Time Series
4. Estimation, validation and forecasting

Josep A. Sanchez-Espigares

Department of Statistics and Operations Research
Universitat Politecnica de Catalunya
Barcelona, Spain

Outline

- Estimation of parameters
- Validation of the model
- Forecasting with ARIMA models

Let $X = \{x_t\}_{t=0}^{N}$ a time series with conditional gaussian distribution

Likelihood of the model:

$$L(\phi, \theta, \sigma_z^2; X) = f_{(X_1, \dots X_T)}(x_1, \dots x_T) = f_1(x_1) \sum_{i=2}^{T} f(x_i | x_{i-1}, \dots, x_1)$$

For example, for an AR(1) model:

$$X_i = \phi X_{i-1} + Z_t \quad Z_t \sim N(0, \sigma_Z^2)$$

$$E(X_i | X_{i-1}, \dots, X_1) = \phi X_{i-1}$$

$$V(X_i | X_{i-1}, \dots, X_1) = \sigma_Z^2$$

$$L(\phi, \sigma_z^2; X) = f_1(x_1) \sum_{i=2}^{N} f\left(\frac{x_i - \phi x_{i-1}}{\sigma_Z}\right)$$

Conditional density: it depends on the distribution considered for the first observation (f_1)

Estimation Criteria

Maximum likelihood estimation

$$\hat{\Lambda}_{ML} = (\hat{\phi}, \hat{\theta}, \hat{\sigma}_z^2) = argmax\left(L(\phi, \theta, \sigma_z^2; X)\right)$$

Maximum likelihood estimators are **consistent** and **asymptotically eficient and gaussian**

$$\hat{\Lambda}_{ML} \approx N(\Lambda, I_\Lambda^{-1})$$

where $I_\Lambda = E\left(-\frac{\partial^2 logL(\Theta;X)}{\partial\Theta^2}\right)$. This is the Hessian of the objective function $(H(L))$

In particular, $se(\hat{\Lambda}_{ML}) = \sqrt{diag[H(L)^{-1}]}$

Estimation Criteria

Some details of the implementation in R:

1. Expression of the model in State-Space form
2. Start at some initial values for the parameters
3. Calculation of the logLikelihood function for these values
4. Non-linear optimization method (`optim`) to find the Maximum Likelihood estimates
5. The `optim` method gives the values of the Hessian of the function. Its diagonal is used to calculate the standard deviation of the estimates (standard errors)

Significance test for the parameters of the model:

Exemple for a coeficient of the AR-part:

$$\begin{cases} H_0 : \phi_i = 0 \\ H_1 : \phi_i \neq 0 \end{cases}$$

Test statistic:

$$\hat{\phi}_i \approx N(\phi_i, \sigma_{\phi_i}) \Rightarrow \hat{t} = \frac{\phi}{se(\phi_i)} \approx t - Student_{T-k}$$

where $se(\phi_i)$ is the estimate of σ_{ϕ_i} based on the Hessian of the function, and k is the number of total parameters of the model.

For practical purposes: if $|\hat{t}| > 2$, then the parameter is significant.

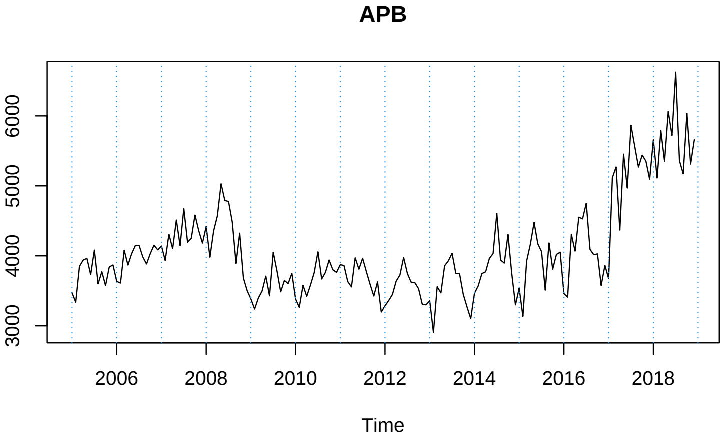
Estimation Criteria

APB: Port traffic in the port of Barcelona (millions of tones)

Source: Ministry of Public Works of Spain (http://www.fomento.gob.es/BE/?nivel=2&orden=04000000)

```
##       Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
## 2005 3473 3338 3850 3940 3963 3733 4082 3600 3773 3575 3842 3870
## 2006 3636 3611 4079 3869 4030 4148 4149 3984 3884 4032 4152 4086
## 2007 4141 3934 4309 4102 4512 4145 4674 4195 4250 4584 4359 4182
## 2008 4418 3981 4357 4570 5030 4792 4776 4482 3891 4324 3682 3504
## 2009 3385 3240 3397 3497 3710 3428 4050 3785 3486 3649 3603 3750
## 2010 3382 3265 3578 3423 3583 3760 4058 3669 3763 3940 3802 3764
## 2011 3874 3862 3633 3556 3971 3810 3965 3776 3593 3425 3628 3197
## 2012 3286 3362 3449 3639 3726 3977 3750 3623 3618 3532 3309 3300
## 2013 3359 2906 3559 3471 3859 3929 4036 3749 3744 3449 3271 3103
## 2014 3464 3569 3747 3772 3962 4033 4607 3942 3896 4305 3749 3300
## 2015 3536 3135 3927 4166 4477 4171 4061 3510 4185 3810 4021 4052
## 2016 3460 3410 4306 4067 4553 4528 4752 4092 4016 4028 3577 3863
## 2017 3674 5116 5270 4368 5454 4969 5865 5564 5268 5440 5353 5094
## 2018 5659 5112 5789 5350 6063 5720 6627 5360 5173 6037 5311 5659
```

APB: Port traffic in the port of Barcelona (millions of tones)



Estimation Criteria

Data: APB

Proposed transformations: $W_t = (1-B)(1-B^{12})log(X_t)$

Proposed model: $ARIMA(0,1,4)(0,1,1)_{12}$

```
(model=arima(serie,order=c(0,1,4),seasonal=list(order=c(0,1,1),period=12)))
```

```
##
## Call:
## arima(x = serie, order = c(0, 1, 4), seasonal = list(order = c(0, 1, 1), period = 12))
##
## Coefficients:
##           ma1      ma2     ma3      ma4     sma1
##       -0.4570  -0.0081  0.0405  -0.1574  -0.9227
## s.e.   0.0823   0.0946  0.0887   0.0861   0.1530
##
## sigma^2 estimated as 0.004536:  log likelihood = 187.27,  aic = -362.54
```

```
cat("\nT-ratios:",round(model$coef/sqrt(diag(model$var.coef)),2))
```

```
##
## T-ratios: -5.55 -0.09 0.46 -1.83 -6.03
```

```
cat("\nSignificant?:",abs(model$coef/sqrt(diag(model$var.coef)))>2)
```

```
##
## Significant?: TRUE FALSE FALSE FALSE TRUE
```

Validation of the model

For all $ARIMA(p,d,q)(P,D,Q)_s$ model, the general expression is:

$$\phi_p(B)\Phi_P(B^S)\left((1-B)^d(1-B^s)^D X_t-\mu\right)=\theta_q(B)\Theta_Q(B^S)Z_t$$

The random part of the model is $Z_t \sim N(0,\sigma_Z^2)$

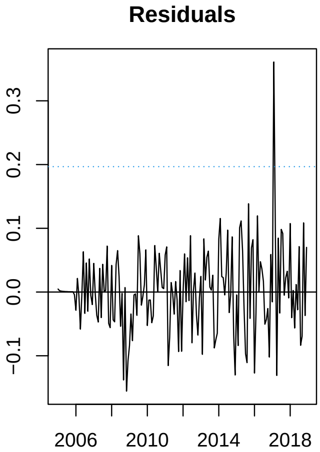
- **Residual Analysis** for Z_t. Premises:
  - Homogeneity of variance (σ_Z^2 constant)
  - Normality ($Z_t \sim Normal$)
  - Independence ($\rho(k)=0 \quad \forall k$)

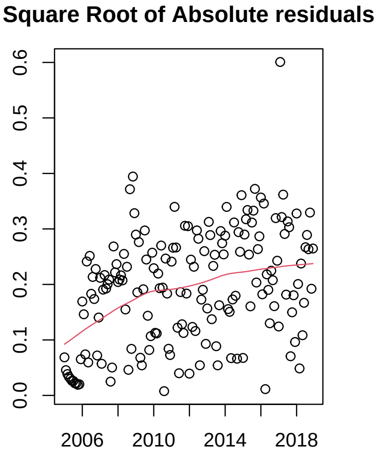
Homogeneity of variance. Tools

- Residuals plot
- Square root of absolute values of the residuals with smooth fit
- ACF/PACF of square of residuals

Validation

```
resid=model$residuals
par(mfrow=c(1,2),mar=c(3,3,3,3))
#Residuals plot
plot(resid,main="Residuals")
abline(h=0)
abline(h=c(-3*sd(resid),3*sd(resid)),lty=3,col=4)
#Square Root of absolute values of residuals (Homocedasticity)
scatter.smooth(sqrt(abs(resid)),main="Square Root of Absolute residuals",
               lpars=list(col=2))
```





Homogeneity of variance. Problems

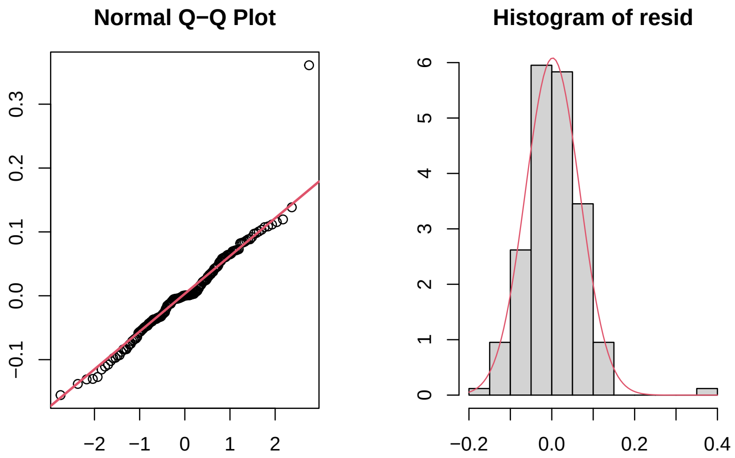
- Outlier observations: Outlier detection and treatment
- Volatility: Models for the variance (GARCH)

Normality. Tools

- Quantile-Quantile plot
- Histogram with theoretical density overlapped
- Shapiro-Wilks test

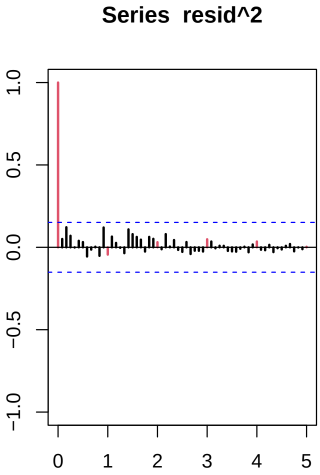
Validation

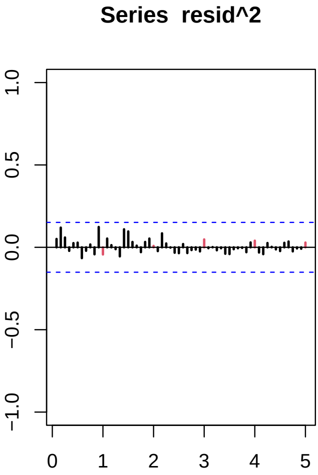
```
par(mfrow=c(1,2),mar=c(3,3,3,3))
#Normal plot of residuals
qqnorm(resid)
qqline(resid,col=2,lwd=2)
##Histogram of residuals with normal curve
hist(resid,breaks=10,freq=F)
curve(dnorm(x,mean=mean(resid),sd=sd(resid)),col=2,add=T)
```



Validation

```
par(mfrow=c(1,2),mar=c(3,3,3,3))
s=12
#ACF & PACF of residuals
par(mfrow=c(1,2))
acf(resid^2,ylim=c(-1,1),lag.max=60,col=c(2,rep(1,s-1)),lwd=2)
pacf(resid^2,ylim=c(-1,1),lag.max=60,col=c(rep(1,s-1),2),lwd=2)
```






```
par(mfrow=c(1,1))
```

Normality. Problems

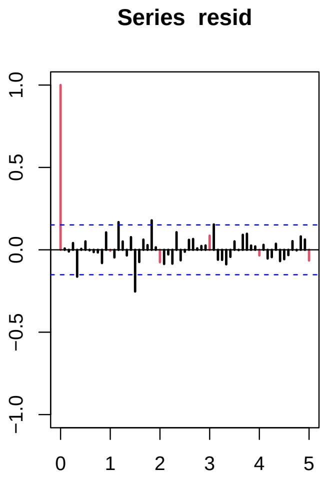
- Outlier observations: Outlier detection and treatment
- Asymmetry or bi-modality: Transformation, outliers or change residuals distribution
- Heavy tails (excess of kurtosi): Volatility models or t-distributions for the residuals

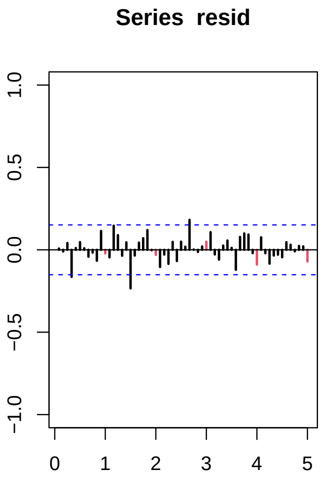
Independence. Tools

- ACF/PACF of residuals
- LJung-Box test
- Durbin-Watson test

Validation

```
par(mfrow=c(1,2),mar=c(3,3,3,3))
s=12
#ACF & PACF of residuals
par(mfrow=c(1,2))
acf(resid,ylim=c(-1,1),lag.max=60,col=c(2,rep(1,s-1)),lwd=2)
pacf(resid,ylim=c(-1,1),lag.max=60,col=c(rep(1,s-1),2),lwd=2)
```






```
par(mfrow=c(1,1))
```

Checking the significance of individual autocorrelation might ignore that the configurantion of all (or a subset) of the lags may be jointly significant.

Ljung-Box Q-statistic: For a lag k test the joint hypothesis that the first k autocorrelations of the residuals are jointly zero:

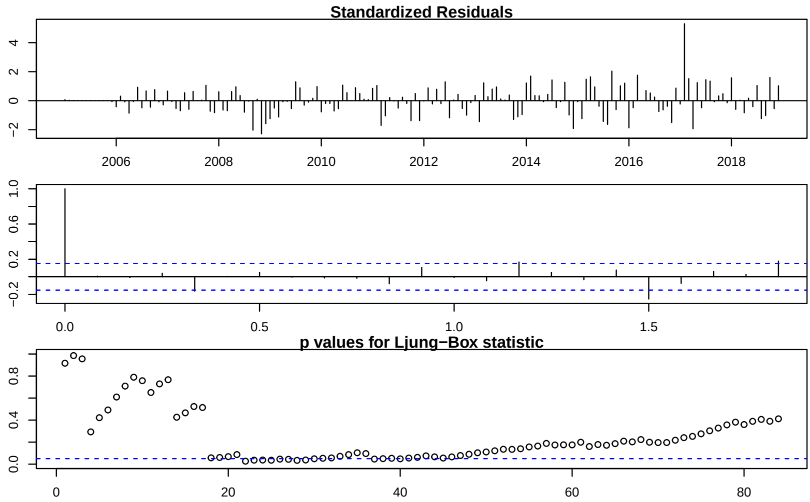
$$H_0 : \rho_Z(1) = \cdots = \rho_Z(k) = 0$$

Test Statistic: $Q = T\sum_{i=1}^{K} \hat{\rho}_Z^2(i)$

Asymptotically, $Q \to \chi_k^2$

Validation

```
par(mfrow=c(1,1),mar=c(3,3,3,3))
#Ljung-Box p-values
par(mar=c(2,2,1,1))
tsdiag(model,gof.lag=7*12)
```



Independence. Problems

- Significant lags: Re-identify or add parameters to the model

Model Selection

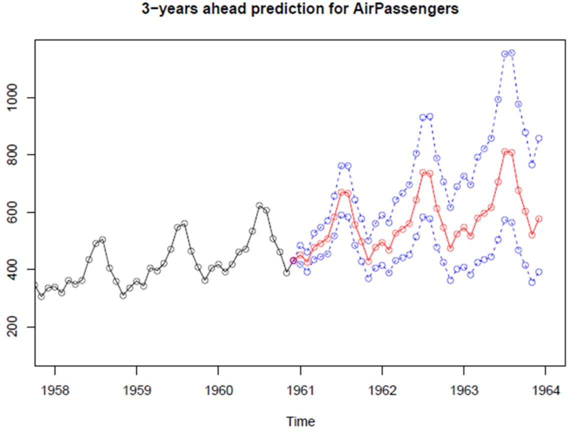
- Information Criterion
  - Balance between godness-of-fit and simplicity of the model
  - Component for the godness-of-fit: $-2 \log L(\phi, \theta, \sigma^2 | X)$
  - Component for the simplicity: $K * par$

$$AIC = -2 \log L(\phi, \theta, \sigma^2 | X) + 2par$$

$$BIC = -2 \log L(\phi, \theta, \sigma^2 | X) + log(N)par$$

$$AICc = -2 \log L(\phi, \theta, \sigma^2 | X) + 2N(p + q + 1)/(N - (p + q) - 2)$$

ARIMA(0,1,1)(0,1,1)$_s$; period $s = 12$; $n.ahead = 36$



Remember: \A times series $\{X_t; t = 0, 1, \cdots\}$ is an **ARMA(p,q)** model if it is **stationary/causal** and **invertible**

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_2 X_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \qquad Z_t \sim WN(0, \sigma_Z^2)$$

with $\phi_p \neq 0$; $\theta_q \neq 0$ and $\sigma_z^2 > 0$

ARMA(p,q) model can be written in concise form as

$$\Phi_p(B) x_t = \Theta_q(B) z_t$$

Objective: To find the best forecast for X_{t+h} at time t using the **h-step-ahead minimum mean square error predictor**, denoted by $\tilde{X}_{t+h|t}$

Questions to help the construction process:

- Which is the **forecasting function**?

  i.e., Which are the **values for** $\tilde{X}_{t+h|t}$ as a function of **h**?

- Which is the **memory of the model**? How are the forecast expressed as a (linear) **function of the previous** (known) observations?

- Which is the **forecasting confidence interval**?

Three equivalent ARMA(p,q) equations:

- **Previous observations** and **random noises**

  $X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$

- **AR** process of **infinite order**

  $X_t = -\pi_1 X_{t-1} - \pi_2 X_{t-2} - \cdots + Z_t$

- **MA** process of **infinite order**

  $X_t = Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \cdots$

Forecasting and Confidence Interval (5/5)

h-step-ahead-forecast

- **Previous observations** and **random noises**

  $$X_{t+h} = \phi_1 X_{t+h-1} + \cdots + \phi_p X_{t+h-p} + Z_{t+h} + \theta_1 Z_{t+h-1} + \cdots + \theta_q Z_{t+h-q}$$

- **AR** process of **infinite order**

  $$X_{t+h} = -\pi_1 X_{t+h-1} - \pi_2 X_{t+h-2} - \cdots + Z_{t+h}$$

- **MA** process of **infinite order**

  $$X_{t+h} = Z_{t+h} + \psi_1 Z_{t+h-1} + \psi_2 Z_{t+h-2} + \cdots$$

Forecasting example. ARMA(2,2) (1/2)

Given the model

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$$

Suppose that **values of past observations are known**, then $Z_t = X_t - \tilde{X}_{t|t-1}$; $Z_{t-1} = X_{t-1} - \tilde{X}_{t-1|t-2}$

Substituting

$$\tilde{X}_{t+1|t} = \phi_1 X_t + \phi_2 X_{t-1} + E(Z_{t+1}) + \theta_1 Z_t + \theta_2 Z_{t-1}$$

$$\tilde{X}_{t+2|t} = \phi_1 \tilde{X}_{t+1|t} + \phi_2 X_t + E(Z_{t+2}) + \theta_1 E(Z_{t+1}) + \theta_2 Z_t$$

$$\tilde{X}_{t+3|t} = \phi_1 \tilde{X}_{t+2|t} + \phi_2 \tilde{X}_{t+1|t} + E(Z_{t+3}) + \theta_1 E(Z_{t+2}) + \theta_2 E(Z_{t+1})$$

$$\ldots$$

$$\tilde{X}_{t+h-3|t} = \phi_1 \tilde{X}_{t+h-1|t} + \phi_2 \tilde{X}_{t+h-2|t} \text{ , } h > 2$$

Forecasting example. ARMA(2,2) (2/2)

In fact,

Given a general **ARMA(p,q)** model, for $h > q$:

h-step-ahead predictor is determined by the difference equations for the autocorrelations (which only depend on the autocorrelation characteristic polynomial)

Non Stationary ARIMA process Forecast. ARIMA(0,1,1) (1/3)

Model:

$$X_t = X_{t-1} + Z_t + \theta Z_{t-1}$$

also defined by

$$(1 - B)X_t = (1 - \theta B)Z_t$$

X_t **can be expressed as**

$$\pi(B) = \frac{1 - B}{1 + \theta B} = 1 - (\theta + 1)B + \theta(\theta + 1)B^2 - \theta^2(\theta + 1)B^3 + \cdots$$

Model:

$$\tilde{X}_{t+1|t} = (\theta + 1)X_t - \theta(\theta + 1)X_{t-1} + \theta^2(\theta + 1)X_{t-2} - \cdot$$

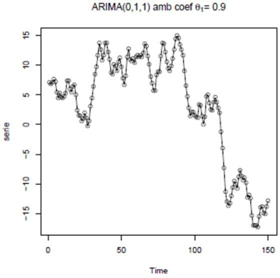
moving θ

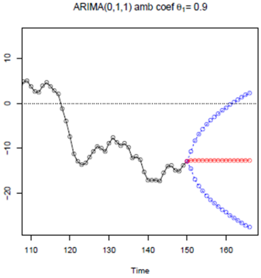
$$\tilde{X}_{t+1|t} = (\theta + 1)X_t - \theta\tilde{X}_{t|t-1}$$

Using the parameter $\lambda = 1 + \theta$. Clearly $\theta = \lambda - 1 = -(1 - \lambda)$

$$\tilde{X}_{t+1|t} = \lambda X_t - (1 - \lambda)\lambda X_{t-1} + (1 - \lambda)^2\lambda X_{t-2} - \cdots = \lambda X_t + (1 - \lambda)\tilde{X}_{t|t-1}$$

Hence, the **forecast is the a linear combination**of the last observation and the forecast obtained in the previous step.





Variance of the forecasting error

Variance of the 1-step-ahead forecasting error

$$Var(e_t(1)) = E\left((X_{t+1} - \tilde{X}_{t+1|t})^2\right) = E(Z_{t+1}^2) = \sigma_Z^2$$

H step forecast for h 2 is

$$Var(e_t(h)) = \sigma_Z^2(1 + \psi_1^2 + \psi_2^2 + \cdots \psi_{h-1}^2)$$

Model:

$$(1-B)(1-B^{12})X_t = (1+\theta B)(1+\theta_{12}B^{12})Z_t$$

also defined by

$$(X_t - X_{t-12}) - (X_{t-1} - X_{t-13}) = Z_t + \theta_1 Z_{t-1} + \theta_{12} Z_{t-12} + \theta_1\theta_{12} Z_{t-13}$$

Hence,

$$X_{t+1} = X_t + (X_{t-11} - X_{t-12} + Z_{t+1} + \theta_1 Z_{t-11} + \theta_1\theta_{12} Z_{t-12})$$

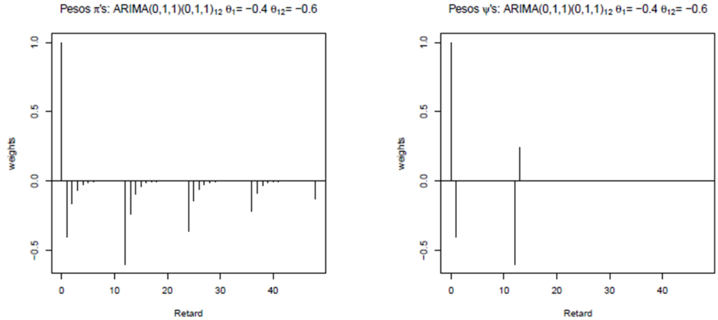
The forecasting function:

$$\tilde{X}_{t+1|t} = X_t + (X_{t-11} - X_{t-12}) + \theta_1 Z_t + \theta_1 2 Z_{t-11} + \theta_1\theta_{12} Z_{t-12}$$

$$\tilde{X}_{t+2|t} = \tilde{X}_{t+1|t} + (X_{t-10} - X_{t-11}) + \theta_{12} Z_{t-10} + \theta_1\theta_{12} Z_{t-11}$$

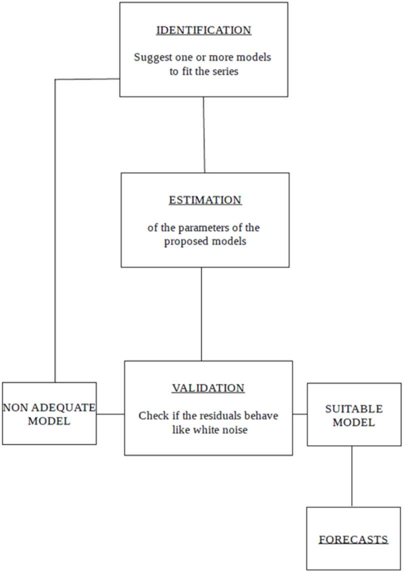
$$\ldots$$

$$\tilde{X}_{t+h|t} = \tilde{X}_{t+h-1|t} + (\tilde{X}_{t+h-12|t} - \tilde{X}_{t+h-13|t})\ ,\ h > 13$$



Weights π and ψ for the model ARIMA(0,1,1)(0,1,1)$_{12}$ for the parameters above

Esquema


IDENTIFICATION
Suggest one or more models
to fit the series
ESTIMATION
of the parameters of the
proposed models
NON ADEQUATE
MODEL
VALIDATION
Check if the residuals behave
like white noise
SUITABLE
MODEL
FORECASTS