# Time Series
# 5. Outliers treatment, Calendar Effects and Intervention Analysis

Josep A. Sanchez-Espigares

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

# Transfer function (Regression with ARMA errors)

ARIMA models with eXogenous variables (ARIMAX)

- $Y_t$ observed series (output)
- $X_{i,t}$ exogenous variables (input)
- $\tilde{Y}_t$ series without the effect of exogenous variables
- Estimate $\beta_i$ with OLS, the residuals are the $\tilde{Y}_t$ series (beware spurious relationships!)

$$Y_t = \sum_{i=1}^{h} \beta_i X_{i,t} + \tilde{Y}_t$$

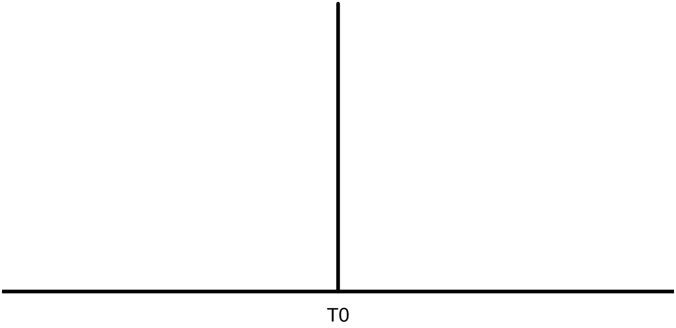
- Estimate an ARMA model for the $\tilde{Y}_t$ series

$$\phi(B)(Y_t - \sum_{i=1}^{h} \beta_i X_{i,t}) = \theta(B) Z_t$$

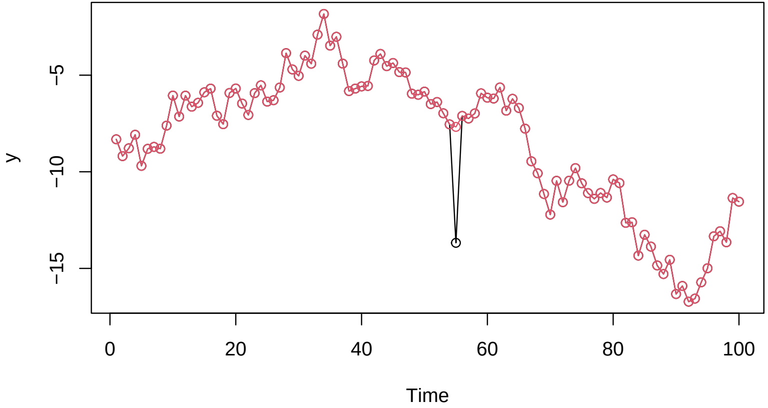
# Additive Outlier (AO)

It affects on one period.

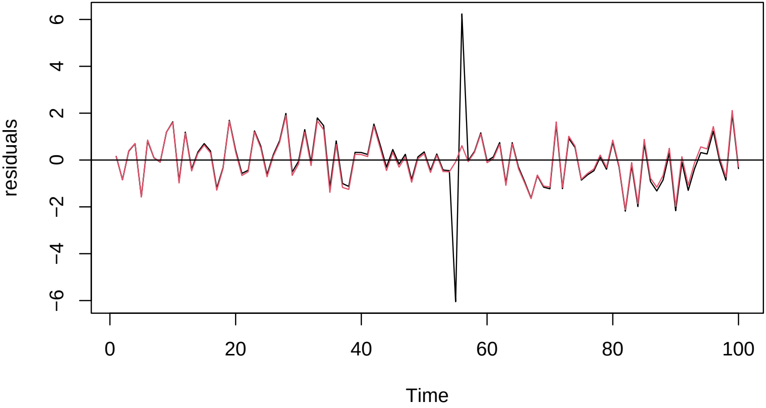
Transfer function: Pulse ($X_t = \mathbf{1}_{t=T0}(t)$)



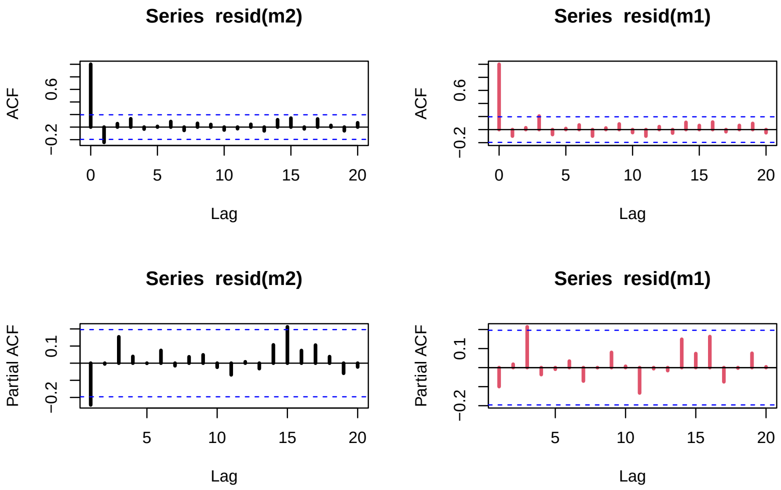
# Example



# Residuals of linear and observed series



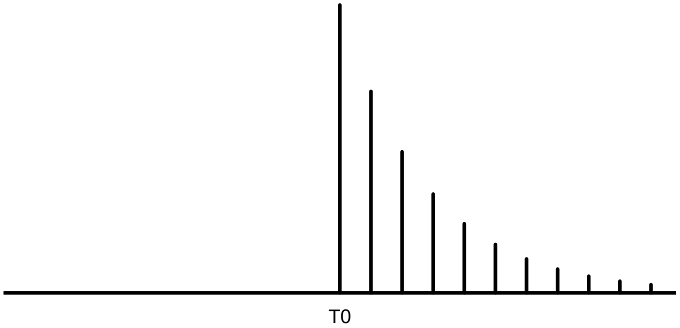
# ACF/PACF of linear and observed series



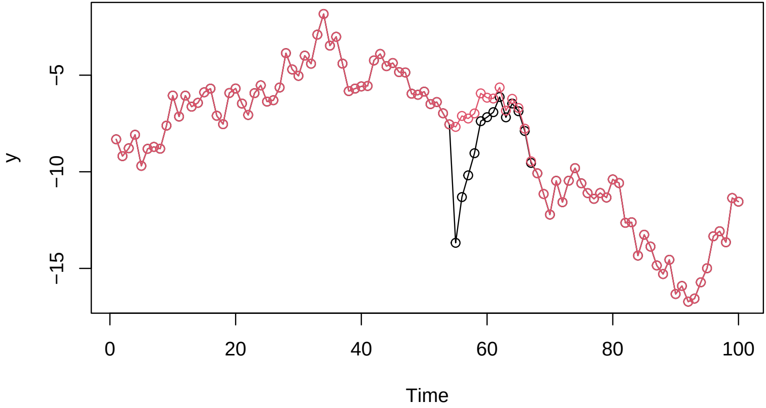
# Transitory Change (TC)

It affects on one period and its effect decreases in the next periods.

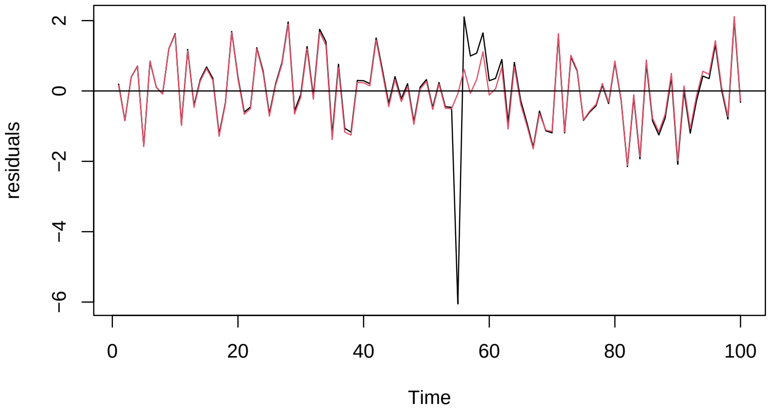
Transfer function: Exponential decreasing with $\delta = 0.7$ ($X_t = \delta^{(t-T0)}\mathbf{1}_{t \geq T0}(t)$)



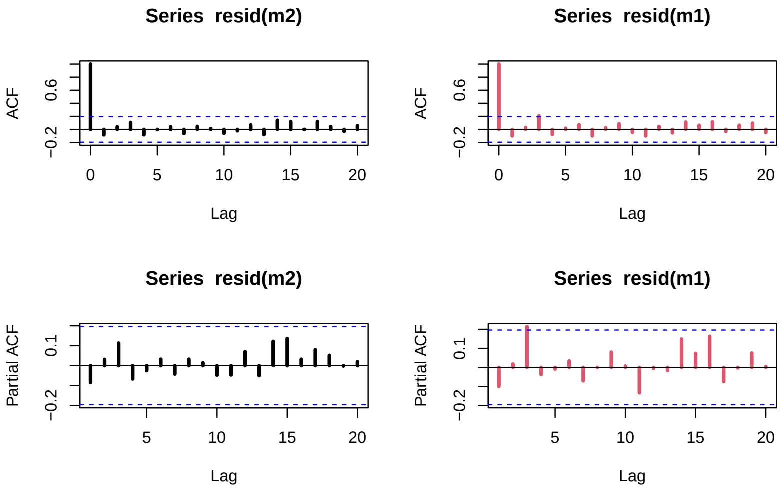
# Example



# Residuals of linear and observed series



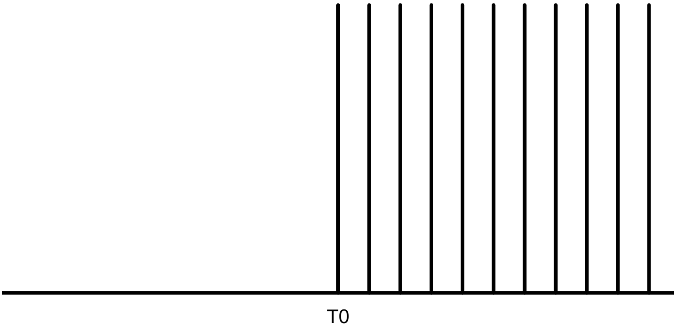
# ACF/PACF of linear and observed series



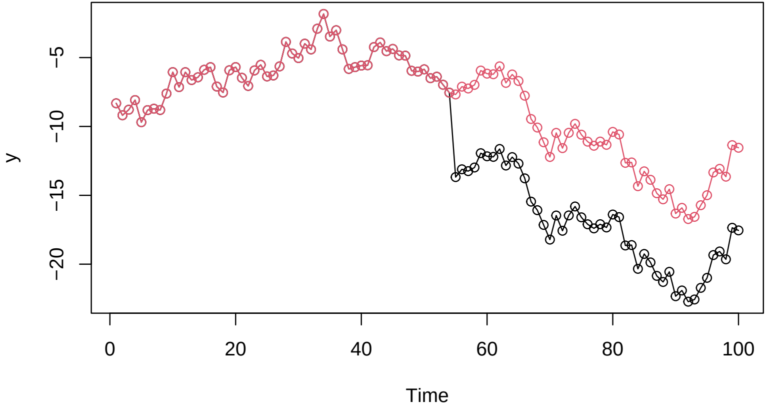
# Level Shift (LS)

It affects on one period and its effect remains in the next periods.

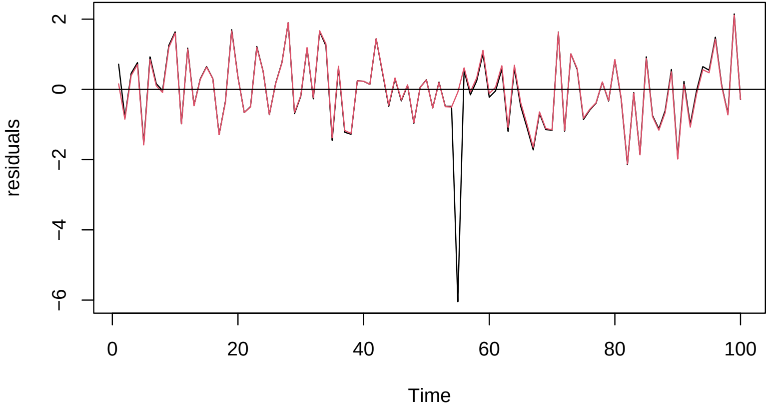
Transfer function: Step ($X_t = \mathbf{1}_{t \geq T0}(t)$)



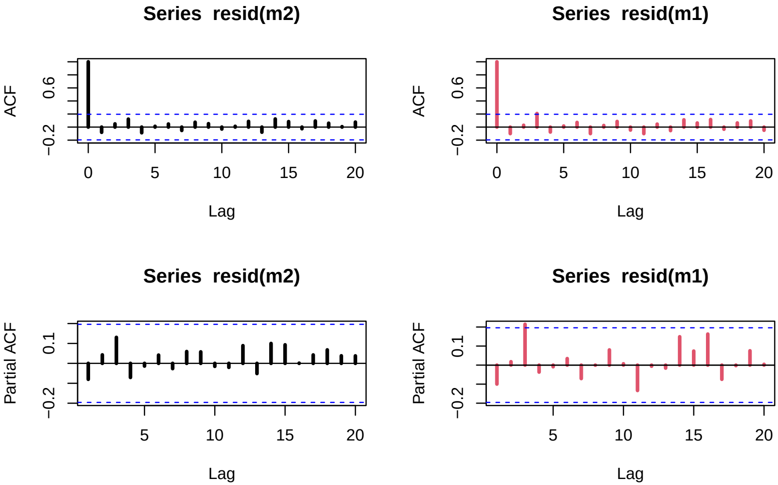
# Example



## Residuals of linear and observed series



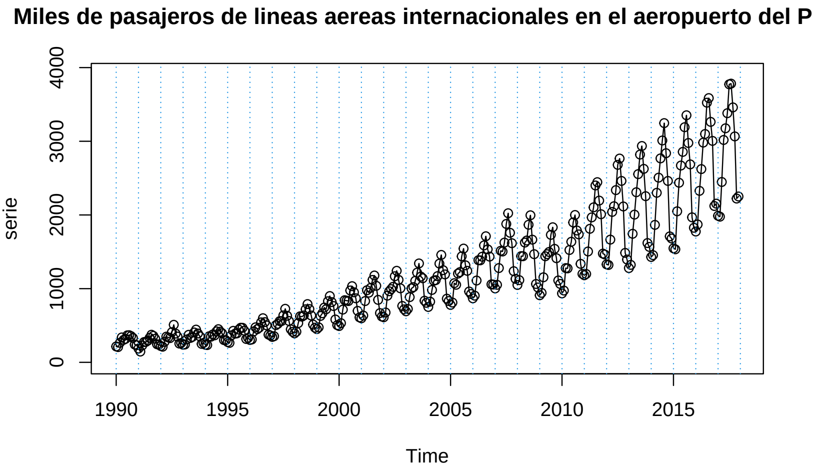
# ACF/PACF of linear and observed series



For the residual with the higher value over a given threshold:

- Detection of outliers, based on a significance test for the 3 types of outliers (AO, TC ans LS)
- Estimation the effect of the most significant type
- Linearized series by removing the outlier
- Repeat the process until all the residuals lie among the threshold

Monthly passengers (in thousands) of international air flights at El Prat (BCN).
Source: Ministry of Public Works of Spain (http://www.fomento.es)



# ARIMA model

```
##
## Call:
## arima(x = lnserie, order = c(0, 1, 1), seasonal = list(order = c(2,
##
## Coefficients:
##           ma1     sar1     sar2
##       -0.3741  -0.6344  -0.4279
## s.e.   0.0566   0.0567   0.0564
##
## sigma^2 estimated as 0.002331:  log likelihood = 516.81,  aic = -102
```

$$(1+0.634B^{12}+0.428B^{24})(1-B)(1-B^{12})\log X_t = (1-0.374B)Z_t$$

$$Z_t \sim N(0, \sigma_Z^2 = 0.00233)$$

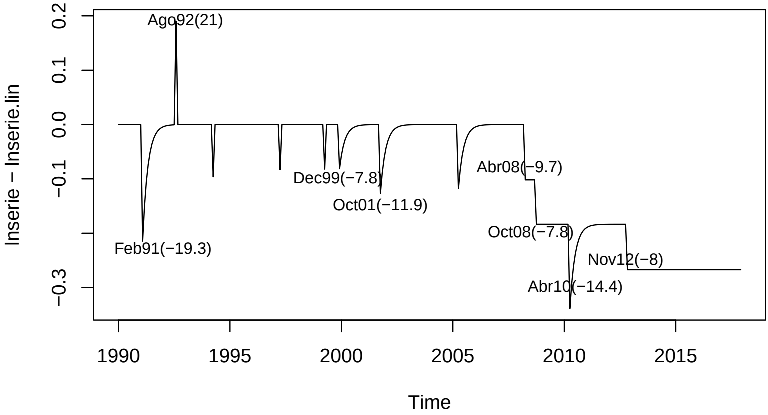
# Outlier detection

```
##     Obs type_detected     W_coeff ABS_L_Ratio
## 1   14             TC -0.21465793    5.997766
## 2   32             AO  0.19056486    6.298607
## 3  244             TC -0.15519774    4.735810
## 4  142             TC -0.12686443    3.960561
## 5   52             AO -0.09618053    3.417006
## 6  220             LS -0.10184648    3.285543
## 7  184             TC -0.11793107    3.889489
## 8   88             AO -0.08320794    3.114094
## 9  112             AO -0.08206789    3.115474
## 10 275             LS -0.08368511    2.874148
## 11 226             LS -0.08167645    2.838597
## 12 120             TC -0.08133733    2.861306
```

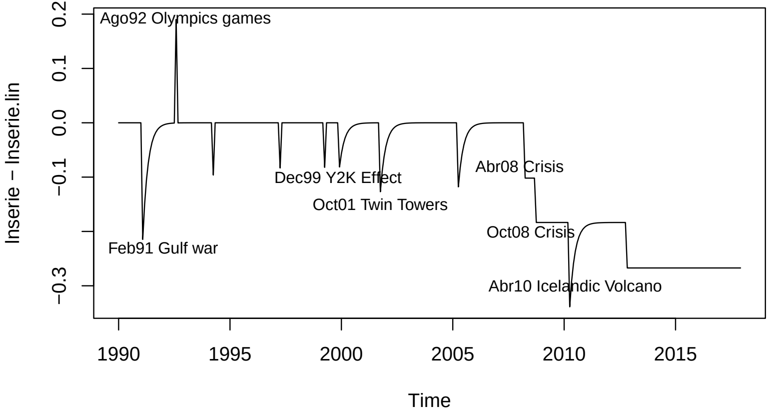
# Chronology

```
##     Obs Type    W_coeff    tStat    Fecha   Eff
## 1   14   TC -0.21465793 5.997766 Feb 1991 -19.3
## 2   32   AO  0.19056486 6.298607 Ago 1992  21.0
## 5   52   AO -0.09618053 3.417006 Abr 1994  -9.2
## 8   88   AO -0.08320794 3.114094 Abr 1997  -8.0
## 9  112   AO -0.08206789 3.115474 Abr 1999  -7.9
## 12 120   TC -0.08133733 2.861306 Dic 1999  -7.8
## 4  142   TC -0.12686443 3.960561 Oct 2001 -11.9
## 7  184   TC -0.11793107 3.889489 Abr 2005 -11.1
## 6  220   LS -0.10184648 3.285543 Abr 2008  -9.7
## 11 226   LS -0.08167645 2.838597 Oct 2008  -7.8
## 3  244   TC -0.15519774 4.735810 Abr 2010 -14.4
## 10 275   LS -0.08368511 2.874148 Nov 2012  -8.0
```

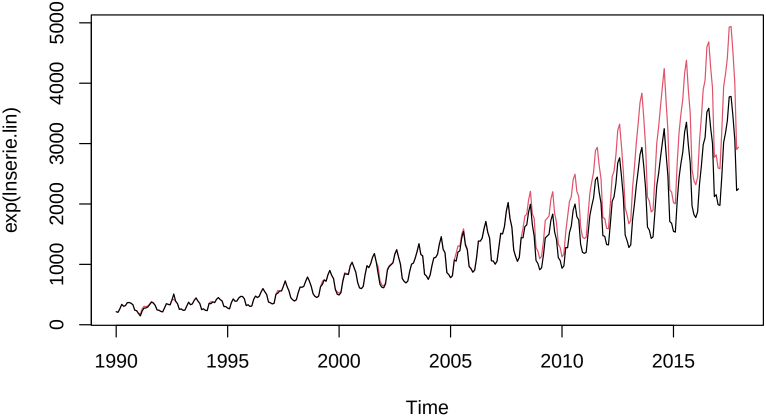
# Outlier Effects



# Outlier Effects



# Comparison of observed and linearized series



# Linearized series Model

```
##
## Call:
## arima(x = lnserie.lin, order = c(0, 1, 1), seasonal = list(order = c
##     period = 12))
##
## Coefficients:
##           ma1     sar1     sar2
##       -0.4635  -0.5759  -0.3781
## s.e.   0.0577   0.0569   0.0554
##
## sigma^2 estimated as 0.001272:  log likelihood = 615.35,  aic = -122
```

$$\log Xlin_t = \log X_t - \sum_{i=1}^{m} \omega_i \mathbf{1}_{t=t_i}^{Type}(i)$$

$$(1 + 0.576B^{12} + 0.378B^{24})(1 - B)(1 - B^{12}) \log Xlin_t = (1 - 0.464B)Z_t$$

$$Z_t \sim N(0, \sigma_Z^2 = 0.00127)$$

# Calendar Effects

- These calendar effects are only applied for monthly series
- Each month has the same number of days (except February in leap years!)
- There are some configurations in the month that can affect the phenomenon measured by the time series (Calendar effects)
- These configurations are known, because of the Calendar
- Only two situations considered: Easter and Trading Days

# Calendar Effects. Easter effect

- Easter some year happens in April, some other in March and sometimes between the two months.
- If the series is affected by the Easter week, predictions would fail to reflect those changes
- We change the series in order to deal with an ideal situation: "Half Easter time will always happens in both months"
- So, for those years with Easter totally in March, we move half of the effect to April, and vice versa.

# Calendar Effects. Easter effect: $Ea_t$ auxiliar series

```
##        Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
## 1990  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 1991  0.00  0.00  0.50 -0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 1992  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 1993  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 1994  0.00  0.00  0.17 -0.17  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 1995  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 1996  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 1997  0.00  0.00  0.50 -0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 1998  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 1999  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2000  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2001  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2002  0.00  0.00  0.50 -0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2003  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2004  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2005  0.00  0.00  0.50 -0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2006  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2007  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2008  0.00  0.00  0.50 -0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2009  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2010  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2011  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2012  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2013  0.00  0.00  0.50 -0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2014  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2015  0.00  0.00 -0.17  0.17  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2016  0.00  0.00  0.50 -0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
## 2017  0.00  0.00 -0.50  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
```

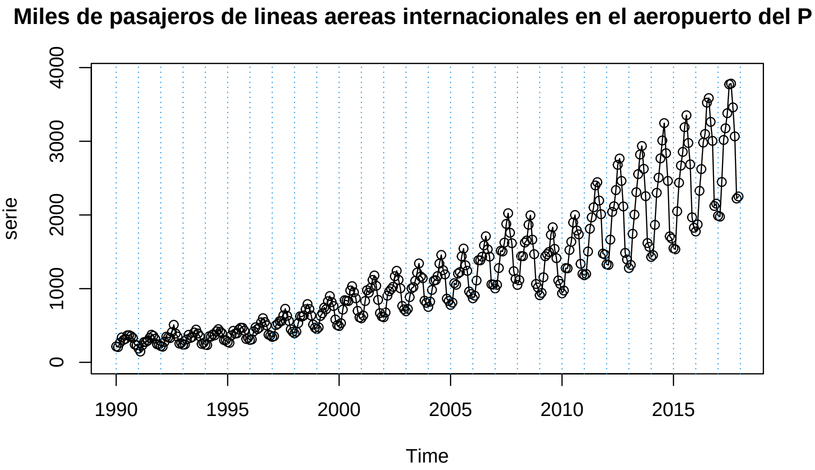
# Calendar Effects. Trading Days effect

- When looking a month across the years, the proportion of tranding days and weekend days changes
- If the series is affected by the number of Trading Days, predictions would fail to reflect those changes
- We change the series in order to deal with an ideal situation: "The proportion of Trading Days/Weekends in all months will be always 5/2"
- So, for those months with a proportion of trading days vs. weekends that does not fit the condition, we add/substract days to impose this condition

# Calendar Effects. Trading Days effect: $Td_t$ auxiliar series

```
##       Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
## 1990  3.0  0.0 -0.5 -1.5  3.0 -1.5 -0.5  3.0 -5.0  3.0  2.0 -4.0
## 1991  3.0  0.0 -4.0  2.0  3.0 -5.0  3.0 -0.5 -1.5  3.0 -1.5 -0.5
## 1992  3.0 -2.5 -0.5  2.0 -4.0  2.0  3.0 -4.0  2.0 -0.5 -1.5  3.0
## 1993 -4.0  0.0  3.0  2.0 -4.0  2.0 -0.5 -0.5  2.0 -4.0  2.0  3.0
## 1994 -4.0  0.0  3.0 -1.5 -0.5  2.0 -4.0  3.0  2.0 -4.0  2.0 -0.5
## 1995 -0.5  0.0  3.0 -5.0  3.0  2.0 -4.0  3.0 -1.5 -0.5  2.0 -4.0
## 1996  3.0  1.0 -4.0  2.0  3.0 -5.0  3.0 -0.5 -1.5  3.0 -1.5 -0.5
## 1997  3.0  0.0 -4.0  2.0 -0.5 -1.5  3.0 -4.0  2.0  3.0 -5.0  3.0
## 1998 -0.5  0.0 -0.5  2.0 -4.0  2.0  3.0 -4.0  2.0 -0.5 -1.5  3.0
## 1999 -4.0  0.0  3.0  2.0 -4.0  2.0 -0.5 -0.5  2.0 -4.0  2.0  3.0
## 2000 -4.0  1.0  3.0 -5.0  3.0  2.0 -4.0  3.0 -1.5 -0.5  2.0 -4.0
## 2001  3.0  0.0 -0.5 -1.5  3.0 -1.5 -0.5  3.0 -5.0  3.0  2.0 -4.0
## 2002  3.0  0.0 -4.0  2.0  3.0 -5.0  3.0 -0.5 -1.5  3.0 -1.5 -0.5
## 2003  3.0  0.0 -4.0  2.0 -0.5 -1.5  3.0 -4.0  2.0  3.0 -5.0  3.0
## 2004 -0.5 -2.5  3.0  2.0 -4.0  2.0 -0.5 -0.5  2.0 -4.0  2.0  3.0
## 2005 -4.0  0.0  3.0 -1.5 -0.5  2.0 -4.0  3.0  2.0 -4.0  2.0 -0.5
## 2006 -0.5  0.0  3.0 -5.0  3.0  2.0 -4.0  3.0 -1.5 -0.5  2.0 -4.0
## 2007  3.0  0.0 -0.5 -1.5  3.0 -1.5 -0.5  3.0 -5.0  3.0  2.0 -4.0
## 2008  3.0  1.0 -4.0  2.0 -0.5 -1.5  3.0 -4.0  2.0  3.0 -5.0  3.0
## 2009 -0.5  0.0 -0.5  2.0 -4.0  2.0  3.0 -4.0  2.0 -0.5 -1.5  3.0
## 2010 -4.0  0.0  3.0  2.0 -4.0  2.0 -0.5 -0.5  2.0 -4.0  2.0  3.0
## 2011 -4.0  0.0  3.0 -1.5 -0.5  2.0 -4.0  3.0  2.0 -4.0  2.0 -0.5
## 2012 -0.5  1.0 -0.5 -1.5  3.0 -1.5 -0.5  3.0 -5.0  3.0  2.0 -4.0
## 2013  3.0  0.0 -4.0  2.0  3.0 -5.0  3.0 -0.5 -1.5  3.0 -1.5 -0.5
## 2014  3.0  0.0 -4.0  2.0 -0.5 -1.5  3.0 -4.0  2.0  3.0 -5.0  3.0
## 2015 -0.5  0.0 -0.5  2.0 -4.0  2.0  3.0 -4.0  2.0 -0.5 -1.5  3.0
## 2016 -4.0  1.0  3.0 -1.5 -0.5  2.0 -4.0  3.0  2.0 -4.0  2.0 -0.5
## 2017 -0.5  0.0  3.0 -5.0  3.0  2.0 -4.0  3.0 -1.5 -0.5  2.0 -4.0
```

Monthly passengers (in thousands) of international air flights at El Prat (BCN).
Source: Ministry of Public Works of Spain (http://www.fomento.es)



```
##
## Call:
## arima(x = lnserie, order = c(0, 1, 1), seasonal = list(order = c(2,
##
## Coefficients:
##           ma1     sar1     sar2
##       -0.3741  -0.6344  -0.4279
## s.e.   0.0566   0.0567   0.0564
##
## sigma^2 estimated as 0.002331:  log likelihood = 516.81,  aic = -102
```

$$(1+0.634B^{12}+0.428B^{24})(1-B)(1-B^{12})\log X_t = (1-0.374B)Z_t$$

$$Z_t \sim N(0, \sigma_Z^2 = 0.00233)$$

# Expression of the linearized series

$$Xlin_t = X_t - \omega_{TD} TD_t - \omega_{Ea} Ea_t$$

## Testing for Calendar effects(I)

Model for the original series

```
##           ma1     sar1     sar2
## coef -0.3741 -0.6344 -0.4279
## s.e.  0.0566  0.0567  0.0564
```

```
## SigmaZ^2= 0.00233  AIC= -1025.624  BIC= -1010.513
```

Correction for Trading Days effect

```
##           ma1     sar1     sar2 wTradDays
## coef -0.3280 -0.6605 -0.4424   -0.0021
## s.e.  0.0596  0.0568  0.0556    0.0005
```

```
## SigmaZ^2= 0.0022  AIC= -1042.343  BIC= -1023.454
```

# Testing for Calendar effects(II)

Correction for Easter effect

```
##          ma1    sar1    sar2  wEast
## coef -0.3020 -0.4644 -0.3094 0.0679
## s.e.  0.0587  0.0586  0.0618 0.0082
```

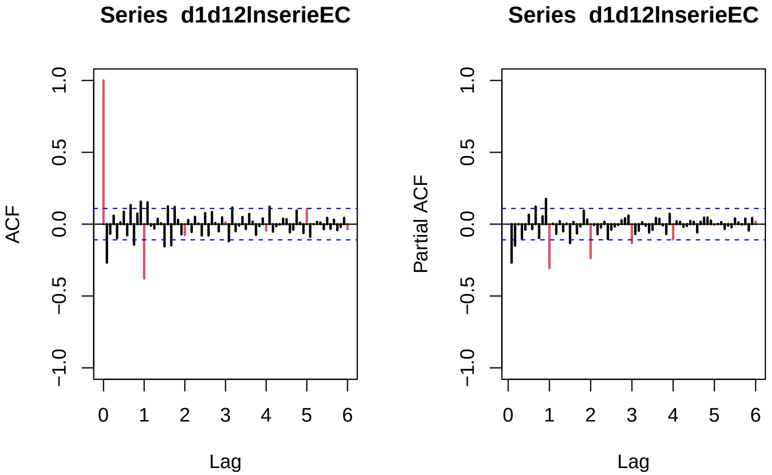
```
## SigmaZ^2= 0.00202  AIC= -1072.923  BIC= -1054.035
```

Correction for both effects

```
##          ma1    sar1    sar2 wTradDays  wEast
## coef -0.2667 -0.4863 -0.3173   -0.0019 0.0641
## s.e.  0.0607  0.0590  0.0608    0.0005 0.0081
```

```
## SigmaZ^2= 0.00193  AIC= -1085.002  BIC= -1062.337
```

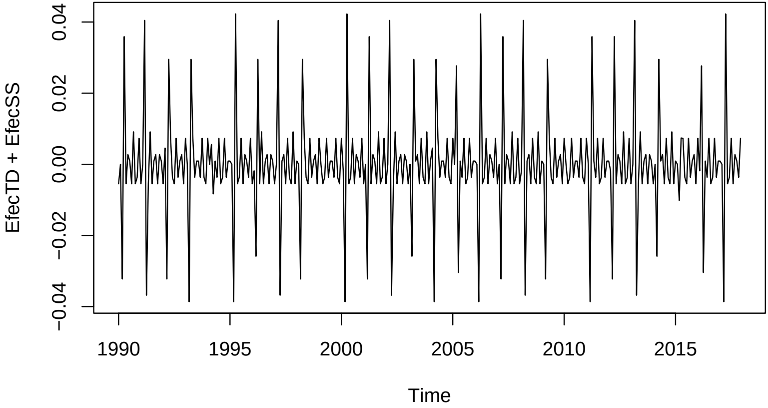
# Identifying the model for the linearized series




##Model for the linearized series

```
##          ma1    sma1 wTradDays  wEast
## coef -0.2492 -0.6136   -0.0018 0.0663
```

# Estimated calendar effects



# Estimated calendar effects

```
##          wTradDays wEast    serie  serieEC
## Jan 2016      -4.0   0.0 1774.953 1761.851
## Feb 2016       1.0   0.0 1870.961 1874.430
## Mar 2016       3.0   0.5 2327.124 2266.267
## Apr 2016      -1.5  -0.5 2622.604 2700.523
## May 2016      -0.5   0.0 2981.745 2978.985
## Jun 2016       2.0   0.0 3098.362 3109.862
## Jul 2016      -4.0   0.0 3523.483 3497.473
## Aug 2016       3.0   0.0 3585.878 3605.860
## Sep 2016       2.0   0.0 3263.092 3275.203
## Oct 2016      -4.0   0.0 3005.086 2982.903
## Nov 2016       2.0   0.0 2121.628 2129.502
## Dec 2016      -0.5   0.0 2153.470 2151.476
## Jan 2017      -0.5   0.0 1988.614 1986.773
## Feb 2017       0.0   0.0 1976.634 1976.634
## Mar 2017       3.0  -0.5 2447.478 2541.289
## Apr 2017      -5.0   0.5 3017.551 2895.413
## May 2017       3.0   0.0 3175.953 3193.651
## Jun 2017       2.0   0.0 3380.555 3393.102
## Jul 2017      -4.0   0.0 3773.045 3745.193
## Aug 2017       3.0   0.0 3781.934 3803.008
## Sep 2017      -1.5   0.0 3459.473 3449.874
```

# Testing for Intervention Analysis

Effect of some factor that could make some change in the series

Example: The new terminal (T1) was inaugurated on june 16th,2009

There is some effect in the series associated to this "intervention"?

We have to propose some transfer function to include to the model

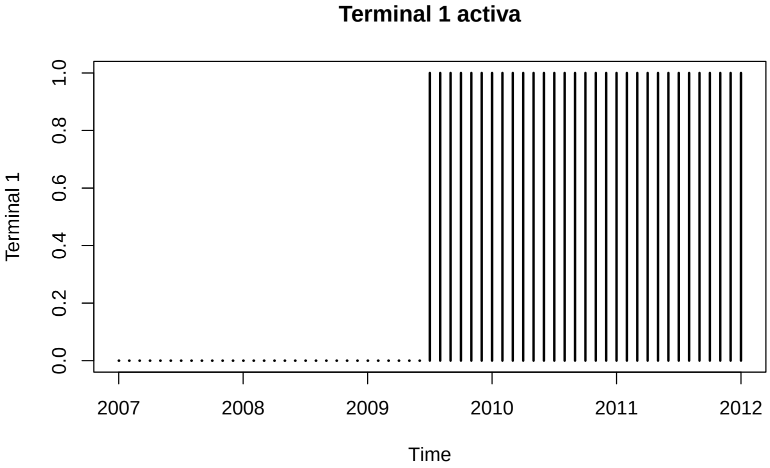
# Transfer function for Intervention Analysis

...

```
##      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 2006   0   0   0   0   0   0   0   0   0   0   0   0
## 2007   0   0   0   0   0   0   0   0   0   0   0   0
## 2008   0   0   0   0   0   0   0   0   0   0   0   0
## 2009   0   0   0   0   0   0   1   1   1   1   1   1
## 2010   1   1   1   1   1   1   1   1   1   1   1   1
## 2011   1   1   1   1   1   1   1   1   1   1   1   1
```

...

In this case, we propose an step function, assuming that the effect of this "intervention" has remained constant since the inauguration



# Testing the Intervention Analysis

We fit the model incorporating the transfer function as an extra "xreg"" term and test its significance

```
##
## Call:
## arima(x = lnserie, order = c(0, 1, 1), seasonal = list(order = c(0,
##     xreg = data.frame(wTradDays, wEast, T1))
##
## Coefficients:
##           ma1     sma1  wTradDays   wEast      T1
##       -0.2486  -0.6138    -0.0018  0.0663  0.0031
## s.e.   0.0613   0.0506     0.0005  0.0075  0.0379
##
## sigma^2 estimated as 0.001875:  log likelihood = 552.88,  aic = -109
```

In this case,

$$H_0 : \text{T1 does not affect the series (non-effective intervention)}$$
$$H_1 : \text{T1 affects the series (effective intervention)}$$

$$|\hat{t}| = |\frac{0.0031}{0.0379}| = 0.08179 < 2 \Rightarrow \text{Non-significant}$$

Box-Jenkins Model:

- AIC $ARIMA(0,1,1)(2,1,0)_{12}$ for $\log X_t = -1025.624$

Calendar Effects:

- AIC $ARIMA(0,1,1)(0,1,1)_{12}$ for $\log X_t^* = \log X_t - [(-0.0018)\,TD_t + 0.0663EA_t] = -1095.756$

Calendar Effects and Intervention Analysis (Effect “T1”):

- AIC $ARIMA(0,1,1)(0,1,1)_{12}$ for $\log X_t^* = \log X_t - [(-0.0018)\,TD_t + 0.0663EA_t + 0.0031\,T1_t] = -1093.76$

# Testing the Intervention Analysis

```
## exp(coef(mod1ECIA)["T1"])-1= 0.003092834
```