Simulación

Ana Xiangning Pereira Ezquerro

Versión 24 abril, 2022

Índice

En este documento se exponen múltiples simulaciones de la selección automática de variables con sus respectivos retardos usando una nueva propuesta. Se mostrarán ejemplos donde la selección automática trabaja sobre un conjunto de variables de las cuales sólo algunas inciden en la variable respuesta y con un retardo concreto (aunque siempre menor o igual a 0). A lo largo de las siguientes secciones se irán complicando escenarios con al finalidad de analizar cómo se comporta la nueva propuesta ante datos simulados e inferir a partir de ellos cómo se comportará en escenarios reales.

Nota: Para generar los datos de las simulaciones se usó del código `arima_simulation.R`, el cual permite generar de forma pseudo-aleatoria series temporales a partir de un proceso ARIMA. Este documento no muestra cómo generar las series (para evitar la aleatoriedad de los resultados), sino que, una vez generadas y guardadas, se cargan directamente de memoria.

1 Simulación de un modelo de regresión dinámica con innovaciones estacionarias

En esta sección veremos cómo se comporta la función de selección automática sobre ejemplos muy básicos donde las innovaciones del modelo son estacionarias:

$$Y_t = \beta_0 + \beta_1 X_{t-r_1}^{(1)} + \beta_2 X_{t-r_2}^{(2)} + \cdots + X_{t-r_p}^{(p)} + \eta_t$$

donde $\eta_t \sim$ ARMA(p,q) (equivalentemente, un ARIMA donde $d = 0$).

1.1 Modelo donde $r_i = 0$ para $i = 1, ..., p$

Supongamos un modelo de regresión dinámica de tres variables regresoras donde todos los retardos son igual a cero ($r_i = 0, \forall i \in [1, p]$). En concreto, nuestro modelo tendrá la forma:

$$Y_t = \beta_0 + \beta_1 X_t^{(1)} + \beta_2 X_t^{(2)} + \beta_3 X_t^{(3)} + \eta_t$$

donde:

- $\eta_t \sim$ ARMA(2, 1), es decir, las innovaciones son estacionarias.
- $X_t^{(1)} \sim$ ARIMA(2, 1, 3) y su coeficiente $\beta_1 = 2.8$.
- $X_t^{(2)} \sim$ ARMA(1, 1, 2) y su coeficiente $\beta_2 = -1.12$.
- $X_t^{(3)} \sim$ ARMA(1, 0, 2) y su coeficiente $\beta_3 = -2.3$.
- El *intercept* es $\beta_0 = 0.8$.

Supongamos otro conjunto de variables (que siguen también un proceso ARIMA) que no van a influir en la variable respuesta:

- $X_t^{(4)} \sim$ ARIMA(1, 1, 3).
- $X_t^{(5)} \sim$ ARIMA(0, 0, 2).
- $X_t^{(6)} \sim$ ARIMA(2, 1, 1).

```
# Cargamos los datos sobre las variables regresoras
load(file='simulaciones/X1 ~ ARIMA(2,1,3).RData')        # X1
load(file='simulaciones/X2 ~ ARIMA(1,1,2).RData')        # X2
load(file='simulaciones/X3 ~ ARIMA(1,0,2).RData')        # X3
load(file='simulaciones/residuals ~ ARIMA(2,0,1).RData') # residuals

# Cargamos las variables independientes
load(file='simulaciones/X4 ~ ARIMA(1,1,3).RData')        # X4
load(file='simulaciones/X5 ~ ARIMA(0,0,2).RData')        # X5
load(file='simulaciones/X6 ~ ARIMA(2,1,1).RData')        # X6
```

Se puede realizar una comprobación de que estas series siguen los procesos ARIMA mencionados. Para chequearlo, consulte el apéndice del documento.

Creamos el modelo y comprobamos la solución final de la función `auto.fit.arima.regression`.

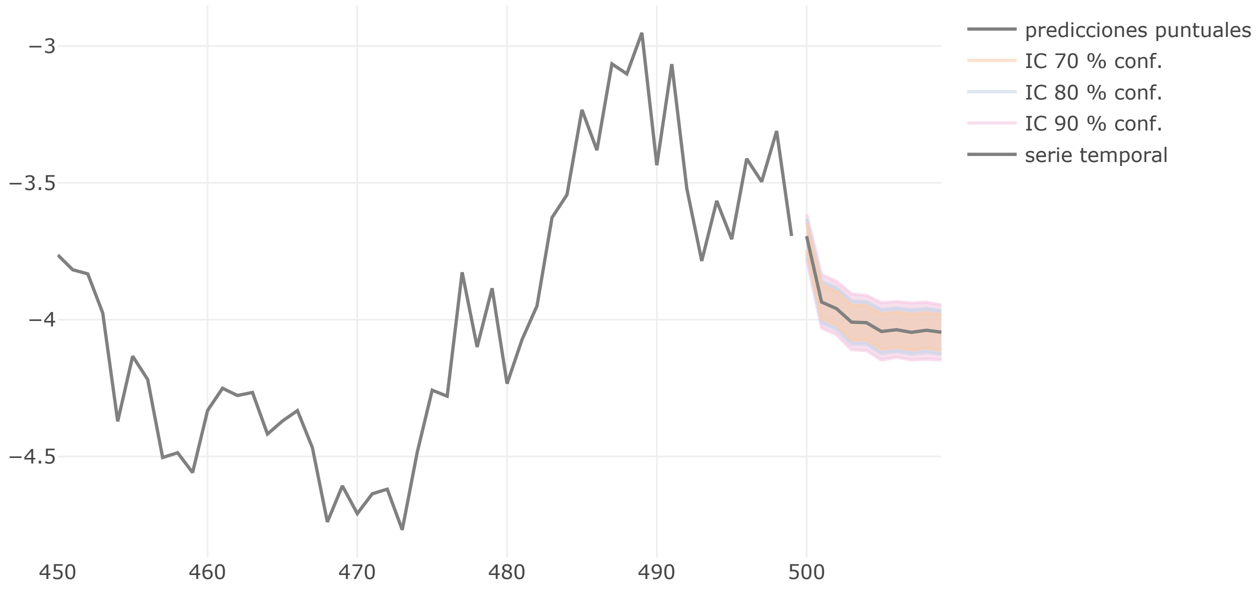
```
beta0 <- 0.8; beta1 <- 2.8; beta2 <- -1.12; beta3 <- -2.3
Y <- beta0 + beta1 * X1$X + beta2 * X2$X + beta3 * X3$X + residuals$X
regresoras <- data.frame(X1=X1$X, X2=X2$X, X3=X3$X, X4=X4$X, X5=X5$X, X6=X6$X)
ajuste <- auto.fit.arima.regression(Y, regresoras, show_info=T)
```

```
----------------------------------------------------------------------
Iteración 1 del algoritmo
----------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-262.66447953697, lag=0]
Se ha probado con la variable X2 [ic=-3.82434068809031, lag=0]
Se ha probado con la variable X3 [ic=-234.82896707977, lag=0]
No se ha podido encontrar un retardo significativo para X4
No se ha podido encontrar un retardo significativo para X5
No se ha podido encontrar un retardo significativo para X6
----------------------------------------------------------------------
Se ha añadido la variable regresora X1 [aicc=-262.66447953697]
----------------------------------------------------------------------
Series: serie
Regression with ARIMA(0,1,4) errors

Coefficients:
          ma1     ma2      ma3     ma4  drift    xreg
      -0.6832  0.5447  -0.4103  0.2145      0  2.9394
s.e.   0.0435  0.0524   0.0527  0.0413      0  0.1322

sigma^2 = 0.03404:  log likelihood = 137.42
AIC=-262.84   AICc=-262.66   BIC=-237.56
----------------------------------------------------------------------
Iteración 2 del algoritmo
----------------------------------------------------------------------
Saltamos X1
Se ha probado con la variable X2 [ic=-565.928310489227, lag=0]
Se ha probado con la variable X3 [ic=-961.548250703332, lag=0]
No se ha podido encontrar un retardo significativo para X4
No se ha podido encontrar un retardo significativo para X5
No se ha podido encontrar un retardo significativo para X6
----------------------------------------------------------------------
Se ha añadido la variable regresora X3 [aicc=-961.548250703332]
----------------------------------------------------------------------
Series: serie
Regression with ARIMA(4,1,0) errors

Coefficients:
         ar1      ar2  ar3     ar4  drift       X3      X1
      0.1983  -0.0983    0  0.1903      0  -2.2307  2.9141
s.e.  0.0438   0.0441    0  0.0443      0   0.0314  0.0646

sigma^2 = 0.0084:  log likelihood = 486.86
AIC=-961.72   AICc=-961.55   BIC=-936.44
```

```
----------------------------------------------------------------------
Iteración 3 del algoritmo
----------------------------------------------------------------------
Saltamos X1
Se ha probado con la variable X2 [ic=-1586.9854751344, lag=0]
Saltamos X3
No se ha podido encontrar un retardo significativo para X4
No se ha podido encontrar un retardo significativo para X5
No se ha podido encontrar un retardo significativo para X6
----------------------------------------------------------------------
Se ha añadido la variable regresora X2 [aicc=-1586.9854751344]
----------------------------------------------------------------------
Series: serie
Regression with ARIMA(2,0,1) errors

Coefficients:
         ar1      ar2     ma1  intercept       X2      X1       X3
      0.3252  -0.4103  0.3479     0.8114  -1.1231  2.8041  -2.2542
s.e.  0.0726   0.0522  0.0771     0.0069   0.0038  0.0041   0.0173

sigma^2 = 0.002401:  log likelihood = 801.64
AIC=-1587.28   AICc=-1586.99   BIC=-1553.56
----------------------------------------------------------------------
Iteración 4 del algoritmo
----------------------------------------------------------------------
Saltamos X1
Saltamos X2
Saltamos X3
No se ha podido encontrar un retardo significativo para X4
No se ha podido encontrar un retardo significativo para X5
No se ha podido encontrar un retardo significativo para X6
No se añaden más variables

----------------------------------------------------------------------
|            Histórico de variables añadidas al modelo                 |
----------------------------------------------------------------------
 var lag                ic
  X1   0   -262.66447953697
  X3   0 -961.548250703332
  X2   0   -1586.9854751344
```

Finalmente realizamos las predicciones puntuales:

```
preds <- forecast_model(ajuste, h=10, mode='bootstrap', level=c(70, 80, 90))
display(plot_forecast(preds, rang=c(450, 510)), name='ejemplo1')
```



1.2 Modelo donde $r_i \leq 0$ para $i = 1, ..., p$

Supongamos un modelo de regresión dinámica parecido al del primer ejemplo, pero donde los retardos sean menores o iguales a 0 (que haya "variedad" en los retardos).

$$Y_t = \beta_0 + \beta_1 X^{(1)}_{t-r_1} + \beta_2 X^{(2)}_{t-r_2} + \beta_3 X^{(3)}_{t-r_3} + \eta_t$$

donde:

- $\eta_t \sim$ ARMA(2, 1).
- $X^{(1)}_t \sim$ ARIMA(2, 1, 3) y su retardo $r_1 = -2$.
- $X^{(2)}_t \sim$ ARMA(1, 1, 2) y su retardo $r_2 = 0$.
- $X^{(3)}_t \sim$ ARMA(1, 0, 2) y su retardo $r_3 = -3$.

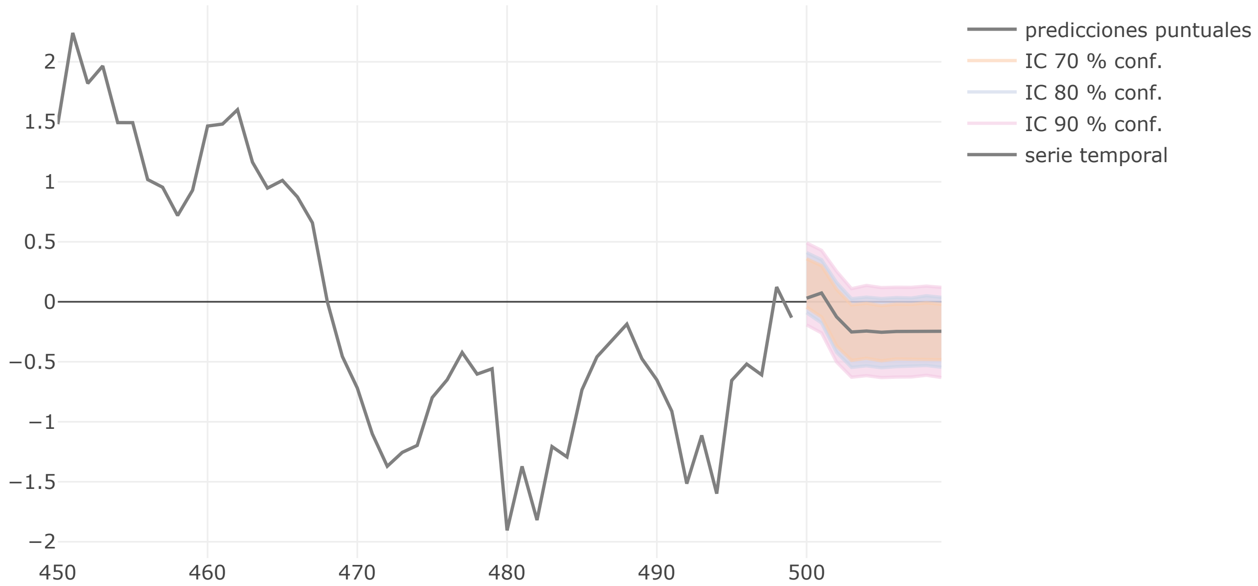
```
# Construimos el modelo
beta0 <- -0.1; beta1 <- 1.2; beta2 <- -4.2; beta3 <- 3.3
r1 <- -2; r3 <- -3
Y <- beta0 + beta1 * lag(X1$X, r1) + beta2 * X2$X + beta3 * lag(X3$X, r3) +
    residuals$X

# Usamos la función
regresoras <- data.frame(X1=X1$X, X2=X2$X, X3=X3$X, X4=X4$X, X5=X5$X, X6=X6$X)
regresoras <- as.data.frame(
    lapply(regresoras, window, start=start(Y)[1], end=end(Y)[1])
)
ajuste <- auto.fit.arima.regression(Y, regresoras, show_info=T,
                                    stationary_method='adf.test')
```

```
----------------------------------------------------------------------
```

```
Iteración 1 del algoritmo
----------------------------------------------------------------------
Se ha probado con la variable X1 [ic=386.152843648257, lag=-2]
Se ha probado con la variable X2 [ic=-45.441956384563, lag=0]
Se ha probado con la variable X3 [ic=3.99621010980833, lag=-3]
Se ha probado con la variable X4 [ic=409.312663479496, lag=-15]
Se ha probado con la variable X5 [ic=412.359106957581, lag=-1]
Se ha probado con la variable X6 [ic=408.856148146608, lag=-5]
----------------------------------------------------------------------
Se ha añadido la variable regresora X2 [aicc=-45.441956384563]
----------------------------------------------------------------------
Series: serie
Regression with ARIMA(0,1,4) errors

Coefficients:
          ma1     ma2      ma3     ma4  drift     xreg
      -0.9118  0.6400  -0.4468  0.1679      0  -4.2976
s.e.   0.0451  0.0572   0.0565  0.0433      0   0.1047

sigma^2 = 0.05237:  log likelihood = 28.81
AIC=-45.62   AICc=-45.44   BIC=-20.56
----------------------------------------------------------------------
Iteración 2 del algoritmo
----------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-143.054141399295, lag=-3]
Saltamos X2
No se ha podido encontrar un retardo significativo para X3
No se ha podido encontrar un retardo significativo para X4
No se ha podido encontrar un retardo significativo para X5
No se ha podido encontrar un retardo significativo para X6
----------------------------------------------------------------------
Se ha añadido la variable regresora X1 [aicc=-143.054141399295]
----------------------------------------------------------------------
Series: serie
Regression with ARIMA(0,1,3) errors

Coefficients:
          ma1     ma2      ma3  drift      X1       X2
      -1.2164  0.6144  -0.3980      0  1.2170  -4.2069
s.e.   0.0439  0.0578   0.0423      0  0.0207   0.0164

sigma^2 = 0.04186:  log likelihood = 77.62
AIC=-143.24   AICc=-143.05   BIC=-118.37
----------------------------------------------------------------------
Iteración 3 del algoritmo
----------------------------------------------------------------------
Saltamos X1
Saltamos X2
Se ha probado con la variable X3 [ic=-145.87682978519, lag=-10]
No se ha podido encontrar un retardo significativo para X4
No se ha podido encontrar un retardo significativo para X5
No se ha podido encontrar un retardo significativo para X6
----------------------------------------------------------------------
Se ha añadido la variable regresora X3 [aicc=-145.87682978519]
```

```
-----------------------------------------------------------------------
Series: serie
Regression with ARIMA(0,1,4) errors

Coefficients:
          ma1     ma2      ma3     ma4  drift       X3       X2      X1
      -1.2622  0.6842  -0.5420  0.1200      0  -0.0566  -4.2076  1.2186
s.e.   0.0468  0.0700   0.0681  0.0452      0   0.1840   0.0144  0.0182

sigma^2 = 0.04139:  log likelihood = 81.1
AIC=-146.19   AICc=-145.88   BIC=-113.04
-----------------------------------------------------------------------
Iteración 4 del algoritmo
-----------------------------------------------------------------------
Saltamos X1
Saltamos X2
Saltamos X3
No se ha podido encontrar un retardo significativo para X4
No se ha podido encontrar un retardo significativo para X5
No se ha podido encontrar un retardo significativo para X6
No se añaden más variables

-----------------------------------------------------------------------
|             Histórico de variables añadidas al modelo                |
-----------------------------------------------------------------------
 var lag                ic
  X2   0  -45.441956384563
  X1  -3 -143.054141399295
  X3 -10  -145.87682978519
```

```r
# Podemos mostrar las predicciones puntuales
preds <- forecast_model(ajuste, h=10, mode='bootstrap', level=c(70, 80, 90))
display(plot_forecast(preds, rang=c(450, 510)), name='ejemplo2')
```



2 Simulación de un modelo de regresión dinámico con errores ARIMA ($d = 1$)

En esta sección consideraremos modelos de regresión dinámica donde las innovaciones no son estacionarias:

$$Y_t = \beta_0 + \beta_1 X^{(1)}_{t-r_1} + \beta_2 X^{(2)}_{t-r_2} + \cdots + X^{(p)}_{t-r_p} + \eta_t$$

donde $\eta_t \sim$ ARIMA(p,d,q).

2.1 Modelo donde $r_i = 0$ para $i = 1, ..., p$

Tomemos el mismo modelo que en el primer ejemplo pero con innovaciones no estacionarias:

$$Y_t = \beta_0 + \beta_1 X^{(1)}_t + \beta_2 X^{(2)}_t + \beta_3 X^{(3)}_t + \eta_t$$

donde:

- $\eta_t \sim$ ARMA(1, 2, 2), es decir, las innovaciones son estacionarias.
- $X^{(1)}_t \sim$ ARIMA(2, 1, 3) y su coeficiente $\beta_1 = 2.8$.
- $X^{(2)}_t \sim$ ARMA(1, 1, 2) y su coeficiente $\beta_2 = -1.12$.
- $X^{(3)}_t \sim$ ARMA(2, 2, 3) y su coeficiente $\beta_3 = -2.3$.
- El *intercept* es $\beta_0 = 0.8$.

```
load(file='simulaciones/X1 ~ ARIMA(2,1,3).RData')
load(file='simulaciones/X2 ~ ARIMA(1,2,1).RData')
load(file='simulaciones/X3 ~ ARIMA(2,2,3).RData')
load(file='simulaciones/X4 ~ ARIMA(1,1,3).RData')
load(file='simulaciones/X5 ~ ARIMA(0,1,2).RData')
load(file='simulaciones/X6 ~ ARIMA(2,1,1).RData')
load('simulaciones/residuals ~ ARIMA(1, 2, 2).RData')
beta0 <- 0.8; beta1 <- -1.3; beta2 <- 7.12; beta3 <- 12.3
Y <- beta0 + beta1 * X1$X + beta2 * X2$X + beta3 * X3$X + 2.1*residuals$X
```

Ajustamos el modelo con las variables originales (no diferenciamos ninguna):

```
regresoras <- data.frame(X1=X1$X, X2=X2$X, X3=X3$X, X4=X4$X, X5=X5$X, X6=X6$X)
ajuste <- auto.fit.arima.regression(Y, regresoras, show_info=T)
```

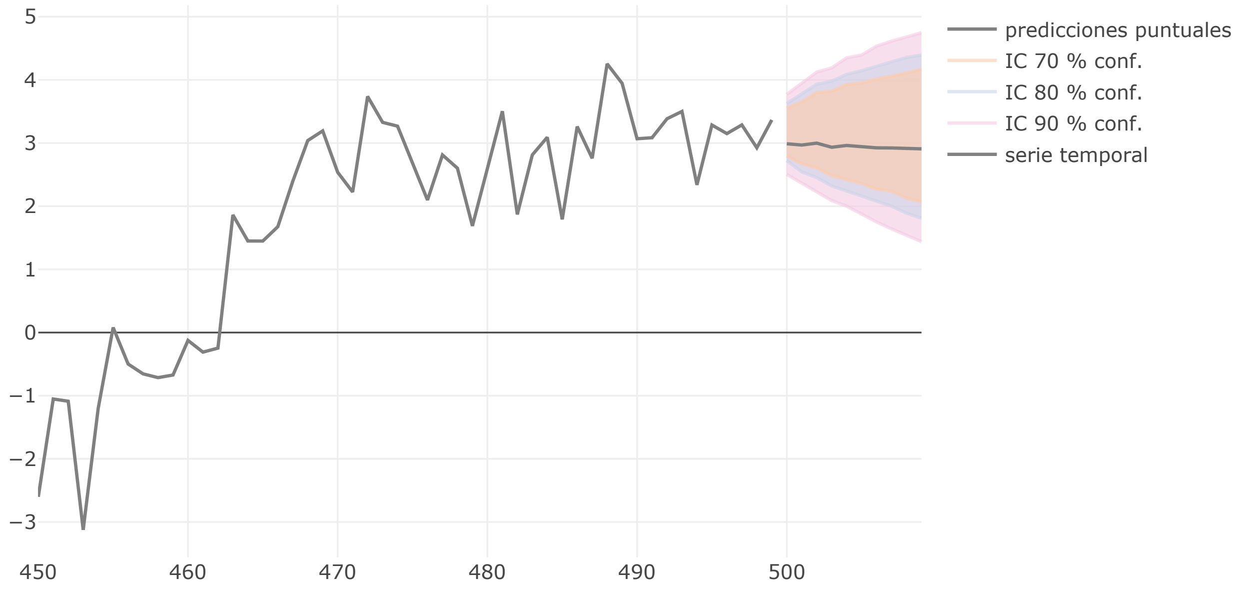
```
----------------------------------------------------------------------
Iteración 1 del algoritmo
----------------------------------------------------------------------
Se ha probado con la variable X1 [ic=1118.61418419575, lag=-2]
Se ha probado con la variable X2 [ic=965.364095020104, lag=0]
Se ha probado con la variable X3 [ic=403.477209059552, lag=0]
Se ha probado con la variable X4 [ic=1140.61900931652, lag=-22]
```

```
No se ha podido encontrar un retardo significativo para X5
Se ha probado con la variable X6 [ic=1142.8955381657, lag=-22]
----------------------------------------------------------------------
Se ha añadido la variable regresora X3 [aicc=403.477209059552]
----------------------------------------------------------------------
Series: serie
Regression with ARIMA(3,2,0) errors

Coefficients:
          ar1  ar2      ar3     xreg
      -0.1818    0  -0.1040  12.5855
s.e.   0.0449    0   0.0451   0.2374

sigma^2 = 0.1352:  log likelihood = -197.7
AIC=403.39   AICc=403.48   BIC=420.05
----------------------------------------------------------------------
Iteración 2 del algoritmo
----------------------------------------------------------------------
No se ha podido ajustar un modelo para X1
Se ha probado con la variable X2 [ic=408.131686276644, lag=-7]
Saltamos X3
Se ha probado con la variable X4 [ic=407.911251927814, lag=-20]
No se ha podido encontrar un retardo significativo para X5
No se ha podido encontrar un retardo significativo para X6
----------------------------------------------------------------------
Iteración 1 del algoritmo
----------------------------------------------------------------------
Se ha probado con la variable X1 [ic=1122.59863963397, lag=-2]
Se ha probado con la variable X2 [ic=965.364548782521, lag=0]
Se ha probado con la variable X3 [ic=405.394093487614, lag=0]
Se ha probado con la variable X4 [ic=1140.61915269641, lag=-22]
No se ha podido encontrar un retardo significativo para X5
Se ha probado con la variable X6 [ic=1142.8956889607, lag=-22]
----------------------------------------------------------------------
Se ha añadido la variable regresora X3 [aicc=405.394093487614]
----------------------------------------------------------------------
Series: serie
Regression with ARIMA(1,1,1) errors

Coefficients:
          ar1     ma1  drift     xreg
      -0.6104  0.4416      0  12.5400
s.e.   0.1545  0.1743      0   0.2353

sigma^2 = 0.1357:  log likelihood = -198.65
AIC=405.31   AICc=405.39   BIC=421.97
----------------------------------------------------------------------
Iteración 2 del algoritmo
----------------------------------------------------------------------
Se ha probado con la variable X1 [ic=397.837724127343, lag=-16]
Se ha probado con la variable X2 [ic=398.68993349132, lag=-7]
Saltamos X3
Se ha probado con la variable X4 [ic=398.797838463971, lag=-20]
No se ha podido encontrar un retardo significativo para X5
```

```
No se ha podido encontrar un retardo significativo para X6
----------------------------------------------------------------------
Se ha añadido la variable regresora X1 [aicc=397.837724127343]
----------------------------------------------------------------------
Series: serie
Regression with ARIMA(1,1,0) errors

Coefficients:
          ar1  drift       X1       X3
      -0.1904      0  -0.2166  12.6284
s.e.   0.0461      0   0.2173   0.2562

sigma^2 = 0.1391:  log likelihood = -194.87
AIC=397.75   AICc=397.84   BIC=414.22
----------------------------------------------------------------------
Iteración 3 del algoritmo
----------------------------------------------------------------------
Saltamos X1
Se ha probado con la variable X2 [ic=399.728185190023, lag=-7]
Saltamos X3
No se ha podido encontrar un retardo significativo para X4
No se ha podido encontrar un retardo significativo para X5
Se ha probado con la variable X6 [ic=399.880910504599, lag=-7]
No se añaden más variables

----------------------------------------------------------------------
|            Histórico de variables añadidas al modelo                |
----------------------------------------------------------------------
 var lag               ic
  X3   0 405.394093487614
  X1 -16 397.837724127343
```

Podemos observar cómo la función ha realizado una diferenciación regular a todas las series (respuesta y regresoras). El resultado final es que las variables $X^{(1)}$ y $X^{(2)}$ intervienen en la variable respuesta, y que los residuos son un ARIMA(2, 0, 0). Obsérvese que los órdenes de los residuos son muy distintos a los que se han usado para generarlos. Etso se debe a las diferenciaciones regulares que se le ha aplicado a todo el conjunto de datos.

```
preds <- forecast_model(ajuste, h=10, mode='bootstrap', levels=c(70, 80, 90))
display(plot_forecast(preds, rang=c(450, 510)), name='ejemplo3')
```



2.2 Modelo con retardos $r < 0$

```
# Creación del modelo con los nuevos residuos
beta0 <- 0.8; beta1 <- 6.8; beta2 <- 1.12; beta3 <- 12.3
r1 <- -2; r3 <- -1
Y <- beta0 + beta1 * lag(X1$X, r1) + beta2 * X2$X +
    beta3 * lag(X3$X, r3) + 1.5*residuals$X

regresoras <- data.frame(X1=X1$X, X2=X2$X, X3=X3$X, X4=X4$X, X5=X5$X, X6=X6$X)
auto.fit.arima.regression(Y, regresoras, show_info=T,
                          stationary_method='adf.test')
```

```
----------------------------------------------------------------------
Iteración 1 del algoritmo
----------------------------------------------------------------------
No se ha podido ajustar un modelo para X1
Se ha probado con la variable X2 [ic=1182.25838911986, lag=-17]
No se ha podido encontrar un retardo significativo para X3
Se ha probado con la variable X4 [ic=1204.2979371295, lag=-21]
No se ha podido encontrar un retardo significativo para X5
Se ha probado con la variable X6 [ic=1180.72267445812, lag=-21]
----------------------------------------------------------------------
Se ha añadido la variable regresora X6 [aicc=1180.72267445812]
----------------------------------------------------------------------
Series: serie
Regression with ARIMA(2,2,2) errors

Coefficients:
      ar1      ar2      ma1     ma2    xreg
        0  -0.4951  -0.7274  0.6013  0.9627
s.e.    0   0.0534   0.0415  0.0399  0.5827
```

```
sigma^2 = 0.6926:  log likelihood = -585.3
AIC=1180.59   AICc=1180.72   BIC=1201.41
----------------------------------------------------------------------
Iteración 2 del algoritmo
----------------------------------------------------------------------
Se ha probado con la variable X1 [ic=891.38318067544, lag=-2]
No se ha podido ajustar un modelo para X2
Se ha probado con la variable X3 [ic=1153.74307220898, lag=-17]
Se ha probado con la variable X4 [ic=1137.13125216193, lag=0]
No se ha podido encontrar un retardo significativo para X5
Saltamos X6
----------------------------------------------------------------------
Se ha añadido la variable regresora X1 [aicc=891.38318067544]
----------------------------------------------------------------------
Series: serie
Regression with ARIMA(2,2,3) errors

Coefficients:
          ar1      ar2  ma1  ma2     ma3      X1      X6
      -0.3134  -0.5280    0    0  0.3532  7.4735  0.1506
s.e.   0.0427   0.0424    0    0  0.0512  0.2553  0.3405

sigma^2 = 0.404:  log likelihood = -439.6
AIC=891.2   AICc=891.38   BIC=915.9
----------------------------------------------------------------------
Iteración 3 del algoritmo
----------------------------------------------------------------------
Saltamos X1
No se ha podido ajustar un modelo para X2
No se ha podido ajustar un modelo para X3
No se ha podido ajustar un modelo para X4
No se ha podido encontrar un retardo significativo para X5
Saltamos X6
----------------------------------------------------------------------
Iteración 1 del algoritmo
----------------------------------------------------------------------
Se ha probado con la variable X1 [ic=1150.51023210728, lag=0]
Se ha probado con la variable X2 [ic=1182.25841396978, lag=-17]
No se ha podido encontrar un retardo significativo para X3
Se ha probado con la variable X4 [ic=1182.7341855372, lag=-21]
No se ha podido encontrar un retardo significativo para X5
Se ha probado con la variable X6 [ic=1180.72275199257, lag=-21]
----------------------------------------------------------------------
Se ha añadido la variable regresora X1 [aicc=1150.51023210728]
----------------------------------------------------------------------
Series: serie
Regression with ARIMA(2,1,3) errors

Coefficients:
          ar1      ar2      ma1  ma2     ma3  drift    xreg
      -0.3705  -0.4041  -0.2188    0  0.3278      0  4.7814
s.e.   0.0665   0.0571   0.0645    0  0.0549      0  0.5019

sigma^2 = 0.6482:  log likelihood = -569.17
```

```
AIC=1150.33   AICc=1150.51   BIC=1175.31
----------------------------------------------------------------------
Iteración 2 del algoritmo
----------------------------------------------------------------------
Saltamos X1
No se ha podido encontrar un retardo significativo para X2
Se ha probado con la variable X3 [ic=1099.88949458802, lag=-15]
Se ha probado con la variable X4 [ic=1099.94524795533, lag=-16]
No se ha podido encontrar un retardo significativo para X5
Se ha probado con la variable X6 [ic=1101.71226357321, lag=0]
----------------------------------------------------------------------
Se ha añadido la variable regresora X3 [aicc=1099.88949458802]
----------------------------------------------------------------------
Series: serie
Regression with ARIMA(2,1,3) errors

Coefficients:
          ar1      ar2      ma1  ma2     ma3  drift       X3      X1
      -0.3684  -0.4173  -0.2316    0  0.3311      0  -0.9663  4.8288
s.e.   0.0663   0.0574   0.0652    0  0.0544      0   0.7114  0.5167

sigma^2 = 0.6464:  log likelihood = -542.82
AIC=1099.64   AICc=1099.89   BIC=1128.47
----------------------------------------------------------------------
Iteración 3 del algoritmo
----------------------------------------------------------------------
Saltamos X1
No se ha podido encontrar un retardo significativo para X2
Saltamos X3
Se ha probado con la variable X4 [ic=1100.20153344874, lag=-16]
No se ha podido encontrar un retardo significativo para X5
Se ha probado con la variable X6 [ic=1101.9535323942, lag=0]
No se añaden más variables

----------------------------------------------------------------------
|            Histórico de variables añadidas al modelo                |
----------------------------------------------------------------------
 var lag               ic
  X1   0 1150.51023210728
  X3 -15 1099.88949458802

Series: serie
Regression with ARIMA(3,0,0) errors

Coefficients:
         ar1  ar2     ar3  intercept      X1       X3
      0.4031    0  0.5754    -4.7075  4.3615  -0.9358
s.e.  0.0320    0  0.0322     1.5786  0.4968   0.7087

sigma^2 = 0.6814:  log likelihood = -557.53
AIC=1127.06   AICc=1127.24   BIC=1151.78
```

3 Comparativa del método de preblanqueado

3.1 Preblanqueado en base al adf.test

Creamos variables regresoras con órdenes d distintos y retardos distintos:

```
# Variables que influyen en el modelo
load(file='simulaciones/X1 ~ ARIMA(2,1,3).RData')
load(file='simulaciones/X2 ~ ARIMA(1,2,1).RData')
load(file='simulaciones/X3 ~ ARIMA(2,2,3).RData')
load(file='simulaciones/X4 ~ ARIMA(1,1,3).RData')
load(file='simulaciones/X5 ~ ARIMA(0,1,2).RData')
load(file='simulaciones/X6 ~ ARIMA(2,1,1).RData')
load(file='simulaciones/residuals ~ ARIMA(2,0,1).RData')
```

```
beta0 <- -0.1; beta1 <- 8.2; beta2 <- -0.5
r1 <- -2; r2 <- -3
Y <- beta0 + beta1 * lag(X1$X, r1) + beta2 * lag(X2$X, r2) + residuals$X
```

Ajustamos un modelo usando como método para chequear estacionariedad la función auto.arima:

```
regresoras <- data.frame(X1=X1$X, X2=X2$X, X3=X3$X, X4=X4$X, X5=X5$X, X6=X6$X)
regresoras <- as.data.frame(lapply(regresoras, window,
                                   start=start(Y)[1], end=end(Y)[1]))
auto.fit.arima.regression(Y, regresoras, show_info=T,
                          stationary_method='auto.arima')
```

```
----------------------------------------------------------------------
Iteración 1 del algoritmo
----------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-1113.09680274809, lag=-2]
Se ha probado con la variable X2 [ic=568.894610494165, lag=-19]
Se ha probado con la variable X3 [ic=566.782532939914, lag=0]
Se ha probado con la variable X4 [ic=565.859291257565, lag=-11]
No se ha podido encontrar un retardo significativo para X5
No se ha podido encontrar un retardo significativo para X6
----------------------------------------------------------------------
Se ha añadido la variable regresora X1 [aicc=-1113.09680274809]
----------------------------------------------------------------------
Series: serie
Regression with ARIMA(2,2,1) errors

Coefficients:
         ar1      ar2      ma1    xreg
      0.1173  -0.4433  -0.6928  8.2396
s.e.  0.0492   0.0446   0.0405  0.0429

sigma^2 = 0.005561:  log likelihood = 561.61
AIC=-1113.22   AICc=-1113.1   BIC=-1092.4
----------------------------------------------------------------------
Iteración 2 del algoritmo
----------------------------------------------------------------------
```

```
Saltamos X1
Se ha probado con la variable X2 [ic=-1205.36965020075, lag=-1]
Se ha probado con la variable X3 [ic=-1061.36648376728, lag=-17]
No se ha podido encontrar un retardo significativo para X4
No se ha podido encontrar un retardo significativo para X5
No se ha podido encontrar un retardo significativo para X6
----------------------------------------------------------------------
Se ha añadido la variable regresora X2 [aicc=-1205.36965020075]
----------------------------------------------------------------------
Series: serie
Regression with ARIMA(2,1,2) errors

Coefficients:
         ar1      ar2  ma1      ma2  drift       X2      X1
      0.1576  -0.2876    0  -0.2664      0  -0.4925  8.2267
s.e.  0.0453   0.0903    0   0.0943      0   0.0033  0.0363

sigma^2 = 0.004142:  log likelihood = 608.78
AIC=-1205.56   AICc=-1205.37   BIC=-1180.79
----------------------------------------------------------------------
Iteración 3 del algoritmo
----------------------------------------------------------------------
Saltamos X1
Saltamos X2
Se ha probado con la variable X3 [ic=-1188.70239442743, lag=-19]
No se ha podido encontrar un retardo significativo para X4
No se ha podido encontrar un retardo significativo para X5
No se ha podido encontrar un retardo significativo para X6
No se añaden más variables

----------------------------------------------------------------------
|            Histórico de variables añadidas al modelo                |
----------------------------------------------------------------------
 var lag                ic
  X1  -2 -1113.09680274809
  X2  -1 -1205.36965020075

Series: serie
Regression with ARIMA(3,0,2) errors

Coefficients:
         ar1      ar2     ar3  ma1      ma2  intercept      X1       X2
      1.1319  -0.4780  0.2963    0  -0.2212    -0.0384  8.2071  -0.4933
s.e.  0.0446   0.1067  0.0871    0   0.1002     0.0684  0.0355   0.0008

sigma^2 = 0.004072:  log likelihood = 614.16
AIC=-1212.31   AICc=-1211.99   BIC=-1179.28
```

Ajustamos un modelo usando como método para chequear estacionariedad el `adf.test`:

```r
regresoras <- data.frame(X1=X1$X, X2=X2$X, X3=X3$X, X4=X4$X, X5=X5$X, X6=X6$X)
regresoras <- as.data.frame(lapply(regresoras, window,
                                   start=start(Y)[1], end=end(Y)[1]))
ajuste <- auto.fit.arima.regression(Y, regresoras, show_info=T,
```

```
                                    stationary_method='adf.test')
```

```
----------------------------------------------------------------------
Iteración 1 del algoritmo
----------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-1110.41152896518, lag=-2]
Se ha probado con la variable X2 [ic=568.722893068277, lag=-19]
Se ha probado con la variable X3 [ic=566.62164073666, lag=0]
Se ha probado con la variable X4 [ic=565.700322494648, lag=-11]
No se ha podido ajustar un modelo para X5
Se ha probado con la variable X6 [ic=568.621344657945, lag=-20]
----------------------------------------------------------------------
Se ha añadido la variable regresora X1 [aicc=-1110.41152896518]
----------------------------------------------------------------------
Series: serie
Regression with ARIMA(2,2,1) errors

Coefficients:
         ar1      ar2      ma1    xreg
      0.1169  -0.4418  -0.6937  8.2407
s.e.  0.0493   0.0447   0.0407  0.0430

sigma^2 = 0.005565:  log likelihood = 560.27
AIC=-1110.54   AICc=-1110.41   BIC=-1089.72
----------------------------------------------------------------------
Iteración 2 del algoritmo
----------------------------------------------------------------------
Saltamos X1
Se ha probado con la variable X2 [ic=-1110.39412183132, lag=0]
Se ha probado con la variable X3 [ic=-1043.14054580125, lag=-16]
No se ha podido encontrar un retardo significativo para X4
No se ha podido encontrar un retardo significativo para X5
No se ha podido encontrar un retardo significativo para X6
----------------------------------------------------------------------
Iteración 1 del algoritmo
----------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-1110.41156101901, lag=-2]
Se ha probado con la variable X2 [ic=568.127977122276, lag=-19]
Se ha probado con la variable X3 [ic=570.100784230106, lag=0]
Se ha probado con la variable X4 [ic=569.550840783176, lag=-11]
Se ha probado con la variable X5 [ic=571.824467932519, lag=-18]
Se ha probado con la variable X6 [ic=572.244774859494, lag=-20]
----------------------------------------------------------------------
Se ha añadido la variable regresora X1 [aicc=-1110.41156101901]
----------------------------------------------------------------------
Series: serie
Regression with ARIMA(2,1,1) errors

Coefficients:
         ar1      ar2      ma1  drift    xreg
      0.1169  -0.4418  -0.6937      0  8.2407
s.e.  0.0493   0.0447   0.0407      0  0.0430

sigma^2 = 0.005565:  log likelihood = 560.27
```

```
AIC=-1110.54   AICc=-1110.41   BIC=-1089.72
----------------------------------------------------------------------
Iteración 2 del algoritmo
----------------------------------------------------------------------
Saltamos X1
Se ha probado con la variable X2 [ic=-1092.10769813569, lag=0]
Se ha probado con la variable X3 [ic=-1057.59220684789, lag=-16]
No se ha podido encontrar un retardo significativo para X4
No se ha podido encontrar un retardo significativo para X5
No se ha podido encontrar un retardo significativo para X6
No se añaden más variables

----------------------------------------------------------------------
|            Histórico de variables añadidas al modelo                |
----------------------------------------------------------------------
 var lag                ic
  X1  -2 -1110.41156101901
```

4 Apéndice

En esta sección se muestra la comprobación con la función `auto.fit.arima` de que las muestras cargadas cumplen con los requisitos mencionados.

Para el primer ejempo, las muestras simuladas eran las siguientes:

```
load(file='simulaciones/X1 ~ ARIMA(2,1,3).RData')        # X1
load(file='simulaciones/X2 ~ ARIMA(1,1,2).RData')        # X2
load(file='simulaciones/X3 ~ ARIMA(1,0,2).RData')        # X3
load(file='simulaciones/residuals ~ ARIMA(2,0,1).RData') # residuals
load(file='simulaciones/X4 ~ ARIMA(1,1,3).RData')        # X4
load(file='simulaciones/X5 ~ ARIMA(0,0,2).RData')        # X5
load(file='simulaciones/X6 ~ ARIMA(2,1,1).RData')        # X6
```

Si la función `auto.fit.arima` y observamos los *outputs*, vemos que siguen el proceso ARIMA anotado:

```
auto.fit.arima(X1$X, show_info=F)
```

```
Series: serie
ARIMA(2,1,3)

Coefficients:
          ar1     ar2  ma1  ma2     ma3
      -0.2926  0.3841    0    0  0.3757
s.e.   0.0450  0.0429    0    0  0.0463

sigma^2 = 0.002731:  log likelihood = 765.93
AIC=-1523.86   AICc=-1523.78   BIC=-1507.01
```

```
auto.fit.arima(X2$X, show_info=F)
```

```
Series: serie
ARIMA(1,1,2)
```

```
Coefficients:
         ar1  ma1     ma2
      0.2659    0  0.4925
s.e.  0.0443    0  0.0397

sigma^2 = 0.002469:  log likelihood = 790.59
AIC=-1575.19   AICc=-1575.14   BIC=-1562.55
```

```
auto.fit.arima(X3$X, show_info=F)
```

```
Series: serie
ARIMA(1,0,2) with non-zero mean

Coefficients:
          ar1  ma1     ma2     mean
      -0.3399    0  0.4990  -0.1648
s.e.   0.0441    0  0.0411   0.0025

sigma^2 = 0.002591:  log likelihood = 780.55
AIC=-1553.09   AICc=-1553.01   BIC=-1536.23
```

```
auto.fit.arima(residuals$X, show_info=F)
```

```
Series: serie
ARIMA(2,0,1) with zero mean

Coefficients:
         ar1      ar2     ma1
      0.3267  -0.4059  0.3421
s.e.  0.0735   0.0525  0.0777

sigma^2 = 0.002422:  log likelihood = 797.37
AIC=-1586.73   AICc=-1586.65   BIC=-1569.87
```

```
auto.fit.arima(X4$X, show_info=F)
```

```
Series: serie
ARIMA(1,1,3)

Coefficients:
         ar1      ma1  ma2     ma3
      0.4303  -0.1705    0  0.4230
s.e.  0.0769   0.0784    0  0.0426

sigma^2 = 0.002462:  log likelihood = 791.72
AIC=-1575.45   AICc=-1575.37   BIC=-1558.6
```

```
auto.fit.arima(X5$X, show_info=F)
```

```
Series: serie
ARIMA(0,0,2) with zero mean

Coefficients:
      ma1      ma2
        0  -0.3412
```

```
s.e.     0   0.0428

sigma^2 = 0.002441:  log likelihood = 794.77
AIC=-1585.54   AICc=-1585.51   BIC=-1577.11
```

```
auto.fit.arima(X6$X, show_info=F)
```

```
Series: serie
ARIMA(2,1,1)

Coefficients:
          ar1     ar2     ma1
      -0.3434  0.3619  0.4754
s.e.   0.0920  0.0432  0.0942

sigma^2 = 0.002508:  log likelihood = 787.34
AIC=-1566.67   AICc=-1566.59   BIC=-1549.82
```

Podemos hacer la misma comprobación con las series del ejemplo 2