



Econometría II

**Modelización del número de pasajeros de líneas aéreas,
desde enero de 1949 hasta diciembre de 1960**

Miguel Jerez y Sonia Sotoca

Universidad Complutense de Madrid

Marzo 2002

Índice

Introducción

Identificación

- Tasa log intermensual
- Cambio en la tasa log interanual

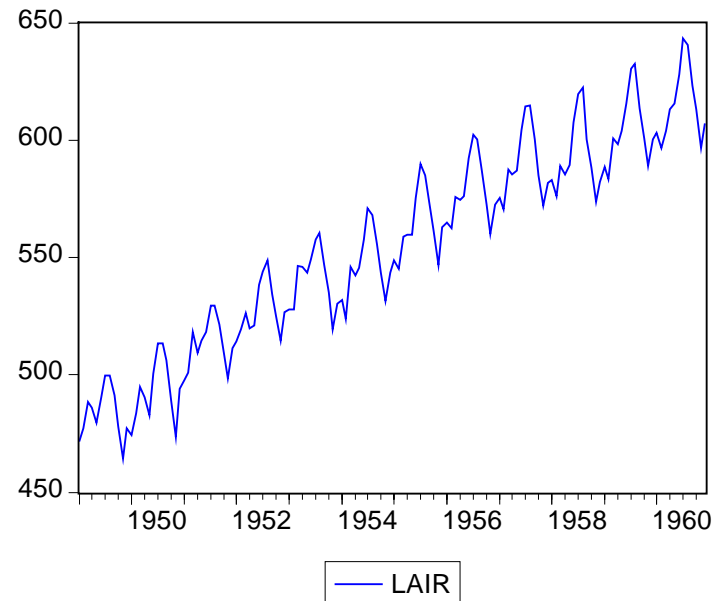
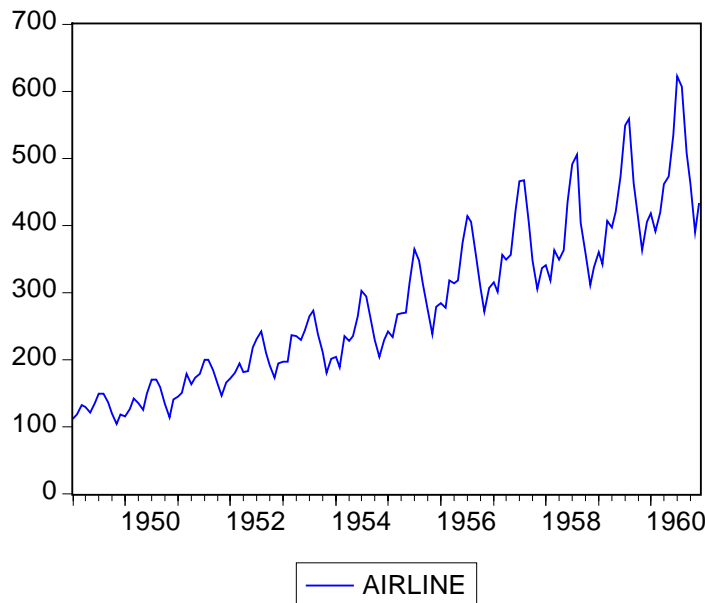
Estimación y diagnóstico

Modelo alternativo

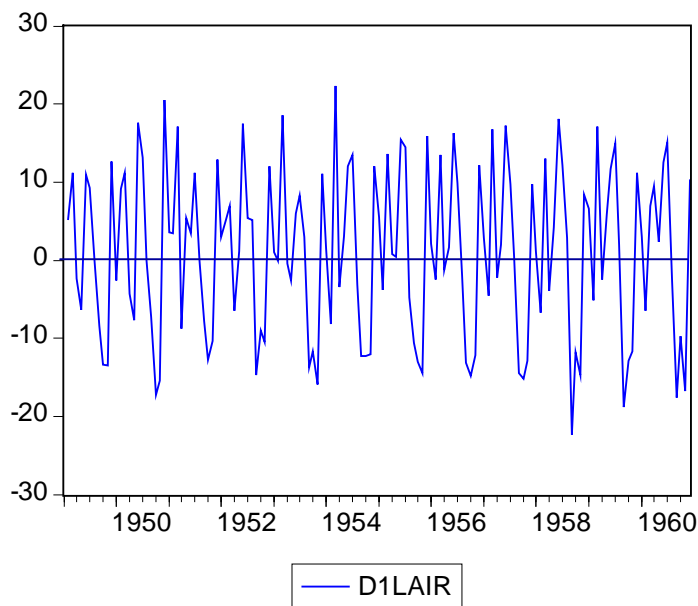
Introducción

Descripción de los datos:

- Número de pasajeros de líneas aéreas, en miles de personas, desde enero de 1949 hasta diciembre de 1960 (144 observaciones).
- Los datos muestran: a) tendencia, b) un ciclo estacional y c) una variabilidad proporcional a la media.
- Esta última característica desaparece al transformar logarítmicamente ($\log x$ 100) los datos.



Identificación (I): Tasa log intermensual



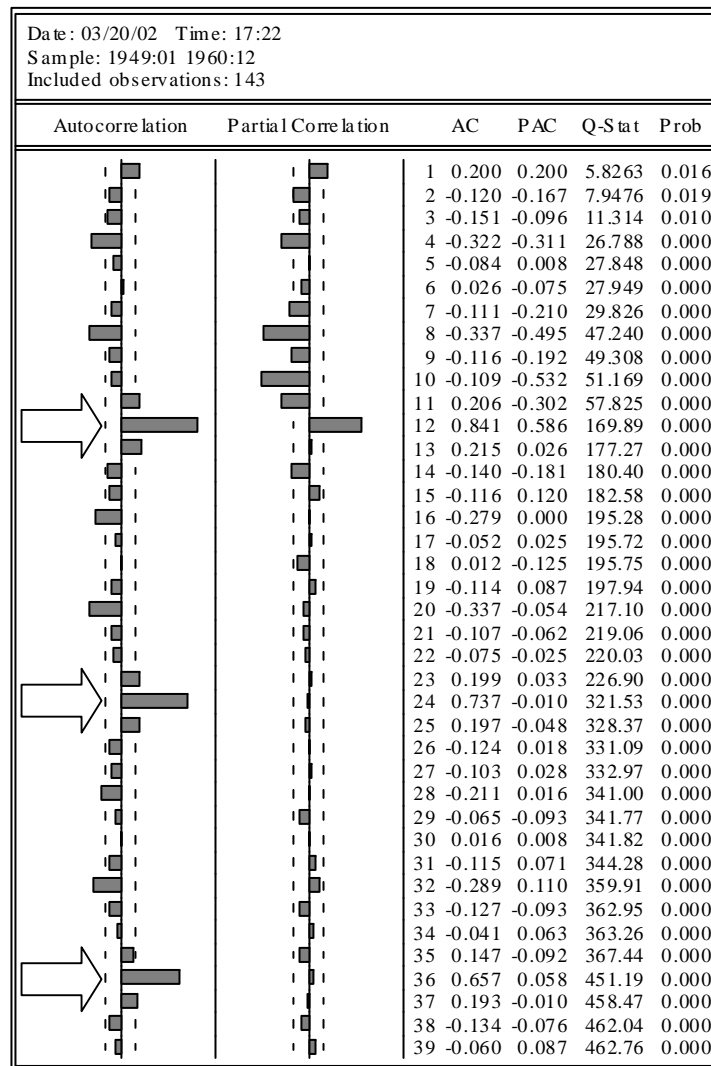
La tasa log intermensual muestra:

- Una pauta periódica en su perfil.
- Autocorrelaciones importantes y persistentes en los retardos múltiplos de doce.

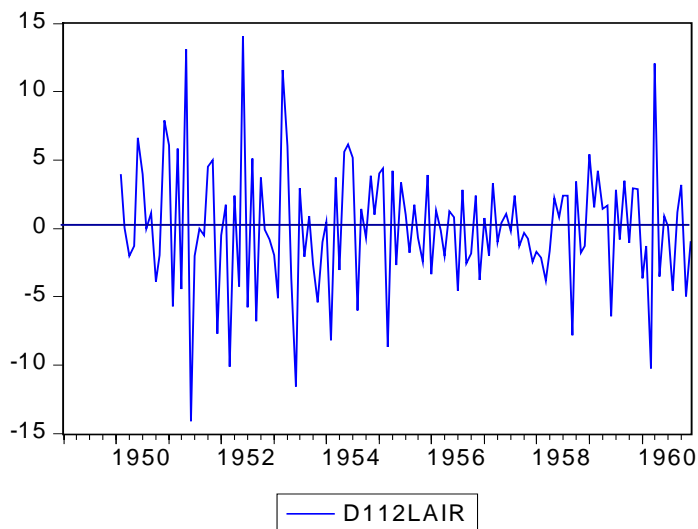
Consecuentemente, se pasa a analizar:

$$z_t = \nabla \nabla_{12} \log(P_t) \times 100$$

Correlogram of D1LAIR

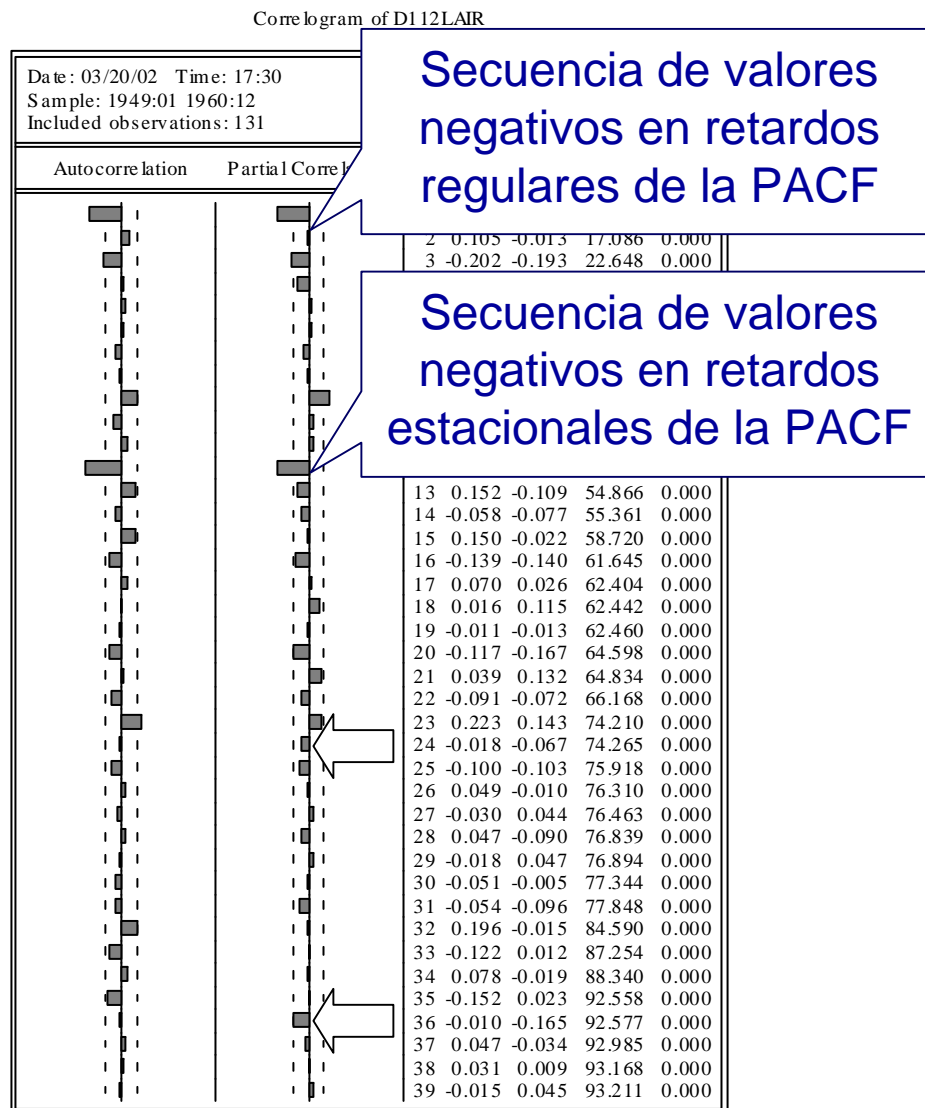


Identificación (II): Cambio en la tasa log interanual



Esta transformación de datos puede interpretarse como “cambio en el mes de la tasa log interanual”. Se observa:

- Posible heterocedasticidad.
- Autocorrelaciones compatibles con una estructura $IMA(1,1) \times IMA(1,1)_{12}$



Estimación y diagnóstico (I)

| Dependent Variable: DLOG(AIRLINE,1,12)*100 | | | | |
|--|-------------|-----------------------|-------------|------------|
| Method: Least Squares | | | | |
| Date: 03/14/02 Time: 11:53 | | | | |
| Sample (adjusted): 1950:02 1960:12 | | | | |
| Included observations: 131 after adjusting endpoints | | | | |
| Convergence achieved after 12 iterations | | | | |
| Backcast: 1949:01 1950:01 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| MA(1) | -0.404858 | 0.080238 | -5.045740 | 0.0000 |
| SMA(12) | -0.631571 | 0.069842 | -9.042835 | 0.0000 |
| R-squared | 0.371098 | Mean dependent var | 0.029088 | |
| Adjusted R-squared | 0.366223 | S.D. dependent var | 4.584829 | |
| S.E. of regression | 3.649988 | Akaike info criterion | 5.442474 | |
| Sum squared resid | 1718.591 | Schwarz criterion | 5.486370 | |
| Log likelihood | -354.4821 | Durbin-Watson stat | 1.933721 | |
| Inverted MA Roots | .96 | .83 -.48i | .83+.48i | .48+.83i |
| | .48 -.83i | .40 | .00+.96i | -.00 -.96i |
| | -.48+.83i | -.48 -.83i | -.83 -.48i | -.83+.48i |
| | -.96 | | | |

El modelo estimado es:

$$\nabla \nabla_{12} \log(P_t) \times 100 = (1 - .40B)(1 - .63B^{12}) \hat{a}_t$$

(.08) (.07)

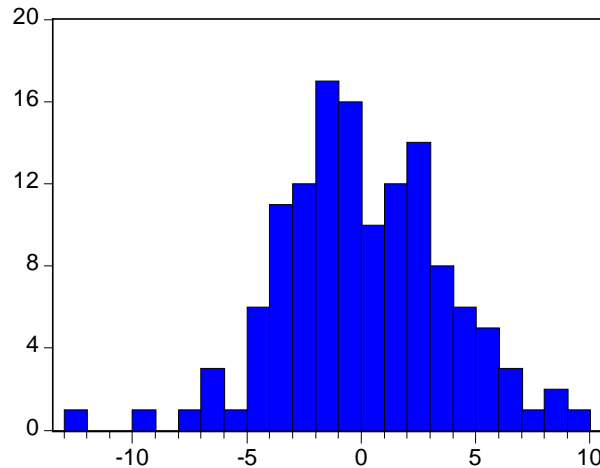
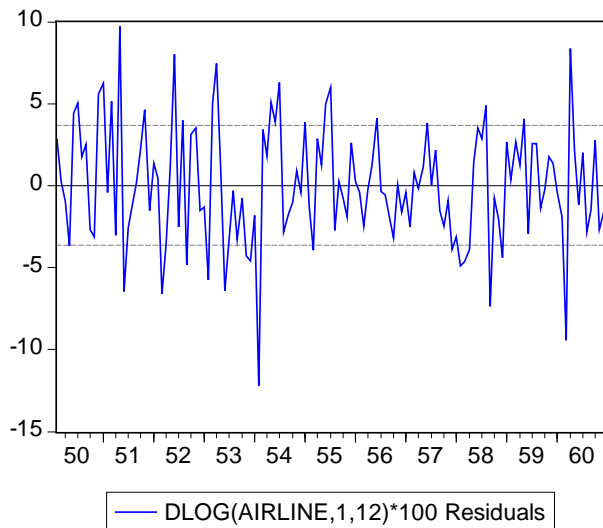
$$\hat{\sigma}_a = 3.65$$

En cuanto a la diagnosis, todos los coeficientes son significativos y las ACFs residuales son compatibles con una estructura de ruido blanco

Correlogram of Residuals

| Date: 03/20/02 Time: 17:39 | | | | | |
|---|---------------------|-----------|--------|--------|-------|
| Sample: 1950:02 1960:12 | | | | | |
| Included observations: 131 | | | | | |
| Q-statistic probabilities adjusted for 2 ARMA term(s) | | | | | |
| Auto correlation | Partial Correlation | AC | PAC | Q-Stat | Prob |
| | | 1 0.030 | 0.030 | 0.1204 | |
| | | 2 0.021 | 0.020 | 0.1805 | |
| | | 3 -0.124 | -0.126 | 2.2844 | 0.131 |
| | | 4 -0.147 | -0.143 | 5.2684 | 0.072 |
| | | 5 0.045 | 0.060 | 5.5496 | 0.136 |
| | | 6 0.060 | 0.052 | 6.0438 | 0.196 |
| | | 7 -0.075 | -0.122 | 6.8403 | 0.233 |
| | | 8 -0.043 | -0.054 | 7.1001 | 0.312 |
| | | 9 0.098 | 0.146 | 8.4610 | 0.294 |
| | | 10 -0.077 | -0.096 | 9.3103 | 0.317 |
| | | 11 0.030 | -0.031 | 9.4384 | 0.398 |
| | | 12 0.006 | 0.045 | 9.4432 | 0.491 |
| | | 13 0.037 | 0.076 | 9.6473 | 0.562 |
| | | 14 0.046 | -0.010 | 9.9597 | 0.619 |
| | | 15 0.045 | 0.032 | 10.258 | 0.673 |
| | | 16 -0.160 | -0.120 | 14.120 | 0.441 |
| | | 17 0.019 | 0.044 | 14.175 | 0.512 |
| | | 18 -0.005 | -0.003 | 14.178 | 0.585 |
| | | 19 -0.109 | -0.140 | 16.012 | 0.523 |
| | | 20 -0.104 | -0.157 | 17.708 | 0.475 |
| | | 21 -0.034 | 0.030 | 17.896 | 0.529 |
| | | 22 -0.025 | -0.031 | 17.993 | 0.588 |
| | | 23 0.220 | 0.132 | 25.775 | 0.215 |
| | | 24 0.043 | -0.011 | 26.077 | 0.248 |
| | | 25 -0.010 | 0.041 | 26.092 | 0.297 |
| | | 26 0.064 | 0.077 | 26.774 | 0.315 |
| | | 27 -0.043 | -0.010 | 27.089 | 0.351 |
| | | 28 -0.070 | -0.107 | 27.915 | 0.363 |
| | | 29 -0.055 | -0.035 | 28.422 | 0.389 |
| | | 30 -0.080 | -0.041 | 29.538 | 0.386 |
| | | 31 -0.048 | -0.050 | 29.937 | 0.417 |
| | | 32 0.116 | 0.027 | 32.292 | 0.354 |
| | | 33 -0.133 | -0.108 | 35.442 | 0.267 |
| | | 34 -0.002 | 0.000 | 35.443 | 0.309 |
| | | 35 -0.060 | -0.084 | 36.102 | 0.326 |
| | | 36 -0.011 | -0.062 | 36.122 | 0.370 |
| | | 37 0.118 | 0.087 | 38.724 | 0.305 |
| | | 38 0.072 | 0.026 | 39.707 | 0.308 |
| | | 39 -0.025 | -0.025 | 39.824 | 0.346 |

Estimación y diagnóstico (II)



| | |
|------------------------|-----------|
| Series: Residuals | |
| Sample 1950:02 1960:12 | |
| Observations 131 | |
| Mean | 0.034246 |
| Median | -0.296196 |
| Maximum | 9.740810 |
| Minimum | -12.20319 |
| Std. Dev. | 3.635760 |
| Skewness | -0.082867 |
| Kurtosis | 3.475030 |
| Jarque-Bera | 1.381619 |
| Probability | 0.501170 |

Asimismo se observa:

- Hay algunos residuos atípicamente elevados.
- El contraste de Jarque Bera no rechaza normalidad.

Consecuentemente, el modelo univariante resulta válido en principio. Sin embargo, puede mejorarse añadiendo variables exógenas como:

- Número de días laborables (L_t) y festivos (F_t) en cada mes de la muestra.
- Mes en que ocurre la Semana Santa (S_t).

ya que estos factores pueden perturbar la pauta estacional y explicar parte de los residuos atípicos.

Modelo alternativo (I): Estimación

| Dependent Variable: DLOG(AIRLINE,1,12)*100 | | | | |
|--|-------------|-----------------------|-------------|------------|
| Method: Least Squares | | | | |
| Date: 03/20/02 Time: 18:41 | | | | |
| Sample (adjusted): 1950:02 1960:12 | | | | |
| Included observations: 131 after adjusting endpoints | | | | |
| Convergence achieved after 13 iterations | | | | |
| Backcast: 1949:01 1950:01 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| D(DIAS_LAB,1,12) | 3.879607 | 1.470812 | 2.637732 | 0.0094 |
| D(DIAS_FEST,1,12) | 4.816347 | 1.493290 | 3.225326 | 0.0016 |
| D(EASTER,1,12) | 2.703436 | 1.020639 | 2.648768 | 0.0091 |
| MA(1) | -0.214428 | 0.087400 | -2.453407 | 0.0155 |
| SMA(12) | -0.606711 | 0.073008 | -8.310191 | 0.0000 |
| R-squared | 0.490330 | Mean dependent var | | 0.029088 |
| Adjusted R-squared | 0.474150 | S.D. dependent var | | 4.584829 |
| S.E. of regression | 3.324712 | Akaike info criterion | | 5.278064 |
| Sum squared resid | 1392.768 | Schwarz criterion | | 5.387804 |
| Log likelihood | -340.7132 | Durbin-Watson stat | | 1.919635 |
| Inverted MA Roots | .96 | .83 -.48i | .83+.48i | .48+.83i |
| | .48 -.83i | .21 | .00+.96i | -.00 -.96i |
| | -.48+.83i | -.48 -.83i | -.83 -.48i | -.83+.48i |
| | -.96 | | | |

El modelo resultante es una regresión con perturbaciones autocorreladas.

Los resultados de la estimación indican que:

- Los coeficientes de las nuevas variables resultan significativos.
- El ajuste ha mejorado. Esta mejora se traduce en: una reducción de a) la desviación típica residual y b) los criterios de información (AIC SBC), así como c) un aumento del R^2 .

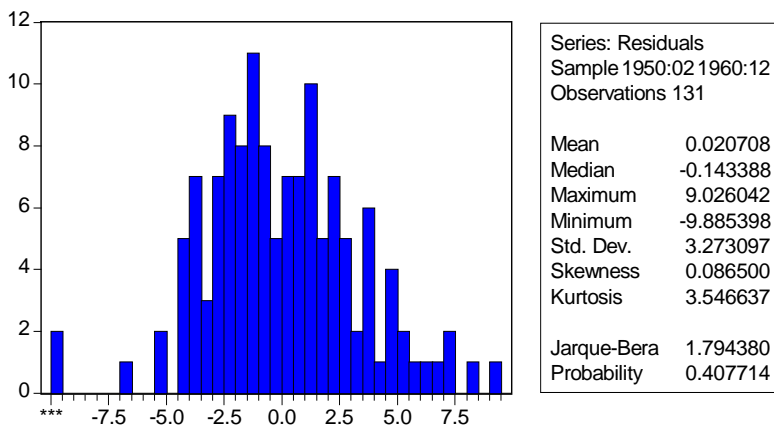
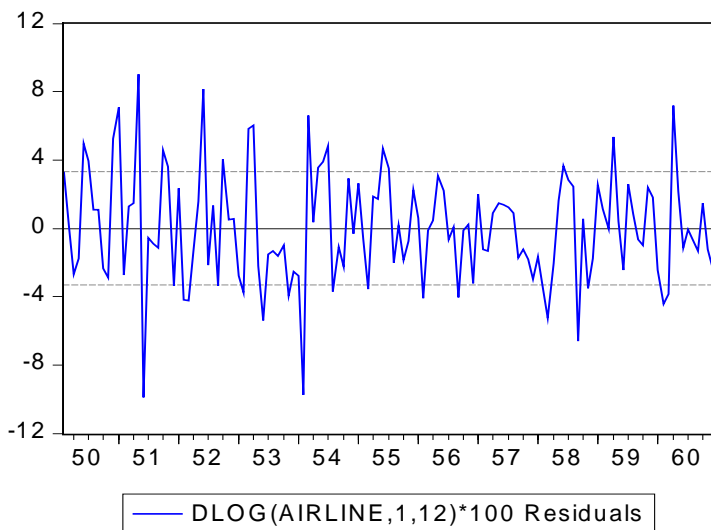
En la forma utilizada para estimar, el modelo puede escribirse como:

$$\nabla \nabla_{12} \log(P_t) \times 100 = \underset{(1.47)}{3.88} \nabla \nabla_{12} L_t + \underset{(1.49)}{4.82} \nabla \nabla_{12} F_t + \underset{(1.02)}{2.70} \nabla \nabla_{12} S_t + \underset{(.09)}{(1 - .21B)} \underset{(.07)}{(1 - .61B^{12})} \hat{a}_t$$

$$\hat{\sigma}_a = 3.32$$

Los diagnósticos residuales no muestran ningún síntoma de mala especificación.

Modelo alternativo (II): Diagnosis residual



Correlogram of Residuals

| Date: 04/08/02 Time: 11:18 Sample: 1950:02 1960:12 Included observations: 131 Q-statistic probabilities adjusted for 2 ARMA term(s) | | | | | | |
|--|---------------------|----|--------|--------|--------|-------|
| Auto correlation | Partial Correlation | AC | PAC | Q-Stat | Prob | |
| | | 1 | 0.034 | 0.034 | 0.1587 | |
| | | 2 | -0.104 | -0.106 | 1.6326 | |
| | | 3 | -0.161 | -0.156 | 5.1793 | 0.023 |
| | | 4 | -0.073 | -0.077 | 5.9028 | 0.052 |
| | | 5 | -0.003 | -0.034 | 5.9040 | 0.116 |
| | | 6 | 0.086 | 0.048 | 6.9467 | 0.139 |
| | | 7 | -0.030 | -0.062 | 7.0759 | 0.215 |
| | | 8 | -0.020 | -0.017 | 7.1322 | 0.309 |
| | | 9 | -0.001 | 0.008 | 7.1323 | 0.415 |
| | | 10 | 0.031 | 0.023 | 7.2674 | 0.508 |
| | | 11 | -0.015 | -0.025 | 7.3021 | 0.606 |
| | | 12 | -0.046 | -0.050 | 7.6166 | 0.666 |
| | | 13 | 0.062 | 0.078 | 8.1893 | 0.696 |
| | | 14 | -0.015 | -0.030 | 8.2225 | 0.768 |
| | | 15 | 0.065 | 0.065 | 8.8522 | 0.784 |
| | | 16 | -0.100 | -0.104 | 10.368 | 0.735 |
| | | 17 | 0.012 | 0.041 | 10.391 | 0.794 |
| | | 18 | 0.014 | 0.016 | 10.419 | 0.844 |
| | | 19 | -0.085 | -0.126 | 11.548 | 0.827 |
| | | 20 | -0.185 | -0.191 | 16.936 | 0.528 |
| | | 21 | 0.023 | 0.002 | 17.018 | 0.589 |
| | | 22 | 0.000 | -0.050 | 17.018 | 0.652 |
| | | 23 | 0.163 | 0.086 | 21.309 | 0.440 |
| | | 24 | 0.143 | 0.114 | 24.644 | 0.314 |
| | | 25 | -0.062 | -0.036 | 25.277 | 0.336 |
| | | 26 | 0.021 | 0.125 | 25.353 | 0.387 |
| | | 27 | 0.025 | 0.063 | 25.457 | 0.437 |
| | | 28 | -0.089 | -0.103 | 26.784 | 0.421 |
| | | 29 | -0.072 | -0.044 | 27.659 | 0.429 |
| | | 30 | -0.064 | -0.089 | 28.357 | 0.446 |
| | | 31 | -0.007 | -0.033 | 28.365 | 0.498 |
| | | 32 | 0.102 | 0.043 | 30.191 | 0.456 |
| | | 33 | -0.117 | -0.178 | 32.629 | 0.387 |
| | | 34 | 0.014 | 0.040 | 32.667 | 0.434 |
| | | 35 | -0.118 | -0.114 | 35.195 | 0.365 |
| | | 36 | 0.051 | -0.029 | 35.663 | 0.390 |
| | | 37 | 0.092 | 0.068 | 37.223 | 0.367 |
| | | 38 | 0.072 | 0.025 | 38.197 | 0.370 |
| | | 39 | 0.016 | 0.030 | 38.247 | 0.413 |

Modelo alternativo (III): Interpretación

La formulación anterior es conveniente para estimar. Para otros usos, resulta más clara la expresión equivalente:

$$\log(P_t) \times 100 = 3.88 L_t + 4.82 F_t + 2.70 S_t + \frac{(1 - .21B)(1 - .61B^{12})}{(1 - B)(1 - B^{12})} \hat{a}_t$$

En esta notación, los coeficientes de las variables exógenas tienen un significado claro como *semi-elasticidades*:

- si un mes tiene un día laborable más que otro, cabe esperar que el número de pasajeros sea un 3.88% superior
- si un mes tiene un día festivo más que otro, cabe esperar que el número de pasajeros sea un 4.82% superior
- la Semana Santa produce un aumento estimado del 2.7% en el número de pasajeros en comparación con un mes sin Semana Santa.

Las variables exógenas que hemos utilizado en este ejemplo son un ejemplo de lo que en Econometría se conoce como “efectos calendario”. Estos efectos son muy importantes ya que, además de afectar a muchas variables económicas (típicamente las ligadas al calendario laboral o a los hábitos de consumo, como producción o ventas) tienen la ventaja de resultar exactamente predecibles.