T) + theme_few()

facp_e <- ggPacf(aa_model$residuals, type = "correlation", lag.max = 20,
plot = T) + theme_few()
```

```
## Warning: Ignoring unknown parameters: type
```

```
fac_e
```




```
facp_e
```




```
mean(aa_model$residuals)
```

```
## [1] -0.002315954
```

The results of *auto.arima* imply that the best model is an ARMA(0,3) – i.e., a MA(3):

$$y_t = c + +\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \varepsilon_t, \quad \varepsilon_t \sim wn(0, \sigma^2)$$

Furthermore, the Q-statistic *(Box.test)* seems to indicate that ε_t is truly white noise.

7.2 Cross-validation

Let's now cross-validate or model. This will now be done manually; afterwards, an automatized version from *fpp* shall be presented.

Let $h := 5$; $frac = 0.2$. T is the size of our sample; k is the *training* database. The remainder shall be used for testing purposes.

As we have discovered previously, *auto.arima* yields a **MA(3)** model. It will now be used.

```
h <- 5

frac <- 0.2

T <- length(df$value)

k <- floor((1 - frac) * T)

# Estimating MA(3) with k = 80
fit <- Arima(df$value[1:k], order = c(0, 0, 3))

# Generating predictions from the model
pred <- predict(fit, n.ahead = h)

# Calculating errors between the predicted values of the
# model and the actual values of the testing database

e <- df$value[(k + h)] - pred$pred[h]

e
```

```
## [1] -0.1951299
```

Let's now update our training database iteratively with a for loop.

```
e <- matrix(NA, nrow = 100)

# Updating the model

for (i in k:(T - h)) {

    fit <- Arima(df$value[1:i], order = c(0, 0, 3))

    pred <- predict(fit, n.ahead = h)

    e[i, 1] <- df$value[(i + h)] - pred$pred[h]

}
```

With the matrix e in hands, we can now calculate MSE:

```
mse <- mean(e^2, na.rm = T)
```

This procedure can now be used to compare other models against the model from *auto.arima*.

```
max_p <- 5

max_q <- 5

e <- matrix(NA, nrow = 100, ncol = (max_p + 1) * (max_q + 1))

pred <- vector("list", (max_p + 1) * (max_q + 1))

fit <- vector("list", (max_p + 1) * (max_q + 1))

# Updating the model
for (u in 0:max_q) {

for (j in 0:max_p) {


for (i in k:(T - h)) {

            fit[[(((max_p + 1) * j) + u + 1)]] <- Arima(df$value[1:i],
order = c(j, 0, u))

# fit <- append(fit, Arima(df$value[1:i], order = c(j,0,u)))

# pred <- append(pred, predict(fit[[(j+u)]], n.ahead = h))

            pred[[(((max_p + 1) * j) + u + 1)]] <- predict(fit[[(((max_p +
                1) * j) + u + 1)]], n.ahead = h)

            e[i, (((max_p + 1) * j) + u + 1)] <- df$value[(i +
                h)] - pred[[(((max_p + 1) * j) + u + 1)]]$pred[h]

        }

    }

}

mse <- matrix(NA, nrow = ((max_p + 1) * (max_q + 1)), ncol = 1)



mse <- colMeans(e^2, na.rm = T)

mse
```

```
##  [1] 0.1357466 0.1354001 0.1368083 0.1374243 0.1376508 0.1441940 0.1347115
##  [8] 0.1269779 0.1347789 0.1373465 0.1398588 0.1436175 0.1313779 0.1315448
## [15] 0.1435805 0.1355649 0.1421277 0.1335153 0.1316765 0.1333955 0.1400838
## [22] 0.1427467 0.1473227 0.1347447 0.1320856 0.1333025 0.1354734 0.1341742
## [29] 0.1380676 0.1357880 0.1346228 0.1382810 0.1319484 0.1308446 0.1382417
## [36] 0.1327046
```

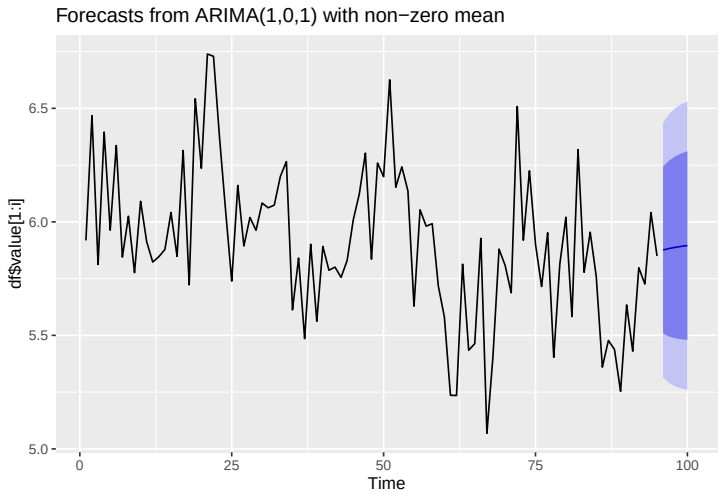
```
optimal_index <- which.min(mse)

cv_model <- fit[[optimal_index]]

summary(cv_model)
```

```
## Series: df$value[1:i]
## ARIMA(1,0,1) with non-zero mean
##
## Coefficients:
##          ar1      ma1    mean
##       0.8253  -0.4888  5.9125
## s.e.  0.0814   0.1118  0.0815
##
## sigma^2 estimated as 0.08209:  log likelihood=-14.72
## AIC=37.43   AICc=37.88   BIC=47.65
##
## Training set error measures:
##                       ME      RMSE       MAE        MPE     MAPE      MASE
## Training set -0.002725722 0.2819456 0.2228183 -0.2756862 3.787114 0.7600473
##                     ACF1
## Training set -0.1279409
```

The cross-validation method constructed above yielded an ARMA(1,1):

$$y_t = c + \phi_1 y_{t-1} + +\theta_1 \varepsilon_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim wn(0, \sigma^2)$$

```
cv_fc <- forecast(cv_model, h = h)

autoplot(cv_fc)
```



7.3 Bootstrapping

Now, let's proceed to *bootstrapping*. It envolves the following steps:

1.
   - Estimate ARMA(p,q)

$$Y_t = c + \sum_{j=1}^{p} \phi_j Y_{t-j} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

   - Calculate the residuals of the regression:

$$\hat{\varepsilon}_t := Y_t - (\hat{c} + \sum_{j=1}^{p} \hat{\phi}_j Y_{t-j} + \sum_{j=1}^{q} \hat{\theta}_j \varepsilon_{t-j})$$

   - If the residuals do not have mean 0, create the centered residuals:

$$\tilde{\varepsilon}_t = \hat{\varepsilon}_t - \frac{1}{t} \sum_{t=1}^{T} \hat{\varepsilon}_t$$

2.
   - Select at random, with replacement, a sample with $T + m$ elements, $m >> 0$:

$$\{\varepsilon_1^*, ..., \varepsilon_{T+m}^*\}$$

   - Create a series $\{Y_t^*\}_{t=1}^{T+m}$:

$$Y_t^* = Y_t, 1 \leq t \leq max(p, q)$$

$$Y_t^* = \hat{c} + \sum_{j=1}^{p} \hat{\phi}_j Y_{t-j} + \sum_{j=1}^{q} \hat{\theta}_j \varepsilon_{t-j}^* + \varepsilon_t^*, max(p, q) < t \leq T + m$$

3.
   - Using the simulated sample $\{Y_t^*\}_{t=1}^{T+m}$, create a forecast for $h > 0$ periods using the estimated coefficients *obtained with the real sample*.
   - This yields a vector of dimension h containing the forecasts in the form:

$$(\hat{Y}_{T+1}^*, ..., \hat{Y}_{T+h}^*)$$

   - Repeat steps 2 and 3 for S times. Create a matrix with the results.
   - This yields a S x h matrix where each row is equal to the aforementioned vector.

We'll use, again, the optimal model from *auto.arima*, MA(3):

$$y_t = c + +\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \varepsilon_t, \quad \varepsilon_t \sim wn(0, \sigma^2)$$

```r
S <- 1000

m <- 100



optimal_p <- aa_model$arma[1]

optimal_q <- aa_model$arma[2]

e_sample <- data.frame(matrix(NA, nrow = S, ncol = (length(df$value) +
    m)))

y_star <- data.frame(matrix(NA, nrow = S, ncol = (length(df$value) +
    m + max(aa_model$arma[1], aa_model$arma[2]))))

arima_star <- data.frame(matrix(NA, nrow = S, ncol = (length(df$value) +
    m + max(aa_model$arma[1], aa_model$arma[2]))))


for (i in 1:S) {

    e_sample[i] <- sample(aa_model$residuals, replace = T, size = (length(df$value) +
        m))

}

for (i in 1:S) {

for (j in ((aa_model$arma[1] + aa_model$arma[2] + 1):(length(df$value) +
        m))) {

        arima_star[i, j] <- (aa_model$coef[4] + (aa_model$coef[1] *
            e_sample[i, j - 1]) + (aa_model$coef[2] * e_sample[i,
            j - 2]) + (aa_model$coef[3] * e_sample[i, j - 3]) +
            e_sample[i, j])

    }

}


y_fixed <- data.frame(matrix(NA, nrow = S, ncol = (aa_model$arma[1] +
    aa_model$arma[2])))

for (i in 1:S) {
    y_fixed[i, 1] <- data.frame(df$value[1])
    y_fixed[i, 2] <- data.frame(df$value[2])
    y_fixed[i, 3] <- data.frame(df$value[3])
}


y_star <- data.frame(y_fixed, arima_star[, -(1:3)])

y_m <- y_star[, -(1:100)]

y_m <- y_m[, -(101:103)]

y_mt <- t(y_m)

y_matrix <- as.matrix(y_m)
```

```r
fc_list <- vector("list", S)

for (i in 1:S) {

    fc_list[[i]] <- forecast(ts(y_matrix[i, ]), model = aa_model,
h = 5)
```

```
}

fc_list[[1]]
```

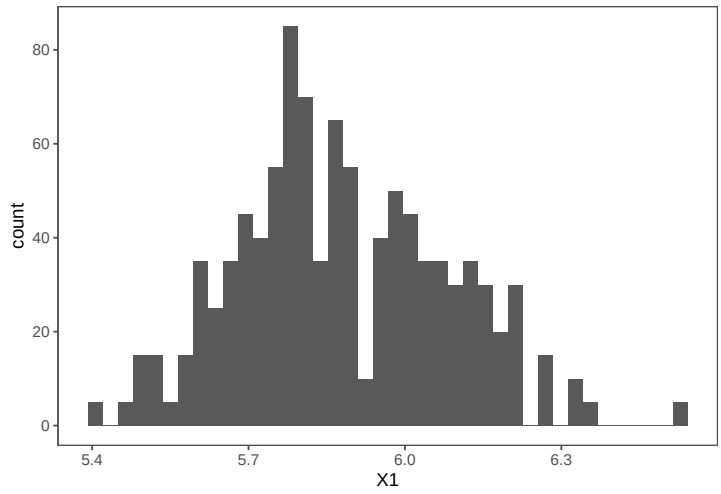
```
##     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 101       5.880900 5.549926 6.211875 5.374719 6.387082
## 102       5.869038 5.532663 6.205413 5.354597 6.383479
## 103       5.841494 5.439567 6.243421 5.226800 6.456189
## 104       5.898184 5.475000 6.321368 5.250980 6.545387
## 105       5.898184 5.475000 6.321368 5.250980 6.545387
```

```
fc_mean <- data.frame(matrix(NA, nrow = S, ncol = 5))

for (i in 1:S) {

    fc_mean[i, ] <- fc_list[[i]]$mean

}
```

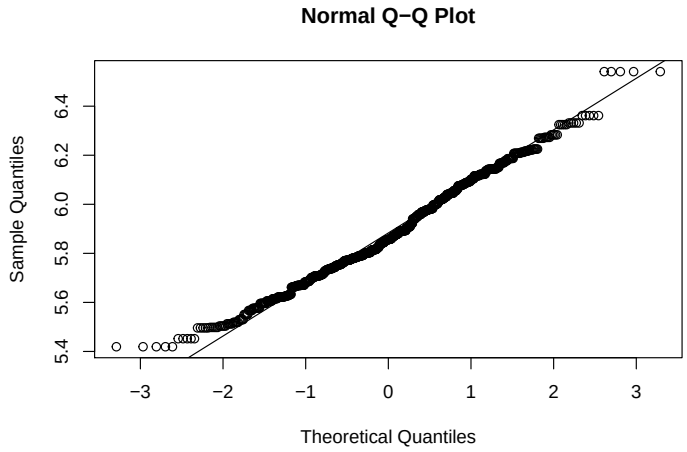
```
head(fc_mean)
```

```
##          X1       X2       X3       X4       X5
## 1 5.880900 5.869038 5.841494 5.898184 5.898184
## 2 5.734428 5.543250 5.786803 5.898184 5.898184
## 3 5.728049 5.722659 5.832225 5.898184 5.898184
## 4 5.844103 5.943213 5.958997 5.898184 5.898184
## 5 5.550780 5.518844 5.732659 5.898184 5.898184
## 6 5.742853 5.863462 5.848949 5.898184 5.898184
```

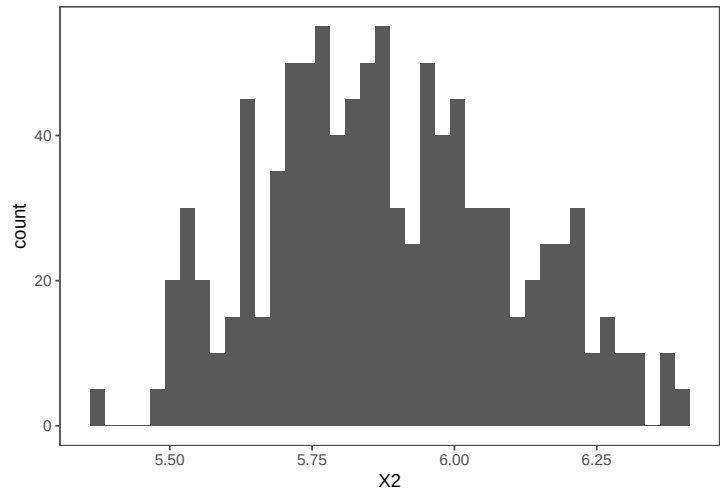
```
hist_x1 <- ggplot(data = fc_mean, aes(x = X1)) + geom_histogram(bins = 40) +
    theme_few()
hist_x1
```



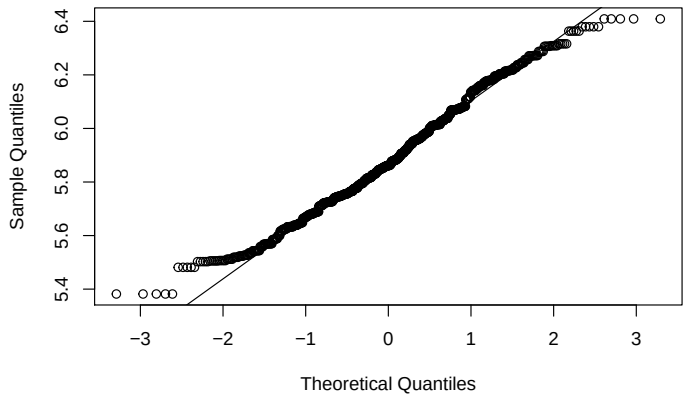

```
qq_x1 <- qqnorm(fc_mean$X1)
qqline(fc_mean$X1)
```



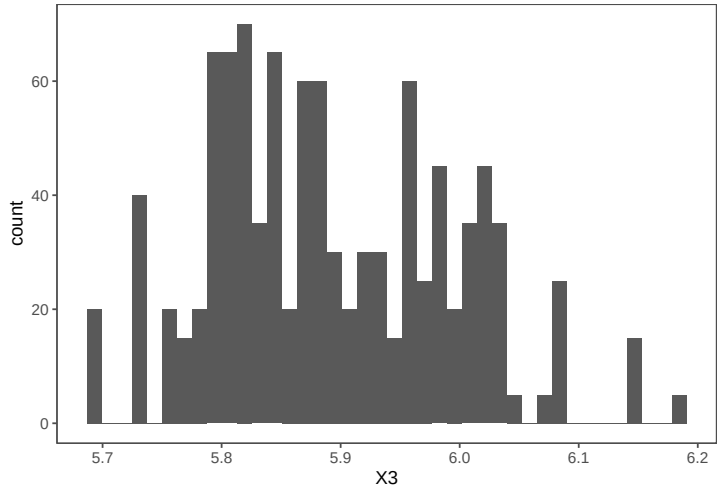

```
hist_x2 <- ggplot(data = fc_mean, aes(x = X2)) + geom_histogram(bins = 40) +
    theme_few()
hist_x2
```



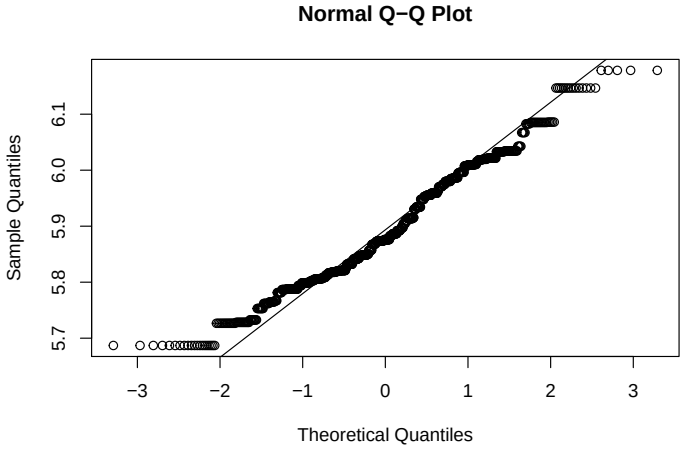

```
qq_x2 <- qqnorm(fc_mean$X2)
qqline(fc_mean$X2)
```



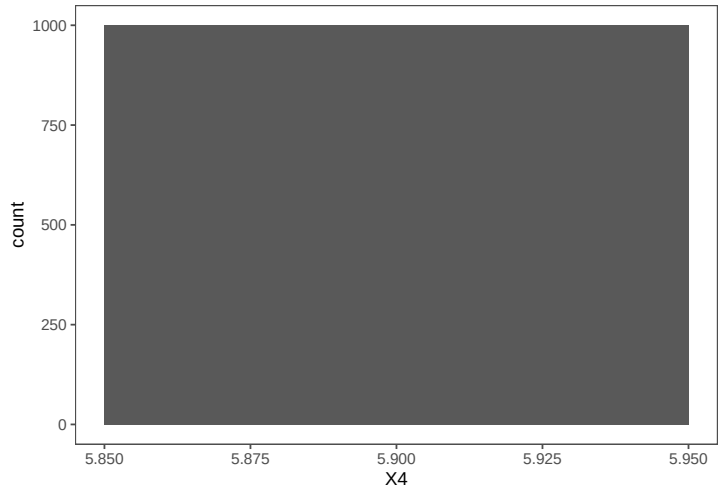
```
hist_x3 <- ggplot(data = fc_mean, aes(x = X3)) + geom_histogram(bins = 40) +
    theme_few()
hist_x3
```



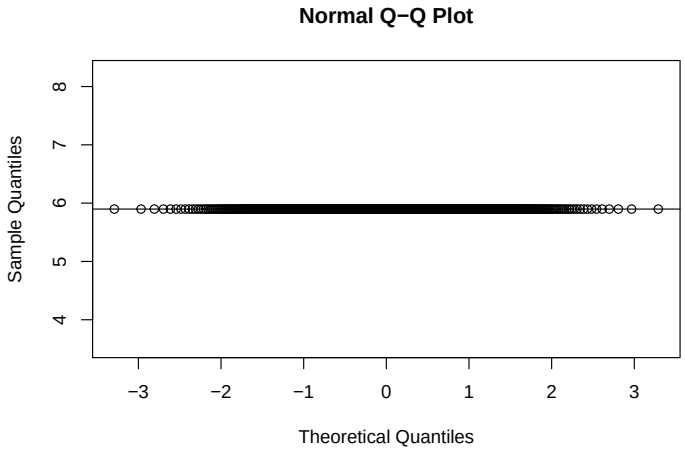

```
qq_x3 <- qqnorm(fc_mean$X3)
qqline(fc_mean$X3)
```



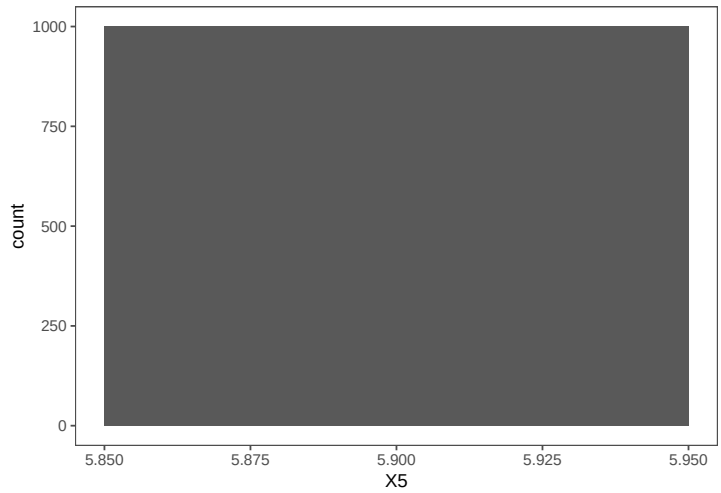

```
hist_x4 <- ggplot(data = fc_mean, aes(x = X4)) + geom_histogram(bins = 40) +
    theme_few()
hist_x4
```



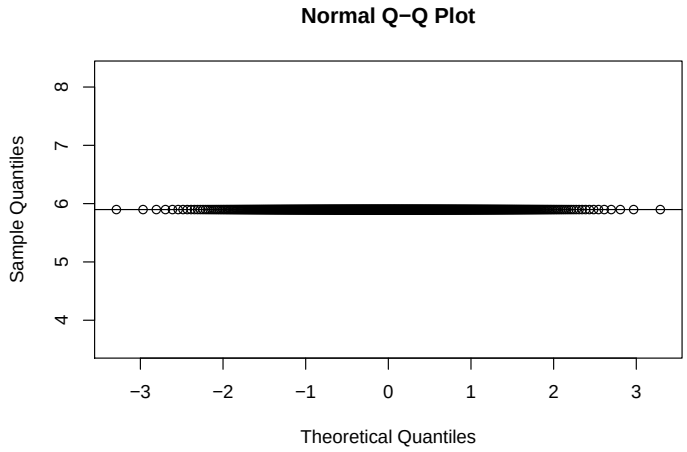

```
qq_x4 <- qqnorm(fc_mean$X4)
qqline(fc_mean$X4)
```




```
hist_x5 <- ggplot(data = fc_mean, aes(x = X5)) + geom_histogram(bins = 100) +
    theme_few()
hist_x5
```



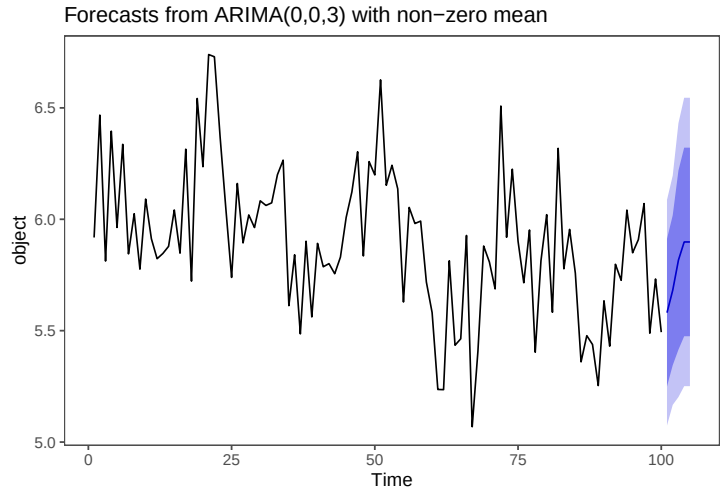
```
qq_x5 <- qqnorm(fc_mean$X5)
qqline(fc_mean$X5)
```



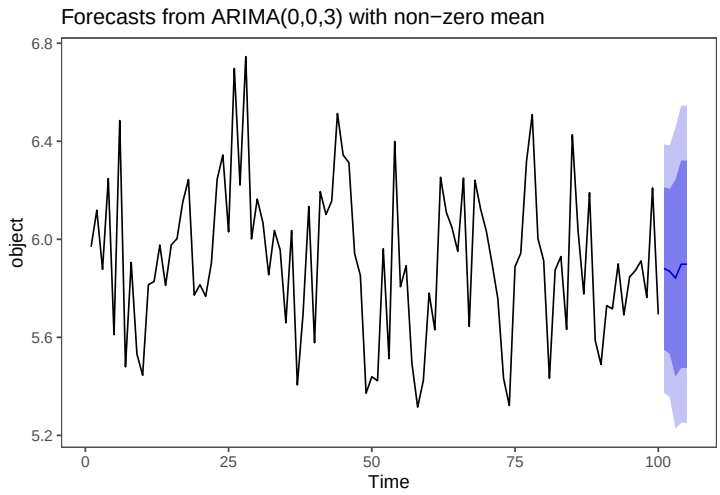
The results show that, from $h \geq 4$, the predicted value is the mean of the series.

Now, some forecasting plots:

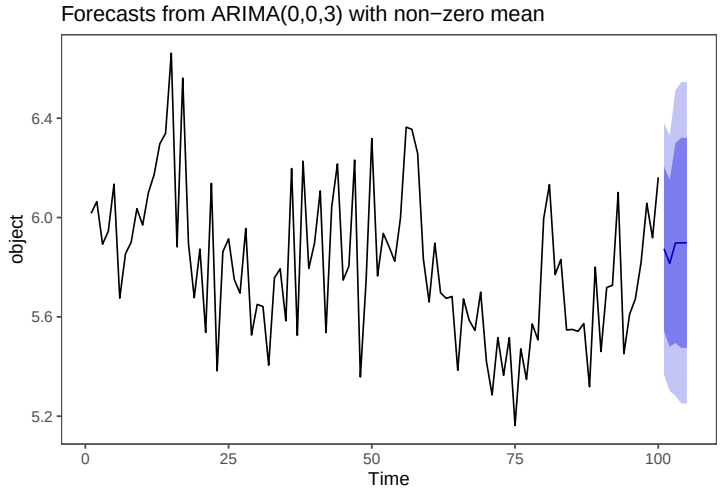
```
fc <- forecast(df$value, model = aa_model, h = h)

autoplot(fc) + theme_few()
```



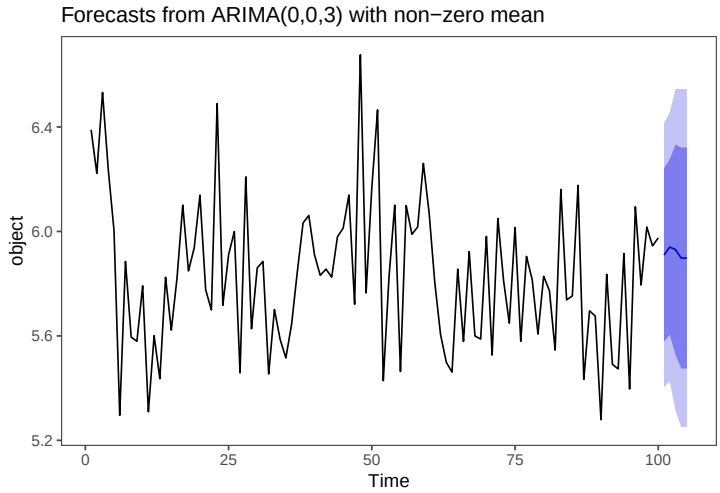

```
autoplot(fc_list[[1]]) + theme_few()
```




```
autoplot(fc_list[[66]]) + theme_few()
```




```
autoplot(fc_list[[796]]) + theme_few()
```



Chapter 8

Time series decomposition

In this chapter, we'll begin classifying time series according to some criteria:

- Does the time series gravitates around a certain mean?
- Does it present some sort of trend or seasonality?
- Does it have some sort of structrual break?
- Does it have constant variance?
- How much "memory" does the ts have?

8.1 Decomposing a time series

We can decompose a time series in the following way:

$$X_t \equiv f_t + Y_t,$$

where f_t is *deterministic* and Y_t is *stochastic.*

It is comprised of two steps:

1. Split the deterministic component f_t
2. Model the statistical properties of the stochastic component Y_t

The intuition behind this procedure is to transform a time series into a stationary and ergodic process – after all, this is a necessary condition to "apply Econometrics"!

8.1.1 Seasonality and trend

In general, the deterministic component f_t of a process can be decomposed between:

1. **Deterministic trend** (f_t). It is usually some function of time, e.g., $t, t^2, log\, t, exp\{t\}$.
2. **Seasonality** (s_t). Usually a *periodic function of time.*

$$X_t \equiv f_t + s_t + Y_t$$

The additive trend and seasonality model is the most common in practice:

$$X_t = f_t + s_t + Y_t$$

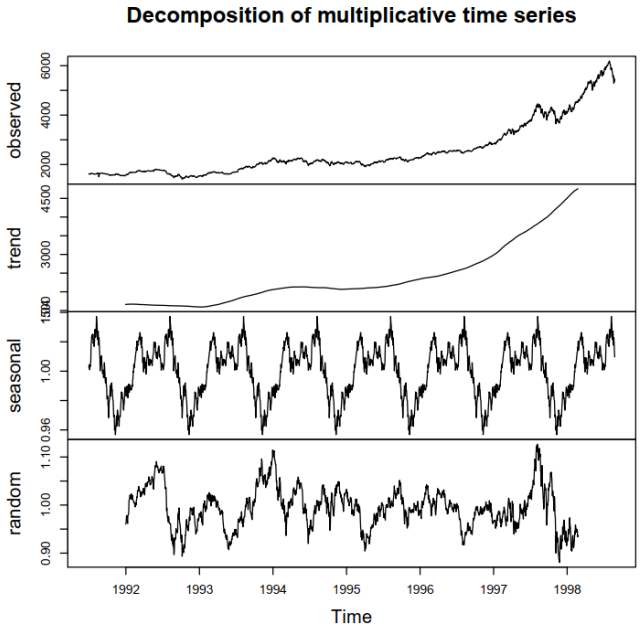
But, in some cases, it makes sense to postulate a multiplicative model:

$$X_t = f_t s_t Y_t$$

Or even a mixed form:

$$X_t = f_t + s_t Y_t$$

Note that, if the model is multiplicative, it is additive in *log*. Here is a graphical example of a decomposition:



8.1.2 Parametric deterministic trend

The most straightforward way to model a deterministic trend is with a *parametric specification* – i.e., a known function of time with a finite number of unknown parameters:

$$f_t = f(t, \gamma), \quad \gamma \in \mathbb{R}^k$$

Here are some frequently used examples: $\gamma, \quad \gamma t, \quad \gamma_0 + \gamma t, \quad \gamma_0 exp\{\gamma t\}$.

When we *take first differences* (or log first differences), *we are removing stochastic tendencies* (more on that on Chapter **??**).

8.1.3 De-Trending

The function f_t may be a polynomial or another more complex known function with unknown parameters. We *regress* X_t *on the function* f_t and find the estimator $\hat{\gamma}$, which we use to find the residuals:

$$\hat{Y}_t + s_t = X_t - f(t, \hat{\gamma})$$

We now have a *consistent estimator* for the stochastic and seasonality components[1]:

$$\hat{Y}_t \to_p Y_t, \text{ given that } \hat{\gamma} \to_p \gamma$$

8.1.4 Seasonality

Definition 8.1.1. *Regular and periodic movements in a process are called* ***seasonal movements*** *or* ***seasonality.***

Economic time series usually present some sort of seasonality – for example, during Holidays or weekends. Retail sales are usually higher at the end of the year, higher volatility in the stock market at the beginning of the week.

A simple way to estimate the seasonality component of the process is by constructing a *dummy*. Remember that a dummy for an event A is:

$$d_{A,t} = \begin{cases} 1 & , \text{ if } A \text{ happened in period } t \\ 0 & , \text{ otherwise} \end{cases}$$

With this, we can estimate the seasonality of the time series via the following parametric model:

$$\hat{s}_t = \hat{c} + \sum_{i=1}^{j-1} \hat{\delta}_i d_{A_i t},$$

where j represents the *size of the seasonality cycle* (e.g., for monthly data, $j = 12$).

Our estimate of the *stochastic* component of the series can now be obtained by:

$$\hat{Y}_t = X_t - f(t, \hat{\gamma}) - \hat{s}_t$$

8.1.5 Non-parametric decomposition

There are a number of *non-parametric* ways to detrend or remove seasonality of a process – i.e., that do not involve adjusting and estimating a model with given parameters.

Some examples are:

- HP filter
- Holt-Winters
- Baxter-King
- Christiano and Fitzgerald
- Butterworth (allows for structural breaks)
- Moving averages
- Exponential smoothing

[1] This is not an obvious conclusion. More on that on Chapter **??**.

8.1.6 Hodrick Prescott (HP) filter

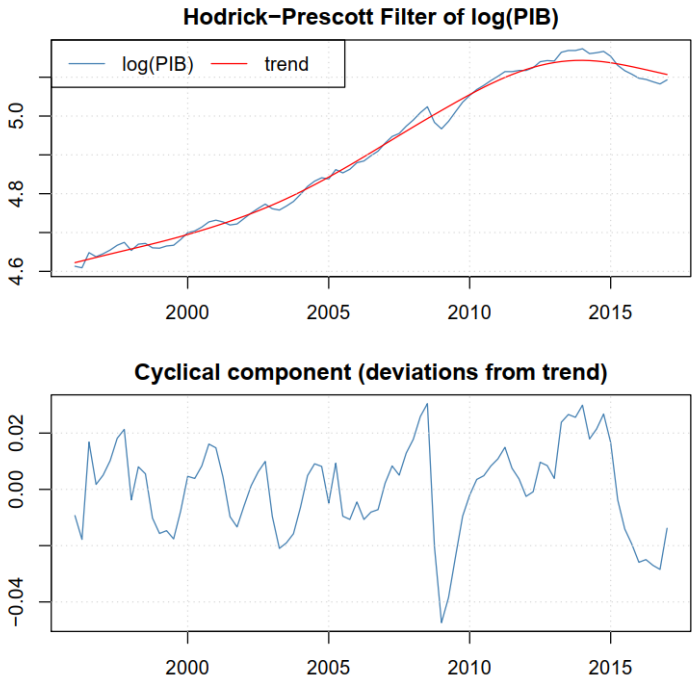
The idea behind the HP filter is to find a trend that is well adjusted to the observed time series. It is constructed by *weighting deviations from a purely linear trend:*

$$\min \left\{ \sum_{t=1}^{T} (X_t - f_t)^2 + \lambda \sum_{t=2}^{T-1} [(f_{t+1} - f_t) - (f_t - f_{t-1})]^2 \right\}$$

In other words, we're looking for $f_1, ..., f_T$ that solves the expression above for some $\lambda \geq 0$.

- If $\lambda = 0$, we have $f_t = X_t, \forall t$.
- If $\lambda = \infty$, f_t is a straight line.
- We usually solve the expression with $\lambda \in (0, \infty)$. For quarterly data, authors usually suggest $\lambda = 1600$.

Here's an example of a fitted HP filter with $\lambda = 1600$:



8.1.7 The *other* moving average

This is one of the first non-parametric methods of time series decomposition.

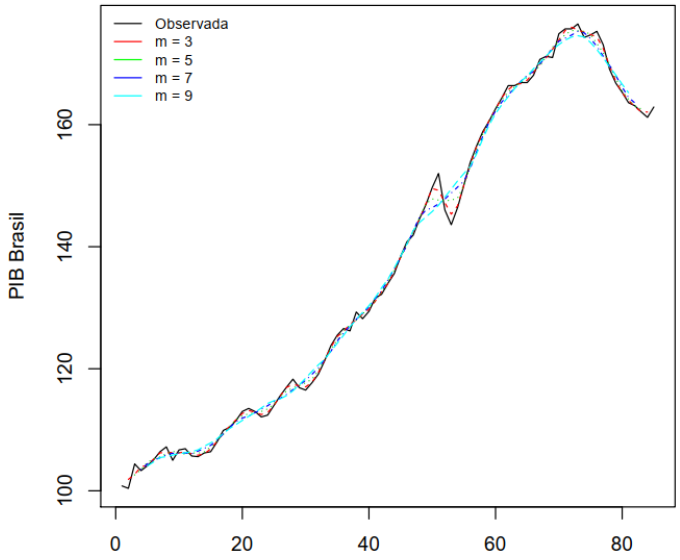
Definition 8.1.2. *A **simple symmetric moving average of order** m is given by:*

$$f_t = \frac{1}{m} \sum_{i=-k}^{k} Y_{t+i},$$

where $m = 2k + 1$.

The trend is, then, the mean of m periods centered around Y_t, hence the name *moving average.*

Here's an example of a trend estimated by a moving average:



The moving average does not need to weigh all observations uniformally evenly. We can calculate, for example, a *weighted moving average:*

$$f_t = \frac{1}{\sum_{i=-k}^{k} \omega_i} \sum_{i=-k}^{k} \omega_i Y_{t+i},$$

where $m = 2k + 1$.

It doesn't even have to be symmetric in regards to period t – i.e., it can be non-centered:

$$f_t = \frac{1}{\sum_{i=a}^{b} \omega_i} \sum_{i=a}^{b} \omega_i Y_{t+i},$$

where $a \leq b, (a, b) \in \mathbb{Z}$.

8.1.8 Parametric *vs.* Non-parametric functions

Non-parametric estimates and deterministic functions that depend on other variables z_t *cannot be easily extrapolated.* For example, with an HP filter, future observations review the most recent previous ones.

For a simple forecast, it is usual to postulate a parametric deterministic function *that allows for extrapolation.* To estimate sensibility to shocks, the non-parametric estimators are very common – e.g. potential GDP, NAIRU.

A protocol

To analyze a time series, we'll proceed in three steps:

1. Split the deterministic from the stochastic component – more art than science!
    - Clean the time series of "NA", trend over time and seasonality.
2. Look for structural breaks (More on that on Chapter **??**).

- When there are structural breaks, we will split the time series in two and jointly estimate two different models.

3. After cleaning up the time series, we shall check if the stochastic is stochastic or not by inspecting the ACF and the PACF.
   - If the ACF swiftly decreases, the stochastic component is probably stationary and ergodic – which means we can estimate it properly.
   - If the ACF decreases slowly, the stochastic component probably *won't be ergodic.* In that case, ARMA models won't work so well. More on that on Chapter **??**.

8.2 Estimation of non-ergodic processes

Consider this model:

$$Y_t = \alpha + \delta t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

Y_t satisfies all classical hypothesis of OLS regression, with usual F, t statistics. In case ε_t is not Gaussian, we need another technique to find the OLS distributions of $\hat{\alpha}, \hat{\delta}$.

$$\sqrt{(T)}(\hat{\beta} - \beta) = [(1/T)\sum X_t X_t']^{-1}[(1/\sqrt{T})\sum X_t \varepsilon_t]$$

The second term converges in distribution to $\mathcal{N}(0, \sigma^2)$. However, we cannot say that the first term converges in probability to a positive definite matrix Q^{-1}! The OLS estimators $\hat{\alpha}, \hat{\delta}$ *have different rates of convergence.* We can account for this by multiplying $\hat{\alpha}$ by $\sqrt{T}$ and $\hat{\delta}$ by $T^{3/2}$. This will be the idea behind the following proposition, presented without proof[2]

Proposition 8.2.1. *Let Y_t be generated according to a simple trend process $Y_t = \alpha + \delta t + \varepsilon_t$, where ε_t is iid white noise, $\mathbb{E}(\varepsilon_t^2) = \sigma^2$ and $\mathbb{E}(\varepsilon_t^4) < \infty$. Then*

$$\begin{bmatrix} \sqrt{T}(\hat{\alpha}_T - \alpha) \\ T^{3/2}\left(\hat{\delta}_T - \delta\right) \end{bmatrix} \xrightarrow{L} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}^{-1} \right)$$

Furthermore, we call $\hat{\delta}_t$ *superconsistent*:

$$\hat{\delta}_t \to_p \delta, \quad T(\hat{\delta}_t - \delta) \to_p 0$$

Given this proposition, we can use the usual t, F tests. This can be achieved by multiplying the statistic by the smallest rate of convergence among estimators[3].

8.3 Inference with linear dependence

As we have shown in the previous section, OLS estimation with deterministic parameters is convergent (or superconvergent) even with the series at hand is not ergodic. F, t tests are valid, given a good estimate of the errors of the model. We will now present a theorem that postulates a LLN for series with dependence without proof[4].

[2] See Hamilton (1994), 16.1-2.
[3] More on that on Hamilton (1994), 16.2.
[4] Prova no Material Complementar do Problema 5.

Theorem 8.3.1. ***Law of Large Numbers for Time Series with Dependence.*** *Let Y_t be a second order stationary time series where $\mathbb{E}(Y_t) := \mu$and $\mathbb{E}(Y_t - \mu)(Y_{t-j} - \mu) := \gamma_j$. Assume that $\sum_{j=0}^{\infty} |\gamma_j| < \infty$. Then, the sample mean $\bar{Y}_T := \frac{1}{T}\sum_{t=1}^{T} Y_t$ meets the following conditions:*

1. $\bar{Y}_t \to \mu$
2. $\lim_{T\to\infty} T\mathbb{E}\left(\bar{Y}_t - \mu\right)^2 = \sum_{j=-\infty}^{\infty} \gamma_j$

This means, essentially, that when a time series satisfies our condition for ergodicity (absolute sum of autocorrelations is finite), the sample mean is consistent for the population mean. Furthermore, (2) implies that, if we want the asymptotic variance to converge to a non-zero real number, we can multiply it by T – which suggests the estimator $\frac{1}{T}\sum_{t=-\infty}^{\infty} \hat{\gamma}_j$.

Some issues arise. First, this estimator is very different from OLS, and it is not feasible – after all, we have only so many (finite) observations. We now have two alternatives:

1. Model the ergodic series of the errors as an ARMA to isolate a white noise.
2. Modify our estimate of the variance of the erros in such a way to enable us to account for its dependence.

The second option is enabled by the *Newey-West estimator.* Approximating $\frac{1}{T}\sum_{t=-\infty}^{\infty} \hat{\gamma}_j$ by $\frac{1}{T}\sum_{t=-q}^{q} \hat{\gamma}_j$, we can arrive at the estimator. If we define $q := T^{1/4}$:

$$\tilde{s} := \gamma_0 + 2\sum_{v=1}^{q}\left[1 - \frac{v}{q+1}\right]\hat{\gamma}_v$$

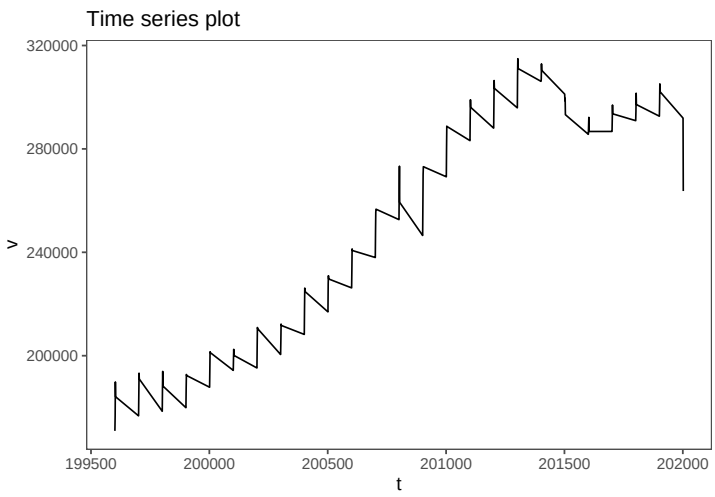
This is a consistent estimator that is robust to heteroskedasticity and dependence.

Chapter 9

Problem 5: Forecasting GDP

In this problem, we'll be forecasting GDP in the short term and creating some models of GDP growth in the long run. This presents some challenges, namely those related to *ergodicity* and *stationarity*.

```
pplot <- ggplot(data = pib, aes(x = t, y = v)) + geom_line() +
    ggtitle("Time series plot") + theme_few()
pplot
```



As we have downloaded the *pure* quarterly data, it presents *seasonality* and an upwards trend. This implies that the *time series will not be stationary*. Therefore, we need to employ methods that circumvent this issue and assure us that we can continue modelling the series as an ARMA(p,q).

9.1 Decomposing the time series

We will now assume that we can decompose the time series in three distinct elements in an additive model:

$$X_t = f_t + s_t + Y_t$$

, where f_t denotes the trend of the ts, s_t denotes seasonality, Y_t is stochastic. We also assume that f_t, s_t are *deterministic*.

9.1.1 Trend

First, we'll construct a *parametric* model of the trend. Let's assume that f_t can be modelled by a linear form:

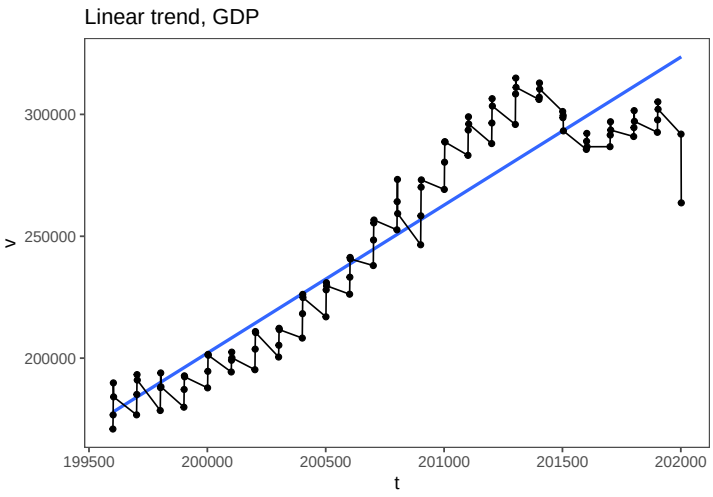
$$f_t = \gamma_0 + \gamma * t$$

```
linear_trend <- lm(v ~ t, data = pib)

summary(linear_trend)
```

```
##
## Call:
## lm(formula = v ~ t, data = pib)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -59917 -10577  -2046  11571  33717
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.194e+07  4.644e+05  -25.71   <2e-16 ***
## t            6.070e+01  2.313e+00   26.25   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16200 on 96 degrees of freedom
## Multiple R-squared:  0.8777, Adjusted R-squared:  0.8764
## F-statistic: 688.8 on 1 and 96 DF,  p-value: < 2.2e-16
```

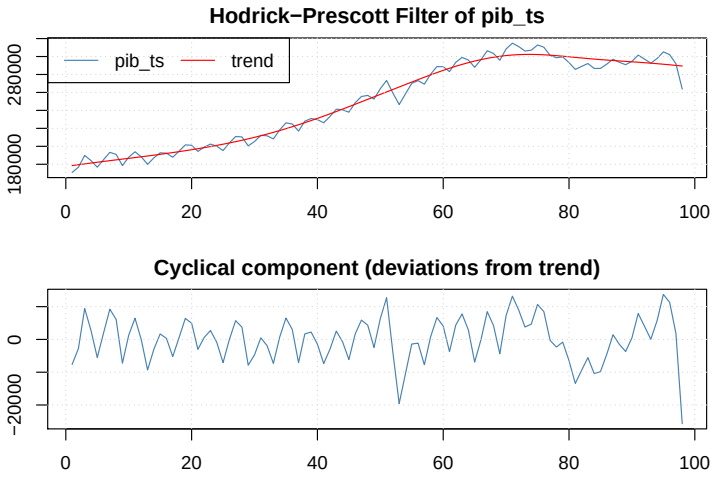
```
ggplot(data = pib, aes(x = t, y = v)) + stat_smooth(method = "lm",
se = F) + geom_line() + geom_point() + theme_few() + ggtitle("Linear trend, GDP")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Another way to find f_t is via a *non-parametric* process. For this, we'll use an HP filter and a moving average.

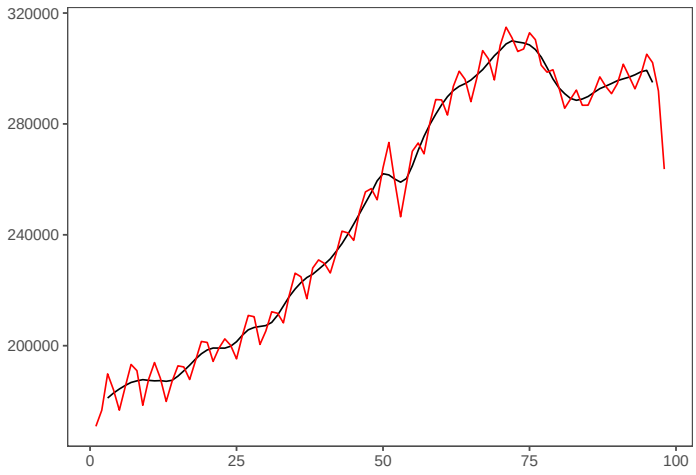
```
pib_ts <- ts(pib$v)

hp_trend <- hpfilter(pib_ts, freq = 1600, type = "lambda")

plot(hp_trend)
```



Now, a moving average.

```
pib_ma <- ma(pib$v, order = 4)

autoplot(pib_ma, color = "blue") + geom_line(data = pib, aes(x = 1:length(pib$t),
y = v), color = "red") + theme_few()
```

```
## Warning: Use of `pib$t` is discouraged. Use `t` instead.

## Warning: Removed 4 row(s) containing missing values (geom_path).
```



9.1.2 Seasonality

We can now create a function for s_t. This will be done with dummies:

$$D_i = 1, i = t$$

$$D_i = 0 \, otherwise$$

```r
tri <- c(NA)

tri1 <- c(1, 2, 3, 4)

i = 1

while (i < 25) {

    tri <- append(tri, tri1)

    i = i + 1

}

tri <- tri[-1]
tri <- c(tri, 1, 2)

length(tri)
```

```
## [1] 98
```

```r
pib <- data.frame(pib, tri)

names(pib)[1] <- "t"
names(pib)[2] <- "v"
names(pib)[3] <- "tri"

dummies <- data.frame(matrix(NA, nrow = length(pib$t), ncol = 4))

for (j in 1:4) {

    dummies[j] <- as.numeric(pib$tri == j)

}

hp_fitted <- hp_trend[2]

hp_fitted <- hp_fitted$trend

detrend <- pib$v - hp_fitted

pib <- data.frame(pib, dummies, detrend)

names(pib) <- c("t", "v", "tri", "X1", "X2", "X3", "X4", "detrend")

head(pib)
```

```
##           t         v tri X1 X2 X3 X4   detrend
## 18   199601 170920.0   1  1  0  0  0 -7639.363
## 40   199602 176708.8   2  0  1  0  0 -2784.369
## 62   199603 189844.3   3  0  0  1  0  9422.159
## 84   199604 184112.9   4  0  0  0  1  2773.147
## 106 199701 176732.2   1  1  0  0  0 -5513.319
## 128 199702 185109.5   2  0  1  0  0  1968.942
```

```r
dummy_lm <- lm(detrend ~ X2 + X3 + X4, data = pib)

summary(dummy_lm)
```

```
##
## Call:
## lm(formula = detrend ~ X2 + X3 + X4, data = pib)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24604.3  -2770.1    904.6   2781.6   9672.2
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -5885       1079  -5.454 3.97e-07 ***
## X2              4775       1526   3.129  0.00234 **
## X3             11476       1542   7.443 4.63e-11 ***
## X4              7580       1542   4.916 3.73e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5396 on 94 degrees of freedom
## Multiple R-squared:  0.3854, Adjusted R-squared:  0.3658
## F-statistic: 19.65 on 3 and 94 DF,  p-value: 5.704e-10
```

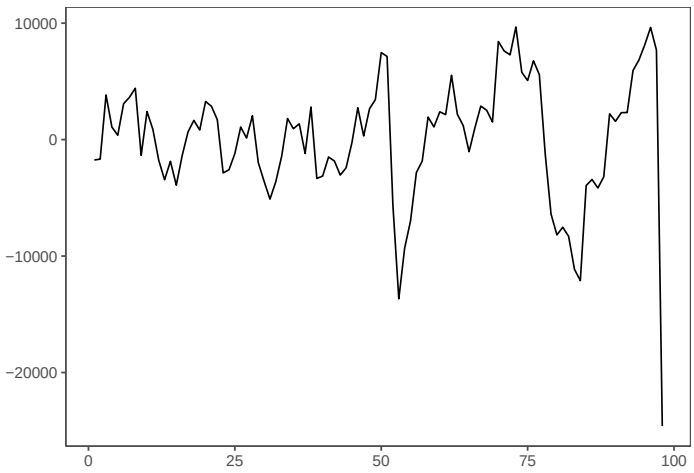
9.1.3 Y_t

We'll now use the HP-fitered version of f_t and the dummy approach to s_t.

```
yt <- as.vector(pib$v) - (hp_fitted + dummy_lm$fitted.values)

mean(yt)
```

```
## [1] 2.970866e-13
```

```
autoplot(yt) + theme_few()
```




```
y <- data.frame(1:98, yt)

names(y) <- c("t", "yt")

y
```

9.2 Identifying and estimating ARMA(p,q) for Y_t

We are now in a position to identify and estimate the best model for our time series Y_t.

Applying the function *auto.arima* from the package *forecast* to identify and estimate the model:

```
aa_model <- auto.arima(y$yt, num.cores = 24, max.d = 0, stepwise = F)

summary(aa_model)
```

```
## Series: y$yt
## ARIMA(2,0,2) with zero mean
##
## Coefficients:
##           ar1     ar2     ma1     ma2
##       -0.5799  0.3799  1.6394  0.7070
## s.e.   0.1463  0.1453  0.1191  0.1165
##
## sigma^2 estimated as 16014834:  log likelihood=-951.01
## AIC=1912.01   AICc=1912.66   BIC=1924.94
##
## Training set error measures:
##                       ME     RMSE      MAE      MPE     MAPE      MASE
## Training set -166.6391 3919.333 2469.197 32.42135 93.29931 0.9554301
##                       ACF1
## Training set -0.001381943
```

```
print("t-values: ")
```

```
## [1] "t-values: "
```

```
aa_t <- matrix(NA, nrow = aa_model$arma[1] + aa_model$arma[2])

for (i in c(1:(aa_model$arma[1] + aa_model$arma[2]))) {

    aa_t[i] <- aa_model$coef[i]/sqrt(aa_model$var.coef[i, i])

}

aa_t <- data.frame(aa_t)

aa_t
```

```
##        aa_t
## 1 -3.965137
## 2  2.614792
## 3 13.768478
## 4  6.067555
```

```
aa_q <- Box.test(aa_model$residuals, lag = aa_model$arma[1] +
    aa_model$arma[2])
aa_q
```

```
##
##  Box-Pierce test
##
## data:  aa_model$residuals
## X-squared = 0.22, df = 4, p-value = 0.9944
```

```
criteria <- matrix(NA, nrow = 1, ncol = 3)

aa_criteria <- data.frame("AR(2)*", aa_model$aic, aa_model$bic)
```

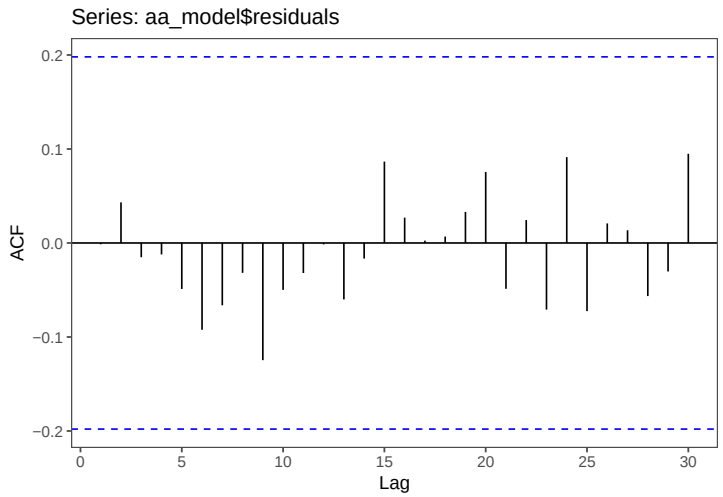
```
names(aa_criteria) <- c("Model", "AIC", "BIC")

aa_criteria
```

```
##     Model      AIC      BIC
## 1 AR(2)* 1912.011 1924.936
```

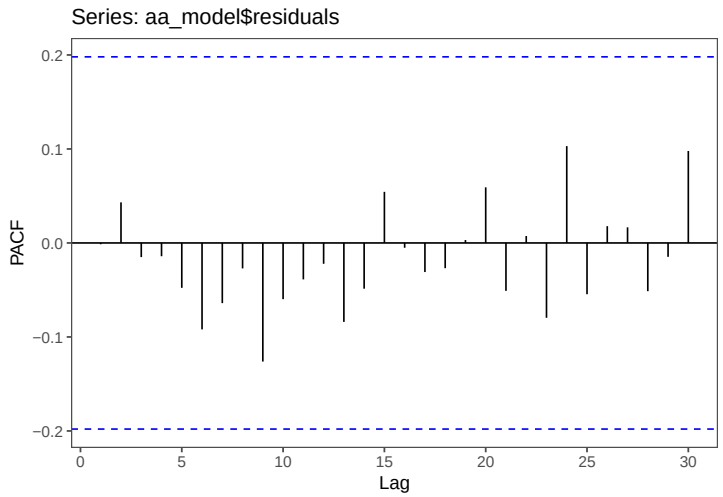
```
fac_e <- ggAcf(aa_model$residuals, type = "correlation", lag.max = 30,
plot = T) + theme_few()

facp_e <- ggPacf(aa_model$residuals, type = "correlation", lag.max = 30,
plot = T) + theme_few()
```

```
## Warning: Ignoring unknown parameters: type
```

```
fac_e
```



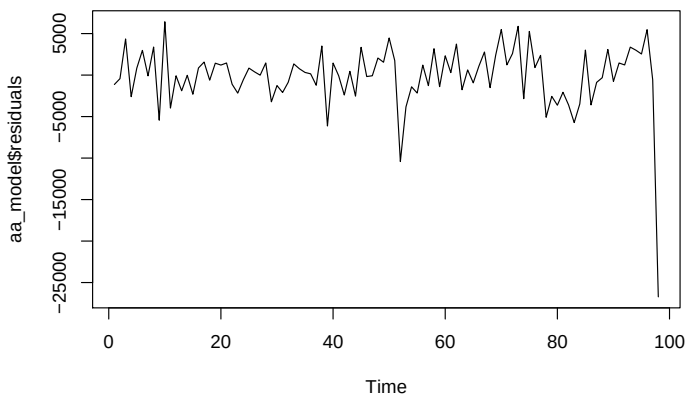

```
facp_e
```




```
mean(aa_model$residuals)
```

```
## [1] -166.6391
```

```
plot(aa_model$residuals)
```



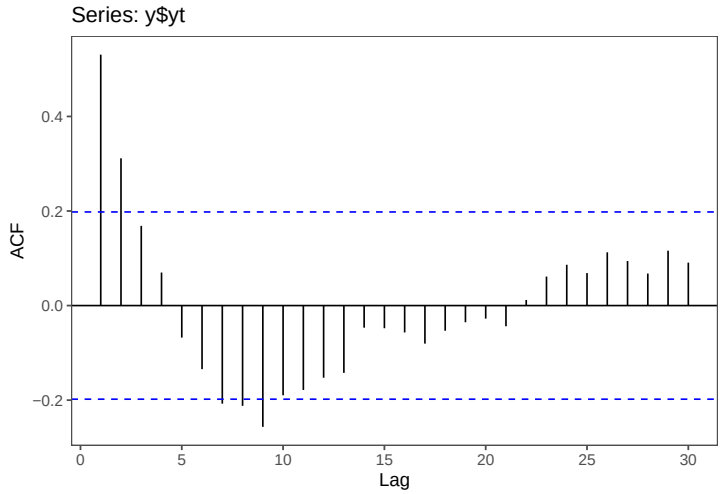

```
facst <- ggAcf(y$yt, type = "correlation", lag.max = 30, plot = T) +
    theme_few()
faclt <- ggAcf(y$yt, type = "correlation", lag.max = 5000, plot = T) +
    theme_few()

facpst <- ggPacf(y$yt, type = "correlation", lag.max = 30, plot = T) +
    theme_few()
```

```
## Warning: Ignoring unknown parameters: type
```

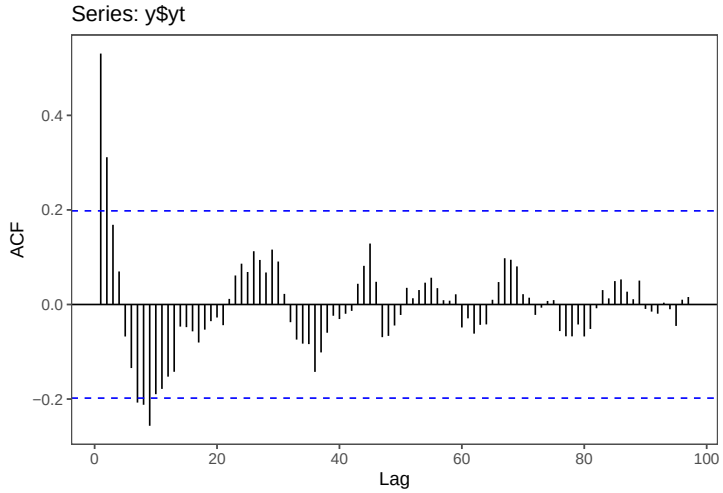
```
facplt <- ggPacf(y$yt, type = "correlation", lag.max = 5000,
plot = T) + theme_few()
```

```
## Warning: Ignoring unknown parameters: type
```

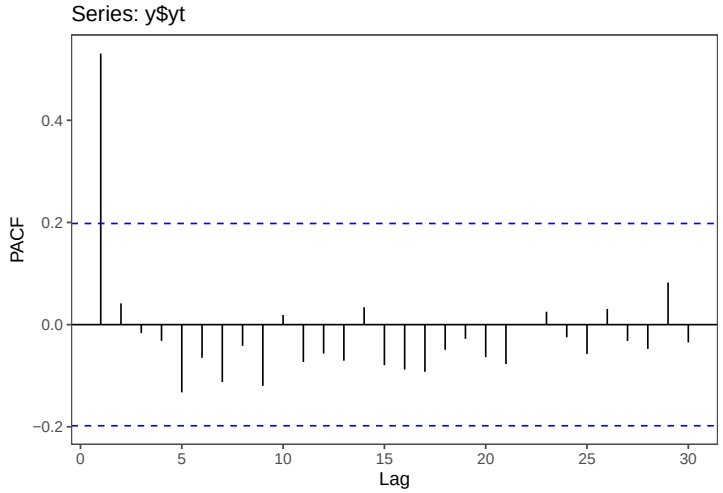
```
facst
```



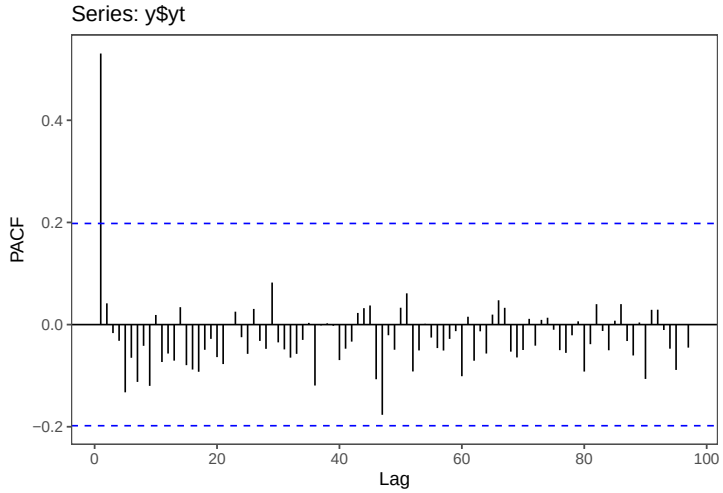

```
faclt
```




```
facpst
```



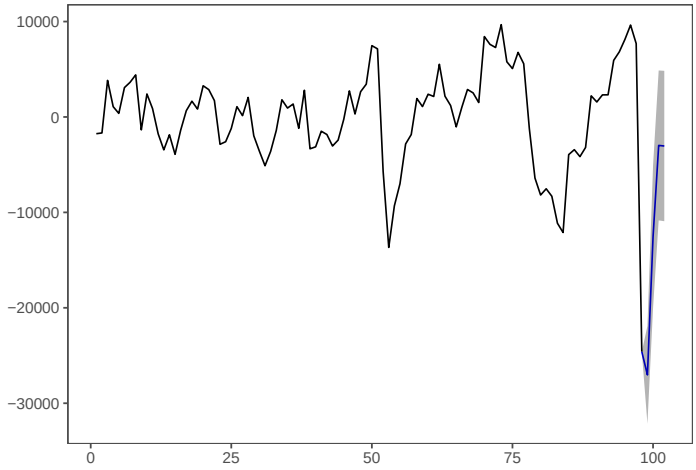

```
facplt
```



The results of *auto.arima* imply that the best model is an ARMA(2,0) – i.e., an AR(2):

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim wn(0, \sigma^2)$$

```
fc <- forecast(y$yt, model = aa_model, h = 4)

autoplot(fc) + theme_few()
```



9.3 Long term GDP growth

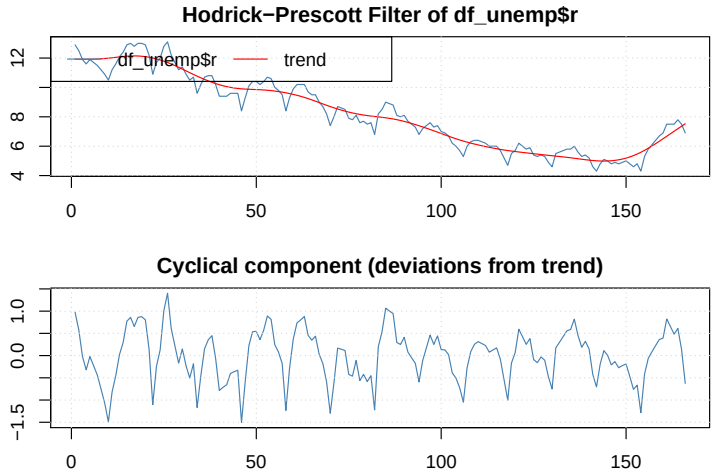
```
unemp <- read_excel("C:/Users/William/Downloads/tabela2176.xlsx")
```

```
unemp1 <- as.numeric(unemp[11, ])
```

```
## Warning: NAs introduced by coercion
```

```
unemp2 <- unemp1[2:(length(unemp1) - 2)]

unemp <- unemp2

df_unemp <- data.frame(1:length(unemp), unemp)

names(df_unemp) <- c("t", "r")

df_unemp
```

```
hp_unemp <- hpfilter(df_unemp$r, freq = 1600, type = "lambda")

plot(hp_unemp)
```




```
nairu <- hp_unemp$trend

nairu
```

```
t6162 <- get_sidra(6612, variable = 9318, category = 90707, period = c("200202",
"200203", "200204", "200301", "200302", "200303", "200304",
"200401", "200402", "200403", "200404", "200501", "200502",
"200503", "200504", "200601", "200602", "200603", "200604",
"200701", "200702", "200703", "200704", "200801", "200802",
"200803", "200804", "200901", "200902", "200903", "200904",
"201001", "201002", "201003", "201004", "201101", "201102",
"201103", "201104", "201201", "201202", "201203", "201204",
"201301", "201302", "201303", "201304", "201401", "201402",
"201403", "201404", "201501", "201502", "201503", "201504"))
```

```
## Considering all categories once 'classific' was set to 'all' (default)
```

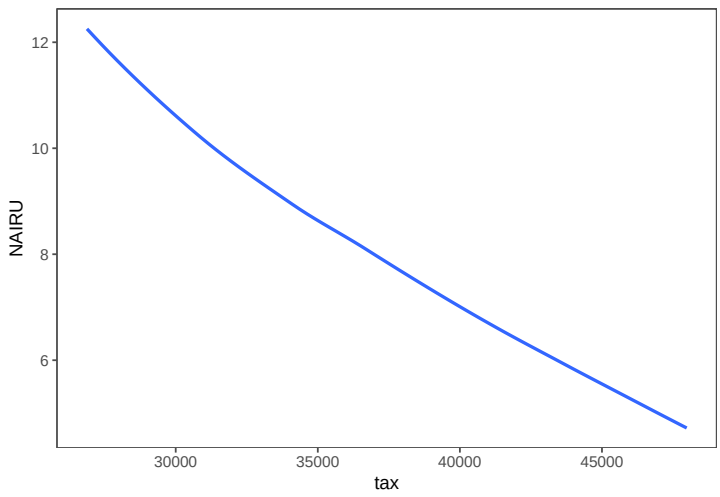
```
View(t6162)

tax <- t6162[(t6162$`Setores e subsetores (Código)` == 90706),
    ]

tax2 <- tax[, c(5, 13)]

names(tax2)[1] <- "t"
names(tax2)[2] <- "r"

tax <- tax2

trimestra <- c(NA)

i <- 0
while (i < length(unemp)) {

    media <- (unemp[i] + unemp[i + 1] + unemp[i + 2])/3
    trimestra <- append(trimestra, media)

    i <- i + 3
}

nairu_3m <- trimestra

df <- data.frame(nairu_3m[-1], tax)

df
```

```r
names(df) <- c("NAIRU", "t", "tax")

growth_lm <- lm(NAIRU ~ tax, data = df)

summary(growth_lm)
```

```
##
## Call:
## lm(formula = NAIRU ~ tax, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.1801 -0.3400 -0.0709  0.3167  1.6633
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.110e+01  4.471e-01   47.19   <2e-16 ***
## tax         -3.479e-04  1.184e-05  -29.37   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5991 on 52 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9431, Adjusted R-squared:  0.942
## F-statistic: 862.5 on 1 and 52 DF,  p-value: < 2.2e-16
```

```r
ggplot(data = df, aes(x = tax, y = NAIRU)) + stat_smooth(nethod = "lm",
se = F) + theme_few()
```

```
## Warning: Ignoring unknown parameters: nethod

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```



Chapter 10

Integrated Processes

10.1 Motivation

Up to this point, we have focused exclusively on stationary processes. Henceforth, we'll deal with some classes of non-stationary processes. Informally, we can say that the class of non-stationary processes is much more extensive than the class of stationary ones – in particular, ARMA processes. Most of the time, we will try to "stationarize" a non-stationary process.

10.2 Non-stationary processes

Definition 10.2.1. *A **random walk process** is given by:*

$$Y_t = Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma^2),$$

with $Y_0 = 0$.

Note that a random walk is *an AR(1) with* $\phi = 1$, which implies that *it is not stationary.*

Definition 10.2.2. *A **random walk process with drift** is given by:*

$$Y_t = \delta + Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma^2)$$

The other case defined earlier is a pure random walk ($\delta = 0$). Furthermore, the initial condition can be $Y_0 = c \neq 0$. Defining a generic random walk:

Definition 10.2.3. *A **generic random walk process** is given by:*

$$Y_t = \delta + Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma^2),$$

with $Y_0 = x$*, which can be deterministic (c) or random (*f_x*).*

10.2.1 Recursive representation of a random walk

Beginning at an arbitrary period t, we can recursively substitute until $t = 0$:

$$\begin{aligned} Y_t &= \delta + Y_{t-1} + \varepsilon_t \\ &= \delta + \delta + Y_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\ &= \delta + \delta + \delta + Y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\ &= \\ &= \delta + \delta + \cdots + \delta + Y_0 + \varepsilon_1 + \cdots + \varepsilon_{t-1} + \varepsilon_t, \end{aligned}$$

which gives us an alternate representation of the random walk:

$$Y_t = c + \delta t + \sum_{i=1}^{t} \varepsilon_t \tag{10.1}$$

It is now evident that the process *is not stationary* – the influence of previous innovations is *permanent!*

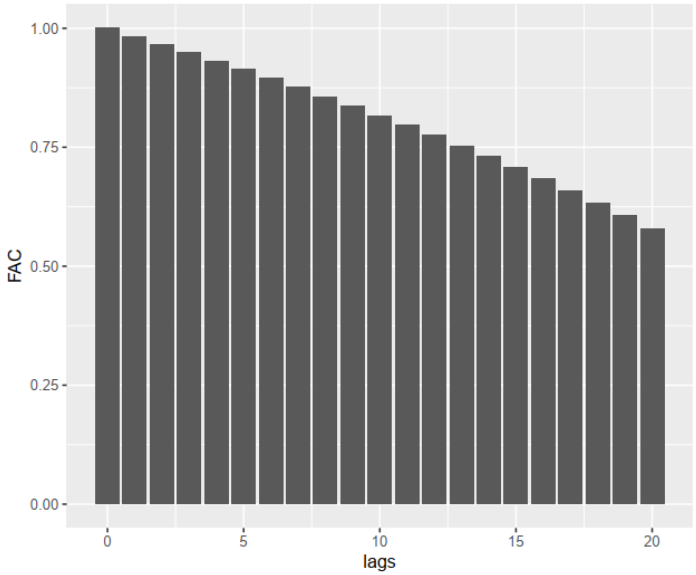
10.2.2 Moments of a random walk

Using the alternate representation of 10.1, it is simple to derive the moments of the random walk:

$$\begin{aligned} \mu_t &:= \mathbb{E}(Y_t) = c + \delta t \\ \gamma_{j,t} &:= Cov(Y_t, Y_{t-j}) = (t-j)\sigma^2, \quad j \geq 0 \\ \rho_{j,t} &:= \frac{\gamma_{j,t}}{\sqrt{\gamma_{0,t}\gamma_{0,t-j}}} = \frac{t-j}{\sqrt{t(t-j)}} = \sqrt{\frac{t-j}{t}} = \sqrt{1-\frac{j}{t}} \end{aligned}$$

This means that, as t grows, $\rho_{j,t}$ gets closer to 1 for fixed j. For fixed t, the correlation falls approximately linearly with j.

Here's a typical random walk ACF:



10.3 Unit root processes

The random walk is an example of the more general class of *unit root processes.*

Definition 10.3.1. *An **unit root process** follows the following form:*

$$Y = c + \delta t + u_t,$$

where c, δ, u_t have an ARMA(p,q) representation:

$$\Phi_p(L)u_t = \Theta_q(L)\varepsilon_t, \quad \varepsilon_t \sim (0, \sigma^2)$$

If all roots of $\Theta_p(L)$ are outside of the unit circle, then u_t is weakly stationary.

10.3.1 Integrated processes of order 1

Now, suppose that one of these roots is *unity*, and the rest is outside of the unit circle:

$$\Phi_p(L) = (1 - L)(1 - \lambda_2 L)(1 - \lambda_3 L)\dots(1 - \lambda_p L),$$

where λ_i is the inverse of the root.

Then, we have:

$$(1 - L)u_t = (1 - \lambda_2 L)^{-1}\dots(1 - \lambda_p L)^{-1}\Theta_q(L)\varepsilon_t =: \Psi(L)\varepsilon_t$$

This yields:

$$\begin{aligned}(1 - L)Y_t &= (1 - L)c + (1 - L)\delta t + (1 - L)u_t \\ &= 0 + \delta + (1 - L)u_t \\ &= \delta + \Psi(L)\varepsilon_t\end{aligned}$$

Equivalently, we can define:

$$\Delta Y_t := (1 - L)Y_t = Y_t - Y_{t-1} = \delta + \Psi(L)\varepsilon_t$$

This form highlights a very important result. If the process has *one* unit root, by *taking differences* we arrive at a stationary process! This process is called integrated of order 1.

Definition 10.3.2. *An **integrated process of order 1, or I(1),** is a stochastic process whose difference is stationary.*

Using this nomenclature, stationary processes are called integrated of order zero, or I(0).

10.3.2 Integrated process of order d

A process may have *more than one unit root.*

Definition 10.3.3. *An **integrated process of order d, or I(d),** is a stochastic process whose d-th difference is stationary:*

$$\Delta^d Y_t = (1 - L)^d Y_t = u_t, \quad u_t \sim I(0)$$

$$Y_t \sim I(d) \implies \Delta^d Y_t \sim I(0)$$

10.3.3 ARIMA(p,d,q) processes

We now combine this concept with the ARMA(p,q) class.

Definition 10.3.4. *An **integrated autoregressive moving average model of (p,d,q) order, or ARIMA(p,d,q),** is a process that, after taking d differences, becomes a stationary ARMA(p,q):*

$$\Phi_p(L)(1-L)^d Y_t = c + \Theta_q(L)\varepsilon_t; \quad \varepsilon_t \sim \left(0, \sigma^2\right),$$

where p is the autoregressive order (aside from the unit roots), q is the order of the moving average and the roots of $\Phi_p(L)$ are outside the unit circle.

10.4 Stochastic or deterministic trend?

Let's compare two non-stationary processes:

$$Y_t = c + \delta_t + u_t, \quad u_t \sim I(0)$$
$$Z_t = \delta + Z_{t-1} + u_t, \quad t \geq 1, Z_0 = c$$

The first model has a deterministic linear trend. The second one has a linear *stochastic* trend. Both processes clearly have *linear* trends:

$$\mathbb{E}(Y_t) = \mathbb{Z}_{\approx} = c + \delta t$$

However, the variances of these processes are very different:

$$Var(Y_t) = \sigma^2, \quad Var(Z_t) = t\sigma^2$$

Clearly, the first one is stationary (constant variance), while the second has an ever increasing variance. From an economic point of view, this has a fundamental implication: is the impact of the shock *permanent* or *temporary?* e.g. Is the impact of a recession on GDP temporary or permanent?

10.4.1 Detrending by taking first differences

If the trend is stochastic, the correct procedure is to take first differences of the series to obtain a stationary process:

$$\Delta Z_t = Z_t - Z_{t-1} = \delta + u_t$$

However, if we take first differences of a process with a *deterministic trend*:

$$\Delta Y_t = \delta + u_t - u_{t-1} = \delta + (1-L)u_t$$

The result, while still stationary, induces a unit root in the MA portion of the process. This implies that the MA($\cdot$) will no longer be invertible – which hinders estimation! (See Chapter 4).

10.4.2 Detrending by parametric decomposition

If the trend is deterministic, the correct procedure is to adjust a parametric linear tendency, given that $Y_t - c - \delta t = u_t$. However, if we adjust a linar trend in a process with stochastic trend, *we do not obtain stationarity*:

$$Z_t - c - \delta t = \sum_{i=1}^{t} u_t,$$

which is a process whose variance grows with t.

10.4.3 How to distinguish between trends?

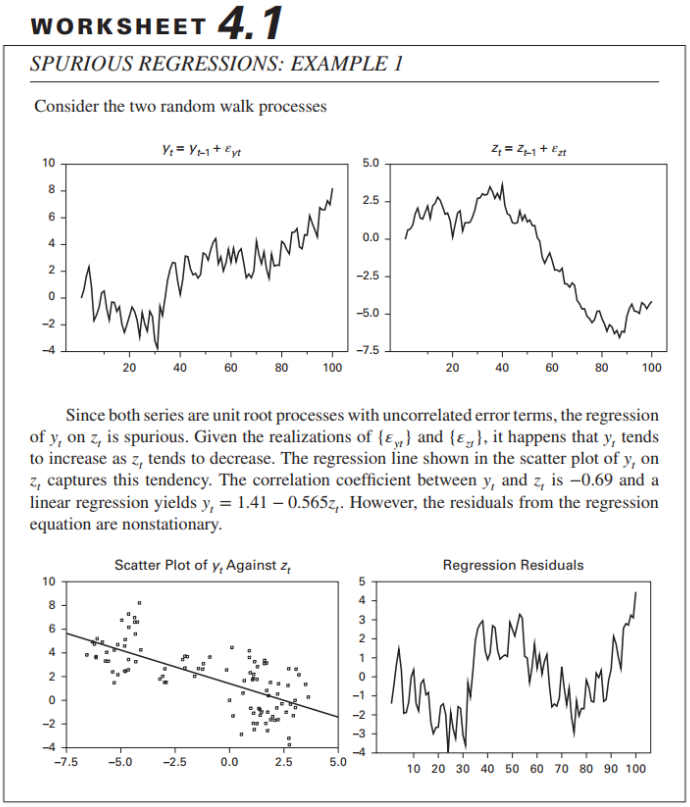
This is a major practical issue. In practice, it is very hard to distinguish a pure random walk and an AR(1) with $\phi = 0.99$! Most tests for unit root detection are low-powered, have unconventional distributions and vary wildly according to the formulation of the null hypothesis (non-pivotal). More on this on Chapter **??**.

Chapter 11

Problem 6: Spurious regressions and estimating unit root processes

11.1 Spurious regressions

A **spurious regression** has high R^2, t, notwithstanding the fact that the results appear to have no economic meaning. In other words, the estimated parameters are significant while the population parameters are equal to zero. This result happens essentially because the residuals of the time series at hand *are not stationary.* This implies that the errors will have permanent effects and will accumulate over time, which causes the R^2 to approach unity and the variance of the error to explode.



11.2 Unit roots

As has been discussed during the lecture, when the population model is a *random walk*:

$$Y_t = 1 * Y_{t-1} + \varepsilon_t, \quad \varepsilon \sim wn(0, \sigma^2),$$

it happens that the series *is not ergodic* (nor is it stationary). Therefore, the usual asymptotic properties *do not hold*, as all innovations have *permanent effects*.

$$Y_t = c + \delta t + \sum_{i=1}^{t} \varepsilon_i$$

The random walk, defined above, is an example of the more general class of *unit root processes*:

$$Y_t = c + \delta t + u_t,$$

u_t has an ARMA(p,q) representation:

$$\Phi_p(L)u_t = \Theta_q(L)\varepsilon_t, \quad \varepsilon \sim wn(0, \sigma^2)$$

Suppose that one of the roots of $\Theta_p(L)$ is equal to 1:

$$\Theta_p(L) = (1 - [1]L)(1 - \lambda_2 L)...(1 - \lambda_p L)$$

$$(1 - L)u_t = (1 - \lambda_2 L)^{-1}...(1 - \lambda_p L)^{-1}\Theta_q(L)\varepsilon_t =: \Psi(L)\varepsilon_t$$

We can now rewrite this as:

$$(1 - L)Y_t = (1 - L)c + (1 - L)\delta t + (1 - L)u_t$$

$$(1 - L)Y_t = \delta + \Psi(L)\varepsilon_t$$

Defining ΔY_t, we have:

$$\Delta Y_t := (1 - L)Y_t = Y_t - Y_{t-1} = \delta + \Psi(L)\varepsilon_t$$

With this concept, we can define ARIMA(p,d,q) processes:

$$\Phi_p(L)(1 - L)^d Y_t = c + \Theta_q(L)\varepsilon_t, \quad \varepsilon_t \sim wn(0, \sigma^2)$$

11.2.1 Hypothesis testing

It turns out that testing for unit root presence presents some challenges. Under the null ($a_1 = 1$), its distribution *is not standard* and does not present the usual asymptotic properties. We can circumvent this issue with the use of numeric methods, such as the *Monte Carlo simulation*. It will now be employed, following Dickey and Fuller (1979).

First, some notation:

We begin with a simple model:

$$Y_t = a_1 y_{t-1} + \varepsilon_t$$

$$\Delta Y_t = \gamma y_{t-1} + \varepsilon_t, \quad \gamma := a_1 - 1$$

Dickey and Fuller constructed the following regression equations:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t \tag{11.1}$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t \tag{11.2}$$

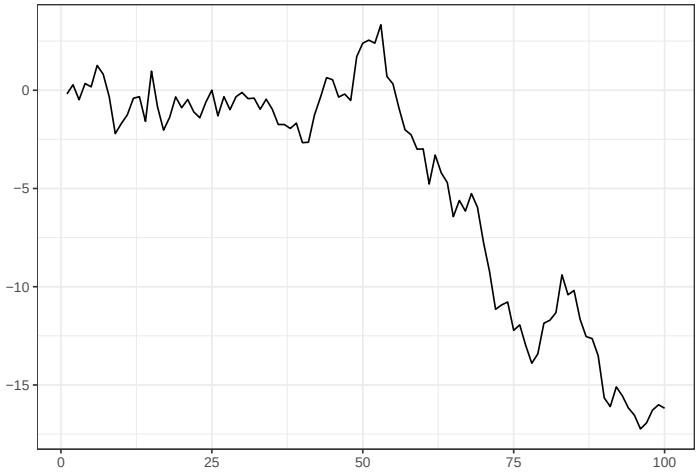
$$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t \tag{11.3}$$

(1) has no intercept and represents a simple random walk. (2) includes an intercept. (3) also adds a deterministic time trend.

Note that the critical values of the t-statistics are *different* between these regressions. This will now be shown with our Monte Carlo simulation.

For equation (1):

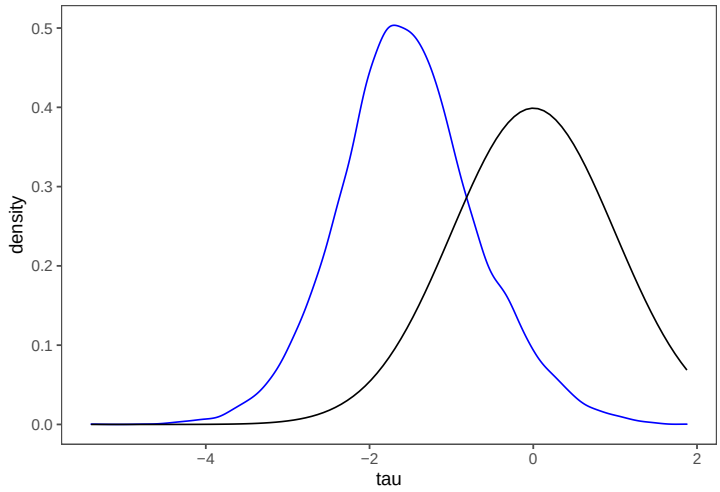
```
set.seed(76345)

# length of ts
T <- 100

# Loops
S <- 10000

e <- rnorm(T, 0, 1)

# As y_{t+1} = y_t + \varepsilon_t, we can write this as an MA(\infty) model. y_t = \sum \varepsilon_i. This is now done with

y <- cumsum(e)

y <- vector()

y[1] = e[1]

for (i in (2:T)) {

    y[i] <- y[(i-1)] + e[i]

}

y <- as.ts(y)

autoplot(y) + theme_bw()
```




```
# Taking the first difference

y_diff <- diff(y)

# Regression of I(1) model

reg <- dynlm(y_diff ~ 0 + L(y,1)) # no lag (x1 = 0)

reg <-summary(reg)

reg$coefficients[1,3] # t value
```

```
## [1] 1.110447
```

```
# Loop

tau <- vector()

k <- 1

delta <- 0

for (i in 1:S) {

  e <- rnorm(T,0,1)
  y = k + delta*seq(1:T) + cumsum(e)
  y <- as.ts(y)

  y_diff <- diff(y)

  reg <- summary(dynlm(y_diff ~ 1 + L(y, 1)))

  tau[i] <- reg$coefficients[2,3]

}

tau.df <- data.frame(tau)

ggplot(data = tau.df, aes(x = tau)) + geom_density(color = "blue") + stat_function(fun = dnorm, n = 101, args = c(mean = 0, s
```




```
jarque.bera.test(tau) # We reject H0 at the 1% significance level.
```

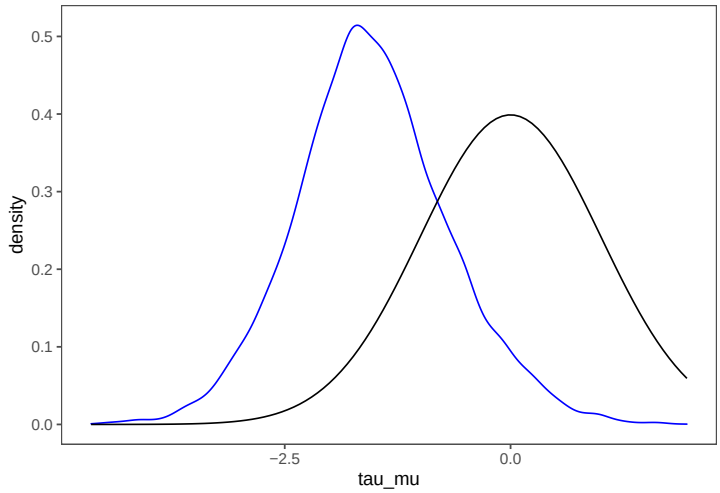
```
##
##  Jarque Bera Test
##
## data:  tau
## X-squared = 86.138, df = 2, p-value < 2.2e-16
```

```
tau.ci <- quantile(tau, c(0.01, 0.05, 0.1))

tau.ci
```

```
##         1%         5%        10%
## -3.494124 -2.880240 -2.567557
```

For equation (2) – with an intercept:

```
# Loop

tau_mu <- vector()

for (i in 1:S) {

    e <- rnorm(T, 0, 1)
    y <- cumsum(e)
    y <- as.ts(y)

    y_diff <- diff(y)

    reg <- summary(dynlm(y_diff ~ 1 + L(y, 1)))

    tau_mu[i] <- reg$coefficients[2, 3]

}

tau_mu.df <- data.frame(tau_mu)

ggplot(data = tau_mu.df, aes(x = tau_mu)) + geom_density(color = "blue") +
    stat_function(fun = dnorm, n = 101, args = c(mean = 0, sd = 1)) +
    theme_few()
```




```
jarque.bera.test(tau_mu)  # We reject HO at the 1% significance level.
```

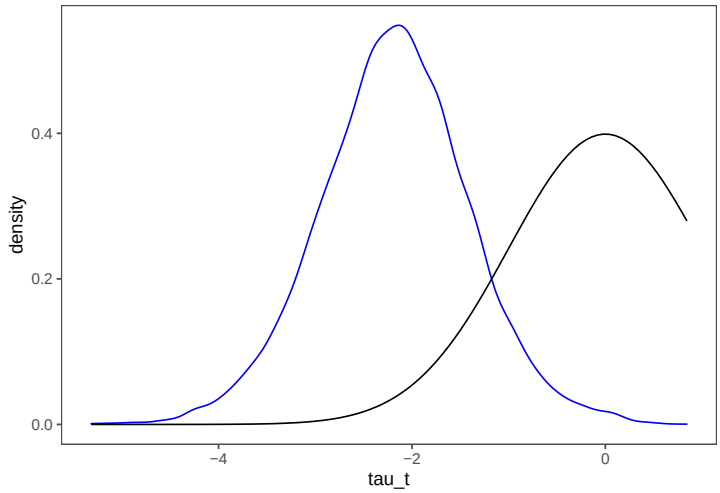
```
##
##  Jarque Bera Test
##
## data:  tau_mu
## X-squared = 96.894, df = 2, p-value < 2.2e-16
```

```
tau_mu.ci <- quantile(tau_mu, c(0.01, 0.05, 0.1))

tau_mu.ci
```

```
##         1%         5%        10%
## -3.494281 -2.895757 -2.577903
```

And finally, for equation (3) – with an intercept and a deterministic time trend:

```
# Loop

tau_t <- vector()
time <- c(1:T)

for (i in 1:S) {

    e <- rnorm(T, 0, 1)
    y <- cumsum(e)
    y <- as.ts(y)

    y_diff <- diff(y)

    reg <- summary(dynlm(y_diff ~ 1 + L(y, 1) + time[-1]))  # removed 1 dimension for no. of obs.

    tau_t[i] <- reg$coefficients[2, 3]

}

tau_t.df <- data.frame(tau_t)

ggplot(data = tau_t.df, aes(x = tau_t)) + geom_density(color = "blue") +
    stat_function(fun = dnorm, n = 101, args = c(mean = 0, sd = 1)) +
    theme_few()
```




```
jarque.bera.test(tau_t)  # We reject H0 at the 1% significance level.
```

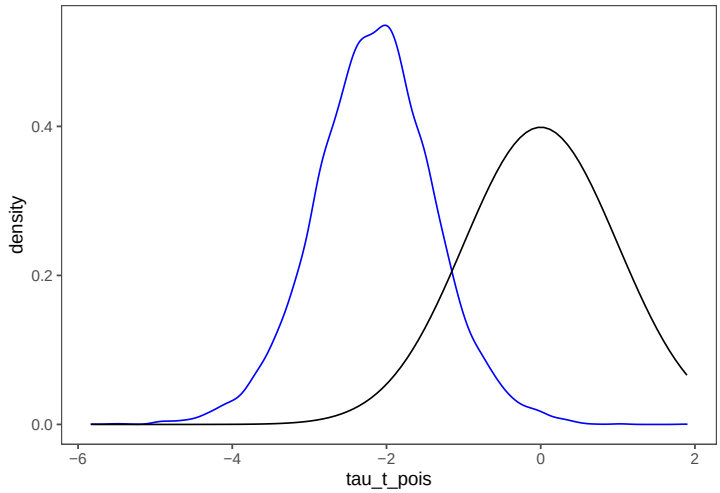
```
##
##  Jarque Bera Test
##
## data:  tau_t
## X-squared = 58.223, df = 2, p-value = 2.275e-13
```

```
tau_t.ci <- quantile(tau_t, c(0.01, 0.05, 0.1))

tau_t.ci
```

```
##         1%         5%        10%
## -4.065067 -3.462447 -3.158227
```

Let's now change the distributions of the errors for equation (3):

```
# Loop

tau_t_pois <- vector()
time <- c(1:T)

for (i in 1:S) {

    e <- rpois(T, 1)
    y <- cumsum(e)
    y <- as.ts(y)

    y_diff <- diff(y)

    reg <- summary(dynlm(y_diff ~ 1 + L(y, 1) + time[-1]))  # removed 1 dimension for no. of obs.

    tau_t_pois[i] <- reg$coefficients[2, 3]

}

tau_t_pois.df <- data.frame(tau_t_pois)

ggplot(data = tau_t_pois.df, aes(x = tau_t_pois)) + geom_density(color = "blue") +
    stat_function(fun = dnorm, n = 101, args = c(mean = 0, sd = 1)) +
    theme_few()
```




```
jarque.bera.test(tau_t_pois)  # We reject H0 at the 1% significance level.
```

```
##
##  Jarque Bera Test
##
## data:  tau_t_pois
## X-squared = 120.61, df = 2, p-value < 2.2e-16
```

```
tau_t_pois.ci <- quantile(tau_t_pois, c(0.01, 0.05, 0.1))

tau_t_pois.ci
```

```
##         1%         5%        10%
## -4.085753 -3.436365 -3.127525
```

11.3 Applying the Dickey-Fuller test for GDP

Loading the data from the previous problem:

```
pib <- get_sidra(6612, variable = 9318, category = 90707, period = "all")
```

```
## Considering all categories once 'classific' was set to 'all' (default)
```

```
pib_limpo <- pib[(pib$`Setores e subsetores (Código)` == 90707),
    ]

pib <- pib_limpo

pib_limpo2 <- pib[, c(5, 13)]

pib <- pib_limpo2

names(pib)[1] <- "t"
names(pib)[2] <- "v"

pib$t <- as.numeric(pib$t)

names(pib)
```

```
## [1] "t" "v"
```

```
head(pib)
```

```
##          t        v
## 18  199601 170920.0
## 40  199602 176708.8
## 62  199603 189844.3
## 84  199604 184112.9
## 106 199701 176732.2
## 128 199702 185109.5
```

```
tail(pib)
```

```
##           t        v
## 2042 201901 292647.6
## 2064 201902 297748.9
## 2086 201903 305150.9
## 2108 201904 302108.7
## 2130 202001 291912.5
## 2152 202002 263699.7
```

```
pib <- ts(pib$v)
```

```
# Choosing the correct model with auto.arima

aa_pib <- auto.arima(pib, stepwise = F)

summary(aa_pib)
```

```
## Series: pib
## ARIMA(2,1,2) with drift
##
## Coefficients:
##           ar1      ar2     ma1    ma2     drift
##       -0.0092  -0.9764  0.2721  0.956  864.9855
## s.e.   0.0233   0.0171  0.0455  0.117  537.6284
##
## sigma^2 estimated as 23455415:  log likelihood=-961.03
## AIC=1934.06   AICc=1934.99   BIC=1949.51
##
## Training set error measures:
##                     ME    RMSE      MAE        MPE     MAPE      MASE       ACF1
## Training set 64.56963 4692.48 3219.806 0.05416274 1.322955 0.5203698 0.01746157
```

auto.arima yields an ARIMA(2,1,2) model with a drift parameter:

$$\Delta y_t = a_0 + a_1 \Delta y_{t-1} + a_2 \Delta y_{t-2} + a_3 \Delta \varepsilon_{t-1} + a_4 \Delta \varepsilon_{t-2} + \varepsilon_t$$

Let's decompose this process. First, we'll perform the Dickey-Fuller test on the GDP ts.

```
adf.test(pib)
```

```
## Warning in adf.test(pib): p-value greater than printed p-value
##
```

```
##  Augmented Dickey-Fuller Test
##
## data:  pib
## Dickey-Fuller = 0.12028, Lag order = 4, p-value = 0.99
## alternative hypothesis: stationary
```

As we have not been able to reject H_0, it follows that the ts includes (at least one) unit root. Now, let's find the optimal model for the regression.

```
max_p <- 5

max_q <- 5

max_d <- 2

models1 <- vector("list", (max_p + 1) * (max_q + 1))

models2 <- vector("list", (max_p + 1) * (max_q + 1))

# Updating the model

fit1 <- vector("list", (max_p + 1) * (max_q + 1))

fit2 <- vector("list", (max_p + 1) * (max_q + 1))

model_info1 <- data.frame(matrix(NA, nrow = ((max_p + 1) * (max_q +
    1)), ncol = 3))




# Updating the model

for (u in 0:max_q) {

    for (j in 0:max_p) {


        fit1[[(((max_p + 1) * j) + u + 1)]] <- Arima(pib, order = c(j,
            1, u))

        model_info1[(((max_p + 1) * j) + u + 1), 1:2] <- c(fit1[[(((max_p +
            1) * j) + u + 1)]]$aic, fit1[[(((max_p + 1) * j) +
            u + 1)]]$bic)




    }
}

names(model_info1) <- c("AIC", "BIC", "void")



which.min(model_info1$AIC)
```

```
## [1] 23
```

```
which.min(model_info1$BIC)
```

```
## [1] 15
```

```
fit1[[which.min(model_info1$AIC)]]
```

```
## Series: pib
```

```
## ARIMA(3,1,4)
##
## Coefficients:
##           ar1      ar2      ar3     ma1     ma2     ma3     ma4
##       -0.9688  -0.9834  -0.9685  1.4659  1.3755  1.2705  0.5008
## s.e.   0.0305   0.0210   0.0270  0.1344  0.1682  0.1821  0.1484
##
## sigma^2 estimated as 21987482:  log likelihood=-956.26
## AIC=1928.53   AICc=1930.17   BIC=1949.13
```

```
fit1[[which.min(model_info1$BIC)]]
```

```
## Series: pib
## ARIMA(2,1,2)
##
## Coefficients:
##           ar1      ar2     ma1     ma2
##       -0.0084  -0.9757  0.2769  0.9673
## s.e.   0.0236   0.0175  0.0408  0.1130
##
## sigma^2 estimated as 23692560:  log likelihood=-962.3
## AIC=1934.6   AICc=1935.26   BIC=1947.47
```

```
# For I(2)

model_info2 <- data.frame(matrix(NA, nrow = max_d * ((max_p +
    1) * (max_q + 1)), ncol = 3))

for (u in 0:max_q) {

    for (j in 0:max_p) {


        fit2[[(((max_p + 1) * j) + u + 1)]] <- Arima(pib, order = c(j,
            2, u))

        model_info2[(((max_p + 1) * j) + u + 1), 1:2] <- c(fit2[[(((max_p +
            1) * j) + u + 1)]]$aic, fit2[[(((max_p + 1) * j) +
            u + 1)]]$bic)

    }
}

names(model_info2) <- c("AIC", "BIC", "void")

which.min(model_info2$AIC)
```

```
## [1] 24
```

```
which.min(model_info2$BIC)
```

```
## [1] 16
```

```
fit2[[which.min(model_info2$AIC)]]
```

```
## Series: pib
## ARIMA(3,2,5)
##
## Coefficients:
```

```
##           ar1      ar2      ar3     ma1     ma2      ma3      ma4      ma5
##       -0.9700  -0.9835  -0.9674  0.4748  0.0162  -0.0121  -0.6534  -0.3766
## s.e.   0.0315   0.0218   0.0277  0.1511  0.1211   0.1508   0.1090   0.1793
##
## sigma^2 estimated as 22148502:  log likelihood=-946.73
## AIC=1911.47   AICc=1913.56   BIC=1934.55
```

```
fit2[[which.min(model_info2$BIC)]]
```

```
## Series: pib
## ARIMA(2,2,3)
##
## Coefficients:
##           ar1      ar2      ma1     ma2      ma3
##       -0.0054  -0.9795  -0.6754  0.6720  -0.7903
## s.e.   0.0248   0.0162   0.1121  0.0786   0.1482
##
## sigma^2 estimated as 23858414:  log likelihood=-951.82
## AIC=1915.63   AICc=1916.58   BIC=1931.02
```

Chapter 12

Structural breaks

Informally, a structural break is a kink, or literally a break, in the time series that is observable by plotting the time series. Most of the time we have a vague idea of what happened there by visually inspecting; other times, we know *exactly* what happened – for example, the stock market on Joesley Day (2017). In other cases, however, it is very difficult to visually identify a structural break, especially when the process has a stochastic trend. This is an important issue: if we don't take into account the break, the ACF and the unit root tests may point towards non-stationarity!

Structural breaks are actually connected to the *instability* of a given process over time. In economics in particular, structural breaks are closely related to Lucas' critique: "the parameters of an econometric model depend on the rules of the game". In a given time series, it is usually possible to identify a number of strucutral breaks. What will happen if the rules of the game? Econometrics is not able to answer this question.

12.1 Known break

Most of the time, in practice, we know the time at with the structural break happened. Denote it by T_1.

$$Y_t = \begin{cases} M_1\left(\mathrm{W}_t\right) + u_{1t}; & t = 1, \ldots, T_1 \\ M_2\left(\mathrm{W}_t\right) + u_{2t}; & t = T_1 + 1, \ldots, T, \end{cases}$$

where $M_1(\cdot), M_2(\cdot)$ are models with regressors W_t – which can include a constant, a trend, lags of the dependent variable, exogenous regressors, etc.

12.1.1 Linear parametric modelling

We, then, postulate a parametric model – in the sense that it presents a finite number of unknown parameters:

$$M_1\left(\mathrm{W}_t\right) = M\left(\mathrm{W}_t; \beta_1\right); \quad M_2\left(\mathrm{W}_t\right) = M\left(\mathrm{W}_t; \beta_2\right)$$

In this case, testing for the presence of the break is equivalent to testing the following null:

$$H_0 : \beta_1 = \beta_2$$

For the moment, we have only considered linear models of the form:

$$Y_t = \begin{cases} Z_t'\beta_0 + W_t'\beta_1 + u_{1t}; & t = 1, \dots, T_1 \\ Z_t'\beta_0 + W_t'\beta_2 + u_{2t}; & t = T_1 + 1, \dots, T \end{cases}$$

We can merge this model in a single expression.

Definition 12.1.1. *A **linear parametric model of known structural break** follows the following form:*

$$Y_t = Z_t'\beta_0 + (W_t'\beta_1 + u_{1t})\mathbb{1}_{\{t \leq T_1\}} + (W_t'\beta_2 + u_{2t})\mathbb{1}_{\{t \geq T_1\}}$$

where $\mathbb{1}_{\{A\}}$ is the indicator function of event A, β_1 and β_2 are $(q \times 1)$ parameters that will be tested for the break before and after T_1, $\beta_0(r \times 1)$ are a priori stable parameters that won't be tested for a break, $W_t(q \times 1)$ are the regressors of the unstable parameters, $Z_t(r \times 1)$ are the regressors of the stable parameters.

12.1.2 OLS estimation of a model with known structural break

We can estimate this model by usual methods, such as OLS or MLE, given that it it a simple regression of the form

$$Y_t = X_t'\beta + v_t,$$

in which $\mathbb{1}_{\{A\}}$ is the indicator function of event A, $X_t := (Z_t', W_t'\mathbb{1}_{\{t \leq T_1\}}, W_t'\mathbb{1}_{\{t \geq T_1\}})$, $\beta := (\beta_0', \beta_1', \beta_2')'$ is a $(k \times 1)$ vector, $k := r + 2q$, $v_t := u_{1t}\mathbb{1}_{\{t \leq T_1\}} + u_{2t}\mathbb{1}_{\{t > T_1\}}$.

Let $\hat{\beta}$ be the OLS estimator of β. We know from Econometrics I that it has the following form:

$$\hat{\beta} = (X'X)^{-1}(X'Y),$$

where $X_{T \times k} = (X_1, ..., X_T)'$ and $Y_{T \times 1} = (Y_1, ..., Y_T)'$. $\hat{V}$ is an estimator for the variance of the estimator, $Var(\hat{\beta})$, given by:

$$\hat{V} = T(X'X)^{-1}\hat{\Omega}(X'X)^{-1},$$

where $\hat{\Omega}$ is a consistent estimator for:

$$\Omega = \mathbb{V}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t v_t\right)$$

12.1.3 Estimating covariance with Newey-West

As we have seen in Problem **??**, autocorrelation of the residuals is a big obstacle for inference. We can correct this in one of two ways: (i) guaranteeing that the error is truly white noise, with ARMA modelling; (ii) using robust errors to heteroskedasticity and autocovariance. The latter procedure will be presented now.

Definition 12.1.2. *The **Newey-West estimator** for covariance is given by:*

$$\widehat{\Omega} = \sum_{j=-m}^{m}\left(1 - \frac{|j|}{m+1}\right)\Gamma_j$$

where $\Gamma_j := \frac{1}{T-j}\sum_{t=j+1}^{T} X_t X_{t-j}'\widehat{v}_t\widehat{v}_{t-j}$ for $j \geq 0$ and $\Gamma_j = \Gamma_{-j}'$ for $j < 0$. Furthermore,

$$\widehat{v}_t = Y_t - X_t'\widehat{\beta}$$

The maximum lag is suggested by authors as being of the $T^{1/4}$ order (See Hamilton (1994), 10.5).

Under usual asymptotical conditions, we have:

$$\hat{V}^{-1/2}(\hat{\beta} - \beta) \to_d \mathcal{N}(0, I_k),$$

where I_k is an identity matrix of size k.

Note that, to test $H_0 : \beta_1 - \beta_2 = 0$, we'll use the statistic $\hat{\beta}_1 - \hat{\beta}_2$, since, under the null, we have:

$$P_T := \left(\widehat{V}_1 + \widehat{V}_2\right)^{-1/2} \left(\widehat{\beta}_1 - \widehat{\beta}_2\right) \xrightarrow{d} N\left(0, I_q\right),$$

where $\hat{V}_1, \hat{V}_2$ are $(q \times q)$ submatrices of the full matrix $\hat{V}_{(k\times k)}$ for β_1, β_2.

12.1.4 Wald test for structural break

The test statistic is $W_T = P_T' P_T$, or:

$$W_T = \left(\widehat{\beta}_1 - \widehat{\beta}_2\right)' \left(\widehat{V}_1 + \widehat{V}_2\right)^{-1} \left(\widehat{\beta}_1 - \widehat{\beta}_2\right)$$

This means that:

$$W_T \to_d \chi_q^2$$

Note that $W_T := W_T(T_1)$, *since partitioning the time series is only possible with previous knowledge of* T_1. Then, we reject H_0 of no structural break if $W_T > c_\alpha$, where $\mathbb{P}(\chi_q^2 > c_\alpha) = \alpha$.

12.1.5 Break in the variance of a time series

The break can also occur on the conditional variance of the model. Suppose that we have a general model

$$Y_t = X_t \beta + u_t,$$

where, again, X_t can contain constants, deterministic trends, autoregressive parameters. Let's construct a simple case in which u_t, assumed to be uncorrelated serially, has the following variance:

$$\mathbb{V}(u_t) = \begin{cases} \sigma_1^2 & , t = 1, \ldots, T_1 \\ \sigma_2^2 & , t = T_1 + 1, \ldots, T \end{cases}$$

We would like to test $H_0 : \sigma_1^2 = \sigma_2^2$. It is possible to estimate the model with OLS using the entire sample to obtain:

$$\widehat{u}_t = Y_t - X_t' \widehat{\beta}, \quad t = 1, \ldots, T$$

This yields estimators for σ_1^2 and σ_2^2:

$$s_1^2 = \frac{1}{T_1 - k} \sum_{t=1}^{T_1} \widehat{u}_t^2$$

$$s_2^2 = \frac{1}{T - T_1 - k} \sum_{t=T_1+1}^{T} \widehat{u}_t^2$$

Assuming a normal distribution for u_t

We know from Econometrics I that:

$$(T_1 - k)\,\frac{s_1^2}{\sigma_1^2} \sim \chi^2_{T_1-k}$$
$$(T - T_1 - k)\,\frac{s_2^2}{\sigma_2^2} \sim \chi^2_{T-T_1-k}$$

Therefore, we have:

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim \frac{\chi^2_{T_1-k}/\left(T_1-k\right)}{\chi^2_{T-T_1-k}/\left(T-T_1-k\right)} =: F_{T_1-k,T-T_1-k}$$

where $F(a,b)$ is a F distribution with a, b degrees of freedom. We reject H_0 at the level of significance α if:

$$W \equiv \frac{s_1^2}{s_2^2} > c_\alpha, \text{ where } \mathbb{P}\left(F_{T_1-k,T-T_1-k} > c_\alpha\right) = \alpha$$

General distribution

We now need an asymptotic argument. Define $s^2 := \left(s_1^2, s_2^2\right)'$ e $\sigma^2 := \left(\sigma_1^2, \sigma_2^2\right)'$ and $K := \mathbb{V}\left(s^2\right) = \mathbb{E}\left[\left(s^2-\sigma^2\right)\left(s^2-\sigma^2\right)'\right]$. From the CLT for *iid* vectors, we have:

$$K^{-1/2}\left(s^2-\sigma^2\right) \xrightarrow{d} N\left(0, I_2\right)$$

From the Continuous Mapping Theorem, this yields:

$$\frac{\left(s_1^2-s_2^2\right)-\left(\sigma_1^2-\sigma_2^2\right)}{\sqrt{K_1+K_2}} \xrightarrow{d} \mathcal{N}(0,1)$$

where $K_1 := \frac{\kappa_1}{T_1}, K_2 := \frac{\kappa_2}{T-T_1}, \kappa_1 := \mathbb{V}\left(\sqrt{T_1}s_1^2\right)\ \kappa_2 := \mathbb{V}\left(\sqrt{T-T_1}s_2^2\right)$.

12.1.6 Estimating the variance of the sample variance

We will use again the Newey-West to arrive at a robust estimate for the variance of the sample variance – after all, we assumed that the *errors weren't correlated, not their squares!*

$$\widehat{\kappa}_1 := \sum_{j=-m}^{m}\left(1-\frac{|j|}{m+1}\right)a_j$$

where $a_j = \frac{1}{T_1-j}\sum_{t=j+1}^{T_1}\left(\widehat{u}_t^2 - s_1^2\right)\left(\widehat{u}_{t-j}^2 - s_1^2\right)$

Analogously for κ_2, we have:

$$\widehat{\kappa}_2 := \sum_{j=-m}^{m}\left(1-\frac{|j|}{m+1}\right)b_j$$

where $b_j = \frac{1}{T-T_1-j}\sum_{t=j+T_1+1}^{T}\left(\widehat{u}_t^2 - s_2^2\right)\left(\widehat{u}_{t-j}^2 - s_2^2\right)$

12.1.7 Test statistic for variance break

We can finally construct our test statistic. Under the null $H_0 : \sigma_1^2 - \sigma_2^2 = 0$:

$$W_T \equiv \frac{s_1^2 - s_2^2}{\sqrt{\widehat{K_1} + \widehat{K_2}}} \xrightarrow{d} \mathcal{N}(0,1)$$

Here, we can consider one- or two-sided tests. In other words, at a given level of significance α we reject H_0...

- ...in favor of $H_1 : \sigma_1^2 \neq \sigma_2^2$ if $|W_T| > c_{\alpha/2}$
- ...in favor of $H_1 : \sigma_1^2 > \sigma_2^2$ if $W_T < -c_\alpha$
- ...in favor of $H_1 : \sigma_1^2 < \sigma_2^2$ if $W_T > c_\alpha$,

where $\mathbb{P}(Z > c_\alpha) = \alpha$ and $Z \sim \mathcal{N}(0,1)$.

12.2 Unknown break

The biggest limitation of the method presented in the previous section is its dependence on the knowledge of the *exact moment* T_1 *of the break.* It is possible that we do not know T_1, but know that it is in roughly in the middle of the sample (e.g., between 20% and 80%):

$$T_1 \in \mathcal{T} = \{t_1, \ldots, t_2\}, \quad 1 < t_1 < t_2 < T$$

In this case, we can calculate the Wald statistic for each $t \in \mathcal{T}$ and take the maximum of the statistics as our test statistic:

$$W_T^* \equiv \max_{t \in \mathcal{T}} W_T(t)$$

This is known as the "supremum" test, and has a non-standard limiting distribution (Andrews (1993)). Under H_0 :

$$W_T^* \to_d \mathcal{W}$$

The critical values of $\mathcal{W}$ are available on a table available in this paper. In particular, these critical values $c = c(\alpha, k, \pi_1, \pi_2)$ depend on:

- α: the level of significance;
- k: the number of parameters to be tested;
- $\pi_1 = t_1/T, \pi_2 = t_2/T$ through $\lambda = \dfrac{\pi_2(1-\pi_1)}{\pi_1(1-\pi_2)}$.

We reject H_0 if:

$$W_T^* > c(\alpha, k, \pi_1, \pi_2).$$

Chapter 13

Problem 7: Testing for structural breaks

13.1 Testing for structural breaks

If we know the date of the suspected break, we can apply a Chow test. It is implemented by fitting the same ARMA model for $t \leq T_1$, $t > T_1$. If both models are sufficiently similar, we cannot reject the null of no structural break. This was discussed in the last Chapter.

$$F = \frac{(SSR - SSR_1 - SSR_2)/n}{(SSR_1 + SSR_2)/(T - 2n)}$$

If the coefficients are equal, $SSR_1 + SSR_2 = SSR \implies F = 0$.

The model can also be constructed using dummies (See 12.1.1). For endogenous structural breaks, we can perform a supremum test (See 12.2).

13.1.1 Parameter instability

It is possible that there is no single date of break, but rather a gradual change which renders the parameters unstable. e.g., climate change.

In this case, it is best to estimate the model recursively and plot the coefficients over time. If the coefficients present significant trends and deviations, then we can suspect that there's a structural break. It is possible to, at each step, forecast the error $e_t(1) = y_{t+1} - \mathbb{E}_t y_{t+1}$. If the model is well specified and fitted, then the sum of these errors should not differ significantly from zero. This is the intuition behind the $CUSUM_N$ test.

$$CUSUM_N = \sum_{i=n}^{N} \frac{e_i(1)}{\sigma_e}, \quad N = n, ..., T-1$$

where T denotes the date of the last observation and σ_e is the *estimated standard deviation of the forecast errors.* Note that σ_e is created using all $T - n$ forecast errors (Enders, p. 106).

13.2 Implementing the tests

```
library(readxl)
library(ggplot2)
library(forecast)
library(dynlm)
library(ggthemes)
library(strucchange)
```

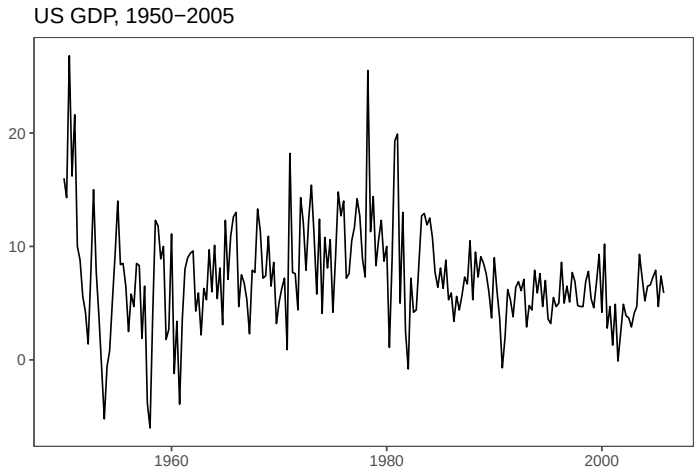
```
##
## Attaching package: 'strucchange'
```

```
## The following object is masked from 'package:stringr':
##
##     boundary
```

```
library(lmtest)
library(car)
library(dplyr)

df <- read_excel("G:/My Drive/FGV EESP/4o SEMESTRE/Econometria II/QuantEconEESP/QuantEconEESP/pibeua_real.xlsx")
```

```
## New names:
## * `` -> ...1
```

```
series <- ts(df$`Cresimento percentual`[13:236], start = c(1950,
    1), end = c(2005, 4), frequency = 4)  # 1950-2005

autoplot(series) + theme_few() + ggtitle("US GDP, 1950-2005")
```




```
# We can clearly see a reduction in variance during the 80s.

df$observ <- 1:length(df$`PIB nominal`)

# Suppose that the break happens at time t = 153 (Q1, 1985).

df$d <- as.numeric(df$observ > 152)

m1 <- lm(series ~ df$d[13:236])

summary(m1)
```

```
##
## Call:
```

```
## lm(formula = series ~ df$d[13:236])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.2379  -2.0571  -0.1379   2.2871  18.5621
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8.2379     0.3728  22.098  < 2e-16 ***
## df$d[13:236]  -2.5057     0.6088  -4.116 5.43e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.411 on 222 degrees of freedom
## Multiple R-squared:  0.0709, Adjusted R-squared:  0.06672
## F-statistic: 16.94 on 1 and 222 DF,  p-value: 5.435e-05
```

```
m2 <- dynlm(series ~ df$d[13:236] + L(series, 1) + L(series,
    1) * df$d[13:236])

summary(m2)
```

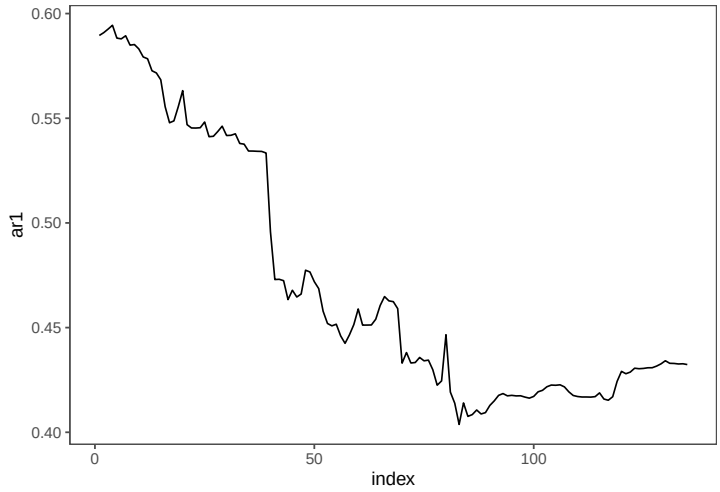
```
##
## Time series regression with "ts" data:
## Start = 1950(2), End = 2005(4)
##
## Call:
## dynlm(formula = series ~ df$d[13:236] + L(series, 1) + L(series,
##     1) * df$d[13:236])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.2567  -2.1410  -0.0335   2.0335  17.7119
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  4.76430    0.63168   7.542 1.22e-12 ***
## df$d[13:236]                -0.40507    1.40339  -0.289    0.773
## L(series, 1)                 0.41421    0.06436   6.436 7.63e-10 ***
## df$d[13:236]:L(series, 1) -0.17495    0.21438  -0.816    0.415
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.033 on 219 degrees of freedom
## Multiple R-squared:  0.2209, Adjusted R-squared:  0.2103
## F-statistic:  20.7 on 3 and 219 DF,  p-value: 7.576e-12
```

```
# ARIMA model

arima_unr <- arima(series, order = c(1, 0, 0))

arima_r1 <- arima(series[1:152], order = c(1, 0, 0))
```

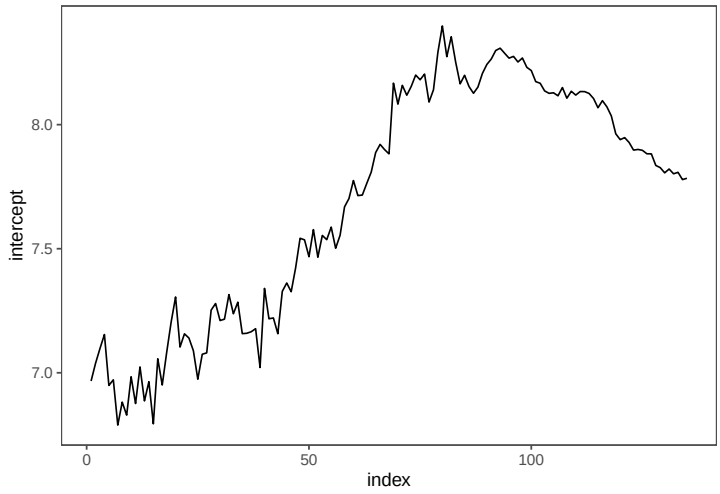
```
arima_r2 <- arima(series[153:length(series)], order = c(1, 0,
    0))

ssr_unr <- sum(arima_unr$residuals^2)

ssr_r1 <- sum(arima_r1$residuals^2)

ssr_r2 <- sum(arima_r2$residuals^2)

# We will now define the Chow test for the null H0: \beta_m1
# - \beta_m2 = 0 (no structural break)

chow <- function(SSR_unr, SSR_r1, SSR_r2, t, n) {

    ((SSR_unr - SSR_r1 - SSR_r2)/n)/((SSR_r1 + SSR_r2)/(t - 2 *
        n))

}


chow(SSR_unr = ssr_unr, SSR_r1 = ssr_r1, SSR_r2 = ssr_r2, t = length(series),
    n = length(arima_unr$coef))  # T statistic (n, T - 2n).
```

```
## [1] 3.736208
```

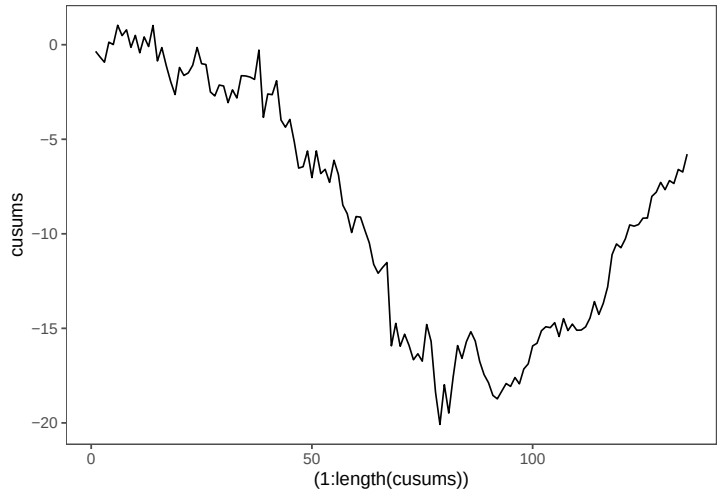
```
# Now, suppose that we do not know when the break happened.

t0 = 45
tf = 180  # Boundaries for the process.

models = list(NA)

coefs = matrix(NA, nrow = length(t0:tf), ncol = 2)

forecasts = list(NA)

ci = data.frame(matrix(NA, nrow = length(t0:tf), ncol = 5))

e = matrix(NA, nrow = length(t0:tf), ncol = 1)


# 1. Plotting coefficients.

for (i in (1:(tf - t0))) {

    models[[i]] = arima(series[1:(i + t0)], order = c(1, 0, 0))

    coefs[i, ] = models[[i]]$coef

    forecasts[[i]] = forecast(series[1:(i + t0)], model = models[[i]],
        h = 1)

    e[i, ] = forecasts[[i]]$mean - series[(i + t0 + 1)]

}

coefs = data.frame(coefs)

coefs = data.frame(coefs, (1:length(coefs$X1)))

df.coefs = na.omit(data.frame(coefs))

names(df.coefs) = c("ar1", "intercept", "index")

ggplot(df.coefs, aes(x = index, y = ar1)) + geom_line() + theme_few()
```




```
ggplot(df.coefs, aes(x = index, y = intercept)) + geom_line() +
    theme_few()
```




```
# 2. Cusum test.

cusums = matrix(NA, nrow = length(e), ncol = 1)

e = na.omit(e)

for (i in 1:(length(e))) {

    cusums[i, ] = sum(e[1:i])/sd(e)

}


cusums = na.omit(cusums)

df.cusums = data.frame(cusums)

ggplot(df.cusums, aes(x = (1:length(cusums)), y = cusums)) +
    geom_line() + theme_few()
```

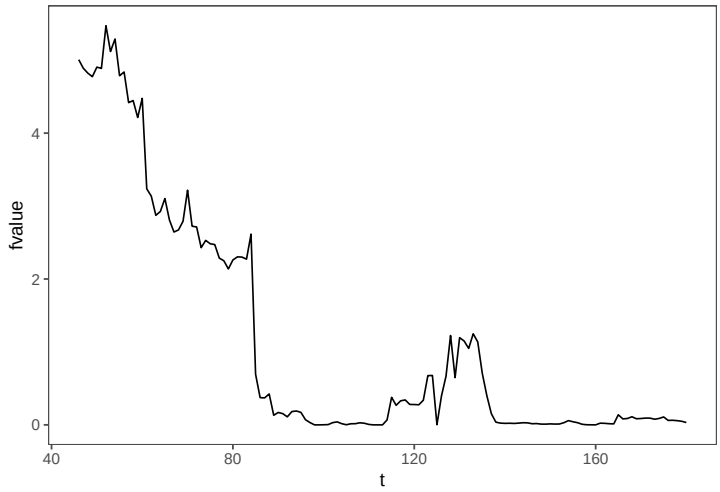



```
# 3. Iterative F-tests.

models_unr = list(NA)

models_r = list(NA)


# dummy <- df$d[13:236]

num_series = as.numeric(series)

f_values = matrix(NA, nrow = length(num_series), ncol = 2)

hyp = c(0, -1, 0, 1)

rhs = 0

length((t0:tf))
```

```
## [1] 136
```

```
dummies = data.frame(matrix(NA, ncol = 224, nrow = 224))

for (i in (13:236)) {
    dummies[i] = as.numeric(df$observ[13:236] >= i)
}


for (i in (1:(tf - t0))) {

    adummy <- rep(0, 1)
    j <- 0
    k <- t0 + i
    while (j <= length(num_series)) {
        if (j > k) {
            # nÃ£o sei se isso deveria ser maior ou igual ou sÃ³ maior
            adummy[j] <- 1
        } else {
            adummy[j] <- 0
        }
        j <- j + 1
    }
    models_unr[[i]] <- dynlm(num_series ~ adummy + lag(num_series) +
        lag(num_series) * adummy)
    f_values[i, 1] = linearHypothesis(models_unr[[i]], hyp, rhs)$F[1]
    f_values[i, 2] = linearHypothesis(models_unr[[i]], hyp, rhs)$F[2]  #eu acho que funcionou?????

}
```

```r
df.f_values <- data.frame(f_values)

df.f_values <- df.f_values$X2$

df.f_values <- na.omit(df.f_values)

df.f_values <- data.frame(df.f_values, ((t0 + 1):tf))

names(df.f_values) <- c("fvalue", "t")

ggplot(df.f_values, aes(x = t, y = fvalue)) + geom_line() + theme_few()
```




```r
linearHypothesis(models_unr[[10]], hyp, rhs)
```

```
## Linear hypothesis test
##
## Hypothesis:
## - adummy  + adummy:lag(num_series) = 0
##
## Model 1: restricted model
## Model 2: num_series ~ adummy + lag(num_series) + lag(num_series) * adummy
##
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    220 3673.8
## 2    219 3595.2  1    78.591 4.7874 0.02973 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Chapter 14

Conditional Heteroskedasticity Models

In this chapter, we'll explore heteroskedasticity models. First, we'll motivate the subject with a discussion on some financial stylized facts.

14.1 Motivation

Plotting a financial time series suggests some sort of exponential upwards trend.

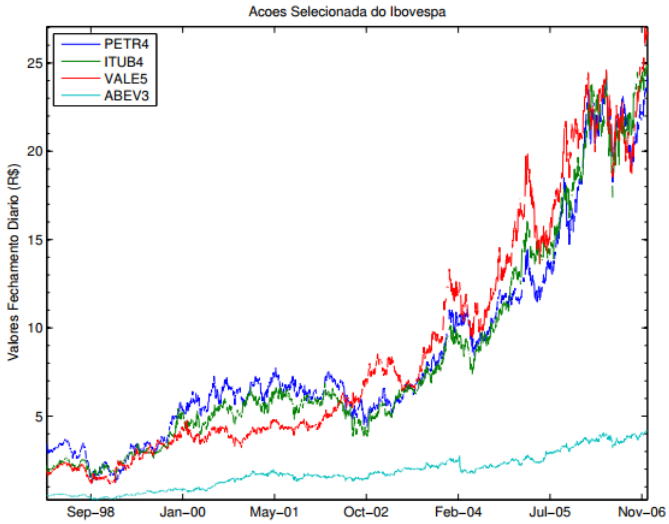


Figure 14.1: Prices of assets over time

After detrending, its ACF decays linearly and its PACF has a significant lag at L = 1.

When we take first differences, the autocorrelation functions suggest no correlation structure – i.e., an ARIMA(0,1,0).

When working with financial data, we usually model returns or log-returns instead of price:

$$R_t := \frac{p_t - p_{t-1}}{p_{t-1}}, \qquad r_t := log\left(\frac{p_t}{p_{t-1}}\right) \tag{14.1}$$

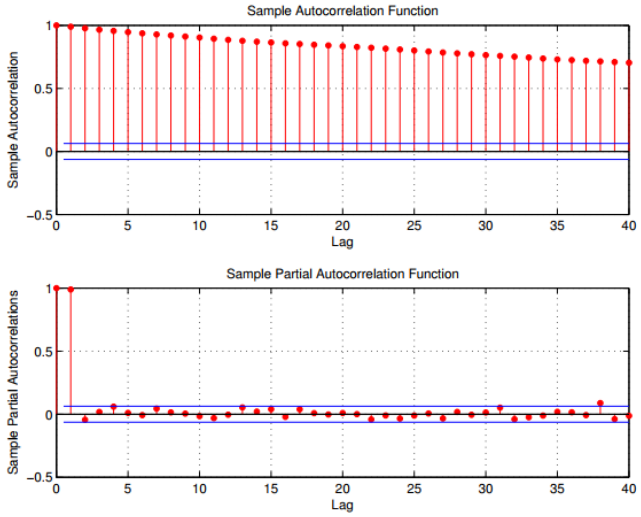


Figure 14.2: ACF and PACF of prices

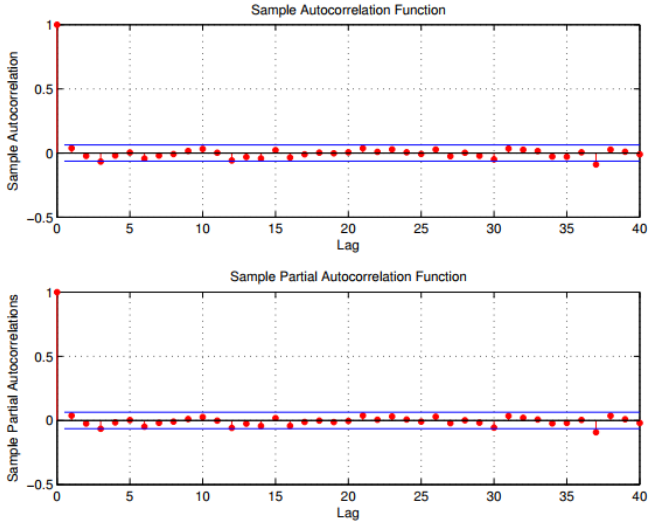


Figure 14.3: First differences

In practice, using R_t or r_t is negligible for small values, because $r_t = log(1 + R_t) \approx R_t$. This result comes from a Taylor expansion around 1. We'll use, then, log-returns as $r_t = \Delta logp_t = logp_T - log_{t-1}$. This has the advantage of always being summable, while regular returns are not – e.g., falling 10% in t and rising 10% in $t + 1$ does not yield the same starting value.

14.1.1 Stylized facts on financial time series

Let $\mathcal{F}_{t-1} := r_{t-1}, r_{t-2}, ...$ and $\mu := \mathbb{E}(r_t)$.

Fact 14.1.1. *Financial returns usually present $E(r_t|\mathcal{F}_{t-1}) = \mu$ – i.e., its past has no statistically significant explanatory power.*

Fact 14.1.2. *In practice, financial returns usually present $\mathbb{E}(r_t) = \mu > 0$.*

If $\mu = 0$, there would be no incentive for investing (no risk premium).

Fact 14.1.3. *Financial returns are usually highly leptokurtic – i.e., their distributions have longer tails when compared to a Normal distribution.*

Kurtosis is defined as $\kappa := \mathbb{E}(\frac{r_t - \mu}{\sigma})^4$ and can be interpreted as the propensity for outliers. The Normal distribution has $\kappa = 3$. Distributions that have $\kappa > 3$ are called leptokurtic.

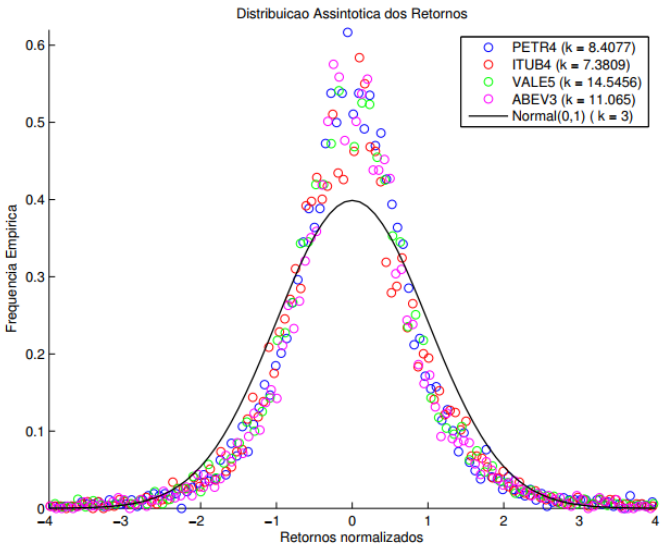


Figure 14.4: Empirical distribution of returns

Definition 14.1.1. ***Weak market efficiency hypothesis.*** *Current prices reflect all available information contained in past prices.*

Definition 14.1.2. ***Semi-strong market efficiency hypothesis.*** *Prices reflect all public information.*

Definition 14.1.3. ***Strong market efficiency hypothesis.*** *Prices reflect all information, public or private.*

Note that, if the strong hypothesis is true, we should not punish insider trading! It is, therefore, most certainly false.

How can we relate the weak market efficiency hypothesis with fact 14.1.1? Nowadays, there's much evidence *against* all three of these hypothesis using non-linear models. Notwithstanding this evidence, no one can consistently beat the market. This empirical contradiction is explored by Fama and Shiller, winners of the 2013 Nobel Prize in Economics.

Applying ARMA modelling to *volatility*, however, yields different results, suggesting some sort of correlation structure.

Fact 14.1.4. ***Volatility clustering.*** *High volatility at time t raises the probability of high volatility at $t + 1$.*

This is also valid for alternate measures of volatility, such as $\mathbb{E}(r_t - \mu)$ or $Var(r_t)$. It can be interpreted as motivation for autoregressive models of conditional heteroskedasticity, ARCH.

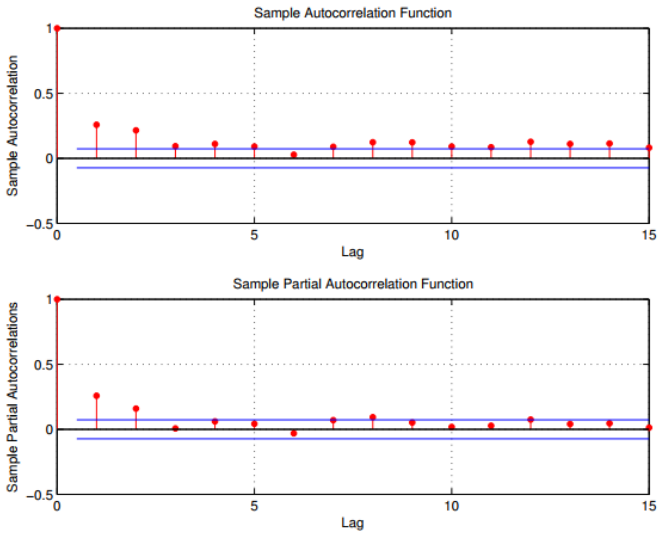


Figure 14.5: Correlation structure of the variance

14.2 ARCH

The ARCH (Autoregressive Conditional Heteroskedastic) class of models were proposed by Engle (1982) to model conditional variance. They're usually combined with a model for conditional expectation. For example, an ARIMA(p,d,q) with ARCH(1) innovations is given by:

$$\begin{aligned}\Phi_p(L)(1-L)^d Y_t &= c + \Theta_q(L)\varepsilon_t \\ \varepsilon_t &= \sigma_t u_t \quad u_t \sim iid(0,1) \\ \sigma_t^2 &= \omega + \alpha\varepsilon_{t-1}^2\end{aligned}$$

In financial models we usually have $Y_t = log(p_t) \sim ARIMA(0,1,0)$.

We can generalize this model to an ARCH(s).

Definition 14.2.1. *ARCH(s) is an autoregressive model of conditional heteroskedasticity up to lag s:*

$$\sigma_t^2 = \omega + \sum_{j=1}^{s} \alpha_j \varepsilon_{t-j}^2$$

Note that this model *is not linear in its parameters,* because the variance and mean parameters become mixed. This means that OLS cannot be used for estimation.

Furthermore, because variance shall always be *positive,* it is necessary to impose restrictions on our parameters. It is natural to assume, then, $\omega > 0, \alpha_j > 0, \forall j$. This implies that we'll need some sort of optimization subject to restrictions. Estimation is usually conducted through quasi-maximum likelihood, which means that we'll have to assume some distribution for innovation terms, such as t, Normal etc.

In practice, ARCH(s) models tend to yield high s. An alternative way to model variance is through some sort of ARMA procedure.

Definition 14.2.2. ***GARCH(r,s)*** *is a generalized autoregressive model of conditional het-*

eroskedasticity up to lags r, s:

$$\begin{aligned}\varepsilon_t &= \sigma_t u_t, \quad u_t \sim iid(0,1)\\ \sigma_t^2 &= \omega + \alpha_1\varepsilon_{t-1}^2 + \dots \alpha_s\varepsilon_{t-s}^2 + \beta_1\sigma_{t-1}^2 + \dots + \beta_r\sigma_{t-r}^2\\ &= \omega + \sum_{j=1}^{s}\alpha_j\varepsilon_{t-j}^2 + \sum_{j=1}^{r}\beta_j\sigma_{t-j}^2\end{aligned}$$

GARCH(r,s) works as an ARMA(p,r) for conditional variance, where $p = max\{r, s\}$.[1]

14.3 More stylized facts on financial time series

Fact 14.3.1. *Conditional variance is higher after weekends than during the week.*

Fact 14.3.2. *Volatility is higher during market opening and closing than during the day.*

A way to test these facts is to include dummies in the variance model. This model is called Threshold-GARCH or TARCH.

Definition 14.3.1. ***TARCH(r,s)*** *is a threshold generalized autoregressive model of conditional heteroskedasticity up to lags r, s with $d_t(A)$:*

$$\sigma_t^2 = \omega + \delta d_t(A) + \sum_{j=1}^{s}\alpha_j\varepsilon_{t-j}^2 + \sum_{j=1}^{r}\beta_j\sigma_{t-j}^2,$$

where $d_t(A)$ is a dummy of a given event A.

Fact 14.3.3. ***Leverage effect.*** *Conditional variance increases more when the innovation is negative than when it is positive.*

This indicates a trend of returns being negatively correlated with changes in volatility.

We can test this effect with another dummy, or by constructing an Exponential GARCH.

Definition 14.3.2. ***EGARCH(r,s)*** *is an exponential generalized autoregressive model of conditional heteroskedasticity up to lags r, s, v:*

$$\log\sigma_t^2 = \omega + \sum_{j=1}^{r}\beta_j\log\sigma_{t-j}^2 + \sum_{j=1}^{s}\alpha_j\left[|u_{t-j}| - \mathbb{E}\,|u_{t-j}|\right] + \sum_{j=1}^{v}\gamma_j u_{t-j}$$

For an EGARCH(1,1,1), we have:

$$\log\sigma_t^2 = \omega + \beta\log\sigma_{t-1}^2 + \alpha\left[|u_{t-1}| - \mathbb{E}\,|u_{t-1}|\right] + \gamma u_{t-1}$$

Note that $\mathbb{E}(u_{t-1})$ depends on the distribution of u_t. For a Normally distributed random variable, its value is $\sqrt{2/\pi}$.

If $\gamma < 0$ is significant, then there's evidence of a leverage effect.

ARCH models have contributed to show a number of stylized facts of financial time series. In particular, they yield leptokurtic models of conditional heteroskedasticity, which is consistent with

[1] Details available on Hamilton, Chapter 21.

evidence. However, these models are now more used in Macroeconomomic research. In Finance, they have been abandoned because of their inherent *backward looking* nature. Nowadays, the derivative market is used for risk, because it its forward looking. Implicit and realized volatility, Monte Carlo simulations and bootstrapping are also used to model risk. Furthermore, these models are not very suitable for time-sensitive applications, and are also sentitive to different hypothesis of error distributions.

Chapter 15

Problem 8: ARCH estimation

Chapter 16

ARDL Models

16.1 Motivation

Up to this point, we have been modelling only one time series at a time. However, it is interesting to study not only the temporal structure of a series, but also *relations between series over time.*

Let $\{Y_t\}, \{X_t\}$ be two time series. We could postulate a linear regression model of the form:

$$Y_t = \beta X_t + u_t$$

Does this make sense? How can we interpret and estimate β? These questions will be answered in this chapter.

16.1.1 Static model

Suppose that you are interested in the relationship between gas, $\{G_t\}$, and oil, $\{P_t\}$, prices. We can postulate the following linear model:

$$G_t = \alpha + \beta P_t + u_t$$

What is β? If it is the parameter of the best linear projection of G_t on P_t, then

$$\beta = \frac{Cov(G_t, P_t)}{Var(P_t)}$$

and it can be consistently estimated by OLS as long as both time series are ergodic. What if we define β as the *causal* effect of G_t on P_t?

16.1.2 Distributed Lag model

It might be that gas prices are influenced not only by current oil prices, but also by past prices. In this case, we can model it as:

$$G_t = \alpha + \beta_0 P_t + \beta_1 P_{t-1} + v_t$$

In this case, we have *dynamic effects*, because, if we assume that $\mathbb{E}(u - t|P_t, P_{t-1}) = 0$, then

$$\mathbb{E}\left(G_t \mid P_t, P_{t-1}\right) = \alpha + \beta_0 P_t + \beta_1 P_{t-1}$$

$$\frac{\partial \mathbb{E}\left(G_t\right)}{\partial P_t} = \beta_0; \quad \frac{\partial \mathbb{E}\left(G_{t+1}\right)}{\partial P_t} = \beta_1$$

There can also be an autoregressive component on gas prices, i.e., past prices affecting current gas prices. This yields the following model:

$$G_t = \alpha + \phi G_{t-1} + \gamma_0 P_t + \gamma_1 P_{t-1} + w_t$$

In this case, there will be dynamic effects via P_t and G_t. Generally, this means that the entire oil price time series will have an effect on gas prices, because

$$G_t = \tfrac{\alpha}{1-\phi} + \left(1 + \phi L + \phi^2 L^2 + \ldots\right)\left(\gamma_0 P_t + \gamma_1 P_{t-1} + w_t\right)$$

if $|\phi| < 1$.

16.2 Defining ARDL models

Definition 16.2.1. *A generic distributed lag model with q lags, DL(q), is written in the following form:*

$$Y_t = \alpha + \sum_{j=0}^{q} \theta_j X_{t-j} + \epsilon_t$$

or, using the lag operator $L, \Theta(L) := \theta_0 + \theta_1 L + \cdots + \theta_q L^q$,

$$Y_t = \alpha + \Theta(L) X_t + \epsilon_t$$

This implies that Y_t has serial correlation? Yes, because we can rewrite Y_{t-j} using α, X, ε. Does this imply that X_t has serial correlation? Not necessarily.

Definition 16.2.2. *An autoregressive distributed lag model of order* p, q *is a combination of an AR(p) with a DL(q). It can be written as:*

$$Y_t = \alpha + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=0}^{q} \theta_j X_{t-j} + \epsilon_t$$

or, using the lag operator $L, \Phi(L) := 1 - \phi_1 L - \cdots - \phi_p L^p$,

$$\Phi(L) Y_t = \alpha + \Theta(L) X_t + \epsilon_t$$

Some words of caution are in order.

- The ARDL(p,q) model has a similar structure to an ARMA(p,q), but the MA(.) portion is applied on X_t instead of ε_t.
- ε_t is not always assumed to be white noise processes, as was the case with ARMA models. Generally, we assume stationarity.

16.2.1 Inverting an ARDL model

If all roots of the AR portion of the model, i.e., zeros of $\Phi(L) + 0$, are outside of the unit circle, we can invert the AR portion. This yields

$$Y_t = \Phi^{-1}(L)\alpha + \Phi^{-1}(L)\left(\Theta(L) X_t + \epsilon_t\right)$$

If we further define $\Psi(L) := \Phi^{-1}(L)\Theta(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \ldots$ and remembering that $\Phi^{-1}(L)\alpha = \Phi^{-1}(1)\alpha$, we have

$$Y_t = \Phi^{-1}(1)\alpha + \Psi(L) X_t + \Phi^{-1}(L)\epsilon_t$$

16.2.2 ARDL multipliers

If $\mathbb{E}(\varepsilon_t|X_s) = 0, \forall s, t$, i.e., strict exogeneity holds, then

$$\mathbb{E}\left(Y_t \mid X_t, X_{t-1}, \ldots\right) = \Phi^{-1}(1)\alpha + \Psi(L)X_t$$

This makes it easy to define the impact multiplier on instant $t+s$:

$$\frac{\partial \mathbb{E}\left(Y_{t+s} \mid X_{t+s}, X_{t+s-1}, \ldots\right)}{\partial X_t} = \psi_s; \quad s = 0, 1, 2, \ldots$$

Or yet, the long run impact

$$\sum_{s=0}^{\infty} \frac{\partial \mathbb{E}\left(Y_{t+s} \mid X_{t+s}, X_{t+s-1}, \ldots\right)}{\partial X_t} = \sum_{s=0}^{\infty} \psi_s = \Psi(1)$$

Note that the last equality holds because

$$\Psi(L) := \psi_0 + \psi_1 L + \psi_2 L^2 + \ldots,$$

$$\Psi(1) = \sum_{s=0}^{\infty} \psi_s$$

16.2.3 ARDL with multiple explanatory variables