Time Series
5. Outliers treatment in time series

Josep A. Sanchez-Espigares

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

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

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 β_i with OLS, the residuals are the $\tilde{Y}_t$ series (beware spurious relationships!)

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

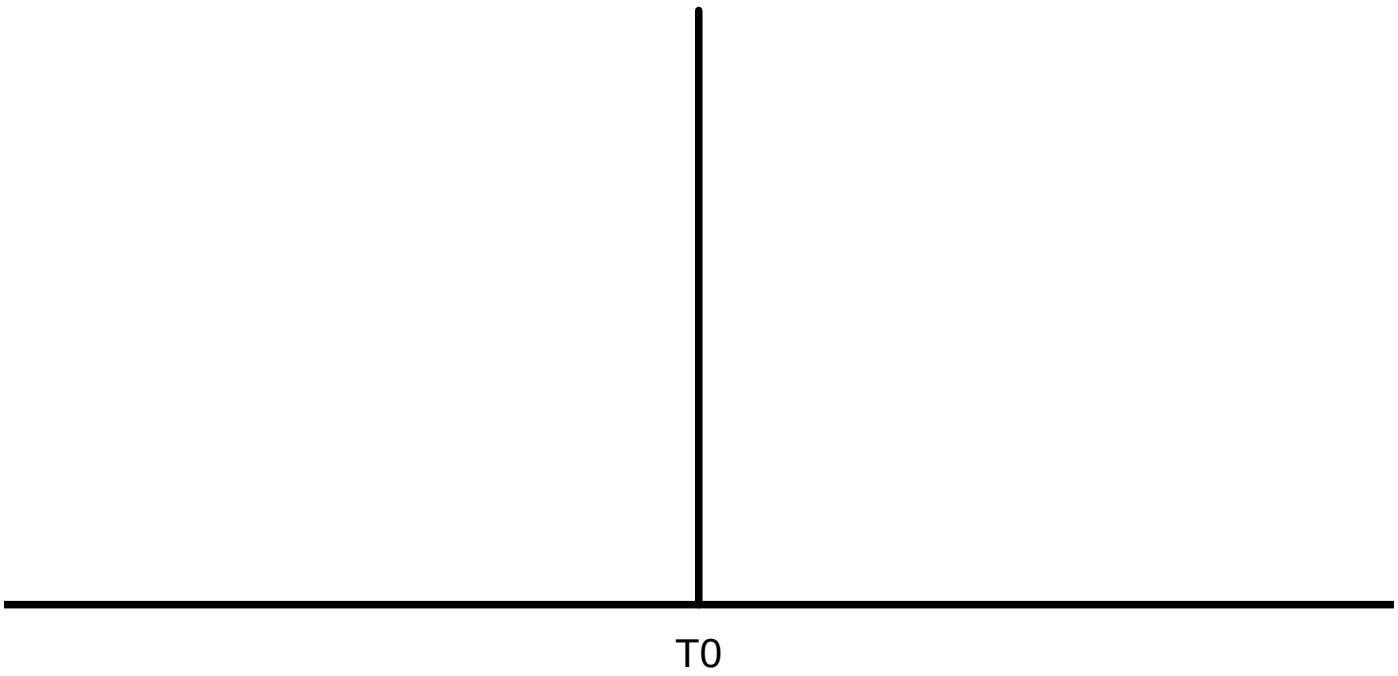
- Estimate an ARMA model for the $\tilde{Y}_t$ series

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

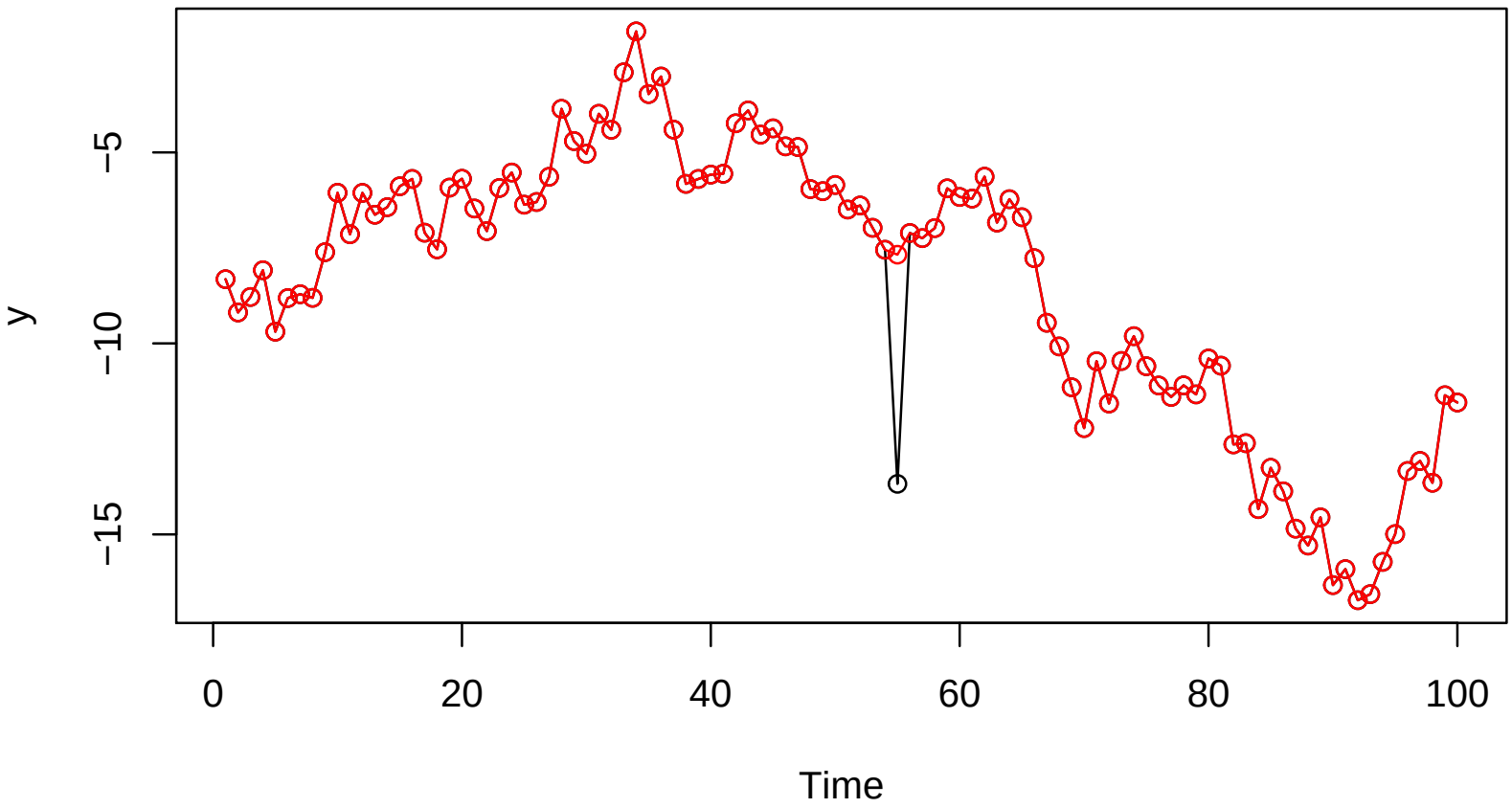
Additive Outlier (AO)

It affects on one period.

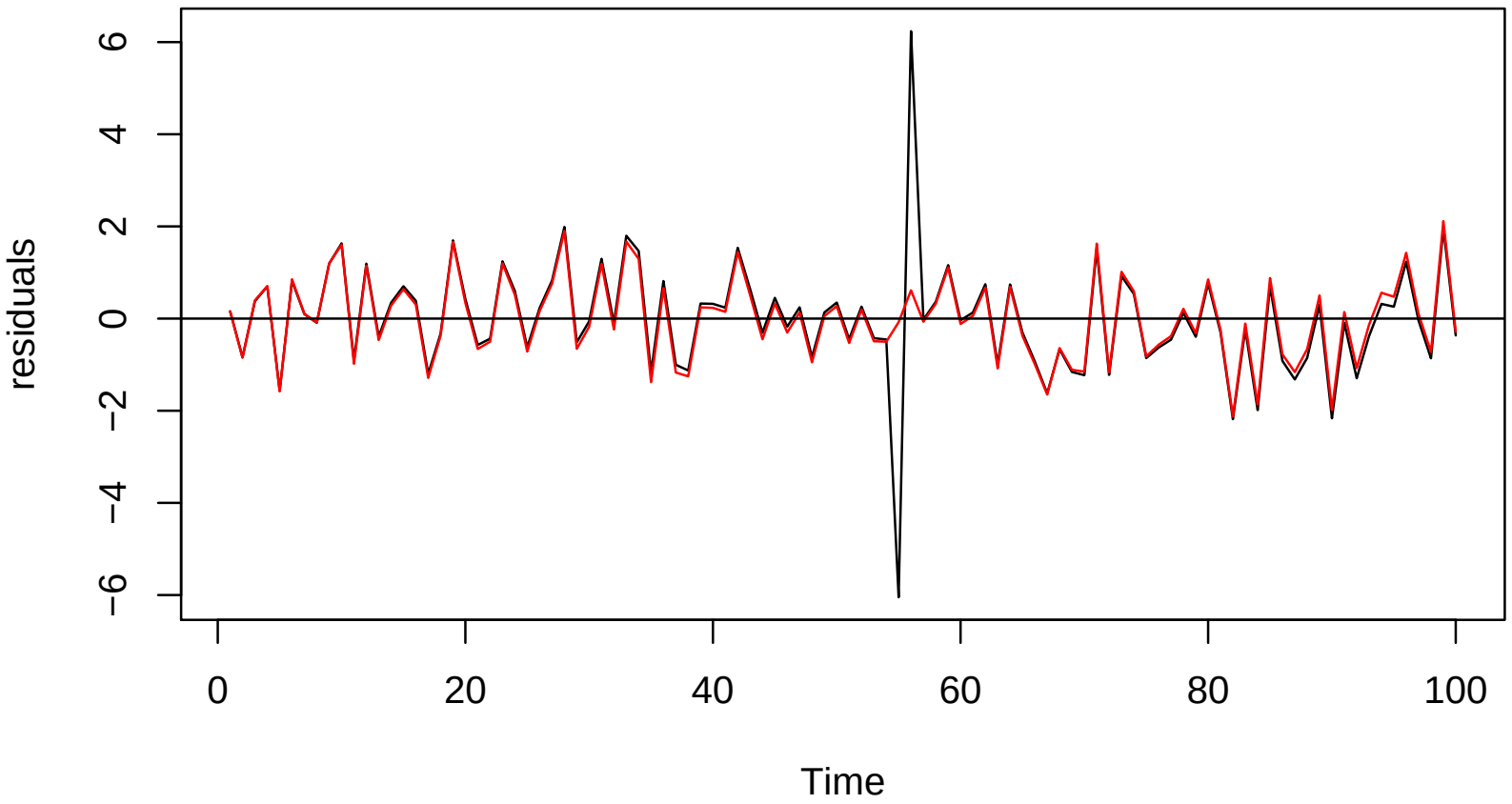
Transfer function: Pulse ($X_t = \mathbf{1}_{t=T0}(t)$)



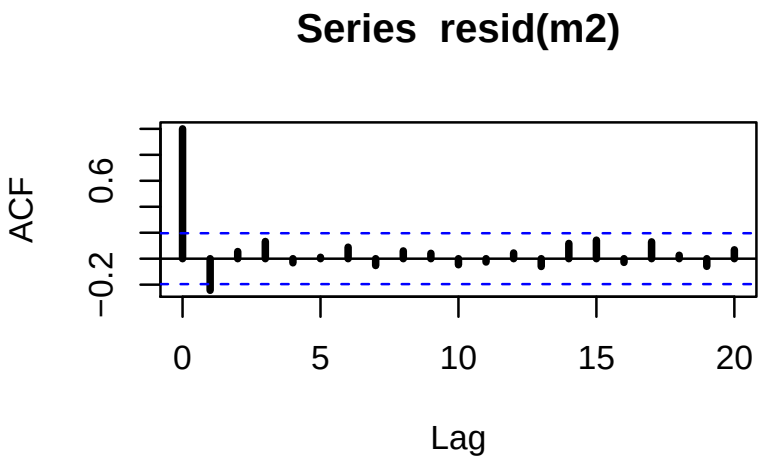
Example

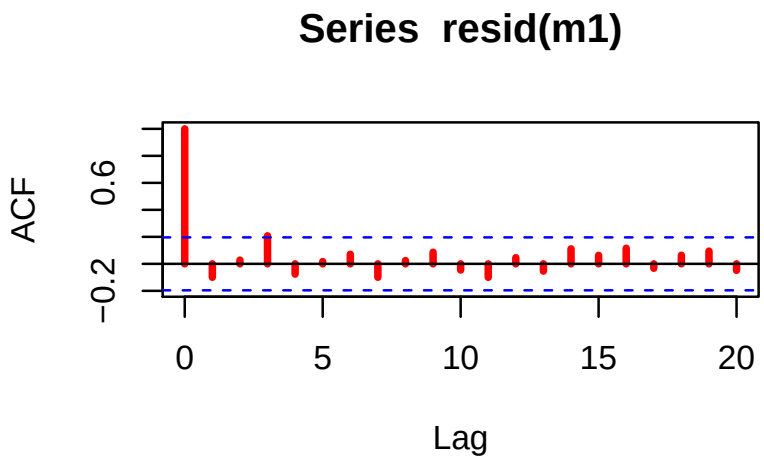


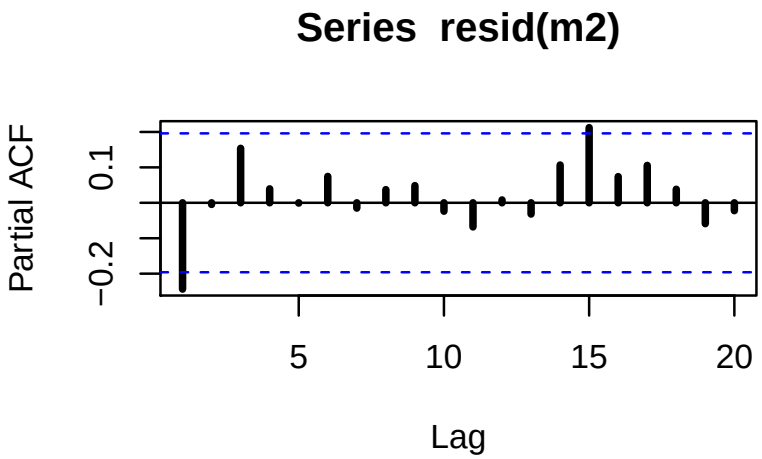
Residuals of linear and observed series

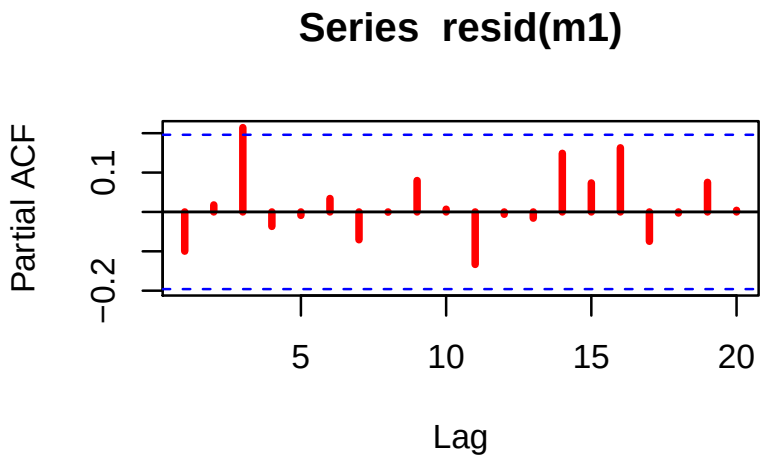


ACF/PACF of linear and observed series





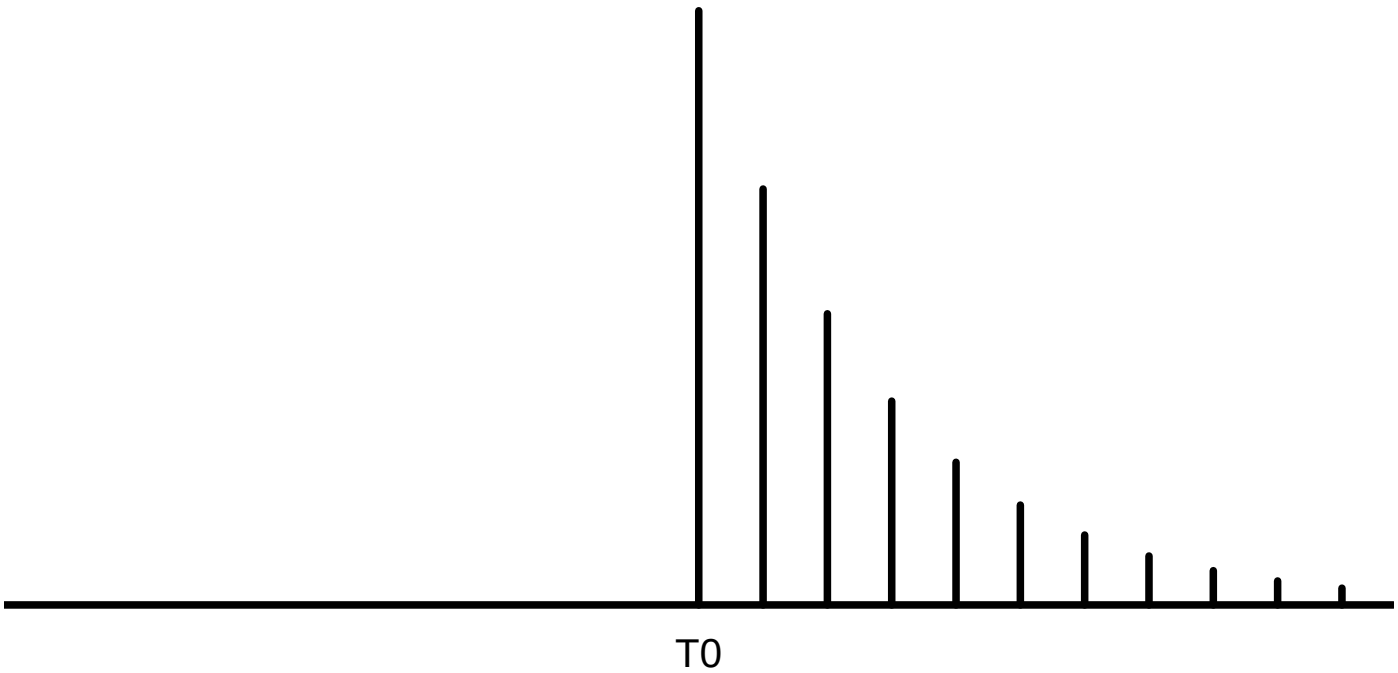




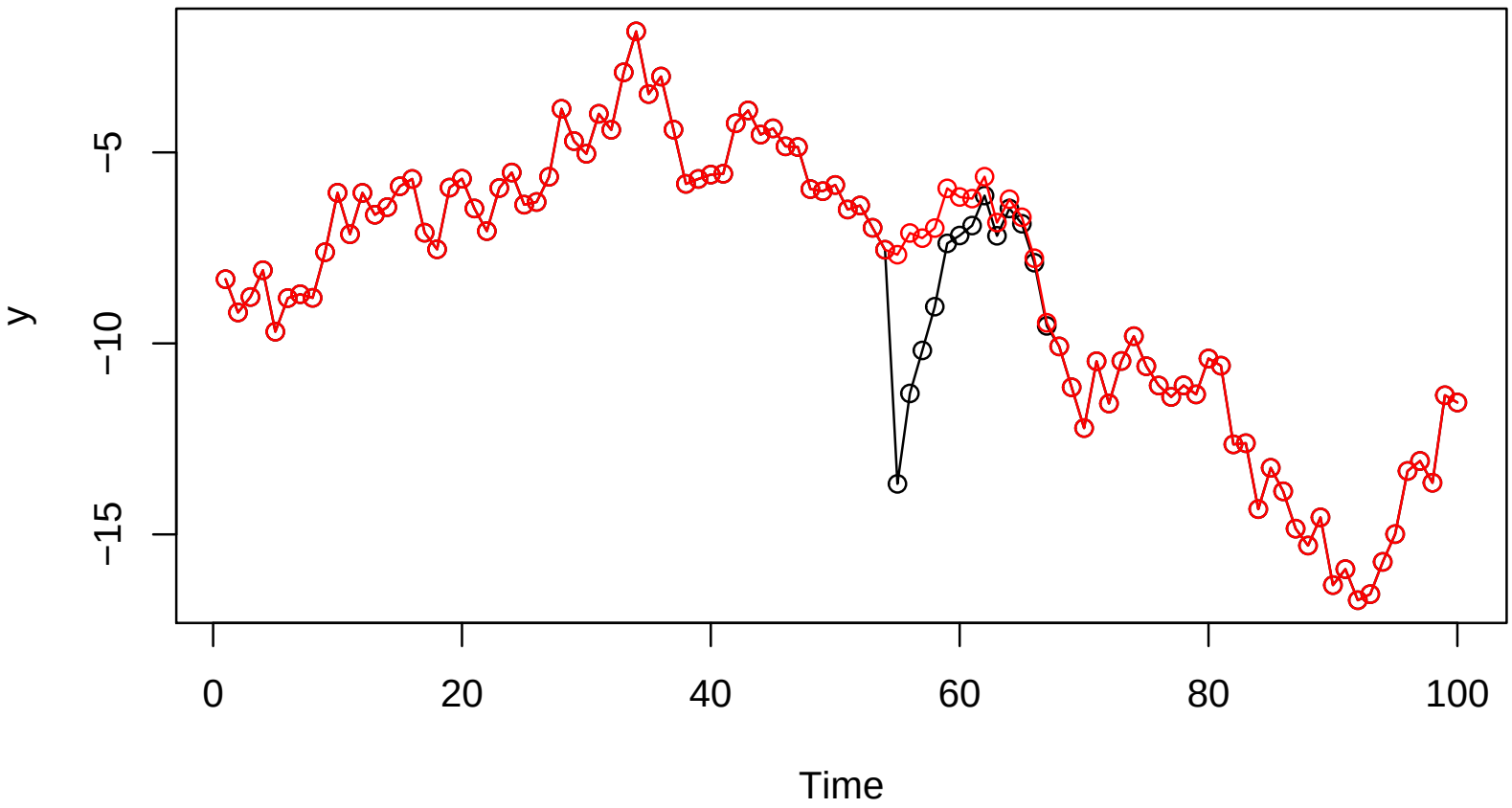
Transitory Change (TC)

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

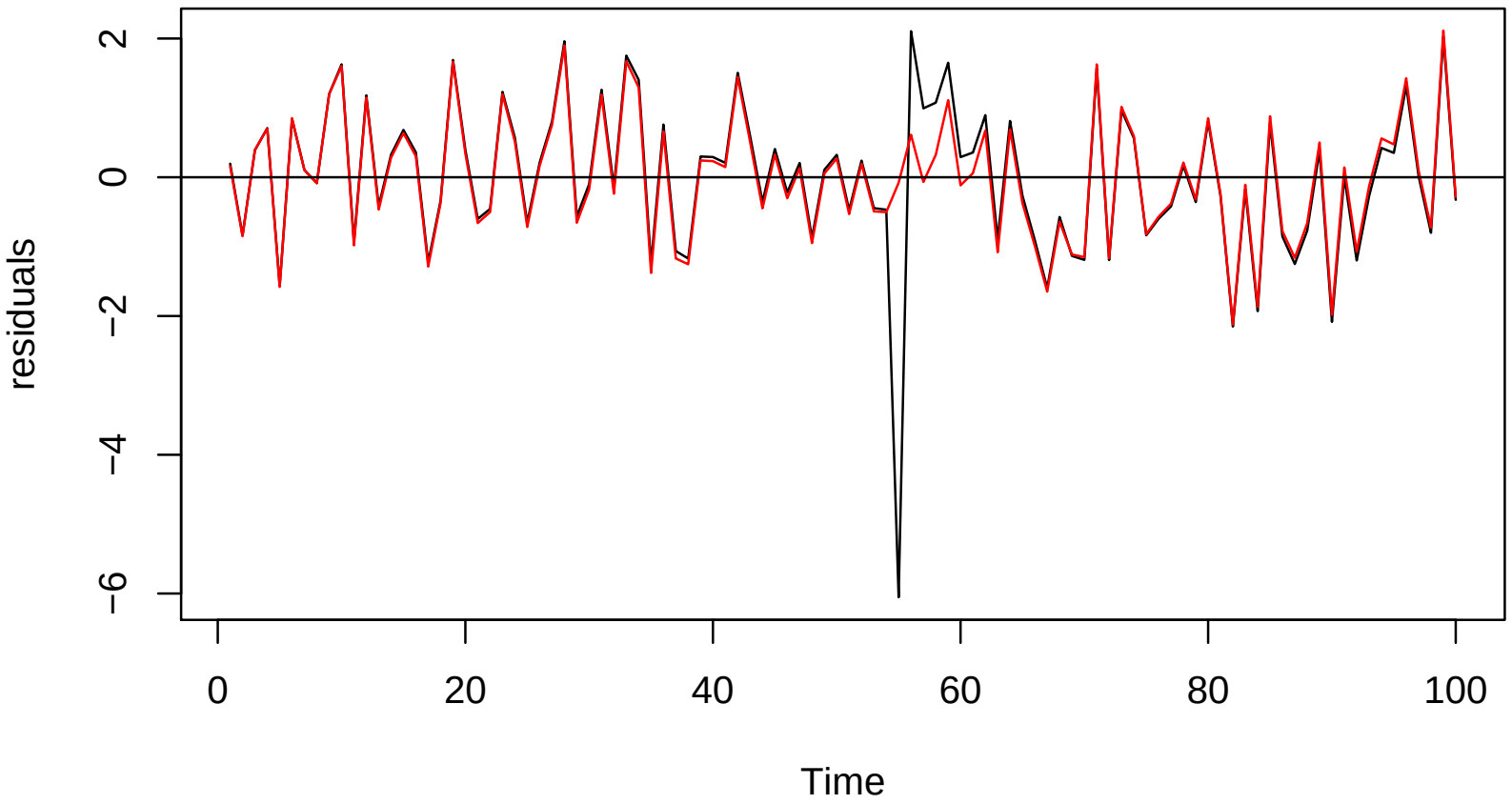
Transfer function: Exponential decreasing with $\delta = 0.7$ ($X_t = \delta^{(t-T0)}\mathbf{1}_{t \geq T0}(t)$)



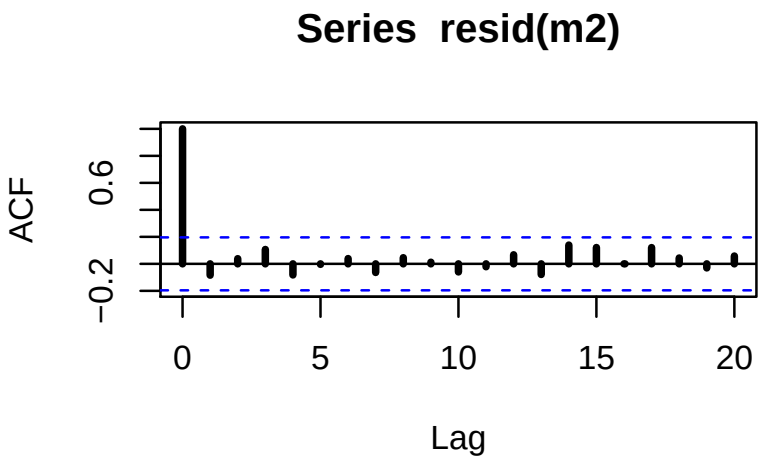
Example

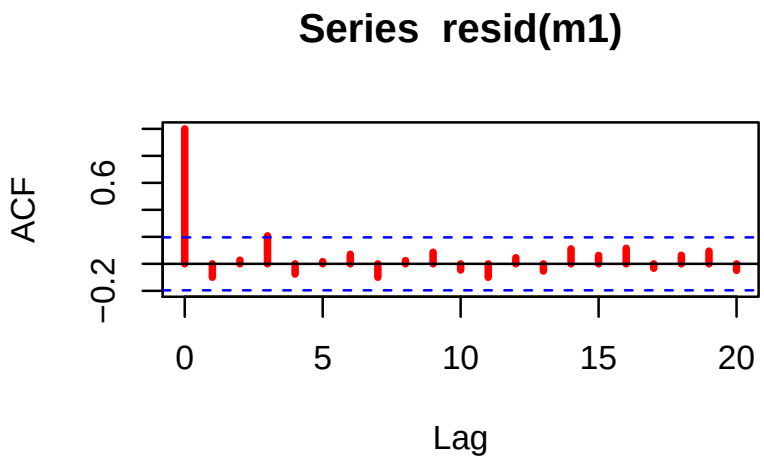


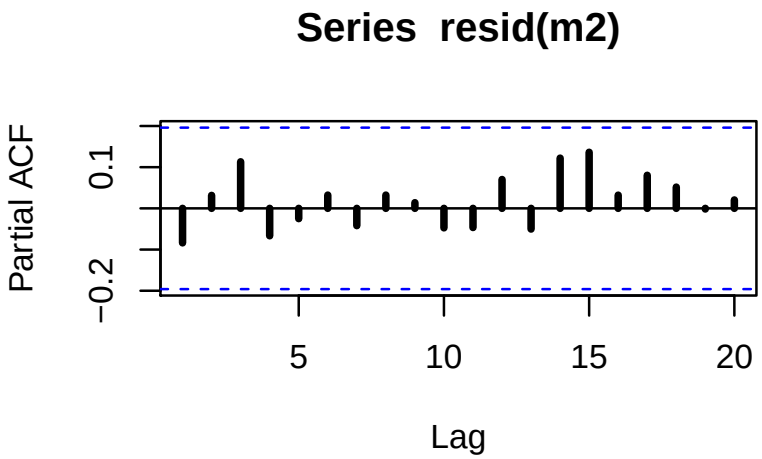
Residuals of linear and observed series

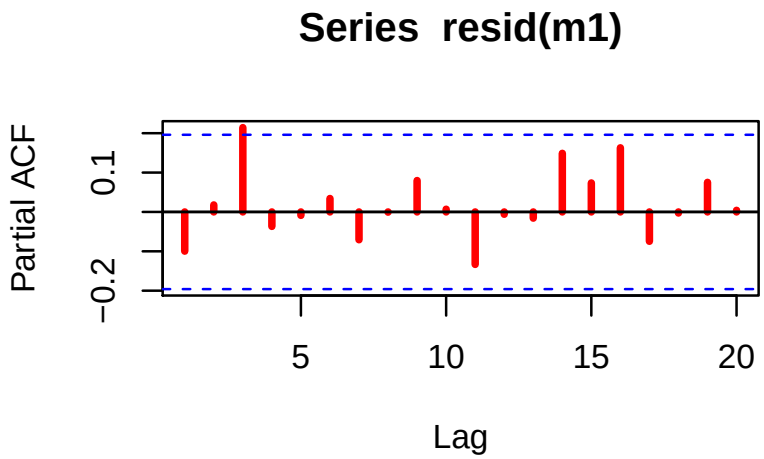


ACF/PACF of linear and observed series





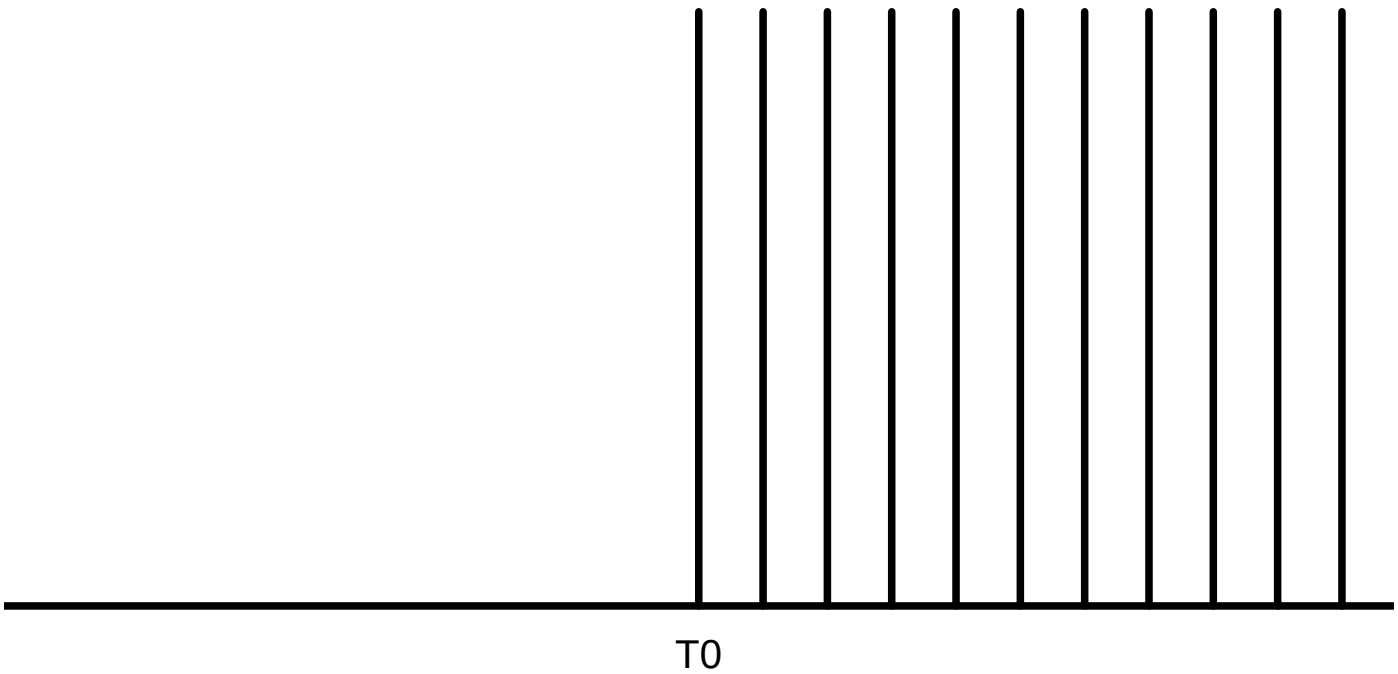




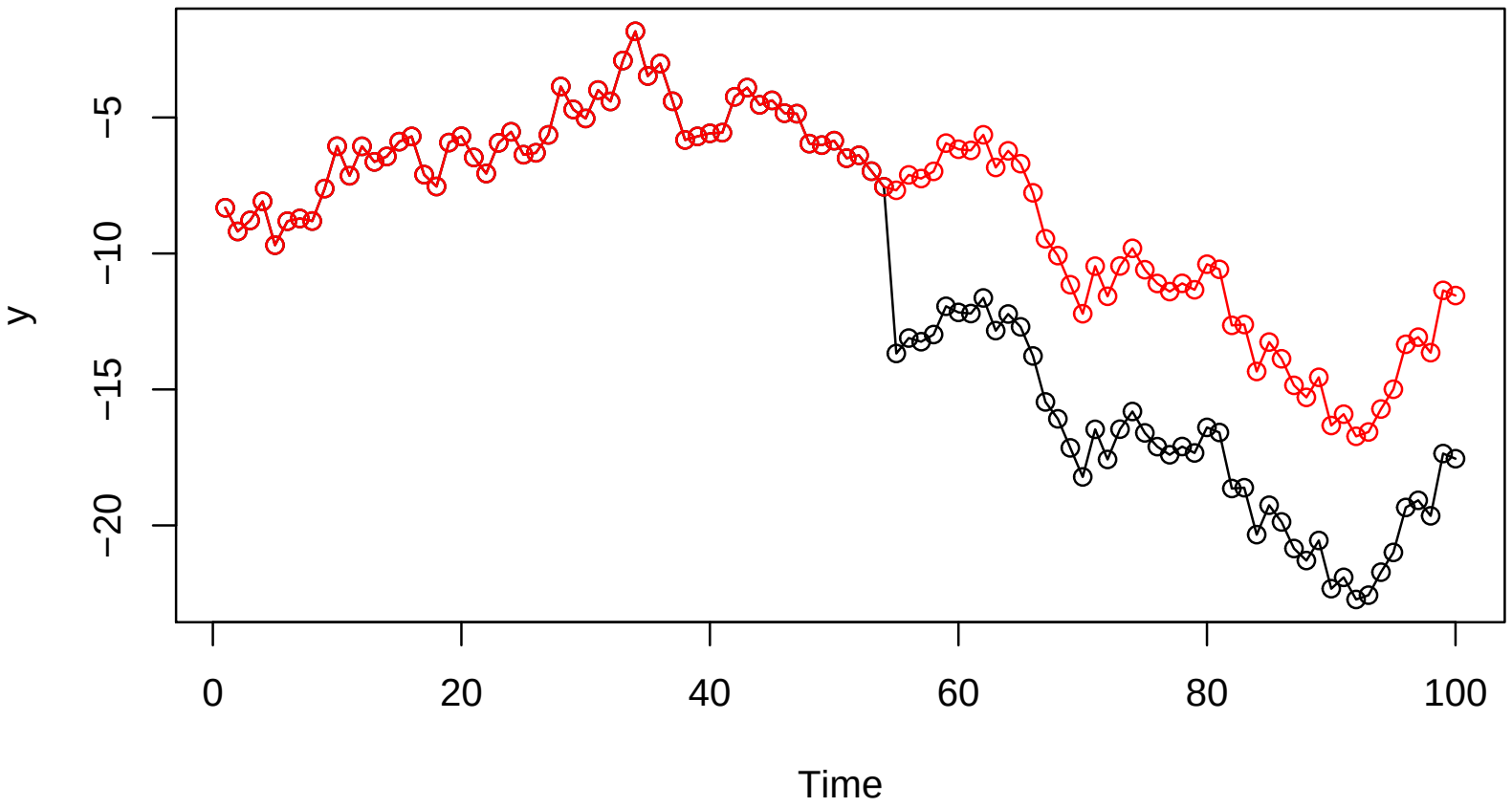
Level Shift (LS)

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

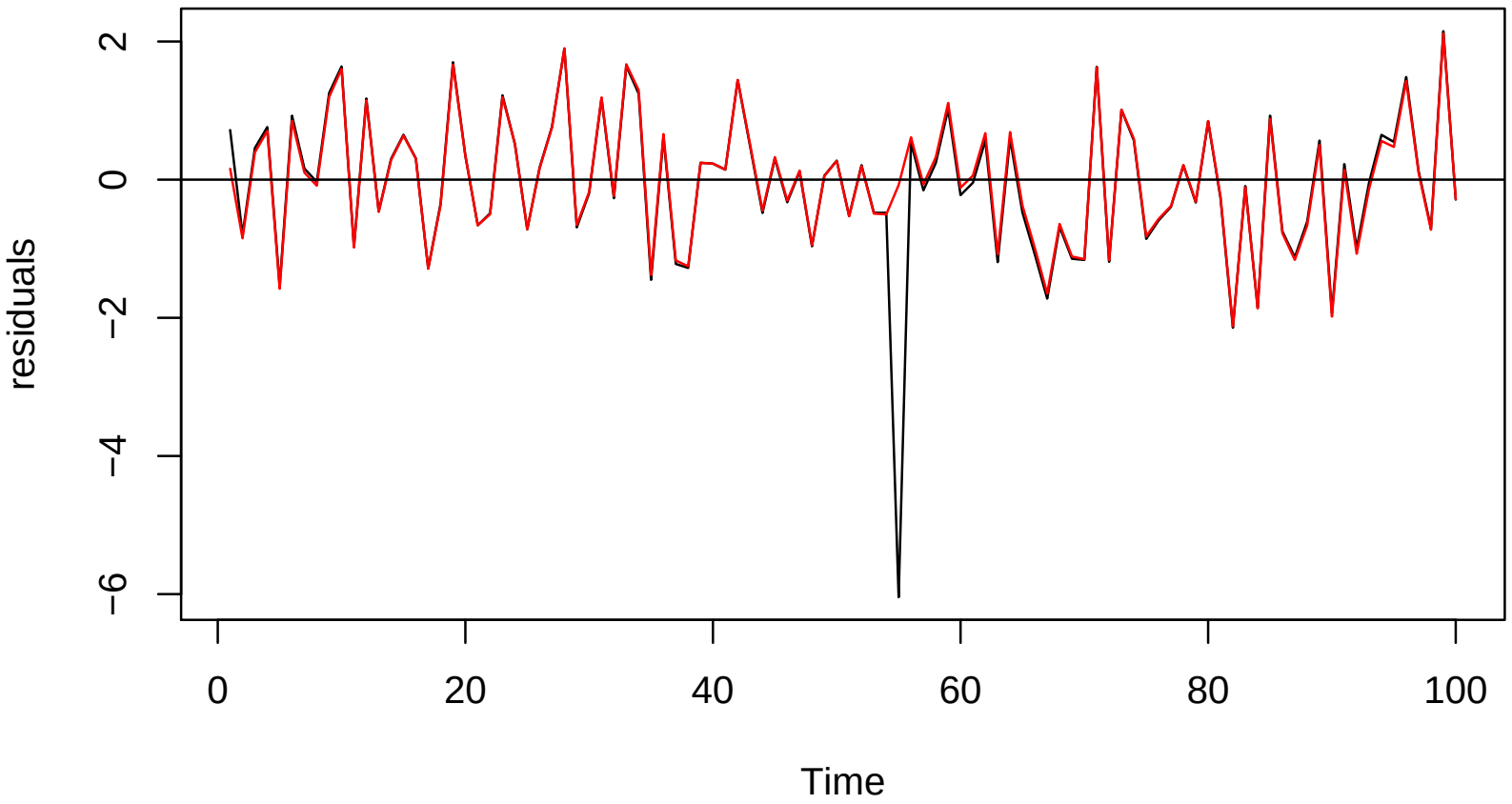
Transfer function: Step ($X_t = \mathbf{1}_{t \geq T0}(t)$)



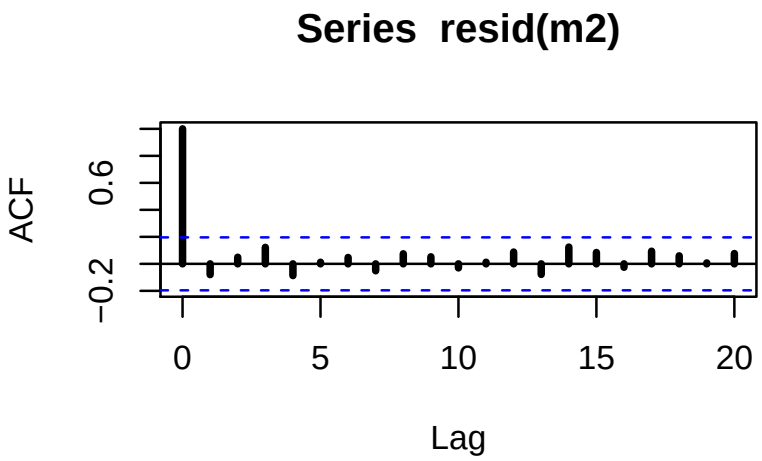
Example

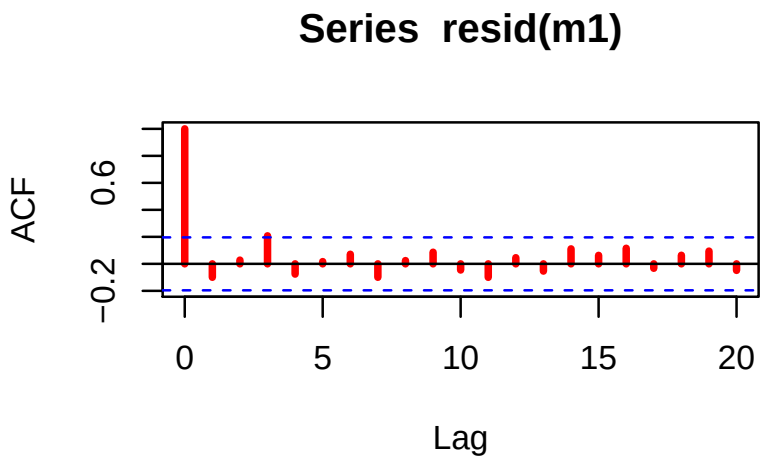


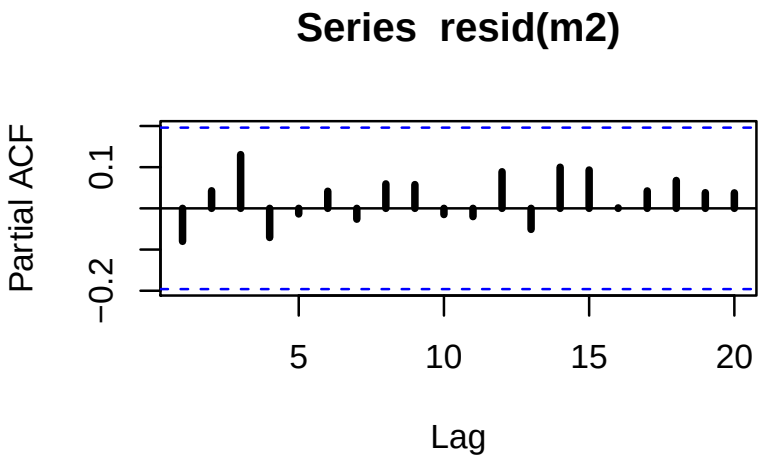
Residuals of linear and observed series

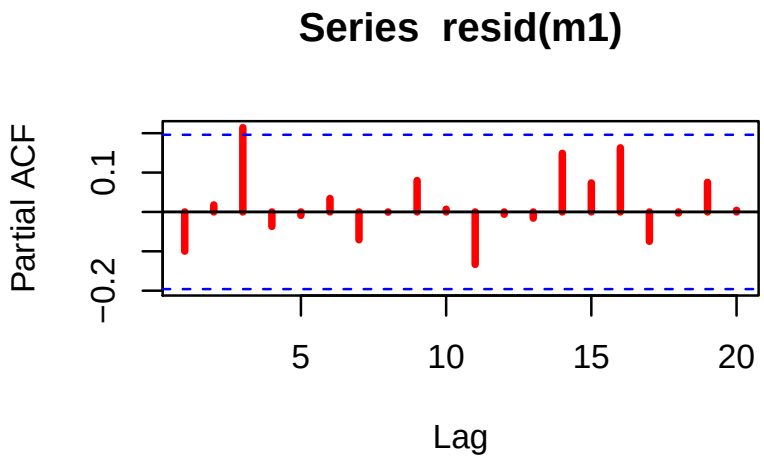


ACF/PACF of linear and observed series









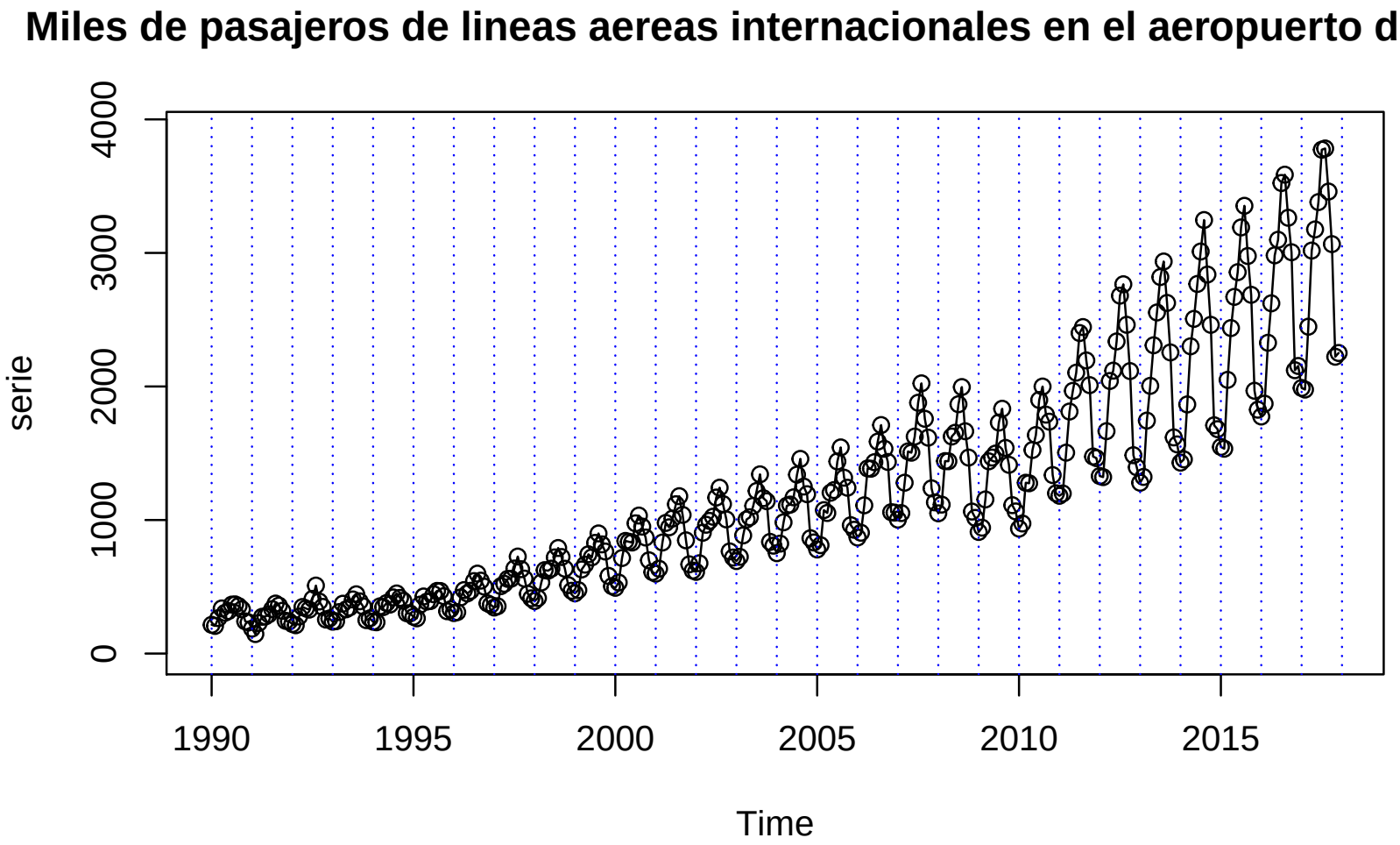
Outlier Treatment

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

Example: AirBCN

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
```

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

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

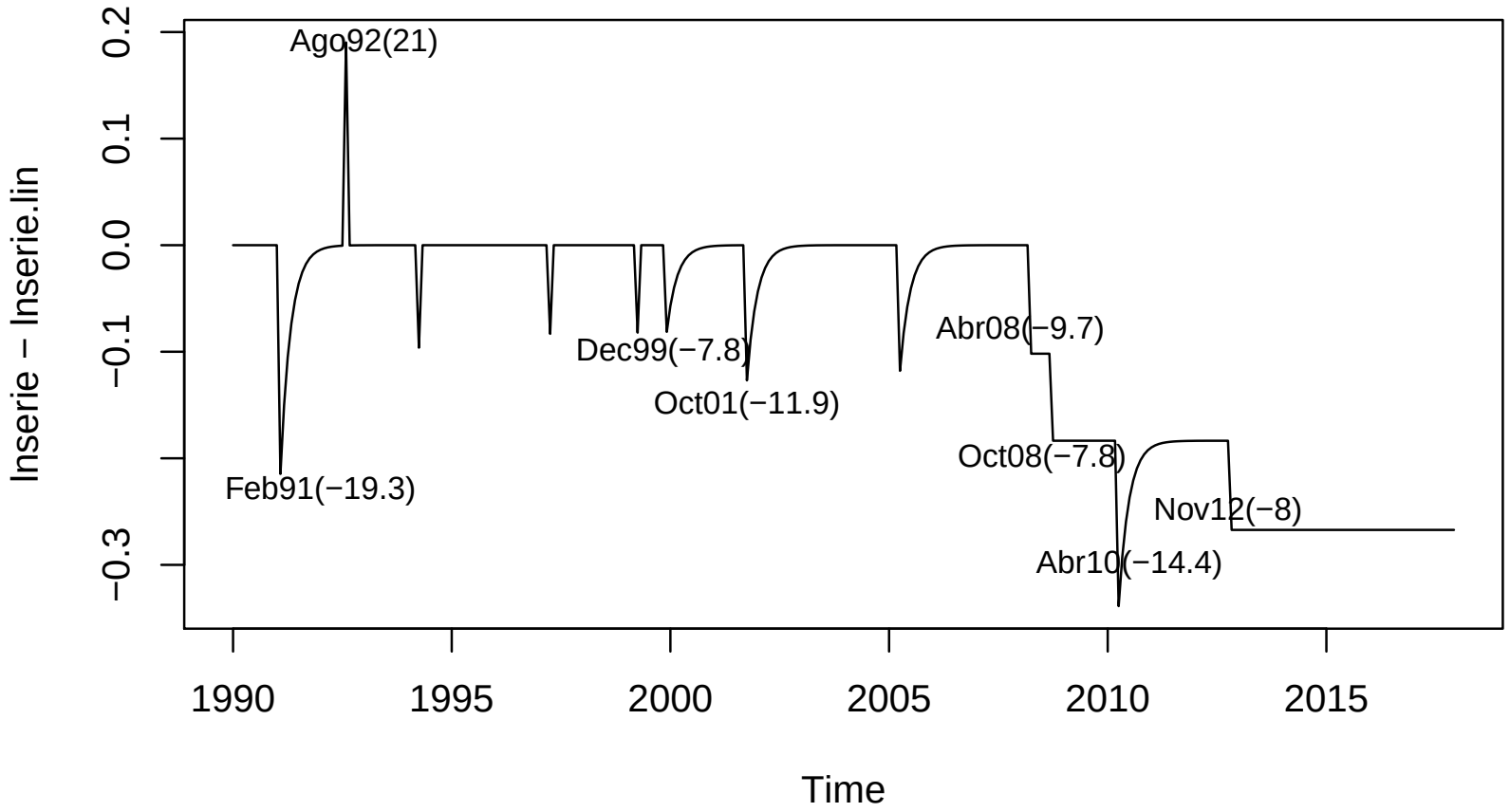
Outlier detection

```
##    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
```

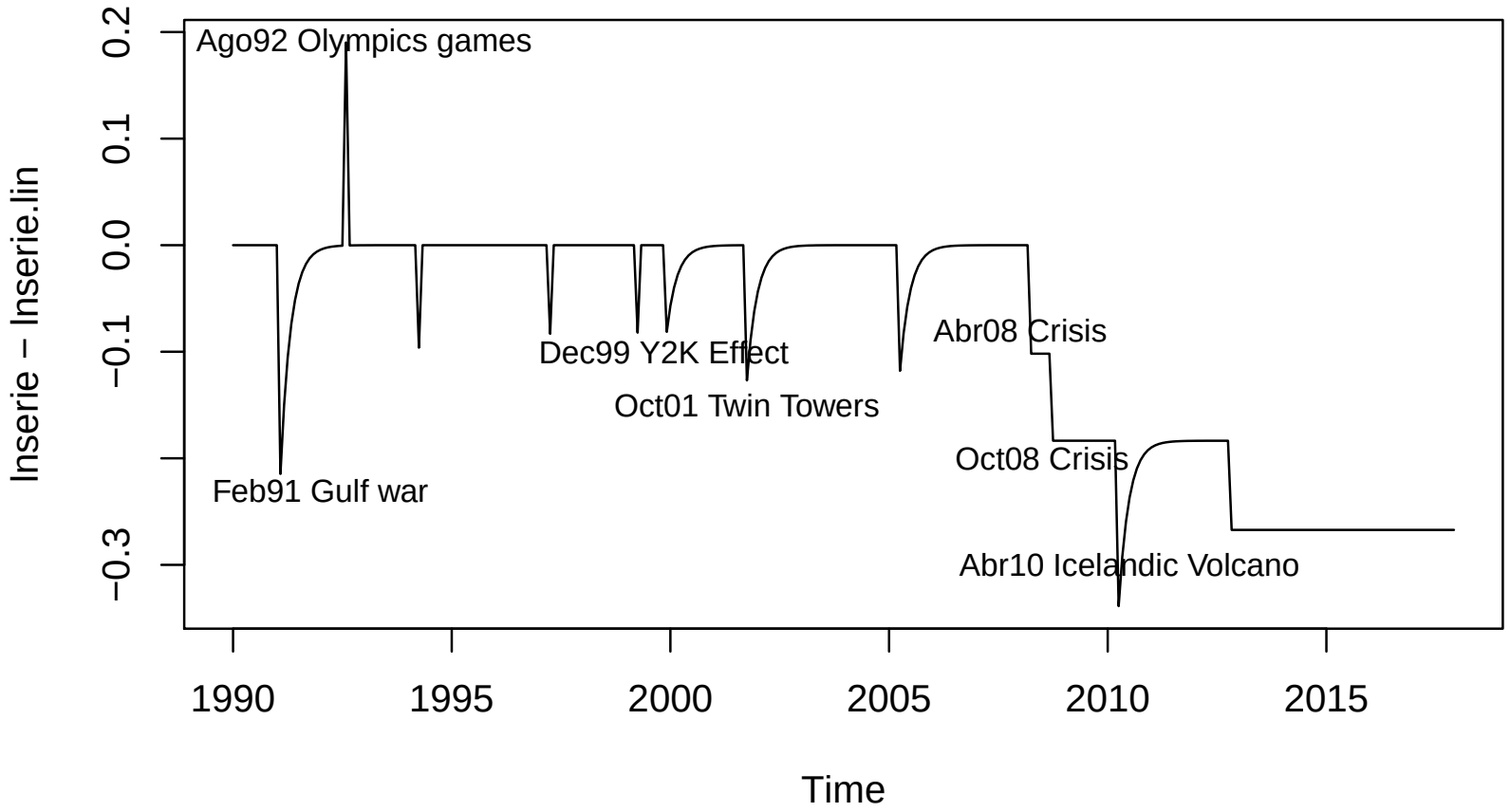
Chronology

```
##    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
```

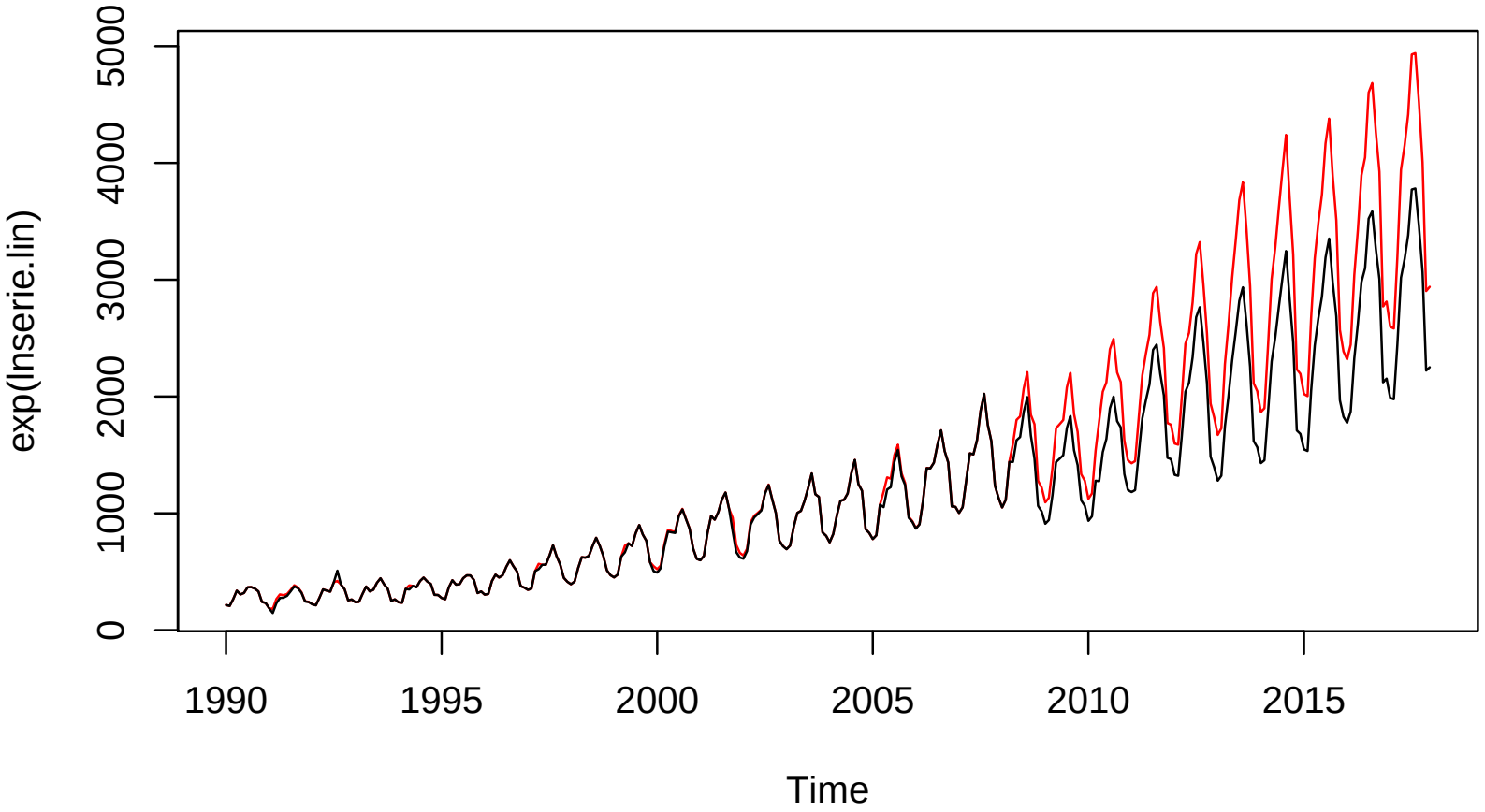
Outlier Effects



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
```

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

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

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