Lecture III.- Univariate Time Series Models

Marcelo Villena, PhD
Santa María University

October, 2022

Outline

1. ARIMA models: modeling the short-term
   - AR Models
   - MA models
   - ARMA models
2. Model selection
3. Application - Short-term Inflation
4. References

ARIMA models: modeling the short-term

- In 1970, George Box and Gwilym Jenkins two engineers with a statistical background, systematized statistical models for the analysis of univariate time series, SEE [1].
- In their seminal work, Box & Jenkings proposed a methodology that take into account the dependence between data.
- Thus, each observation is modeled as a function of the previous values, the time dimension therefore plays a fundamental role in the statistical analysis.

ARIMA models: modeling the short-term

- Box-Jenkins prediction models belong to the family of **linear algebraic models**, which consider a real time series as a probable realization of a certain **stochastic process**.
- These models are known by the generic name of **ARIMA (Auto-regressive Integrated Moving Average)**, which derives from its three components Autoregressive (AR), Integrated (I) Moving Averages (MA). Modeling a time series involves identifying a suitable ARIMA model that fits the series under study, contains the minimum necessary elements to describe the phenomenon and is useful for forecasting.

About the backshift operator

Backshift operator

We define the backshift operator as:

$$Bx_t = x_{t-1} \tag{1}$$

$$B^2x_t = B(Bx_t) = Bx_{t-1} = x_{t-2} \tag{2}$$

Así:

$$B^kx_t = x_{t-k} \tag{3}$$

Thus we have that the first difference can be defined in terms of lags, in other words the backoff operator:

$$\triangle x_t = x_t - x_{t-1} = (1-B)x_t \tag{4}$$

In general:

$$\triangle^d x_t = (1-B)^d x_t \tag{5}$$

Example Autoregressive Process of Order 1: AR(1)

AR(p)

An autoregressive model of order p, often shortened to $AR(p)$, has the form:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \epsilon_t \tag{6}$$

where x_t is a stationary series, and $\phi_1, \phi_2, \ldots, \phi_p$ are constant. If the mean of x_t is μ, then we can replace $x_t - \mu$ in (6)

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + \ldots + \phi_p(x_{t-p} - \mu) + \epsilon_t \tag{7}$$

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \epsilon_t \tag{8}$$

where $\alpha = \mu(1 - \phi_1 - \phi_2 \ldots \phi_p)$

Example Autoregressive Process of Order 1: AR(1)

AR(p)

Using the backward operators AR(p) looks like:

$$(1 - \phi_1 B + \phi_2 B^2 - \ldots - \phi_p B^p) \tag{9}$$

or even more concisely

$$\phi(B)x_t = \epsilon_t \tag{10}$$

Example Autoregressive Process of Order 1: AR(1)

- In an AR(1) process the variable x_t is left only by its past value x_{t-1}:

$$x_t = \phi x_{t-1} + \varepsilon_t \tag{11}$$

- where as we know εt is a white noise process with zero mean and constant variance σ_2, and φ is a parameter pair. To verify that the AR(1) model is stationary we must prove that it is:

Example Autoregressive Process of Order 1: AR(1)

(1) Stationary in mean

$$E(x_t) = E(\phi x_{t-1} + \varepsilon_t) = \phi E(x_{t-1}) \tag{12}$$

- In order for the process to be stationary, the mean must be constant and finite in time, which implies:

$$E(x_t)\,(1-\phi) = 0$$

$$E(x_t) = \frac{0}{1-\phi} \tag{13}$$

- Therefore, for the process to be stationary the pairameter $\phi \neq 0$.

Example Autoregressive Process of Order 1: AR(1)

(2) Stationary in covariance To verify that the AR(1) model is stationary, the variance must be constant and finite in time:

$$\gamma = E(x_t - E(x_t))^2 = E(\phi x_{t-1} + \varepsilon_t - 0)^2 = \phi^2 var(x_{t-1}) + \sigma_2 \quad (14)$$

- Assuming that the process is stationary:

$$E(x_t)^2 = var(x_{t-1}) = var(x_t) = \gamma$$

- From here we have that:

$$\gamma = \phi^2 \gamma + \sigma_2$$

- Therefore:

$$\gamma = \frac{\sigma_2}{1 - \phi^2} \quad (15)$$

- For a process to be stationary, it is necessary that $\phi| < 1$.

Example Autoregressive Process of Order 1: AR(1)

- If it is satisfied that $|\phi| < 1$, then we can represent the AR(1) model as a linear process given by:

$$x_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j} \tag{16}$$

- Equation (16) is called t**he causal stationary solution of the model**. The term causal refers to the fact that x_t does not depend on the future. In fact, by simple substitution,

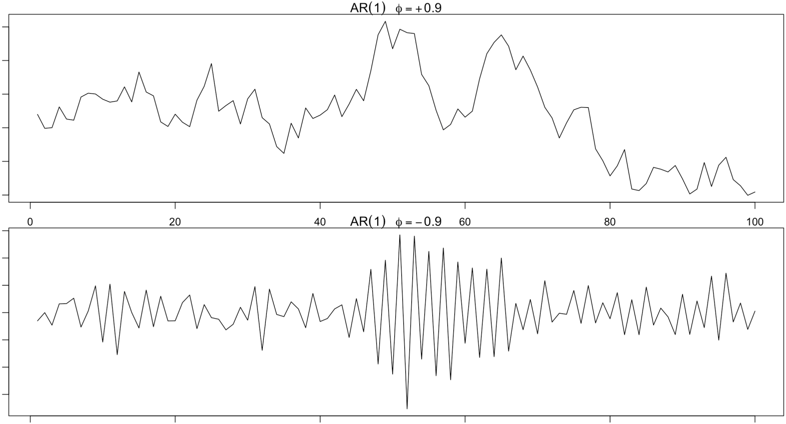
$$\underbrace{\sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}}_{x_t} = \underbrace{\phi \left(\sum_{k=0}^{\infty} \phi^k \epsilon_{t-1-k} \right)}_{x_{t-1}} + \epsilon_t$$

AR(1) model simulation

R Code

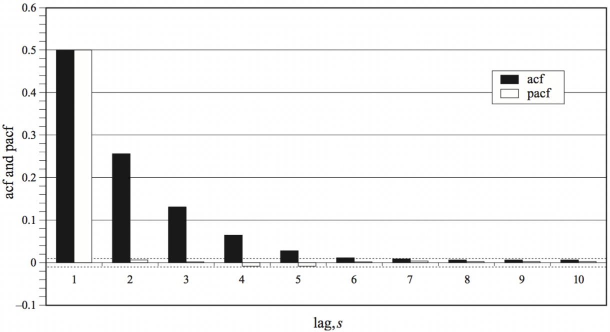
```
par(mar=c(1,1,1,1))
par(mfrow=c(2,1))
plot(arima.sim(list(order=c(1,0,0), ar=.9), n=100), ylab="x",
main=(expression(AR(1)   phi==+.9)))
plot(arima.sim(list(order=c(1,0,0), ar=-.9), n=100), ylab="x",
main=(expression(AR(1)   phi==-.9)))
```

AR(1) model simulation



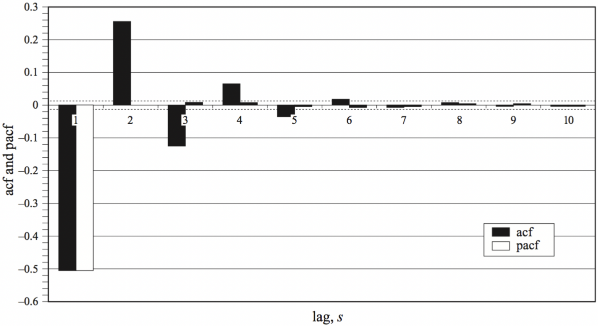
AR(1) model identification

In the case of an AR type process, the correlogram, graphical representation of the autocorrelation function, will have a damped behavior towards zero with all positive values, in case $\theta > 0$, or alternating the sign, starting with negative, if $\theta < 0$.



Sample autocorrelation and partial autocorrelation functions for a more rapidly decaying AR(1) model: $y_t = 0.5y_{t-1} + u_t$

AR(1) model identification



Sample autocorrelation and partial autocorrelation functions for a more rapidly decaying AR(1) model with negative coefficient: $y_t = -0.5y_{t-1} + u_t$

Moving Average - MA (q)

- As an alternative to the autoregressive representation in which the x_t on the left hand side of the equation is assumed to be linearly combined, the q-order moving average model, abbreviated as MA(q), assumes that the white noise ϵ_t usually on the right hand side of the equation, are linearly combined to model the observed data.

ARIMA models: modeling the short-term

Definition: Moving Average - MA (q)

$$x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} \tag{17}$$

where there are q lags of the moving average ϵ_t and $\theta_1 + \theta_2 + \ldots + \theta_q$ are parameters.

Although it is not necessary, we assume that ϵ_t is a white noise series.

ARIMA models: modeling the short-term

Definition: **Moving Average - MA (q)**

We can also write the process $MA(q)$ in the equivalent form:

$$x_t = \theta_t(B)\epsilon_t \tag{18}$$

where θ_t is the moving average operator defined as:

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q \tag{19}$$

Unlike the autoregressive process, the moving average process is stationary for any value of the parameters $\theta_1 + \theta_2 + \ldots + \theta_q$.

Interpretation of the moving average model - MA(q)

- Just as an autoregressive model is intuitively simple to understand, the formulation of a moving average model is often not intuitive. What does it mean that a random variable is explained in terms of errors made in previous periods, where do these errors come from, what is the justification for such a model? In fact, a moving average model can be obtained from an autoregressive model by making successive substitutions.

Interpretation of the moving average model - MA(q)

Assume an $AR(1)$ model, with no independent term:

$$x_t = \phi x_{t-1} + \epsilon_t \tag{20}$$

if we consider $t-1$ instead of$t$the model would be in this case:

$$x_{t-1} = \phi x_{t-2} + \epsilon_{t-1} \tag{21}$$

replacing:

$$x_t = \phi^2 x_{t-2} + \phi \epsilon_{t-1} + \epsilon_t \tag{22}$$

Interpretation of the moving average model - MA(q)

If we now substitute x_{t-2} by its autoregressive expression and so on we arrive at a model of the type:

$$x_t = \epsilon_t + \theta\epsilon_{t-1} + \theta^2\epsilon_{t-2} + \ldots + \theta^q\epsilon_{t-q} \tag{23}$$

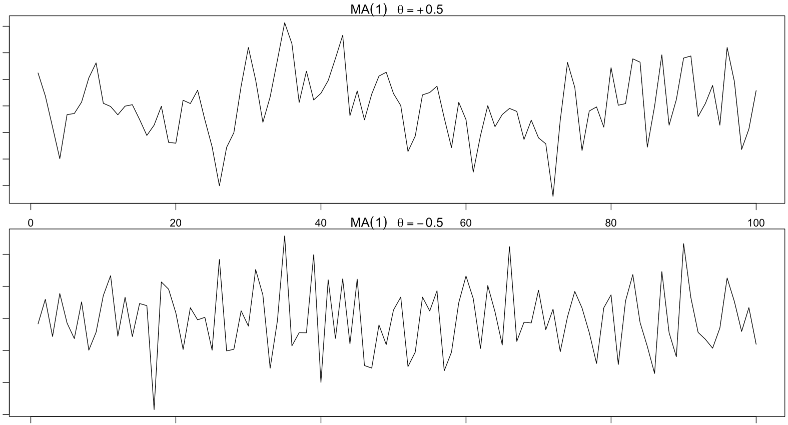
which is the expression, without an independent term, of a moving average model as the one discussed above. In fact, strictly speaking, the passage from one model to the other should be done in reverse, from a MA to an AR, using the general Wold decomposition theorem.

MA(1) model simulation

R Code

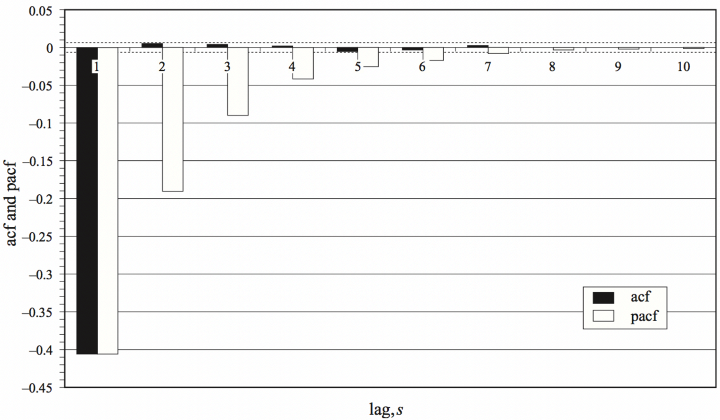
```
par(mfrow = c(2,1))
plot(arima.sim(list(order=c(0,0,1), ma=.5), n=100), ylab="x",
main=(expression(MA(1)   theta==+.5)))
plot(arima.sim(list(order=c(0,0,1), ma=-.5), n=100), ylab="x",
main=(expression(MA(1)   theta==-.5)))
```

MA(1) model simulation



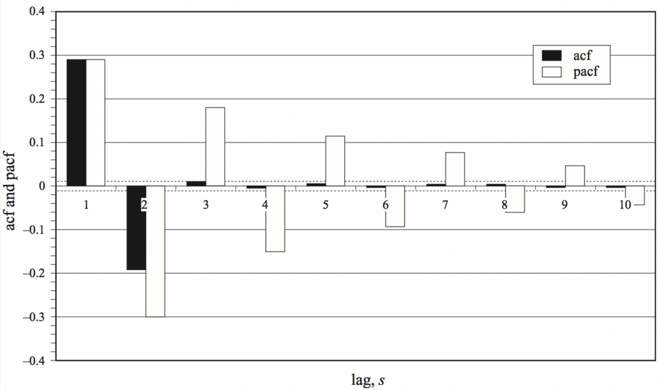
Identification of MA model

For the identification of all the components of the MA model, as we saw for the AR model, we use the autocorrelation function (AFC) and the partial autocorrelation function (PAFC), and thus proceed to the identification of the components, based on the graphs of the different models.



Sample autocorrelation and partial autocorrelation functions for an MA(1) model: $y_t = -0.5u_{t-1} + u_t$

Identification of MA model



Sample autocorrelation and partial autocorrelation functions for an MA(2) model: $y_t = 0.5u_{t-1} - 0.25u_{t-2} + u_t$

ARMA models

Definition: Autoregressive moving average - ARMA (p, q)

A time series $\{x_t, t = 0, \pm 1, \pm 2, \dots\}$ is an ARMA(p, q) process, if it is stationary and

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad (24)$$

The parameters p and $q$are called autoregressive orders and moving averages, respectively.

If x_t has a non-zero mean μ, we establish that $\alpha = \mu(1 - \theta_1, \dots - \theta_q)$ and we can rewrite the model as:

$$x_t = \alpha + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + w_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}. \quad (25)$$

Invertibility

- A time series is invertible if the errors can be inverted in a representation of past observations. Thus, for example, as we have already seen, the AR model is always invertible. In the case of the ARMA model, the roots of the following equations must be analyzed to ensure invertibility.

$$\phi(z) = 1 + \phi_1 z + \phi_2 z^2 + \ldots + \phi_p z^p \tag{26}$$

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q \tag{27}$$

Invertibility

- In particular the ARMA model will be invertible if and only if $\theta(z) \neq 0$ for$|z| \leq 1$ In general, the eigenvalues are the solution of $det(A - \lambda I) = 0$, we see that this is almost the characteristic polynomial of the equations we defined above.
- Therefore, we see that the eigenvalues of A are the inverse of the roots of the characteristic polynomial, and that convergence of the backward iteration occurs when the roots of the characteristic polynomial lie outside the unit circle.

Stationarity and Invertibility

- Wold showed that all stationary stochastic covariance processes could be decomposed as the sum of deterministic and linearly indeterministic processes which were uncorrelated with all lags; that is, if y_t is the stationary covariance, then:

$$y_t = x_t + z_t \tag{28}$$

- where $x_t$is a stationary deterministic process in covariance and z_t is linearly indeterministic, with $Cov(x_t, z_s) = 0$ for all t and s. This result provides a theoretical basis for Box and Jenkins' proposal to model scalar covariance stationary (unseasonalized) processes such as ARMA processes.

ARMA models (p,q)

- As indicated above, when $q = 0$, the model is called the autoregressive model of order p, $AR(p)$, and when $p = 0$, the model is called the moving average model of order q, $MA(q)$.
- It is useful to write ARIMA models using the AR operator and the MA operator described above. In particular, the $ARMA(p, q)$ model can then be written concisely as:

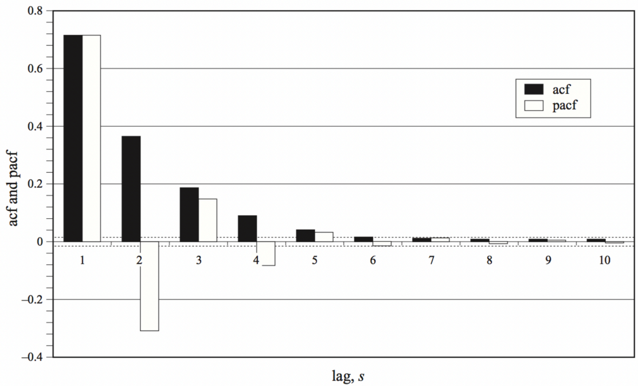
$$\phi(B)x_t = \theta(B)\epsilon_t. \tag{29}$$

- **ARIMA models (p, i, q)** The ARMA model gains its I and becomes ARIMA when it must be integrated to achieve stationarity. The index I will then be the number of times it must be differenced.

ARMA model identification

- The autocorrelation function (AFC) and the partial autocorrelation function (PAFC) are used to identify all the components of the ARMA model, and the seasonal and non-seasonal components are identified separately, based on the graphs of the different models.

ARMA model identification



Sample autocorrelation and partial autocorrelation functions for an ARMA(1, 1) model: $y_t = 0.5y_{t-1} + 0.5u_{t-1} + u_t$

ARMA model identification

Summing up

ACF and PACF properties

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

SARIMA model

- ARIMA models are also capable of modeling a wide range of seasonal data. The so-called SARIMA models, Seasonal ARIMA models, are obtained by including additional seasonal terms in the ARIMA models we have seen so far, as follows:

$$ARIMA(p, d, q)(P, D, Q)m \tag{30}$$

where $m =$ number of periods per season.

- We use uppercase notation for the seasonal parts of the model and lowercase notation for the non-seasonal parts of the model. The seasonal part of the model consists of terms that are very similar to the non-seasonal components of the model, but involve seasonal period regressors.

Statistical evaluation of an ARIMA model

- **Statistical significance of the parameters**: The coefficients obtained in the estimation that are not significantly different from zero, at a significance level of 5%, are not necessary and should be eliminated.
- **Stationarity and invertibility of the estimated model**: For values of the estimated coefficients close to the non-stationarity frontier, it is convenient to carry out a unit root test.
- **Stability of the estimated model**: Even if the parameters are significant, the model can be rejected if there is a strong correlation between the model parameters. This occurs when the correlation coefficient has an absolute value greater than 0.7, then it is convenient to try alternative models.

About Model Selection

- It may happen that several models describe the time series satisfactorily, making it necessary to select the most appropriate model.
- This selection process can be simple or a bit more complex, so it is necessary to use model selection criteria.
- The most common model selection criteria are the **AIC (Akaike Information Criterion)** and the **BIC (Bayesian Information Criterion)** which is a Bayesian extension of the first one.

Information Criteria

Definition

$$AIC = log\hat{\sigma}_k^2 + \frac{n+2k}{n}$$

where $\hat{\sigma}_k^2 = \frac{SSE_k}{n}$, and k is the number of model parameters, n the sample size, and SSE_k is equal to the sum of the squared residuals under the model k ($SSE_k = \sum_{t=1}^{n}(x_t - \bar{x})^2$).

- The value of k that produces the minimum AIC represents the best model. The idea is that minimizing $\hat{\sigma}_k^2$ represents a reasonable objective, except that it decreases monotonically as k increases. Therefore, we should penalize the error variance by a term proportional to the number of parameters.

Information Criteria

Definitions

$$AICc = log\hat{\sigma}_k^2 + \frac{n+k}{n-k-2}$$

$$AICc = log\hat{\sigma}_k^2 + \frac{klogn}{n}$$

- BIC is also known as the **Schwarz Information Criterion (SIC)**. Several simulation studies have verified that BIC is adequate to obtain the correct order in large samples, while AICc tends to be superior in smaller samples where the relative number of parameters is large.

About Model Selection

- Ultimately, one model is better than another if its prediction is better. On the other hand, we will say that **a prediction is better than another when it makes a smaller extra-sampling error.**
- Thus, the accuracy of the methods used to forecast can be measured for example through the loss function: **Mean Square Error (MSE)**, in order to understand which model provides a better out-of-sample forecast over another. That is:

$$\mathbf{MSE} = \frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{x}_t)^2 \tag{31}$$

- where x_t corresponds to the actual value of the series at time t and $\hat{x}$ corresponds to the value predicted by the proposed model at the same instant.

About Model Selection

- Other model selection criteria that consider the extra-sampling error are: i) the Mean Absolute Error (MAD), and ii) Mean Absolute Percentage Error (MAPE).

$$\mathbf{MAD} = \frac{1}{T}\sum_{t=1}^{T}|x_t - \hat{x}_t| \tag{32}$$

$$\mathbf{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|1 - \frac{x_t}{\hat{x}_t}\right| \tag{33}$$

Example CPI

Considering monthly CPI data from January 2013 to date in Chile, obtained from the Central Bank's website, we will try to predict the CPI (original series).

R Code

```
rm(list=ls())
data< −read.csv (“ipc.csv”)
ipc < − ts(data[,2],start = c(2013,1), end=c(2018, 6), frequency
= 12)
plot.ts(ipc, xlab='Years', ylab = “Indice de Precios al
Comsumidor')
```

Example CPI

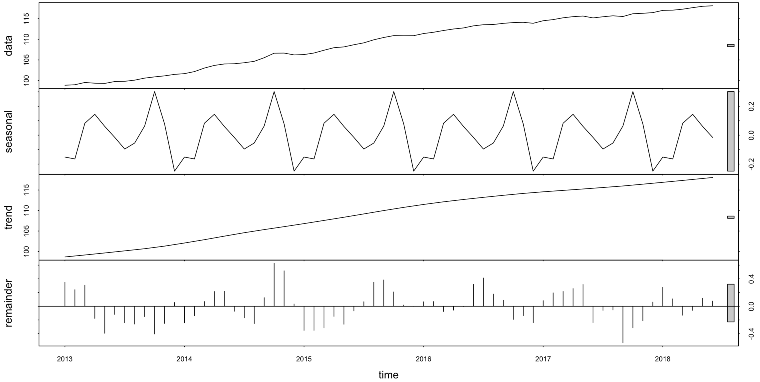
Finding the order of the model. Trend, stationarity, autocorrelation.

R Code

```
# Descomposición
fit < − stl(ipc, s.window="period")
plot(fit)
# Test de raíz unitaria
adf.test(ipc)
adf.test(diff(ipc))
# Función de autocorrelación (AFC) y autocorrelación parcial (PAFC)
acf(diff(ipc),lag=36,lwd=3)
pacf(diff(ipc),lag=36,lwd=3)
```

Example CPI

- **Series decomposition**



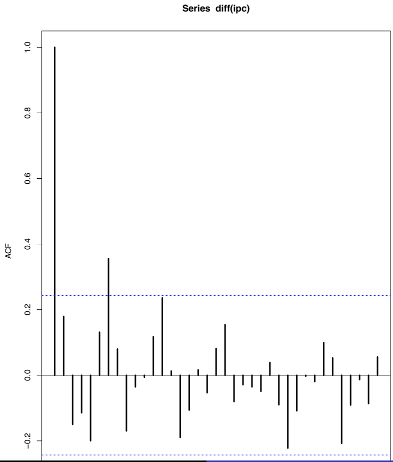
Example CPI - Unit root test

Augmented Dickey-Fuller Test
data: ipc
Dickey-Fuller = -0.11148, Lag order = 4, p-value =0.99
alternative hypothesis: stationary

Augmented Dickey-Fuller Test
data: diff(ipc)
Dickey-Fuller = -5.8024, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary

Example CPI

Autocorrelation Function (AFC) and Partial Autocorrelation Function (PAFC)



Example CPI - Forecast

R Code

```
train.series =ipc [1 : 44]
test.series = ipc [45 : 62]
arima.model=arima(train.series, order=c(0,1,1))
forecast=predict(arima.model, length(test.series)
mse < −sum((forecast$pred-test.series)∧2)/length(test.series)
mad < − sum(abs(forecast$pred-test.series))/length(test.series)
mape < − sum(abs( 1 -
forecast$pred/test.series))/length(test.series)
fit < − auto.arima(ipc)
summary(fit)
plot(fit)
mape < − sum(abs(1 - test.series/f[["mean"]]))/length(test.series)
accuracy(fit)
```

Example CPI - output ARIMA (0, 1, 1)

Call:

arima(x = train.series, order = c(0, 1, 1))

Coefficients:

	ma1
	0.8205
s.e.	0.0906

σ^2 estimated as 0.1029 : $loglikelihood = -12.68, aic = 29.37$

forecast ARIMA (0, 1, 1)

mse [1] 69.80031

Example CPI - Forecast - output ARIMA (0, 1, 1)

```
$pred
Time Series:
Start = 45
End = 54
Frequency = 1
[1] 113.6141 113.6253 113.6292 113.6307 113.6311 113.6313
[7] 113.6314 113.6314 113.6314 113.6314
$se
Time Series:
Start = 45
End = 54
Frequency = 1
[1] 0.3128668 0.6841962 0.9882296 1.2406559 1.4565974
[6] 1.6465783 1.8174943 1.9738906 2.1188485 2.2545301
```

Example CPI - output auto.arima

Series: ipc
ARIMA(0,1,1)(0,0,1)[12] with drift

Coefficients:

	ma1	sma1	drift
	0.2329	0.2483	0.2909
s.e.	0.1443	0.1396	0.0500

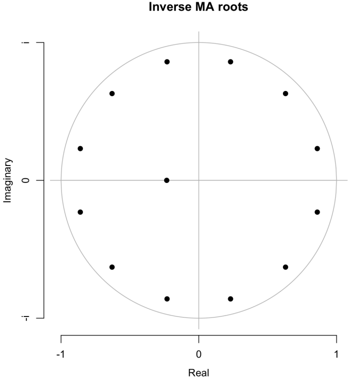
σ^2 estimated as 0.07771 : $loglikelihood = -8.01$
$AIC = 24.02$ $ICc = 24.69$ $BIC = 32.72$

Training set error measures:

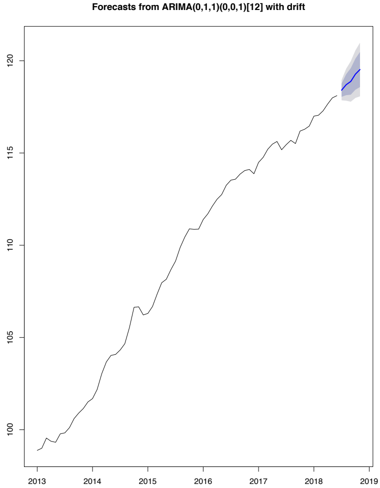
	ME	RMSE	MAE	MPE
Training set	0.00467571	0.2701877	0.2012356	0.005434612

	MAPE	MASE	ACF1
Training set	0.185618	0.05414794	−0.03368001

Example: CPI - nverse MA roots - auto.arima



Example: CPI - Forecast auto.arima



Box-Jenkins modelling procedure

(1) Data preparation involves transformations and differencing. Transformations of the data (such as square roots or logarithms) can help stabilize the variance in a series where the variation changes with the level. This often happens with business and economic data. Then the data are differenced until there are no obvious patterns such as trend or seasonality left in the data. "Differencing" means taking the difference between consecutive observations, or between observations a year apart. The differenced data are often easier to model than the original data.

Box-Jenkins modelling procedure

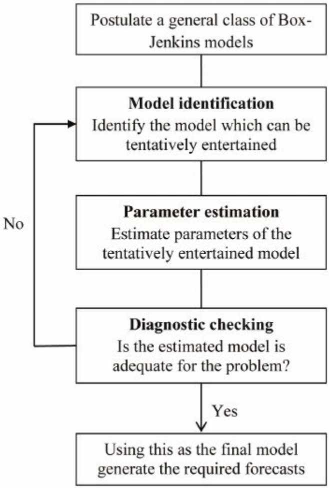
(2) Model selection in the Box-Jenkins framework uses various graphs based on the transformed and differenced data to try to identify potential ARIMA processes which might provide a good fit to the data. Later developments have led to other model selection tools such as Akaike's Information Criterion.

(3) Parameter estimation means finding the values of the model coefficients which provide the best fit to the data. There are sophisticated computational algorithms designed to do this.

Box-Jenkins modelling procedure

(4) Model checking involves testing the assumptions of the model to identify any areas where the model is inadequate. If the model is found to be inadequate, it is necessary to go back to Step 2 and try to identify a better model.
(5) Forecasting is what the whole procedure is designed to accomplish. Once the model has been selected, estimated and checked, it is usually a straight forward task to compute forecasts.

Box-Jenkins modelling procedure



Homework 3

- Calibrate and evaluate the following models for the price of a commodity at your choice:

1. Random walk with drift and without drift.
2. Average of the last 5 years, average of the last 10 years.
3. ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1).
4. AR(1), AR(2), AR(3).
5. α constant, $\psi = 1$ and δ follows a random walk.
6. $\psi = 1, \delta = 0$ and α follows a random walk.

Homework 3

7. α constant,δ and ψ follow random paths with independent innovations.
8. $\delta = 0$, $\alpha$and ψ follow random paths with independent innovations.
9. αconstant, $\delta = 0$ and ψ follows a random walk.
10. α, δ and ψ follow random paths with independent innovations.
11. α constant, $\delta = 0$ and ψ follows an AR(1).
12. α and δ constant, ψ follows an AR(1).

References

- [1] BOX, G.E.P. and G.M. JENKINS (1970) Time series analysis: Forecasting and control, San Francisco: Holden-Day.
- [2] MAKRIDAKIS, S., S.C. WHEELWRIGHT, and R.J. HYNDMAN (1998) Forecasting: methods and applications, New York: John Wiley & Sons.
- [3] PANKRATZ, A. (1983) Forecasting with univariate Box–Jenkins models: concepts and cases, New York: John Wiley & Sons.