# Lección 8.A — Número de viajeros internacionales

Marcos Bujosa

## Número de viajeros internacionales

Guión: P-L08-A-airlinePass.inp

En esta práctica volvemos sobre la primera serie temporal estudiada en el curso: la serie temporal mensual correspondiente al número total de pasajeros (en miles) de vuelos internacionales de una importante aerolínea de EEUU que aparece en manual de Box & Jenkins.

**Objetivo**

1. Identificar un modelo para la serie temporal
2. Pronosticar los datos correspondientes a los meses del último año de la muestra.

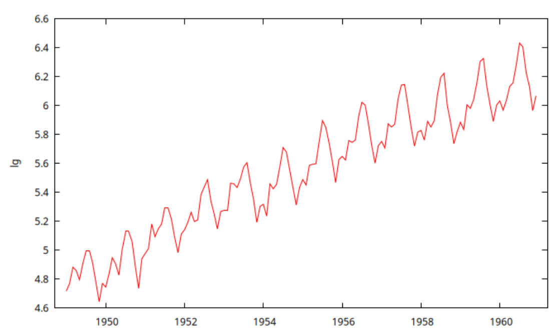
Comencemos cargando los datos:
***Archivo -->Abrir datos -->Archivo de muestra*** y en la pestaña ***Gretl*** seleccione `bjg`.
*o bien teclee en linea de comandos*:

```
open bjg
```

## Actividad 1 - Gráfico de series temporales

Obtenga la figura de la serie temporal en logaritmos `lg`.

```
gnuplot lg --time-series --with-lines  --output="log_AP.png"
```




## Actividad 2 - Identificar un modelo ARIMA para la serie temporal

### Analizando los datos de la serie en logs tras una diferencia regular y otra estacional

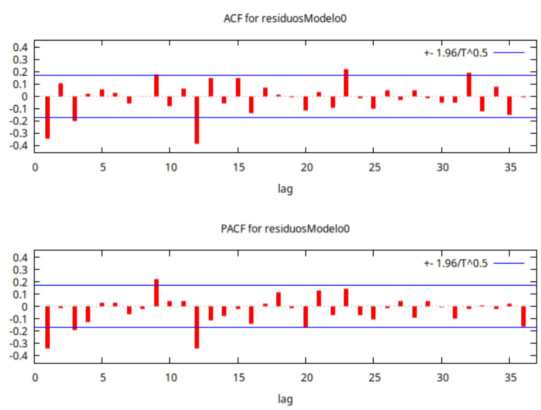
Ajuste un modelo ARIMA con constante sin parte AR ni parte MA, pero indicando una diferencia regular u otra estacional.

```
arima 0 1 0 ; 0 1 0 ; lg
```

```
Model 2: ARIMA, using observations 1950:02-1960:12 (T = 131)
Estimated using least squares (= MLE)
Dependent variable: (1-L)(1-Ls) lg

             coefficient   std. error      z      p-value
  --------------------------------------------------------
  const      0.000290880   0.00400578   0.07261   0.9421

Mean dependent var   0.000291   S.D. dependent var   0.045848
Mean of innovations  0.000000   S.D. of innovations  0.045848
R-squared            0.986863   Adjusted R-squared   0.986963
Log-likelihood       218.4176   Akaike criterion    -432.8353
Schwarz criterion   -427.0849   Hannan-Quinn        -430.4986
```

Y ahora analice el correlograma de los residuos hasta el retardo 36.

```
residuosModelo0 = $uhat
corrgm residuosModelo0 36 --plot="residuosModelo0-ACF-PACF.png"
```




```
Autocorrelation function for residuosModelo0
***, **, * indicate significance at the 1%, 5%, 10% levels
using standard error 1/T^0.5

  LAG      ACF          PACF         Q-stat. [p-value]

    1  -0.3411  ***  -0.3411 ***     15.5957  [0.000]
    2   0.1050       -0.0128         17.0860  [0.000]
    3  -0.2021  **   -0.1927 **      22.6478  [0.000]
    4   0.0214       -0.1250         22.7104  [0.000]
    5   0.0557        0.0331         23.1387  [0.000]
    6   0.0308        0.0347         23.2709  [0.001]
    7  -0.0556       -0.0602         23.7050  [0.001]
    8  -0.0008       -0.0202         23.7050  [0.003]
    9   0.1764  **    0.2256 ***     28.1473  [0.001]
   10  -0.0764        0.0431         28.9869  [0.001]
   11   0.0644        0.0466         29.5887  [0.002]
   12  -0.3866  ***  -0.3387 ***     51.4728  [0.000]
   13   0.1516  *    -0.1092         54.8664  [0.000]
   14  -0.0576       -0.0768         55.3605  [0.000]
   15   0.1496  *    -0.0218         58.7204  [0.000]
   16  -0.1389       -0.1395         61.6452  [0.000]
   17   0.0705        0.0259         62.4045  [0.000]
   18   0.0156        0.1148         62.4421  [0.000]
   19  -0.0106       -0.0132         62.4596  [0.000]
   20  -0.1167       -0.1674 *       64.5984  [0.000]
```

```
21   0.0386        0.1324        64.8338  [0.000]
22  -0.0914       -0.0720        66.1681  [0.000]
23   0.2233  **    0.1429        74.2099  [0.000]
24  -0.0184       -0.0673        74.2652  [0.000]
25  -0.1003       -0.1027        75.9183  [0.000]
26   0.0486       -0.0101        76.3097  [0.000]
27  -0.0302        0.0438        76.4629  [0.000]
28   0.0471       -0.0900        76.8387  [0.000]
29  -0.0180        0.0469        76.8943  [0.000]
30  -0.0511       -0.0049        77.3442  [0.000]
31  -0.0538       -0.0964        77.8478  [0.000]
32   0.1957  **   -0.0153        84.5900  [0.000]
33  -0.1224        0.0115        87.2543  [0.000]
34   0.0777       -0.0192        88.3401  [0.000]
35  -0.1525  *     0.0230        92.5584  [0.000]
36  -0.0100       -0.1649 *      92.5767  [0.000]
```

Fijémonos en los retardos estacionales.

- En la ACF son significativos el 12, pero no el 24 ni el 36.
- En la PACF son significativos el 12 y el 36.

Esto sugiere un truncamiento en la ACF pero no el la PACF. Por tanto, probemos con un MA(1) estacional.

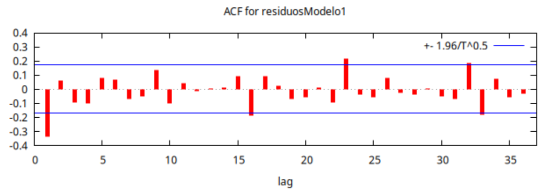
## Probando con un MA(1) estacional

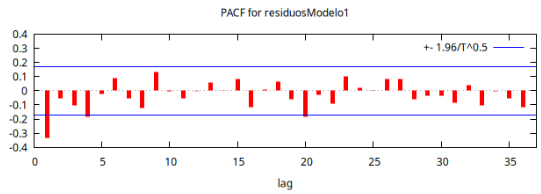
```
arima 0 1 0 ; 0 1 1 ; lg
```

```
Function evaluations: 18
Evaluations of gradient: 6

Model 4: ARIMA, using observations 1950:02-1960:12 (T = 131)
Estimated using AS 197 (exact ML)
Dependent variable: (1-L)(1-Ls) lg
Standard errors based on Hessian

             coefficient    std. error      z      p-value
  --------------------------------------------------------
  const      -0.000211216   0.00159988   -0.1320   0.8950
  Theta_1    -0.602630      0.0785055    -7.676    1.64e-14 ***

Mean dependent var   0.000291   S.D. dependent var   0.045848
Mean of innovations  0.000883   S.D. of innovations  0.039184
R-squared            0.990288   Adjusted R-squared   0.990288
Log-likelihood       235.7851   Akaike criterion    -465.5701
Schwarz criterion   -456.9445   Hannan-Quinn        -462.0652

                        Real  Imaginary    Modulus  Frequency
  -----------------------------------------------------------
  MA (seasonal)
    Root  1           1.6594     0.0000     1.6594     0.0000
  -----------------------------------------------------------
```

Y ahora analice el correlograma de los residuos hasta el retardo 36.

```
residuosModelo1 = $uhat
corrgm residuosModelo1 36 --plot="residuosModelo1-ACF-PACF.png"
```




```
Autocorrelation function for residuosModelo1
***, **, * indicate significance at the 1%, 5%, 10% levels
using standard error 1/T^0.5

  LAG      ACF          PACF         Q-stat. [p-value]

    1  -0.3340  ***  -0.3340 ***     14.9506  [0.000]
    2   0.0622       -0.0555         15.4738  [0.000]
    3  -0.0927       -0.1009         16.6443  [0.001]
    4  -0.1026       -0.1869 **      18.0895  [0.001]
    5   0.0798       -0.0247         18.9703  [0.002]
    6   0.0668        0.0874         19.5929  [0.003]
    7  -0.0720       -0.0520         20.3216  [0.005]
    8  -0.0524       -0.1208         20.7102  [0.008]
    9   0.1372        0.1318         23.3986  [0.005]
   10  -0.1011       -0.0066         24.8709  [0.006]
   11   0.0427       -0.0509         25.1359  [0.009]
   12  -0.0154       -0.0026         25.1707  [0.014]
   13   0.0063        0.0596         25.1765  [0.022]
   14   0.0143        0.0016         25.2068  [0.033]
   15   0.0910        0.0853         26.4505  [0.034]
   16  -0.1889  **   -0.1159         31.8589  [0.010]
   17   0.0915        0.0086         33.1374  [0.011]
   18   0.0247        0.0667         33.2312  [0.016]
   19  -0.0697       -0.0584         33.9860  [0.018]
   20  -0.0569       -0.1836 **      34.4950  [0.023]
   21   0.0111       -0.0261         34.5145  [0.032]
   22  -0.0924       -0.0933         35.8782  [0.031]
   23   0.2183  **    0.0997         43.5690  [0.006]
   24  -0.0349        0.0195         43.7669  [0.008]
   25  -0.0556        0.0023         44.2750  [0.010]
   26   0.0784        0.0827         45.2958  [0.011]
   27  -0.0264        0.0805         45.4123  [0.015]
   28  -0.0395       -0.0618         45.6758  [0.019]
   29   0.0087       -0.0345         45.6886  [0.025]
   30  -0.0518       -0.0332         46.1520  [0.030]
   31  -0.0697       -0.0820         46.9978  [0.033]
   32   0.1866  **    0.0409         53.1233  [0.011]
   33  -0.1789  **   -0.1046         58.8136  [0.004]
   34   0.0715       -0.0033         59.7326  [0.004]
   35  -0.0582       -0.0541         60.3468  [0.005]
   36  -0.0314       -0.1187         60.5276  [0.006]
```

Si nos fijamos en los retardos estacionales, ninguno es significativo ni el la ACF ni en la PACF. Pasemos a la parte regular del modelo.

## Probando con un AR(1) regular (además del MA(1) estacional)

En el correlograma se aprecia que el primer retardo es significativo tanto en la ACF como en la PACF. Con el resto no es fácil detectar una estructura clara; aunque el cuarto retardo de la PACF es significativo.

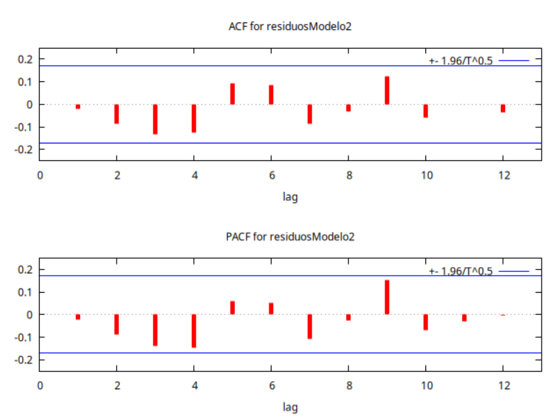
Probemos tentativamente a añadir un polinomio autoregresivo de orden 1.

```
arima 1 1 0 ; 0 1 1 ; lg
```

```
Function evaluations: 25
Evaluations of gradient: 8

Model 6: ARIMA, using observations 1950:02-1960:12 (T = 131)
Estimated using AS 197 (exact ML)
Dependent variable: (1-L)(1-Ls) lg
Standard errors based on Hessian

             coefficient    std. error      z      p-value
  --------------------------------------------------------
  const      -0.000174550   0.00121017   -0.1442   0.8853
  phi_1      -0.339539      0.0821894    -4.131    3.61e-05 ***
  Theta_1    -0.562488      0.0748819    -7.512    5.84e-14 ***

Mean dependent var   0.000291   S.D. dependent var   0.045848
Mean of innovations  0.001038   S.D. of innovations  0.036974
R-squared            0.991359   Adjusted R-squared   0.991292
Log-likelihood       243.7523   Akaike criterion    -479.5046
Schwarz criterion   -468.0038   Hannan-Quinn        -474.8313

                        Real  Imaginary    Modulus  Frequency
  -----------------------------------------------------------
  AR
    Root  1          -2.9452     0.0000     2.9452     0.5000
  MA (seasonal)
    Root  1           1.7778     0.0000     1.7778     0.0000
  -----------------------------------------------------------
```

Y ahora analice el correlograma de los residuos (puesto que ya no hay restos de estacionalidad, basta con mirar los primeros retardos).

```
residuosModelo2 = $uhat
corrgm residuosModelo2 12 --plot="residuosModelo2-ACF-PACF.png"
```



```
Autocorrelation function for residuosModelo2
***, **, * indicate significance at the 1%, 5%, 10% levels
using standard error 1/T^0.5

  LAG      ACF          PACF         Q-stat. [p-value]

    1  -0.0212      -0.0212          0.0600  [0.806]
    2  -0.0874      -0.0879          1.0924  [0.579]
    3  -0.1333      -0.1383          3.5113  [0.319]
    4  -0.1254      -0.1449 *        5.6682  [0.225]
    5   0.0949       0.0613          6.9123  [0.227]
    6   0.0856       0.0514          7.9348  [0.243]
    7  -0.0846      -0.1055          8.9408  [0.257]
    8  -0.0310      -0.0255          9.0773  [0.336]
    9   0.1239       0.1549 *       11.2686  [0.258]
   10  -0.0585      -0.0697         11.7619  [0.301]
   11  -0.0002      -0.0314         11.7619  [0.382]
   12  -0.0359      -0.0032         11.9505  [0.450]
```

Parece que los residuos son ruido blanco. Pero estaba claro el motivo por el que probar con un AR(1) regular en lugar de una MA(1) regular (máxime cuando la PACF tenía un cuarto retardo significativo).

Probemos tentativamente a añadir un polinomio MA de orden 1 (en lugar del polinomio AR).

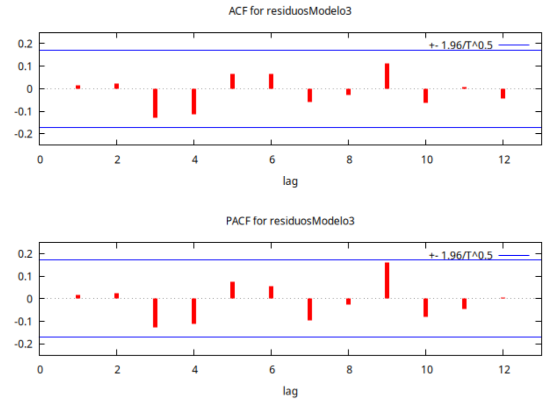
## Probando con un MA(1) regular (además del MA(1) estacional)

```
arima 0 1 1 ; 0 1 1 ; lg
```

```
Function evaluations: 48
Evaluations of gradient: 13

Model 8: ARIMA, using observations 1950:02-1960:12 (T = 131)
Estimated using AS 197 (exact ML)
Dependent variable: (1-L)(1-Ls) lg
Standard errors based on Hessian

             coefficient    std. error       z      p-value
  ----------------------------------------------------------
  const      -0.000162569   0.000974395   -0.1668   0.8675
  theta_1    -0.402057      0.0896975     -4.482    7.38e-06 ***
  Theta_1    -0.557716      0.0731707     -7.622    2.50e-14 ***

Mean dependent var   0.000291   S.D. dependent var   0.045848
Mean of innovations  0.001240   S.D. of innovations  0.036710
R-squared            0.991478   Adjusted R-squared   0.991412
Log-likelihood       244.7104   Akaike criterion    -481.4207
Schwarz criterion   -469.9199   Hannan-Quinn        -476.7475

                        Real  Imaginary    Modulus  Frequency
  -----------------------------------------------------------
  MA
    Root  1           2.4872     0.0000     2.4872     0.0000
  MA (seasonal)
    Root  1           1.7930     0.0000     1.7930     0.0000
  -----------------------------------------------------------
```

Y ahora analice el correlograma de los residuos (puesto que ya no hay restos de estacionalidad, basta con mirar los primeros retardos).

```
residuosModelo3 = $uhat
corrgm residuosModelo3 12 --plot="residuosModelo3-ACF-PACF.png"
```




```
Autocorrelation function for residuosModelo3
***, **, * indicate significance at the 1%, 5%, 10% levels
using standard error 1/T^0.5

  LAG      ACF          PACF         Q-stat. [p-value]

    1   0.0167        0.0167          0.0376  [0.846]
    2   0.0245        0.0242          0.1185  [0.942]
    3  -0.1276       -0.1285          2.3354  [0.506]
    4  -0.1142       -0.1122          4.1246  [0.389]
    5   0.0650        0.0765          4.7088  [0.452]
    6   0.0671        0.0578          5.3369  [0.501]
    7  -0.0580       -0.0967          5.8102  [0.562]
    8  -0.0266       -0.0260          5.9104  [0.657]
    9   0.1142        0.1619 *        7.7730  [0.557]
   10  -0.0637       -0.0807          8.3574  [0.594]
   11   0.0083       -0.0432          8.3673  [0.680]
   12  -0.0431        0.0050          8.6391  [0.733]
```

También este modelo arroja residuos con aspecto de ruido blanco. Comparemos qué modelo parece mejor.

Si nos fijamos en la significatividad de los parámetros, en ambos modelos todos los parámetros son significativos excepto la constate (que omitiremos en el futuro).

Pero para este último modelo, los estadísticos Q de Ljung-Box tienen p-valores más altos, el R-cuadrado corregido es mayor y sus criterios de información más bajos.

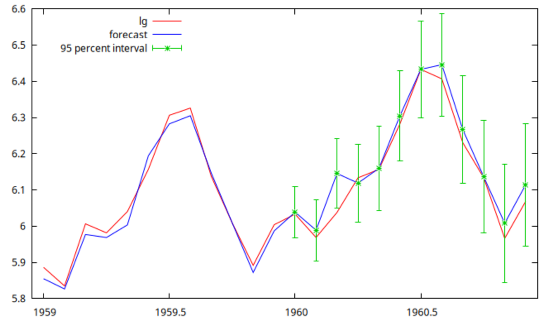
Por tanto, nos quedaremos con este modelo para hacer la predicción (omitiendo la constante).

# Actividad 4 - Previsión para los 12 meses de 1960

Reestimaremos el modelo con datos hasta diciembre de 1959 y haremos la previsión de los meses de 1960.

```
smpl 1949:01 1959:12
arima 0 1 1 ; 0 1 1 ; lg --nc
```

```
dataset addobs 12
fcast 1959:01 1960:12 --plot="prediccion1960.png"
```




### Código completo de la práctica

```
# Los dos primeros comandos son necesarios para que Gretl guarde los resultados de la práctica en el directorio de trabajo
# al ejecutar lo siguiente desde un terminal (use los nombres y ruta que correspondan)
#
# DIRECTORIO="Nombre_Directorio_trabajo"gretlcli -b ruta/nombre_fichero_de_la_practica.inp
#
# Si esto no le funciona en su sistema, comente las siguientes dos líneas y sitúese en el directorio de trabajo de gretl
# que corresponda (configure dicho directorio de trabajo desde la ventana principal de Gretl).

string directory = getenv("DIRECTORIO")
set workdir "@directory"

open bjg

gnuplot lg --time-series --with-lines  --output="log_AP.png"

arima 0 1 0 ; 0 1 0 ; lg

outfile --quiet Modelo0.txt
  arima 0 1 0 ; 0 1 0 ; lg
end outfile

residuosModelo0 = $uhat
corrgm residuosModelo0 36 --plot="residuosModelo0-ACF-PACF.png"

outfile --quiet CorrelogramaModelo0.txt
  corrgm residuosModelo0 36 --quiet
end outfile

arima 0 1 0 ; 0 1 1 ; lg

outfile --quiet Modelo1.txt
  arima 0 1 0 ; 0 1 1 ; lg
end outfile

residuosModelo1 = $uhat
corrgm residuosModelo1 36 --plot="residuosModelo1-ACF-PACF.png"

outfile --quiet CorrelogramaModelo1.txt
  corrgm residuosModelo1 36 --quiet
end outfile

arima 1 1 0 ; 0 1 1 ; lg

outfile --quiet Modelo2.txt
  arima 1 1 0 ; 0 1 1 ; lg
end outfile

residuosModelo2 = $uhat
corrgm residuosModelo2 12 --plot="residuosModelo2-ACF-PACF.png"

outfile --quiet CorrelogramaModelo2.txt
```

```
  corrgm residuosModelo2 12 --quiet
end outfile

arima 0 1 1 ; 0 1 1 ; lg

outfile --quiet Modelo3.txt
  arima 0 1 1 ; 0 1 1 ; lg
end outfile

residuosModelo3 = $uhat
corrgm residuosModelo3 12 --plot="residuosModelo3-ACF-PACF.png"

outfile --quiet CorrelogramaModelo3.txt
  corrgm residuosModelo3 12 --quiet
end outfile

smpl 1949:01 1959:12
arima 0 1 1 ; 0 1 1 ; lg --nc

dataset addobs 12
fcast 1959:01 1960:12 --plot="prediccion1960.png"
```