Simulaciones

Ana Xiangning Pereira Ezquerro

Versión 08 septiembre, 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 la 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 el 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 errores estacionarios

En esta sección veremos cómo se comporta la función de selección automática sobre ejemplos muy básicos donde los errores del modelo son estacionarios:

$$Y_t = \beta_0 + \beta_1 X^{(1)}_{t-r_1} + \beta_2 X^{(2)}_{t-r_2} + \cdots + X^{(p)}_{t-r_p} + \eta_t, \qquad \eta_t \sim \text{ARMA(p,q)}, \quad r_i \geq 0 \text{ para } i = 1, ..., p$$

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. En concreto, nuestro modelo tendrá la forma:

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

donde:

- $\eta_t \sim$ ARMA(2,1), por tanto, los errores son estacionarios.
- $X^{(1)}_t \sim$ ARIMA(2, 1, 3) y su coeficiente $\beta_1 = 2.8$.
- $X^{(2)}_t \sim$ ARIMA(1, 1, 2) y su coeficiente $\beta_2 = -1.12$.
- $X^{(3)}_t \sim$ ARMA(1, 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:

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

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

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

Selección de variables y ajuste del modelo: 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 <- cbind(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)
```

```
Se ha probado con la variable X1 [ic=-1135.19120552446, lag=0]
Se ha probado con la variable X2 [ic=-286.986760465024, lag=0]
Se ha probado con la variable X3 [ic=-640.026014781463, lag=0]
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=-185.777061985974, lag=0]
Se ha añadido la variable regresora X1 [aicc=-1135.19120552446, lag=0]
Series: serie
Regression with ARIMA(4,1,1) errors

Coefficients:
         ar1     ar2      ar3      ar4      ma1    xreg
      0.3013  0.2546  -0.1522  -0.0784  -0.9035  2.7997
s.e.  0.0378  0.0342   0.0335   0.0347   0.0218  0.0485

sigma^2 = 0.0184:  log likelihood = 574.65
AIC=-1135.31   AICc=-1135.19   BIC=-1101.03
-------------------------------------------------------------------------------------
Se ha probado con la variable X2 [ic=-1396.16397504914, lag=0]
Se ha probado con la variable X3 [ic=-2213.72165014961, 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=-2213.72165014961, lag=0]
Series: serie
Regression with ARIMA(0,1,3) errors

Coefficients:
          ma1  ma2      ma3     X1       X3
      -0.5907    0  -0.1460  2.7962  -2.2890
s.e.   0.0274    0   0.0269  0.0350   0.0505

sigma^2 = 0.006201:  log likelihood = 1111.89
AIC=-2213.78   AICc=-2213.72   BIC=-2189.3
-------------------------------------------------------------------------------------
Se ha probado con la variable X2 [ic=-3163.18780840121, lag=0]
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=-2213.72165025041, lag=-3]
Se ha añadido la variable regresora X2 [aicc=-3163.18780840121, lag=0]
Series: serie
Regression with ARIMA(2,0,1) errors

Coefficients:
          ar1     ar2     ma1  intercept      X1       X3       X2
      -0.1989  0.4054  0.4443     0.8056  2.7982  -2.2719  -1.1084
```

```
s.e.   0.0702  0.0302  0.0746     0.0034  0.0092   0.0320   0.0108

sigma^2 = 0.002375:  log likelihood = 1589.67
AIC=-3163.33   AICc=-3163.19   BIC=-3124.15
------------------------------------------------------------------------------------
No se ha podido encontrar un retardo significativo para X4
Se ha probado con la variable X5 [ic=-3163.18780355806, lag=-8]
No se ha podido encontrar un retardo significativo para X6
No se añaden más variables
------------------------------------------------------------------------------------
|                Histórico de variables añadidas al modelo (ndiff=0)               |
------------------------------------------------------------------------------------
 var lag                ic
  X1   0 -1135.19120552446
  X3   0 -2213.72165014961
  X2   0 -3163.18780840121
------------------------------------------------------------------------------------
Series: serie
Regression with ARIMA(2,0,1) errors

Coefficients:
          ar1     ar2     ma1  intercept      X1       X3       X2
      -0.1989  0.4054  0.4443     0.8056  2.7982  -2.2719  -1.1084
s.e.   0.0702  0.0302  0.0746     0.0034  0.0092   0.0320   0.0108

sigma^2 = 0.002375:  log likelihood = 1589.67
AIC=-3163.33   AICc=-3163.19   BIC=-3124.15
```

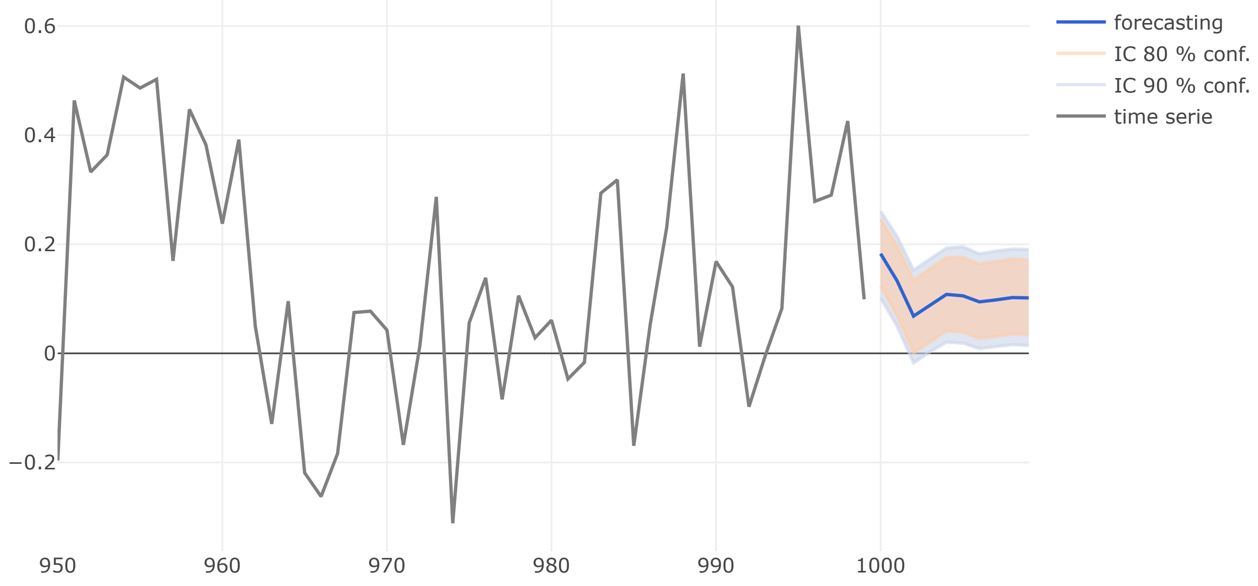
En el *output* de la función vemos cuál ha sido el proceso de selección de variables regresoras:

1. En la primera iteración se añade la variable $X_t^{(1)}$ con retardo nulo (algo que es correcto teniendo en cuenta cómo se ha creado el modelo en 1) y obteniendo un AICc=-1135.19120552446.
2. En la segunda iteración se introduce la variable $X_t^{(3)}$ con un retardo nulo y mejorando el AICc del modelo anterior (que sólo contaba con la variable $X_t^{(1)}$) con un AICc=-2213.72165014961.
3. En la tercera iteración se añade la variable $X_t^{(2)}$ con retardo nulo y mejorando el criterio de información del modelo anterior (que tenía las variables $X_t^{(1)}$ y $X_t^{(1)}$) con un AICc=-3163.18780840121.
4. En la siguiente iteración no se encuentran correlaciones significativas con ningún retardo para las variables $X_t^{(4)}$ y $X_t^{(6)}$, por lo que éstas no se pueden añadir al modelo. Para la varialbe $X_t^{(5)}$, se encuentra un retardo significativo en $k = -8$, pero no se mejora el AICc del modelo anterior (con las tres variables regresoras), por lo que se detiene el bucle para añadir variables.

Como el modelo resultante de haber añadido de forma iterativa todas las variables tiene errores estacionarios (siguen un ARIMA(2,0,1), lo que coincide con la simulación realizada), se considera como modelo válido para definir la relación entre la variable respuesta Y y el conjunto de variables regresoras seleccionadas ($X_t^{(1)}$, $X_t^{(2)}$ y $X_t^{(3)}$). Podemos observar cómo el valor del *intercept* y de los coeficientes de regresión se aproximan bastante a los coeficientes seleccionados al construir de forma artificial la variable regresora Y en @ref(eq:ejemplo1).

Predicción: Finalmente realizamos las predicciones puntuales:

```
preds <- forecast_model(Y, regresoras, ajuste, h=10, mode='bootstrap')
display(plot_forecast(preds, rang=c(950, 1009)), name='ejemplo1')
```



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

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

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

donde:

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

```
# Construimos el modelo
beta0 <- -0.6; beta1 <- 1.7; beta2 <- -2.2; beta3 <- 1.3
r1 <- 2; r3 <- 3
Y <- beta0 + beta1 * lag(X1$X, -r1) + beta2 * X2$X + beta3 * lag(X3$X, -r3) +
    residuals$X
```

Selección de variables y ajuste del modelo:

```
regresoras <- cbind(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,
                                    stationary_method='adf.test')
```

```
Se ha probado con la variable X1 [ic=-989.555684098221, lag=-2]
Se ha probado con la variable X2 [ic=-1156.68486061937, lag=0]
Se ha probado con la variable X3 [ic=-802.904991565362, lag=-3]
Se ha probado con la variable X4 [ic=-580.329461807603, lag=-1]
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=-1156.68486061937, lag=0]
Series: serie
Regression with ARIMA(4,1,0) errors

Coefficients:
          ar1      ar2      ar3     ar4     xreg
      -0.4245  -0.3382  -0.1395  0.1144  -2.2607
s.e.   0.0319   0.0344   0.0344  0.0320   0.0789

sigma^2 = 0.01767:  log likelihood = 584.39
AIC=-1156.77   AICc=-1156.68   BIC=-1127.5
------------------------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-2171.66958134745, lag=-2]
Se ha probado con la variable X3 [ic=-1561.28469740141, lag=-3]
No se ha podido encontrar un retardo significativo para X4
Se ha probado con la variable X5 [ic=-1156.6848606207, lag=-21]
No se ha podido encontrar un retardo significativo para X6
Se ha añadido la variable regresora X1 [aicc=-2171.66958134745, lag=-2]
Series: serie
Regression with ARIMA(3,0,0) errors

Coefficients:
         ar1     ar2      ar3  intercept       X2      X1
      0.3022  0.3526  -0.1198    -0.6039  -2.2393  1.7268
s.e.  0.0318  0.0313   0.0319     0.0065   0.0200  0.0171

sigma^2 = 0.006229:  log likelihood = 1092.89
AIC=-2171.79   AICc=-2171.67   BIC=-2137.62
------------------------------------------------------------------------------------
Se ha probado con la variable X3 [ic=-3108.15443209894, lag=-3]
Se ha probado con la variable X4 [ic=-2171.66942313677, lag=-15]
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=-3108.15443209894, lag=-3]
Series: serie
Regression with ARIMA(0,0,4) errors

Coefficients:
         ma1     ma2  ma3     ma4  intercept       X2      X1      X3
      0.2498  0.3360    0  0.1589    -0.5947  -2.1868  1.6949  1.3083
s.e.  0.0304  0.0302    0  0.0300     0.0033   0.0105  0.0089  0.0320

sigma^2 = 0.002377:  log likelihood = 1562.15
AIC=-3108.3   AICc=-3108.15   BIC=-3069.26
------------------------------------------------------------------------------------
No se ha podido encontrar un retardo significativo para X4
Se ha probado con la variable X5 [ic=-3108.15429253356, lag=-8]
No se ha podido encontrar un retardo significativo para X6
```

```
No se añaden más variables
------------------------------------------------------------------------------------
|                Histórico de variables añadidas al modelo (ndiff=0)               |
------------------------------------------------------------------------------------
 var lag                ic
  X2   0 -1156.68486061937
  X1  -2 -2171.66958134745
  X3  -3 -3108.15443209894
------------------------------------------------------------------------------------
Series: serie
Regression with ARIMA(0,0,4) errors

Coefficients:
         ma1     ma2  ma3     ma4  intercept       X2      X1      X3
      0.2498  0.3360    0  0.1589    -0.5947  -2.1868  1.6949  1.3083
s.e.  0.0304  0.0302    0  0.0300     0.0033   0.0105  0.0089  0.0320

sigma^2 = 0.002377:  log likelihood = 1562.15
AIC=-3108.3   AICc=-3108.15   BIC=-3069.26
```

De nuevo, observando el *output* de la función podemos analizar la selección de variables:

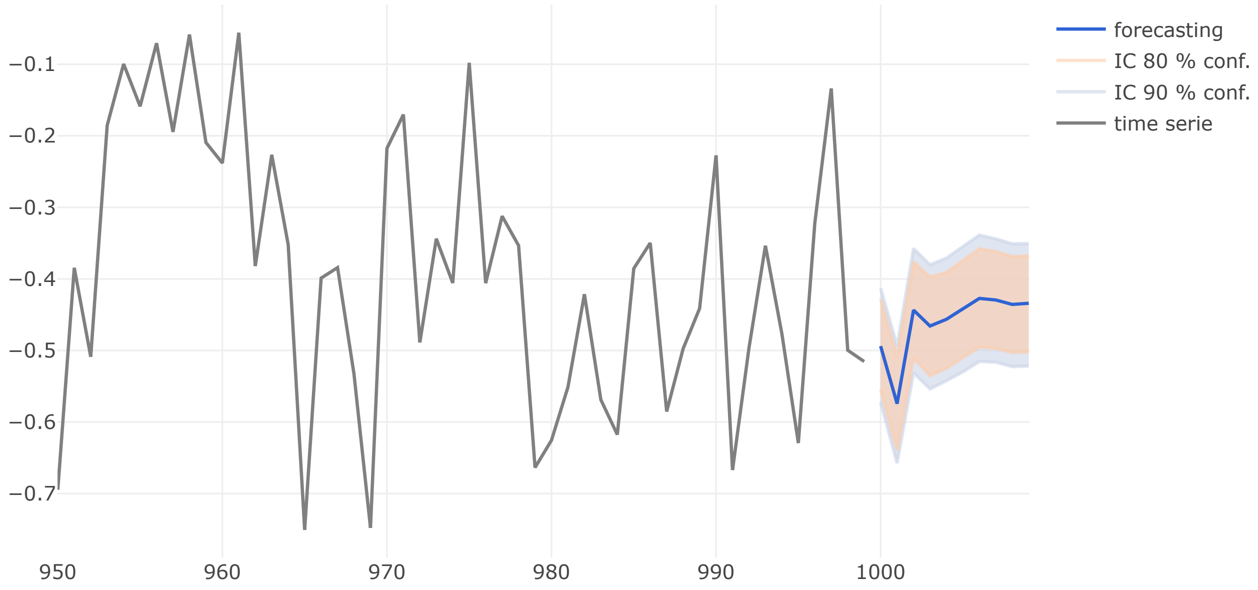
1. Primero se selecciona la variable $X_t^{(2)}$ con retardo nulo, construyendo un modelo de regresión con errores que siguen un ARIMA(4,1,0) y un AICc=-1156.68486061937.
2. En la siguiente iteración se selecciona la variable $X_t^{(1)}$ con un retardo $r_1 = 2$, mejorando el AICc del modelo anterior con un nuevo AICc=-2171.66958134745.
3. En la siguiente iteración se añade la variable $X_t^{(3)}$ con retardo $r_3 = 3$ al construir un modelo de regresión con variables regresoras $X_t^{(1)}$, $X_t^{(2)}$ y $X_t^{(3)}$ con respectivos retardos $r_1 = 2$, $r_2 = 0$ y $r_3 = 3$, consiguiendo un AICc=-3108.15443209894.
4. En la siguiente iteración no se encuentran retardos significativos para $X_t^{(4)}$ y $X_t^{(6)}$, pero sí para $X_t^{(5)}$ con un retardo $r_5 = 8$.. No obstante, añadir esta variable al modelo no supone mejorar el AICc del modelo anterior (sólo se consigue un AICc=-3106.1490878263). Por tanto se detiene la selección de varaibles.

Como el mejor modelo conseguido (el de la tercera iteración) ya tiene errores estacionarios (siguen un ARIMA(0,0,4)), se escoge dicho ajuste para modelizar la dependencia entre Y y las variables regresoras escogidas.

Podemos observar que los valores de los coeficientes de regresión y del *intercept* se aproximan bien a los valores verdaderos que se seleccionaron en la construcción del modelo. No obstante, los errores son modelizables con un ARIMA(0,0,4), no con un ARIMA(2,0,1).

Predicción:

```r
# Podemos mostrar las predicciones puntuales
preds <- forecast_model(Y, regresoras, ajuste, h=10, mode='bootstrap')
display(plot_forecast(preds, rang=c(950, 1009)), name='ejemplo2')
```



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

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

$$Y_t = \beta_0 + \beta_1 X^{(1)}_{t-r_1} + \beta_2 X^{(2)}_{t-r_2} + \cdots + X^{(p)}_{t-r_p} + \eta_t, \qquad \eta_t \sim \text{ARIMA(p,d,q)}$$

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

Tomemos el mismo modelo que en el primer ejemplo pero con errores no estacionarios:

$$Y_t = \beta_0 + \beta_1 X^{(1)}_t + \beta_2 X^{(2)}_t + \beta_3 X^{(3)}_t + \eta_t, \qquad \eta_t \sim \text{ARIMA(1,2,2)} \tag{3}$$

donde el conjunto de variables $\mathcal{X}$ sobre el que se realiza la selección está compuesto por las variables que sí influyen en Y:

- $X^{(1)}_t \sim$ ARIMA(2, 1, 3) y su coeficiente $\beta_1 = -1.3$.
- $X^{(2)}_t \sim$ ARIMA(1, 1, 2) y su coeficiente $\beta_2 = 2.12$.
- $X^{(3)}_t \sim$ ARMA(1, 2) y su coeficiente $\beta_3 = 2.3$.
- El *intercept* es $\beta_0 = 0.8$.

Y las variables que no interfieren en Y (las mismas que en el primer ejemplo).

```
# cargamos únicamente los residuos no estacionarios
load('simulations/residuals ~ ARIMA(1,2,2).RData')
```

```r
# volvemos a generar la variable respuesta
beta0 <- 0.8; beta1 <- -1.3; beta2 <- 2.12; beta3 <- 2.3
Y <- beta0 + beta1 * X1$X + beta2 * X2$X + beta3 * X3$X + 2.1*residuals$X
```

Selección de variables y ajuste del modelo: Ajustamos el modelo con las variables originales (no diferenciamos ninguna):

```r
regresoras <- cbind(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)
```

```
Se ha probado con la variable X1 [ic=59.5855339744035, lag=0]
Se ha probado con la variable X2 [ic=-181.525975728858, lag=0]
Se ha probado con la variable X3 [ic=-202.902218914428, lag=0]
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=199.408749112135, lag=-5]
Se ha añadido la variable regresora X3 [aicc=-202.902218914428, lag=0]
Series: serie
Regression with ARIMA(3,1,2) errors

Coefficients:
         ar1      ar2     ar3      ma1     ma2    xreg
      1.0348  -0.3048  0.2549  -0.7430  0.3264  2.3714
s.e.  0.1407   0.1250  0.0532   0.1408  0.0863  0.0980

sigma^2 = 0.04716:  log likelihood = 108.51
AIC=-203.02   AICc=-202.9   BIC=-168.74
-----------------------------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-417.422002505725, lag=0]
Se ha probado con la variable X2 [ic=-884.596379075807, lag=0]
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=-204.936549154963, lag=-5]
Se ha añadido la variable regresora X2 [aicc=-884.596379075807, lag=0]
Series: serie
Regression with ARIMA(4,1,1) errors

Coefficients:
      ar1     ar2     ar3     ar4     ma1      X3      X2
        0  0.4720  0.2733  0.2332  0.4715  2.3610  2.1784
s.e.    0  0.0323  0.0245  0.0299  0.0309  0.0658  0.0666

sigma^2 = 0.02366:  log likelihood = 449.36
AIC=-884.71   AICc=-884.6   BIC=-850.43
-----------------------------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-1613.80360585469, 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=-1613.80360585469, lag=0]
Series: serie
Regression with ARIMA(2,1,1) errors
```

```
Coefficients:
         ar1     ar2     ma1      X3      X2       X1
      0.2266  0.7606  0.3070  2.3763  2.1588  -1.2784
s.e.  0.0299  0.0295  0.0399  0.0447  0.0391   0.0336

sigma^2 = 0.01131:  log likelihood = 813.96
AIC=-1613.92   AICc=-1613.8   BIC=-1579.64
--------------------------------------------------------------------------------------
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
El modelo global no tiene errores estacionarios
Se intenta ajustar uno que sí los tenga
No se ha podido encontrar un modelo válido con errores estacionarios
Se aplica una diferenciación regular (ndiff=1) y se vuelve a llamar a la función
--------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------
Se ha probado con la variable X1 [ic=61.2589226747805, lag=0]
Se ha probado con la variable X2 [ic=-179.890806223957, lag=0]
Se ha probado con la variable X3 [ic=-201.85732451794, lag=0]
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=201.298523732159, lag=-5]
Se ha añadido la variable regresora X3 [aicc=-201.85732451794, lag=0]
Series: serie
Regression with ARIMA(3,1,0) errors

Coefficients:
          ar1      ar2      ar3    xreg
      -0.7053  -0.4727  -0.1673  2.3747
s.e.   0.0314   0.0355   0.0314  0.0993

sigma^2 = 0.04741:  log likelihood = 105.96
AIC=-201.92   AICc=-201.86   BIC=-177.44
--------------------------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-415.897154793375, lag=0]
Se ha probado con la variable X2 [ic=-624.559112508249, lag=0]
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=-206.141817506691, lag=-5]
Se ha añadido la variable regresora X2 [aicc=-624.559112508249, lag=0]
Series: serie
Regression with ARIMA(0,1,0) errors

Coefficients:
          X3      X2
      2.3654  2.1730
s.e.  0.0557  0.0599

sigma^2 = 0.03099:  log likelihood = 315.29
AIC=-624.58   AICc=-624.56   BIC=-609.9
--------------------------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-1209.59028329543, 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=-1209.59028329543, lag=0]
Series: serie
Regression with ARIMA(0,1,0) errors

Coefficients:
          X3      X2       X1
      2.3776  2.1642  -1.2493
s.e.  0.0414  0.0445   0.0441

sigma^2 = 0.01712:  log likelihood = 608.82
AIC=-1209.63   AICc=-1209.59   BIC=-1190.05
--------------------------------------------------------------------------------------
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
El modelo global no tiene errores estacionarios
Se intenta ajustar uno que sí los tenga
--------------------------------------------------------------------------------------
|                Histórico de variables añadidas al modelo (ndiff=1)                 |
--------------------------------------------------------------------------------------
 var lag               ic
  X3   0  -201.85732451794
  X2   0 -624.559112508249
  X1   0 -1209.59028329543
--------------------------------------------------------------------------------------
Series: serie
Regression with ARIMA(2,0,1) errors

Coefficients:
         ar1     ar2     ma1      X3      X2       X1
      0.2265  0.7606  0.3071  2.3763  2.1588  -1.2784
s.e.  0.0299  0.0295  0.0399  0.0447  0.0391   0.0336

sigma^2 = 0.01131:  log likelihood = 813.96
AIC=-1613.92   AICc=-1613.8   BIC=-1579.64
```

Si volvemos a analizar el *output* de la consola observamos que se ha realizado una diferenciación regular a los datos (dobles líneas horizontales) para conseguir un ajuste en el que los errores fuesen estacionarios.

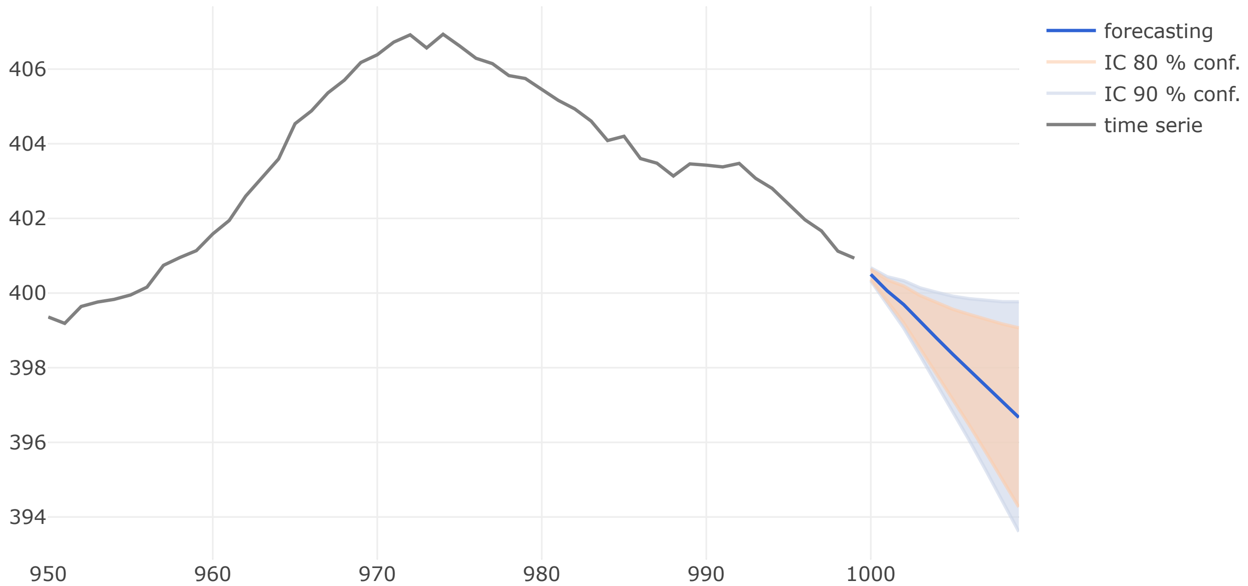
1. En las primeras iteraciones se añaden, en este orden, las variables $X_t^{(3)}$, $X_t^{(2)}$ y $X_t^{(1)}$ con retardos nulos, mejorando el AICc iterativamente hasta alcanzar un valor de AICc=-1613.80360585469. Como el modelo ajustado con estas tres variables regresoras no tiene errores estacionarios (se ajustó un ARIMA(2,1,1) para los residuos), se intenta ajustar un modelo donde el orden de d sea nulo. Como no se puede optimizar ("`No se ha podido encontrar un modelo válido con errores estacionarios...`"), se procede a aplicar una diferenciación regular a todos los datos, tanto variable respuesta como conjunto de posibles variables regresoras, y se vuelve a llamar a la función `auto.fit.arima.regression()` con los datos diferenciados.

2. En la siguiente llamada a la función se consigue añadir al modelo, en este orden, las variables $X_t^{(3)}$, $X_t^{(2)}$ y $X_t^{(1)}$ con retardos nulos, y se ajusta un ARIMA(0,1,0) para los errores de regresión. No obstante, como este ajuste no cumple la condición de errores estacionarios, se intenta ajustar un ARIMA para los errores donde $d = 0$. En este caso sí se consigue optimizar un modelo que respeta dicha condición, obteniendo un ajuste donde los residuos son estacionarios y el AICc=-1613.8.

Obsérvese que habiendo diferenciado los datos, se ha conseguido seleccionar las 3 variables regresoras que realmente tienen una influencia en la construcción de la variable respuesta Y con los retardos correctos.

Predicción: Una vez obtenido el modelo, se pueden realizar predicciones puntuales. Cuando se detecta que el ajuste se corresponde a un ajuste de los datos diferenciados, la función `forecast_model()` lanza un aviso de que se utilizarán los datos en unidades originales para realizar las predicciones.

```
preds <- forecast_model(Y, regresoras, ajuste, h=10, mode='bootstrap')
Se devuelven las predicciones en unidades originales...
display(plot_forecast(preds, rang=c(950, 1009)), name='ejemplo3')
```



2.2 Modelo donde $r_i \geq 0$ para $i = 1, ..., p$

Podemos alterar el ejemplo anterior para que las variables regresoras influyan en Y con cierto retardo.

- La variable $X_t^{(1)}$ se introduce con retardo $r_1 = 2$.
- La variable $X_t^{(3)}$ se introduce con retardo $r_3 = 1$.

```
beta0 <- 0.8; beta1 <- -1.3; beta2 <- 2.12; beta3 <- 2.3
r1 <- 2; r3 <- 1
Y <- beta0 + beta1 * lag(X1$X, -r1) + beta2 * X2$X +
    beta3 * lag(X3$X, -r3) + 1.5*residuals$X
```

Selección de variables y ajuste del modelo:

```r
regresoras <- cbind(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,
                                    stationary_method='adf.test')
```

```
Se ha probado con la variable X1 [ic=-128.730307266973, lag=-2]
Se ha probado con la variable X2 [ic=-353.706447261812, lag=0]
Se ha probado con la variable X3 [ic=-435.193689069136, lag=-1]
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=-435.193689069136, lag=-1]
Series: serie
Regression with ARIMA(3,1,2) errors

Coefficients:
         ar1  ar2     ar3      ma1     ma2    xreg
      0.6985    0  0.2839  -0.4997  0.0790  2.3230
s.e.  0.0440    0  0.0422   0.0558  0.0326  0.0903

sigma^2 = 0.03729:  log likelihood = 223.64
AIC=-435.28   AICc=-435.19   BIC=-405.92
--------------------------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-653.565041178663, lag=-2]
Se ha probado con la variable X2 [ic=-1254.38354028493, lag=0]
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=-438.084375852223, lag=-9]
Se ha añadido la variable regresora X2 [aicc=-1254.38354028493, lag=0]
Series: serie
Regression with ARIMA(3,1,2) errors

Coefficients:
         ar1      ar2     ar3      ma1     ma2      X3      X2
      1.0735  -0.5681  0.4771  -0.6904  0.5167  2.2884  2.1250
s.e.  0.0699   0.0736  0.0450   0.0701  0.0504  0.0530  0.0562

sigma^2 = 0.0162:  log likelihood = 635.27
AIC=-1254.53   AICc=-1254.38   BIC=-1215.38
--------------------------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-2269.80243895166, lag=-2]
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=-2269.80243895166, lag=-2]
Series: serie
Regression with ARIMA(2,1,1) errors

Coefficients:
         ar1     ar2     ma1      X3      X2       X1
      0.2283  0.7589  0.3025  2.2676  2.1484  -1.3085
s.e.  0.0299  0.0294  0.0396  0.0321  0.0280   0.0241
```

```
sigma^2 = 0.005786:  log likelihood = 1141.96
AIC=-2269.92   AICc=-2269.8   BIC=-2235.66
-------------------------------------------------------------------------------------
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
El modelo global no tiene errores estacionarios
Se intenta ajustar uno que sí los tenga
No se ha podido encontrar un modelo válido con errores estacionarios
Se aplica una diferenciación regular (ndiff=1) y se vuelve a llamar a la función
-------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-127.111191500974, lag=-2]
Se ha probado con la variable X2 [ic=-350.746172271828, lag=0]
Se ha probado con la variable X3 [ic=-432.584774965942, lag=-1]
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=-432.584774965942, lag=-1]
Series: serie
Regression with ARIMA(4,1,0) errors

Coefficients:
          ar1      ar2      ar3      ar4    xreg
      -0.7949  -0.6217  -0.2668  -0.0734  2.3192
s.e.   0.0318   0.0397   0.0398   0.0320  0.0907

sigma^2 = 0.03744:  log likelihood = 222.34
AIC=-432.67   AICc=-432.58   BIC=-403.31
-------------------------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-651.671017504022, lag=-2]
Se ha probado con la variable X2 [ic=-1252.41874509484, lag=0]
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=-437.623697388088, lag=-9]
Se ha añadido la variable regresora X2 [aicc=-1252.41874509484, lag=0]
Series: serie
Regression with ARIMA(2,1,2) errors

Coefficients:
      ar1      ar2      ma1     ma2      X3      X2
        0  -0.4898  -0.6156  0.4634  2.2891  2.1249
s.e.    0   0.0430   0.0308  0.0362  0.0531  0.0564

sigma^2 = 0.01629:  log likelihood = 632.25
AIC=-1252.5   AICc=-1252.42   BIC=-1223.15
-------------------------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-1863.55961789977, lag=-2]
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=-1863.55961789977, lag=-2]
Series: serie
```

```
Regression with ARIMA(0,1,0) errors

Coefficients:
          X3      X2       X1
      2.2783  2.1512  -1.3106
s.e.  0.0297  0.0319   0.0316

sigma^2 = 0.008783:  log likelihood = 935.8
AIC=-1863.6   AICc=-1863.56   BIC=-1844.03
--------------------------------------------------------------------------------------
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
El modelo global no tiene errores estacionarios
Se intenta ajustar uno que sí los tenga
--------------------------------------------------------------------------------------
|                Histórico de variables añadidas al modelo (ndiff=1)                  |
--------------------------------------------------------------------------------------
 var lag                ic
  X3  -1 -432.584774965942
  X2   0 -1252.41874509484
  X1  -2 -1863.55961789977
--------------------------------------------------------------------------------------
Series: serie
Regression with ARIMA(2,0,1) errors

Coefficients:
         ar1     ar2     ma1      X3      X2       X1
      0.2283  0.7589  0.3024  2.2677  2.1484  -1.3086
s.e.  0.0299  0.0295  0.0396  0.0321  0.0280   0.0241

sigma^2 = 0.005786:  log likelihood = 1141.96
AIC=-2269.92   AICc=-2269.8   BIC=-2235.66
```

Como podemos observar, el *output* es muy similar al obtenido en el ejemplo anterior. En las primeras iteraciones se añaden las variables $X_t^{(3)}$, $X_t^{(2)}$ y $X_t^{(1)}$, en ese orden, con retardos $r_3 = 1$, $r_2 = 0$ y $r_1 = 1$ (correcto según la definición del modelo), pero al no poder ajustar un modelo con errores estacionarios (se obtiene un modelo con errores ARIMA(2,1,1)), se tienen que diferenciar los datos y volver a llamar a la función.

En la siguiente llamada, se vuelven a añadir las mismas variables con los retardos correctos, pero esta vez sí se consigue ajustar un modelo con errores estacionarios (un ARIMA(2,0,1)) con un AICc=-2269.51.

3 Comparativa del método de preblanqueado

3.1 Con errores estacionarios

```
load(file='simulations/residuals ~ ARIMA(2,0,1).RData') # residuals
beta0 <- -0.1; beta1 <- 3.2; beta2 <- -2.5
r1 <- 2; r2 <- 3
Y <- beta0 + beta1 * lag(X1$X, -r1) + beta2 * lag(X2$X, -r2) + residuals$X
regresoras <- cbind(X1=X1$X, X2=X2$X, X3=X3$X, X4=X4$X, X5=X5$X, X6=X6$X)
```

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

```
ajuste <- auto.fit.arima.regression(Y, regresoras, show_info=T,
                                    stationary_method='auto.arima')
```

```
Se ha probado con la variable X1 [ic=-1088.61845492454, lag=-2]
Se ha probado con la variable X2 [ic=-595.912721666589, lag=-3]
Se ha probado con la variable X3 [ic=-146.378165277729, lag=-5]
No se ha podido encontrar un retardo significativo para X4
Se ha probado con la variable X5 [ic=-142.384016382745, lag=-21]
Se ha probado con la variable X6 [ic=-142.384016382745, lag=-2]
Se ha añadido la variable regresora X1 [aicc=-1088.61845492454, lag=-2]
Series: serie
Regression with ARIMA(0,1,2) errors

Coefficients:
          ma1      ma2    xreg
      -0.4887  -0.1949  3.1621
s.e.   0.0314   0.0316  0.0684

sigma^2 = 0.01904:  log likelihood = 548.33
AIC=-1088.66   AICc=-1088.62   BIC=-1069.13
-------------------------------------------------------------------------------------
Se ha probado con la variable X2 [ic=-3114.07541439909, lag=-3]
Se ha probado con la variable X3 [ic=-1088.61845591082, lag=-6]
Se ha probado con la variable X4 [ic=-1088.61845475434, lag=-9]
No se ha podido encontrar un retardo significativo para X5
Se ha probado con la variable X6 [ic=-1088.61845508187, lag=-8]
Se ha añadido la variable regresora X2 [aicc=-3114.07541439909, lag=-3]
Series: serie
Regression with ARIMA(2,0,1) errors

Coefficients:
          ar1     ar2     ma1  intercept      X1       X2
      -0.1927  0.4046  0.4424    -0.0938  3.1927  -2.4812
s.e.   0.0713  0.0307  0.0757     0.0035  0.0093   0.0109

sigma^2 = 0.002381:  log likelihood = 1564.1
AIC=-3114.19   AICc=-3114.08   BIC=-3080.01
-------------------------------------------------------------------------------------
No se ha podido encontrar un retardo significativo para X3
No se ha podido encontrar un retardo significativo para X4
```

```
Se ha probado con la variable X5 [ic=-3114.07521810619, lag=-8]
No se ha podido encontrar un retardo significativo para X6
No se añaden más variables
-------------------------------------------------------------------------------------
|                Histórico de variables añadidas al modelo (ndiff=0)                |
-------------------------------------------------------------------------------------
 var lag               ic
  X1  -2 -1088.61845492454
  X2  -3 -3114.07541439909
-------------------------------------------------------------------------------------
Series: serie
Regression with ARIMA(2,0,1) errors

Coefficients:
          ar1     ar2     ma1  intercept      X1       X2
      -0.1927  0.4046  0.4424    -0.0938  3.1927  -2.4812
s.e.   0.0713  0.0307  0.0757     0.0035  0.0093   0.0109

sigma^2 = 0.002381:  log likelihood = 1564.1
AIC=-3114.19   AICc=-3114.08   BIC=-3080.01
```

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

```
ajuste <- auto.fit.arima.regression(Y, regresoras, show_info=T,
                                    stationary_method='adf.test')
```

```
Se ha probado con la variable X1 [ic=-1088.61845492454, lag=-2]
Se ha probado con la variable X2 [ic=-595.912721666589, lag=-3]
Se ha probado con la variable X3 [ic=-146.378165277729, lag=-5]
No se ha podido encontrar un retardo significativo para X4
Se ha probado con la variable X5 [ic=-142.384016382745, lag=-21]
Se ha probado con la variable X6 [ic=-142.384016382745, lag=-2]
Se ha añadido la variable regresora X1 [aicc=-1088.61845492454, lag=-2]
Series: serie
Regression with ARIMA(0,1,2) errors

Coefficients:
          ma1      ma2    xreg
      -0.4887  -0.1949  3.1621
s.e.   0.0314   0.0316  0.0684

sigma^2 = 0.01904:  log likelihood = 548.33
AIC=-1088.66   AICc=-1088.62   BIC=-1069.13
-------------------------------------------------------------------------------------
Se ha probado con la variable X2 [ic=-3114.07541439909, lag=-3]
Se ha probado con la variable X3 [ic=-1088.61845591082, lag=-6]
Se ha probado con la variable X4 [ic=-1088.61845475434, lag=-9]
No se ha podido encontrar un retardo significativo para X5
Se ha probado con la variable X6 [ic=-1088.61845508187, lag=-8]
Se ha añadido la variable regresora X2 [aicc=-3114.07541439909, lag=-3]
Series: serie
Regression with ARIMA(2,0,1) errors

Coefficients:
          ar1     ar2     ma1  intercept      X1       X2
```

```
      -0.1927  0.4046  0.4424    -0.0938  3.1927  -2.4812
s.e.   0.0713  0.0307  0.0757     0.0035  0.0093   0.0109

sigma^2 = 0.002381:  log likelihood = 1564.1
AIC=-3114.19   AICc=-3114.08   BIC=-3080.01
-------------------------------------------------------------------------------------
No se ha podido encontrar un retardo significativo para X3
No se ha podido encontrar un retardo significativo para X4
Se ha probado con la variable X5 [ic=-3114.07521810619, lag=-8]
No se ha podido encontrar un retardo significativo para X6
No se añaden más variables
-------------------------------------------------------------------------------------
|                Histórico de variables añadidas al modelo (ndiff=0)                 |
-------------------------------------------------------------------------------------
 var lag                ic
  X1  -2 -1088.61845492454
  X2  -3 -3114.07541439909
-------------------------------------------------------------------------------------
Series: serie
Regression with ARIMA(2,0,1) errors

Coefficients:
          ar1     ar2     ma1  intercept      X1       X2
      -0.1927  0.4046  0.4424    -0.0938  3.1927  -2.4812
s.e.   0.0713  0.0307  0.0757     0.0035  0.0093   0.0109

sigma^2 = 0.002381:  log likelihood = 1564.1
AIC=-3114.19   AICc=-3114.08   BIC=-3080.01
```

3.2 Con errores no estacionarios

```r
load(file='simulations/residuals ~ ARIMA(1,2,2).RData') # residuals
beta0 <- -0.1; beta1 <- 3.2; beta2 <- -2.5
r1 <- 2; r2 <- 3
Y <- beta0 + beta1 * lag(X1$X, -r1) + beta2 * lag(X2$X, -r2) + residuals$X
regresoras <- cbind(X1=X1$X, X2=X2$X, X3=X3$X, X4=X4$X, X5=X5$X, X6=X6$X)
```

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

```r
ajuste <- auto.fit.arima.regression(Y, regresoras, show_info=T,
                                    stationary_method='auto.arima')
```

```
Se ha probado con la variable X1 [ic=-649.682587578322, lag=-2]
Se ha probado con la variable X2 [ic=-434.459021738045, lag=-3]
Se ha probado con la variable X3 [ic=57.7067437758955, lag=-1]
Se ha probado con la variable X4 [ic=53.6412837880618, lag=-9]
Se ha probado con la variable X5 [ic=57.7067437758955, lag=-23]
Se ha probado con la variable X6 [ic=60.6278346700799, lag=-6]
Se ha añadido la variable regresora X1 [aicc=-649.682587578322, lag=-2]
Series: serie
Regression with ARIMA(2,1,3) errors
```

```
Coefficients:
         ar1     ar2      ma1      ma2     ma3    xreg
      0.5931  0.3925  -0.5505  -0.3111  0.2199  3.1695
s.e.  0.1656  0.1644   0.1633   0.1478  0.0367  0.0884

sigma^2 = 0.02971:  log likelihood = 331.9
AIC=-649.8   AICc=-649.68   BIC=-615.64
--------------------------------------------------------------------------------------
Se ha probado con la variable X2 [ic=-3025.1467782192, lag=-3]
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
Se ha probado con la variable X6 [ic=-652.50410291203, lag=-8]
Se ha añadido la variable regresora X2 [aicc=-3025.1467782192, lag=-3]
Series: serie
Regression with ARIMA(2,1,1) errors

Coefficients:
         ar1     ar2     ma1      X1       X2
      0.2283  0.7584  0.3004  3.1983  -2.4964
s.e.  0.0299  0.0295  0.0395  0.0163   0.0188

sigma^2 = 0.002584:  log likelihood = 1518.62
AIC=-3025.23   AICc=-3025.15   BIC=-2995.95
--------------------------------------------------------------------------------------
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
No se añaden más variables
El modelo global no tiene errores estacionarios
Se intenta ajustar uno que sí los tenga
No se ha podido encontrar un modelo válido con errores estacionarios
Se aplica una diferenciación regular (ndiff=1) y se vuelve a llamar a la función
--------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-646.897656850379, lag=-2]
Se ha probado con la variable X2 [ic=-430.04365565032, lag=-3]
Se ha probado con la variable X3 [ic=60.6278158177817, lag=-1]
Se ha probado con la variable X4 [ic=53.6412661805395, lag=-9]
Se ha probado con la variable X5 [ic=60.6278158177817, lag=-23]
Se ha probado con la variable X6 [ic=60.6278158177817, lag=-6]
Se ha añadido la variable regresora X1 [aicc=-646.897656850379, lag=-2]
Series: serie
Regression with ARIMA(1,1,3) errors

Coefficients:
          ar1      ma1      ma2     ma3    xreg
      -0.4113  -0.5402  -0.3262  0.2186  3.1688
s.e.   0.1622   0.1597   0.1446  0.0361  0.0882

sigma^2 = 0.02984:  log likelihood = 329.49
AIC=-646.98   AICc=-646.9   BIC=-617.71
--------------------------------------------------------------------------------------
```

```
Se ha probado con la variable X2 [ic=-3087.52651374318, lag=-3]
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
Se ha probado con la variable X6 [ic=-651.270885560183, lag=-8]
Se ha añadido la variable regresora X2 [aicc=-3087.52651374318, lag=-3]
Series: serie
Regression with ARIMA(1,1,2) errors

Coefficients:
          ar1  ma1     ma2     X1       X2
      -0.4461    0  0.3683  3.2043  -2.4945
s.e.   0.0306    0  0.0308  0.0129   0.0165

sigma^2 = 0.002427:  log likelihood = 1548.79
AIC=-3087.59   AICc=-3087.53   BIC=-3063.19
---------------------------------------------------------------------------------------
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
No se añaden más variables
El modelo global no tiene errores estacionarios
Se intenta ajustar uno que sí los tenga
---------------------------------------------------------------------------------------
|                Histórico de variables añadidas al modelo (ndiff=1)                  |
---------------------------------------------------------------------------------------
 var lag                ic
  X1  -2 -646.897656850379
  X2  -3 -3087.52651374318
---------------------------------------------------------------------------------------
Series: serie
Regression with ARIMA(2,0,1) errors

Coefficients:
         ar1     ar2     ma1      X1       X2
      0.2283  0.7584  0.3004  3.1983  -2.4964
s.e.  0.0299  0.0295  0.0395  0.0163   0.0188

sigma^2 = 0.002584:  log likelihood = 1518.62
AIC=-3025.23   AICc=-3025.15   BIC=-2995.95
```

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

```
ajuste <- auto.fit.arima.regression(Y, regresoras, show_info=T,
                                    stationary_method='adf.test')
```

```
Se ha probado con la variable X1 [ic=-649.682587578322, lag=-2]
Se ha probado con la variable X2 [ic=-434.459021738045, lag=-3]
Se ha probado con la variable X3 [ic=57.7067437758955, lag=-1]
Se ha probado con la variable X4 [ic=53.6412837880618, lag=-9]
Se ha probado con la variable X5 [ic=57.7067437758955, lag=-23]
Se ha probado con la variable X6 [ic=60.6278346700799, lag=-6]
Se ha añadido la variable regresora X1 [aicc=-649.682587578322, lag=-2]
Series: serie
```

```
Regression with ARIMA(2,1,3) errors

Coefficients:
         ar1     ar2      ma1      ma2     ma3    xreg
      0.5931  0.3925  -0.5505  -0.3111  0.2199  3.1695
s.e.  0.1656  0.1644   0.1633   0.1478  0.0367  0.0884

sigma^2 = 0.02971:  log likelihood = 331.9
AIC=-649.8   AICc=-649.68   BIC=-615.64
--------------------------------------------------------------------------------------
Se ha probado con la variable X2 [ic=-3025.1467782192, lag=-3]
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
Se ha probado con la variable X6 [ic=-652.50410291203, lag=-8]
Se ha añadido la variable regresora X2 [aicc=-3025.1467782192, lag=-3]
Series: serie
Regression with ARIMA(2,1,1) errors

Coefficients:
         ar1     ar2     ma1      X1       X2
      0.2283  0.7584  0.3004  3.1983  -2.4964
s.e.  0.0299  0.0295  0.0395  0.0163   0.0188

sigma^2 = 0.002584:  log likelihood = 1518.62
AIC=-3025.23   AICc=-3025.15   BIC=-2995.95
--------------------------------------------------------------------------------------
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
No se añaden más variables
El modelo global no tiene errores estacionarios
Se intenta ajustar uno que sí los tenga
No se ha podido encontrar un modelo válido con errores estacionarios
Se aplica una diferenciación regular (ndiff=1) y se vuelve a llamar a la función
--------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------
Se ha probado con la variable X1 [ic=-646.897656850379, lag=-2]
Se ha probado con la variable X2 [ic=-430.04365565032, lag=-3]
Se ha probado con la variable X3 [ic=60.6278158177817, lag=-1]
Se ha probado con la variable X4 [ic=53.6412661805395, lag=-9]
Se ha probado con la variable X5 [ic=60.6278158177817, lag=-23]
Se ha probado con la variable X6 [ic=60.6278158177817, lag=-6]
Se ha añadido la variable regresora X1 [aicc=-646.897656850379, lag=-2]
Series: serie
Regression with ARIMA(1,1,3) errors

Coefficients:
          ar1      ma1      ma2     ma3    xreg
      -0.4113  -0.5402  -0.3262  0.2186  3.1688
s.e.   0.1622   0.1597   0.1446  0.0361  0.0882

sigma^2 = 0.02984:  log likelihood = 329.49
```

```
AIC=-646.98   AICc=-646.9   BIC=-617.71
--------------------------------------------------------------------------------------
Se ha probado con la variable X2 [ic=-3087.52651374318, lag=-3]
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
Se ha probado con la variable X6 [ic=-651.270885560183, lag=-8]
Se ha añadido la variable regresora X2 [aicc=-3087.52651374318, lag=-3]
Series: serie
Regression with ARIMA(1,1,2) errors

Coefficients:
          ar1  ma1     ma2      X1       X2
      -0.4461    0  0.3683  3.2043  -2.4945
s.e.   0.0306    0  0.0308  0.0129   0.0165

sigma^2 = 0.002427:  log likelihood = 1548.79
AIC=-3087.59   AICc=-3087.53   BIC=-3063.19
--------------------------------------------------------------------------------------
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
No se añaden más variables
El modelo global no tiene errores estacionarios
Se intenta ajustar uno que sí los tenga
--------------------------------------------------------------------------------------
|                Histórico de variables añadidas al modelo (ndiff=1)                 |
--------------------------------------------------------------------------------------
 var lag                ic
  X1  -2 -646.897656850379
  X2  -3 -3087.52651374318
--------------------------------------------------------------------------------------
Series: serie
Regression with ARIMA(2,0,1) errors

Coefficients:
         ar1     ar2     ma1      X1       X2
      0.2283  0.7584  0.3004  3.1983  -2.4964
s.e.  0.0299  0.0295  0.0395  0.0163   0.0188

sigma^2 = 0.002584:  log likelihood = 1518.62
AIC=-3025.23   AICc=-3025.15   BIC=-2995.95
```

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 ejemplo, las muestras simuladas eran las siguientes:

```
load(file='simulations/residuals ~ ARIMA(2,0,1).RData') # residuals
```

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  -0.4505  -0.0720  -0.1633  0.4002
s.e.    0   0.0465   0.0321   0.0443  0.0304

sigma^2 = 0.002651:  log likelihood = 1547.33
AIC=-3084.67   AICc=-3084.61   BIC=-3060.13
```

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

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

Coefficients:
          ar1  ma1      ma2
      -0.4570    0  -0.4718
s.e.   0.0317    0   0.0315

sigma^2 = 0.002683:  log likelihood = 1540.68
AIC=-3075.37   AICc=-3075.34   BIC=-3060.65
```

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

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

Coefficients:
         ar1  ma1     ma2
      0.3602    0  0.3508
s.e.  0.0311    0  0.0313

sigma^2 = 0.002282:  log likelihood = 1623.08
AIC=-3240.16   AICc=-3240.14   BIC=-3225.44
```

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

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

```
Coefficients:
          ar1     ar2     ma1
      -0.2356  0.4323  0.4563
s.e.   0.0626  0.0295  0.0668

sigma^2 = 0.002469:  log likelihood = 1584.27
AIC=-3160.53   AICc=-3160.49   BIC=-3140.9
```

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

```
Series: serie
ARIMA(1,0,3) with zero mean

Coefficients:
          ar1     ma1  ma2     ma3
      -0.5708  0.1396    0  0.4028
s.e.   0.0633  0.0650    0  0.0319

sigma^2 = 0.002529:  log likelihood = 1572.27
AIC=-3136.55   AICc=-3136.51   BIC=-3116.92
```

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

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

Coefficients:
          ar1     ar2  ma1     ma2
      -0.0623  0.4527    0  0.4725
s.e.   0.0283  0.0373    0  0.0377

sigma^2 = 0.002489:  log likelihood = 1577.98
AIC=-3147.95   AICc=-3147.91   BIC=-3128.32
```

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

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

Coefficients:
          ar1     ar2     ma1
      -0.2399  0.4477  0.4440
s.e.   0.0597  0.0290  0.0645

sigma^2 = 0.002464:  log likelihood = 1583.7
AIC=-3159.4   AICc=-3159.36   BIC=-3139.78
```

Podemos hacer la misma comprobación con los residuos del ejemplo 2.

```
load('simulations/residuals ~ ARIMA(1,2,2).RData')
auto.fit.arima(residuals$X, show_info=F)
```

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

Coefficients:
```

```
          ar1  ma1     ma2
      -0.4430    0  0.3639
s.e.   0.0303    0  0.0305

sigma^2 = 0.002414:  log likelihood = 1591.83
AIC=-3177.65   AICc=-3177.63   BIC=-3162.93
```