Econometrics II
Notes - Midterm

William Radaic Peron

EESP-FGV

September 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!*

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

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 dependency:

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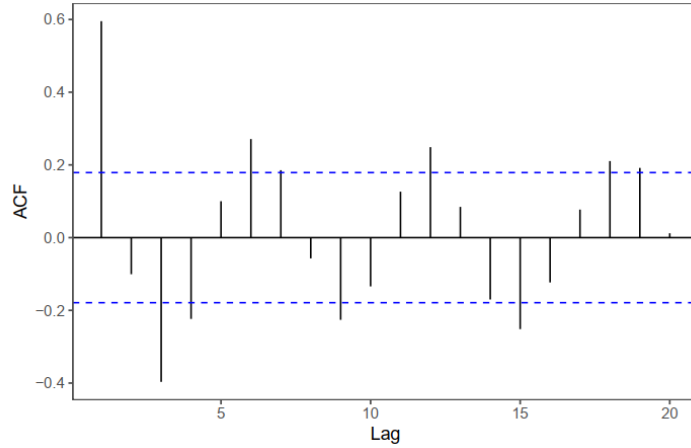
Definition 1.2.1. $\gamma_{j,t} := Cov(Y_t, Y_{t-j})$ is the **autocovariance function** *for a given time series* $\{Y_t, t \in \mathbb{Z}\}$.

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 $\mathbb{E}|Y_1| < \infty$,*

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

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,*

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

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.

$$\mathbb{E}\left(\bar{Z}_T^2\right) = \mathbb{E}\left[\left(\frac{1}{T} \sum_{t=1}^{T} Z_t\right)\left(\frac{1}{T} \sum_{t=1}^{T} Z_t\right)\right] = \frac{1}{T^2} \mathbb{E}\left(\sum_{s=1}^{T} \sum_{t=1}^{T} Z_s Z_t\right)$$

$$= \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E}\left(Z_s Z_t\right) = \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t=1}^{T} \gamma_{s-t} = \frac{1}{T} \sum_{j=-T+1}^{T-1} \frac{T-|j|}{T} \gamma_j$$

$$\leq \frac{1}{T} \sum_{j=-T+1}^{T-1} \frac{T-|j|}{T} |\gamma_j| \leq \frac{1}{T} \sum_{j=-T+1}^{T-1} |\gamma_j| \to 0$$

$\square$

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

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

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:

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

With this result, we can rewrite the model as:

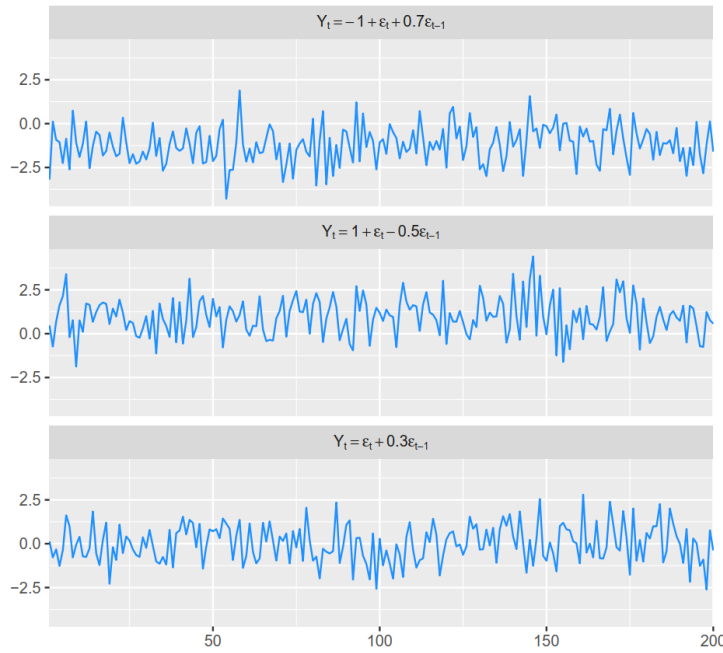$$(Y_t - \mu) = \varepsilon_t + \theta\varepsilon_{t-1}$$

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

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

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.
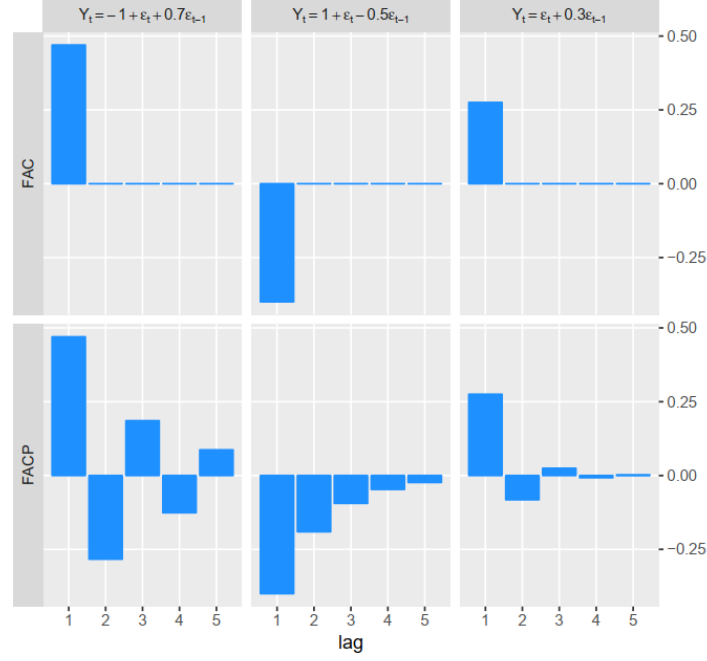
2.3   Generalizing the MA model

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:*

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

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

2.3.1   Moments of an MA(q) process

The expected value of an MA(q) is:

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

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

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

Multiplicando ambos os lados por $(Y_{t-j} - \mu)$, temos

$$(Y_t - \mu)(Y_{t-j} - \mu) = \left(\sum_{k=0}^{q} \theta_k \varepsilon_{t-k}\right)\left(\sum_{k=0}^{q} \theta_k \varepsilon_{t-j-k}\right)$$

$$= \sum_{k=0}^{q} \sum_{\ell=0}^{q} \theta_k \theta_\ell \varepsilon_{t-k} \varepsilon_{t-j-\ell},$$

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

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

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, θ 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:

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

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:*

$$\alpha_j = \begin{cases} \text{Cor}\,(Y_t, Y_{t-1}) =: \rho_1 & j = 1 \\ \text{Cor}\,(Y_t, Y_{t-j} \mid Y_{t-1}, \dots, Y_{t-j+1}) & j \geq 2 \end{cases}$$
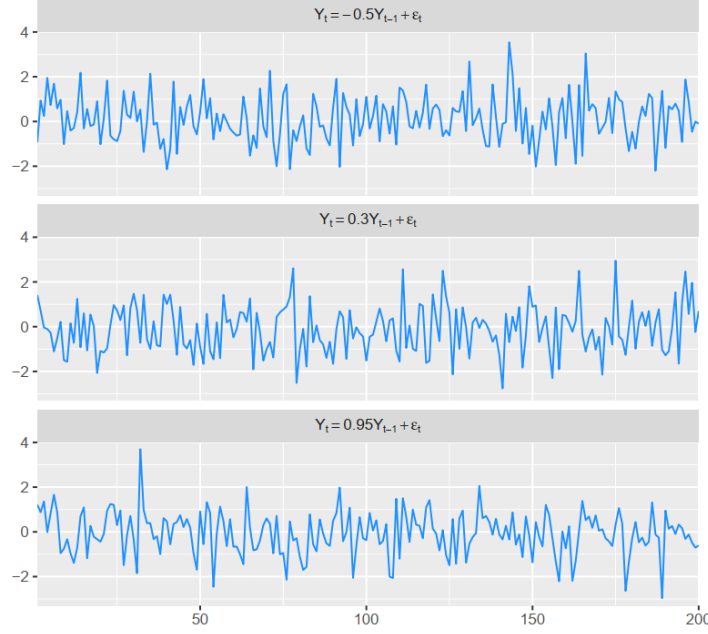
To estimate the ACF of a given time series, we need to use its sample equivalent and a version of the Law of Large Numbers, presented in the previous section, because we're only looking for correlations – i.e., population moments. To estimate the PACF, that is not enough. We're now looking for *partial 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.
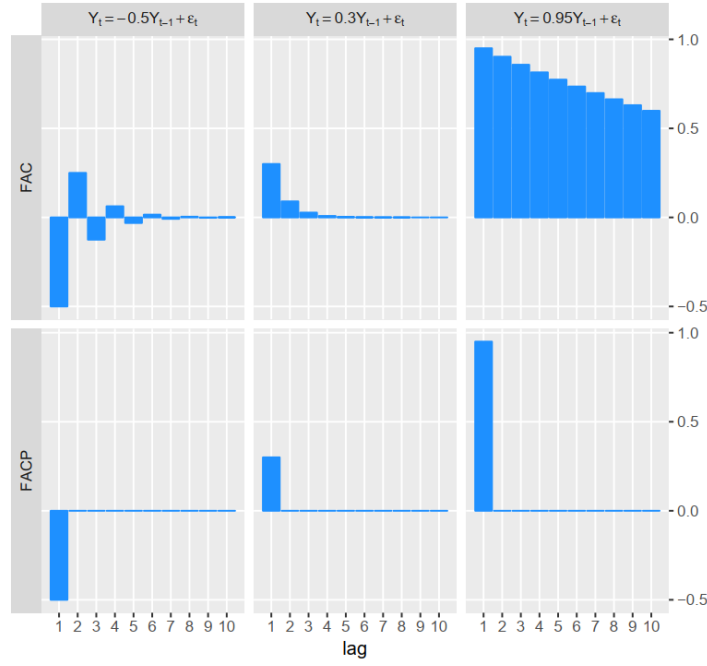
The following plots showcase ACFs and PACFs for AR(1) processes.

2.6.5 Conditions for stationarity

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

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

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

13

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

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:*

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

where $c, \theta_1, \ldots, \theta_p$ 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:

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

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 $\mathbb{E}(\epsilon_t)(Y_t - \mu) = \sigma^2$.

2.7.2 ACF of an AR(p) process

Dividing the previous result by γ_0 yields:

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

Evaluating at $j = 1, 2, \ldots, p-1$ 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 $\rho_1, \rho_2, \ldots, \rho_j$ as functions of $\phi_1, \phi_2, \ldots, \phi_j$. 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:*

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

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:

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

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

$$(L) Y_t = \varepsilon_t$$

---

[2]Review this.

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

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

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]:

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

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:

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

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

$$
(1 - \lambda_1 L)\dots(1 - \lambda_p L) = \varepsilon_t
$$

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

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

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:*

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

where $c, \theta_1, \dots, \theta_p, \phi_1, \dots, \phi_q$ 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:

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

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($\cdot$) 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

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

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($\cdot$) 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(∞)

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

onde

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

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

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
```

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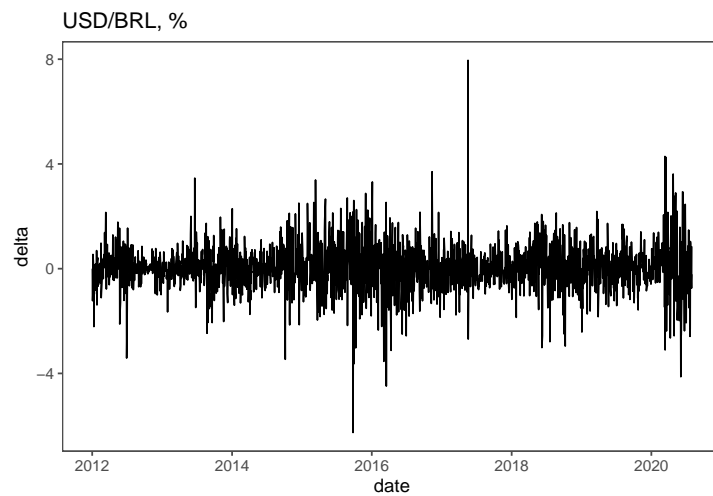
USD/BRL, %


```
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).

USD/BRL, +/-


```
# For delta
```

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

Series df$delta
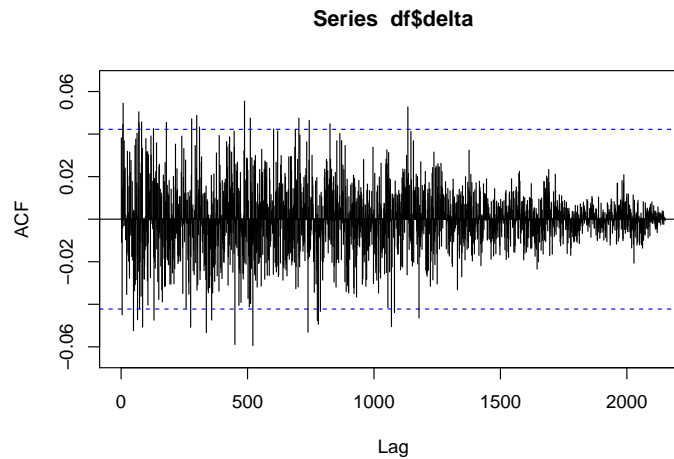

```r
acf_test_values <- acf_delta$acf/sd(acf_delta$acf)

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
```

```r
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
```

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

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

```r
facst
```

Series: df$delta

`faclt`

Series: df$delta

`facpst`

Series: df$delta

22

```
facplt
```


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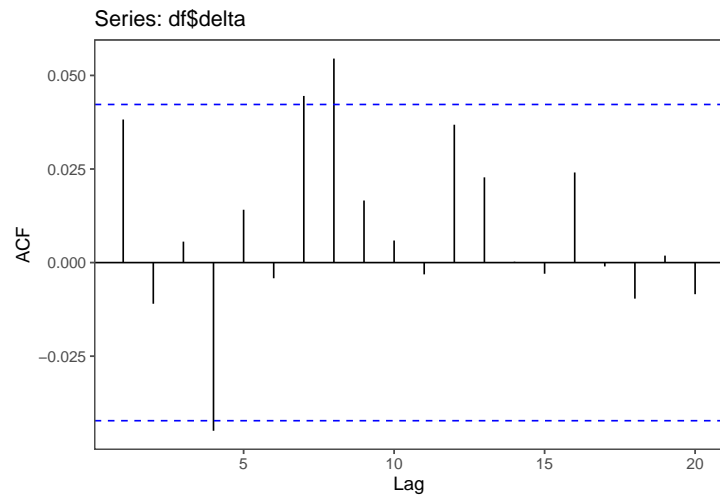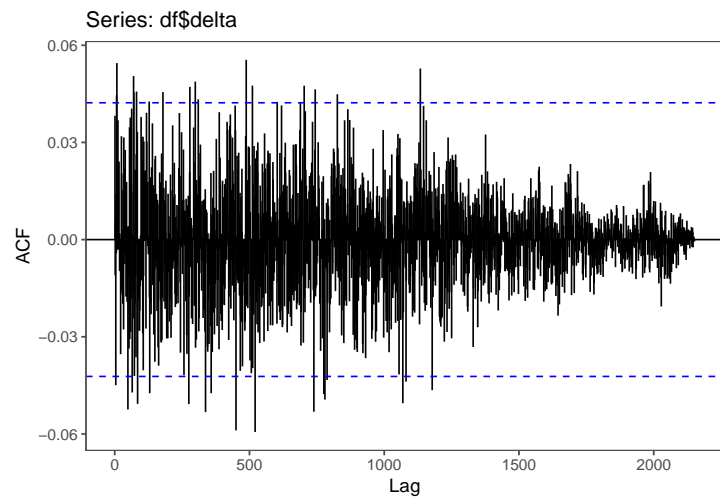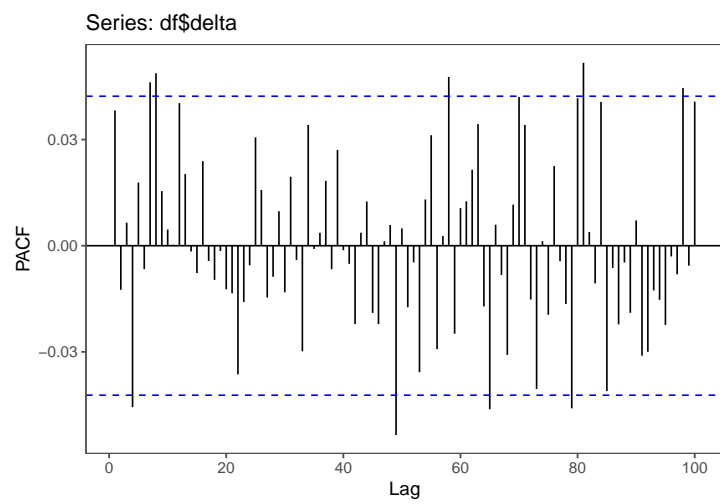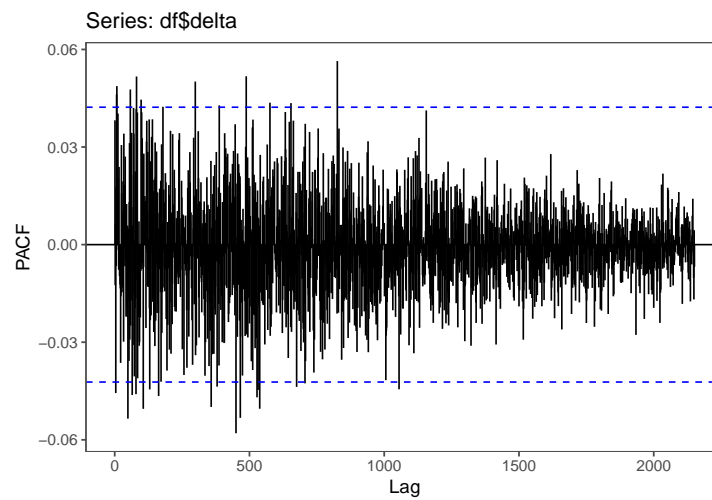
```
facst2 <- ggAcf((df$delta)^2, type = "correlation", lag.max = 20,
    plot = T) + theme_few()
facst2
```


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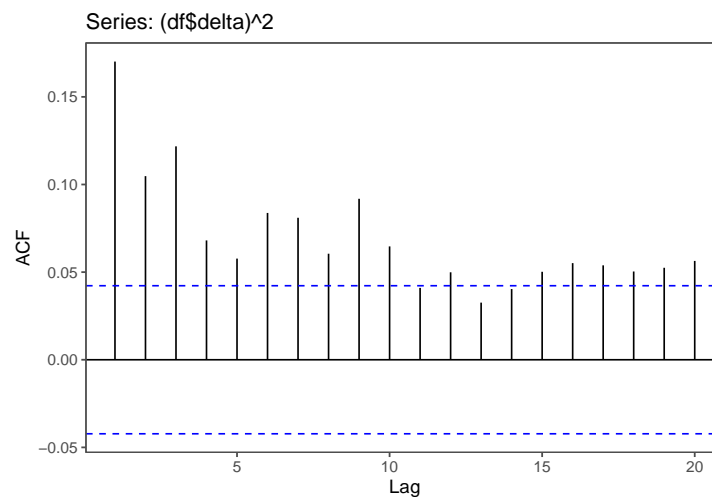
Let's now create our first ARMA models (equivalent to ARIMA with 2nd argument = 0). We'll begin with the first hypothesis: $\mathbb{P}(+) = \mathbb{P}(-)$. Modelling this with an AR(1), we have:

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

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
```
```r
plot(AR1sign)
```

Inverse AR roots
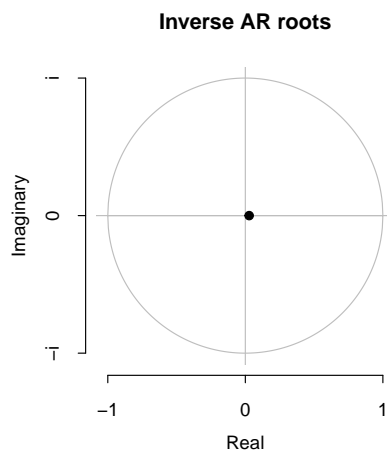

```r
tsdisplay(AR1sign$residuals)
```

AR1sign$residuals
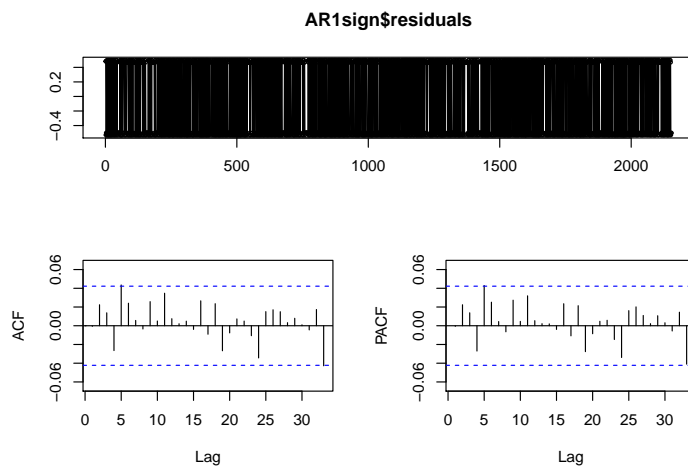
With the results of the summary, we can now apply a hypothesis test for our first question.[1]

$$H_0 : \beta = 0$$
$$H_1 : \beta \neq 0$$

---

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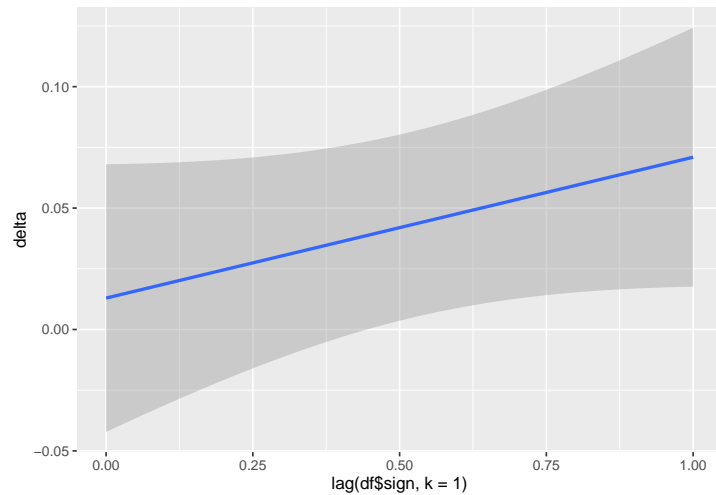
$$\Delta_{t+1} = \alpha + \beta_1 \Delta_t + \beta_2 Sign_t + \varepsilon, \qquad \varepsilon \sim wn(0, \sigma^2)$$

```
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)
```

Inverse AR roots
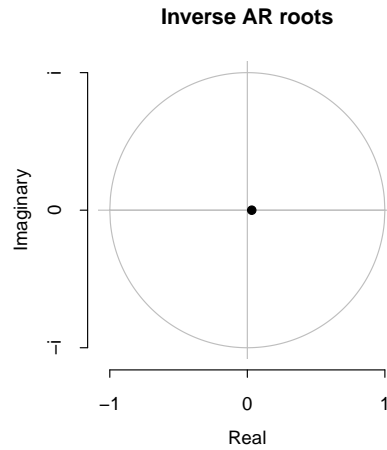

```
tsdisplay(AR1delta$residuals)
```

AR1delta$residuals
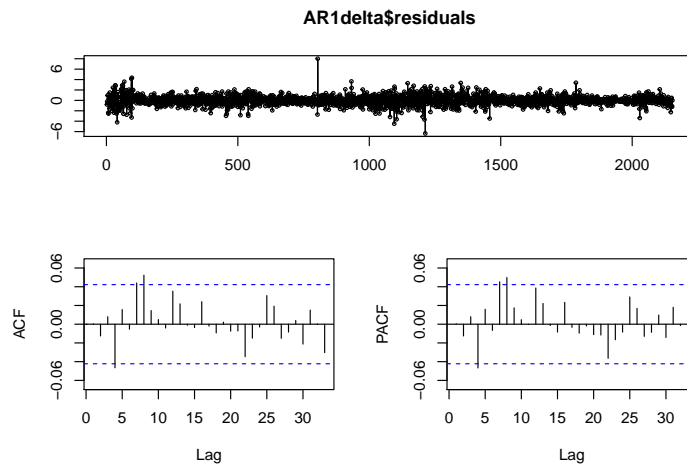
The last hypothesis in the problem refers to the variance:

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

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

```
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
```

27

```
## 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
```

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

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

```r
plot(AR1var)
```


Inverse AR roots
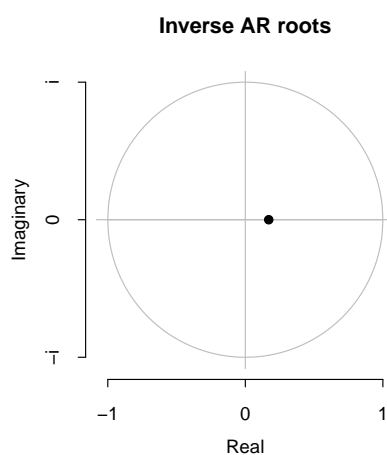
```r
tsdisplay(AR1var$residuals)
```


AR1var$residuals
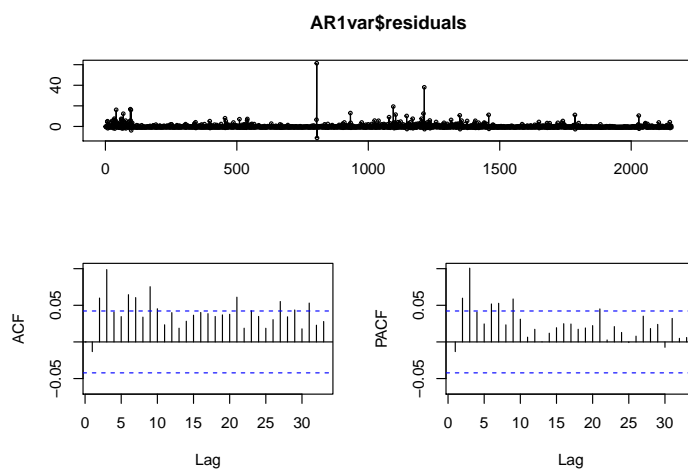
Now, let's run *auto.arima.*

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

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

28

```
## 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)
```
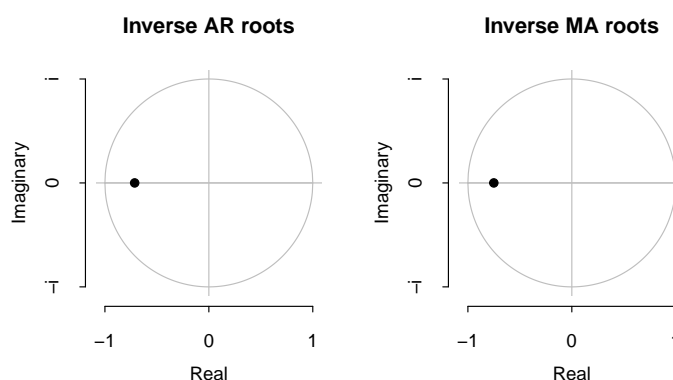


```
tsdisplay(aadelta$residuals)
```
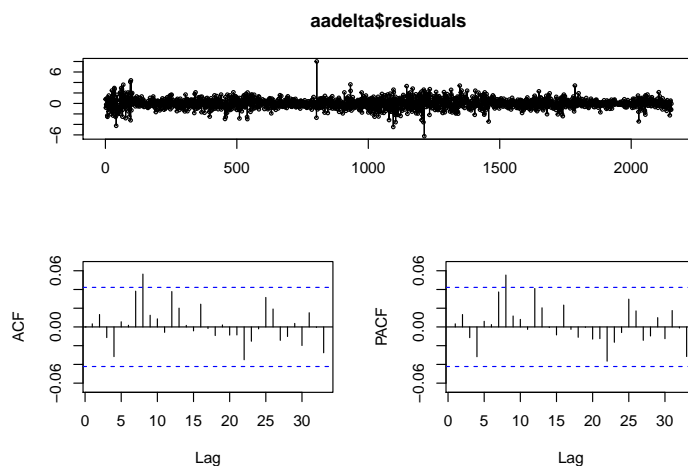


```
aasign <- auto.arima(df$sign, stepwise = F)
summary(aasign)
```

```
## 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
```

```r
plot(aasign)
```

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

No AR or MA roots


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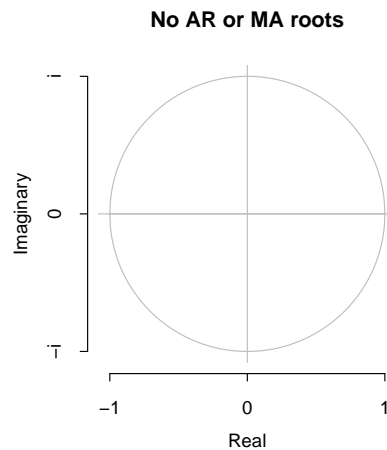
aasign$residuals
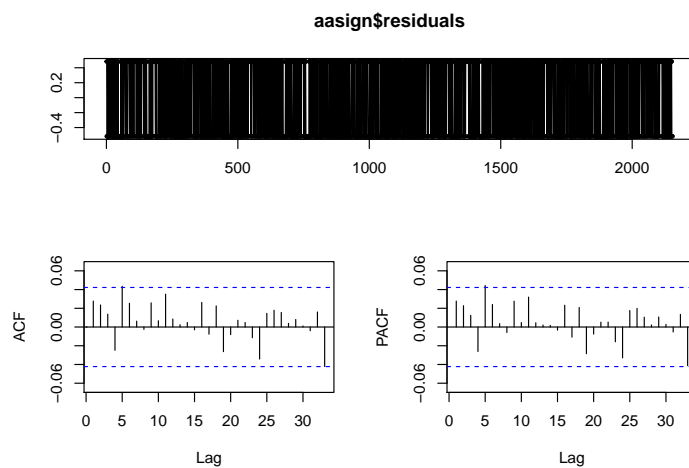

```r
aavar <- auto.arima((df$delta)^2, stepwise = F)
summary(aavar)
```

```
## 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
```

```r
plot(aavar)
```

Inverse MA roots


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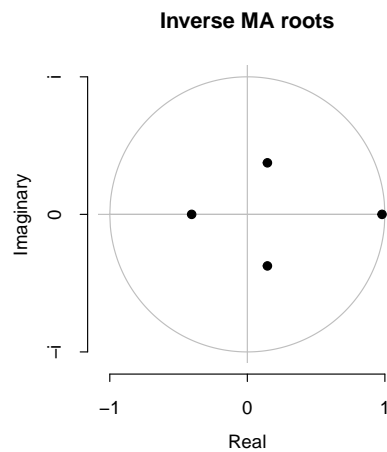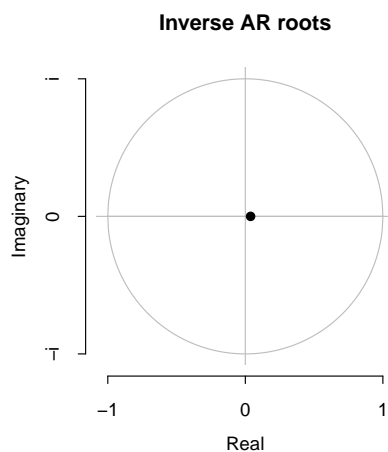
aasign$residuals
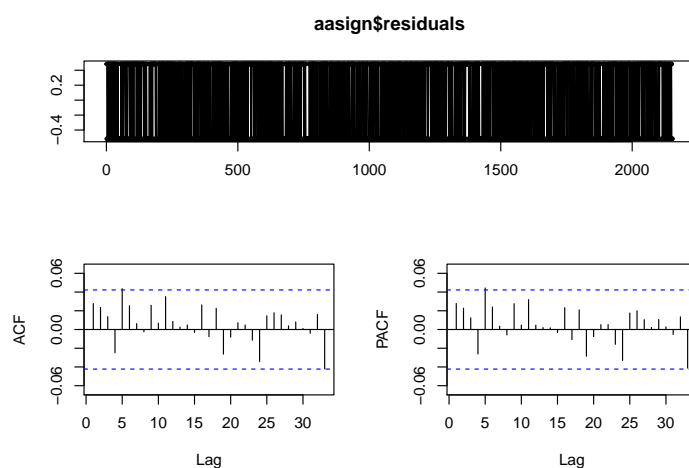

```r
aardelta <- auto.arima(df$delta, max.q = 0, stepwise = F)
summary(aardelta)
```

31

```
## 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
```

```r
plot(aardelta)
```

Inverse AR roots


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

aardelta$residuals
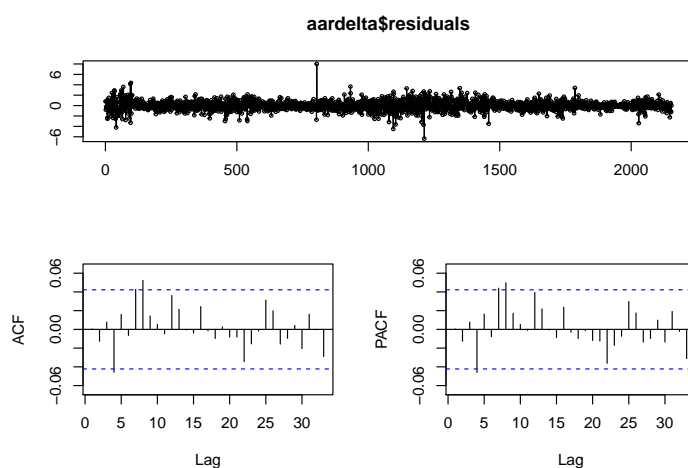

```r
aarsign <- auto.arima(df$sign, max.q = 0, stepwise = F)
summary(aarsign)
```

32

```
## 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
```

```r
plot(aarsign)
```

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

No AR or MA roots
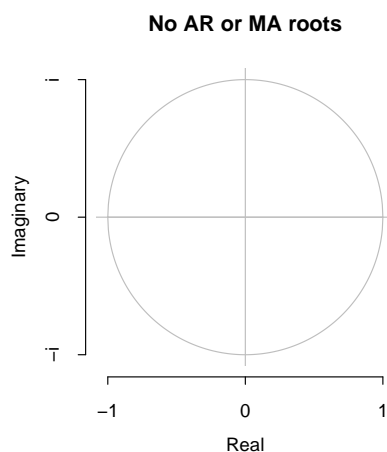

```r
tsdisplay(aarsign$residuals)
```

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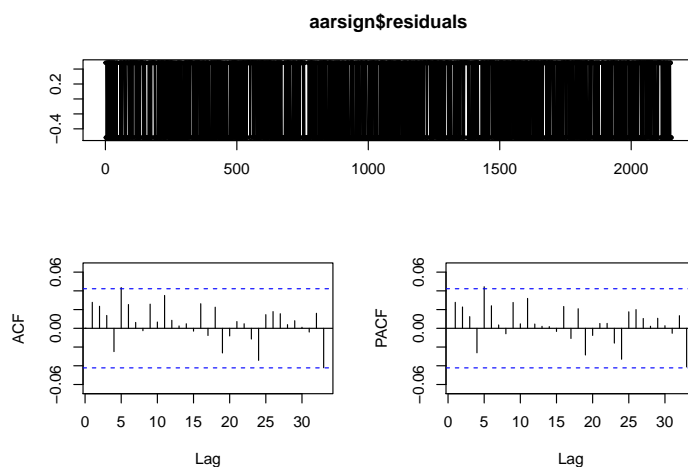
```
aarvar <- auto.arima((df$delta)^2, max.q = 0, stepwise = F)
summary(aarvar)
```

```
## 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
```

```
plot(aarvar)
```

Inverse AR roots
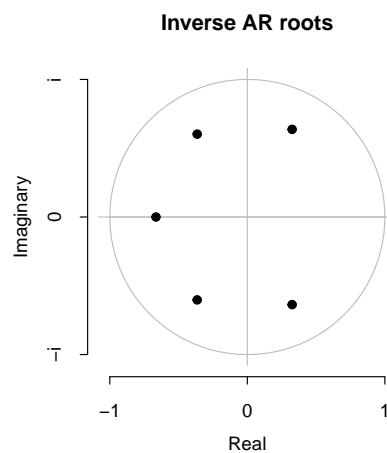

```
tsdisplay(aarvar$residuals)
```

34

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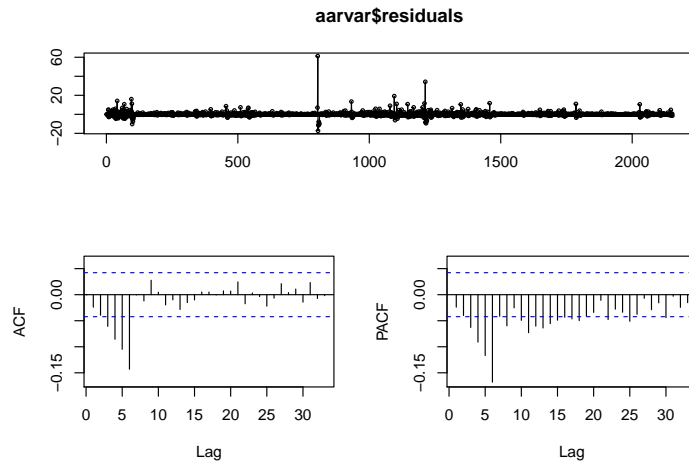
$$\Delta_{t+1} = c + \beta\Delta_t + \varepsilon$$

```r
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
```

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

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