

Análisis de una serie de tiempo de datos real

Transporte aéreo

Series de tiempo. Curso 2009-2010

Leyenda Rodríguez, María

1. Introducción

En este trabajo se analizará la serie de tiempo número de pasajeros en el transporte aéreo español (Enero de 1990 - Enero de 2010, 241 meses) (Figura 1).

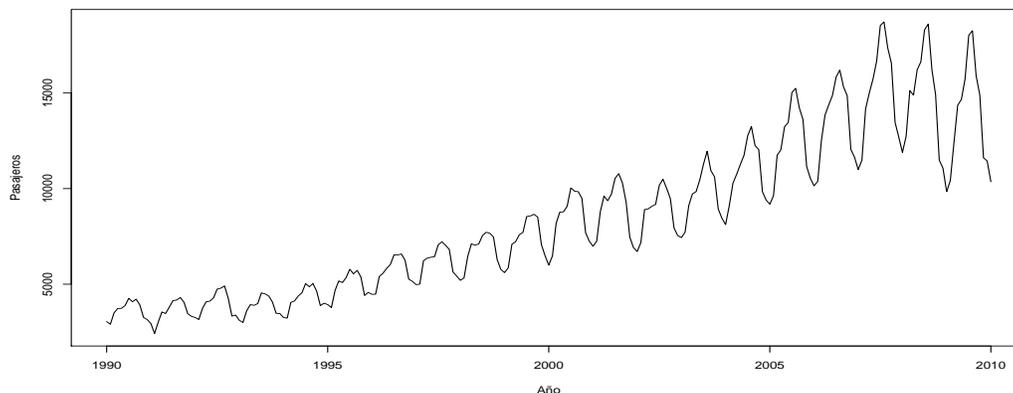


Figura 1: Gráfico de secuencia de la serie de tiempo número de pasajeros en el transporte aéreo español

Estos datos fueron extraídos de la página web del Instituto Nacional de Estadística aunque la fuente de información es Dirección General de Aviación Civil del Ministerio de Fomento.

El objetivo de este trabajo será construir un modelo Box-Jenkins (ARIMA) para modelizar dicha serie y realizar predicciones; para ello dividiremos la serie original en dos series:

- pasajeros1: Es parte de la serie de tiempo original que contiene los datos correspondientes al intervalo de tiempo comprendido entre Enero de 1990 y Diciembre de 2008(Figura 2).

Ésta serie será utilizada para seleccionar y ajustar un modelo Box-Jenkins (ARIMA). Además, en base a dicho modelo, se realizarán predicciones para el número de pasajeros en el transporte aéreo español correspondientes a los próximos 13 meses (Enero de 2009- Enero de 2010).

- pasajeros2: Es parte de la serie de tiempo original que contiene los datos correspondientes al intervalo de tiempo comprendido entre Enero de 2009 y Enero de 2010 (Figura 3).

Ésta serie será utilizada para compararla con las predicciones obtenidas.

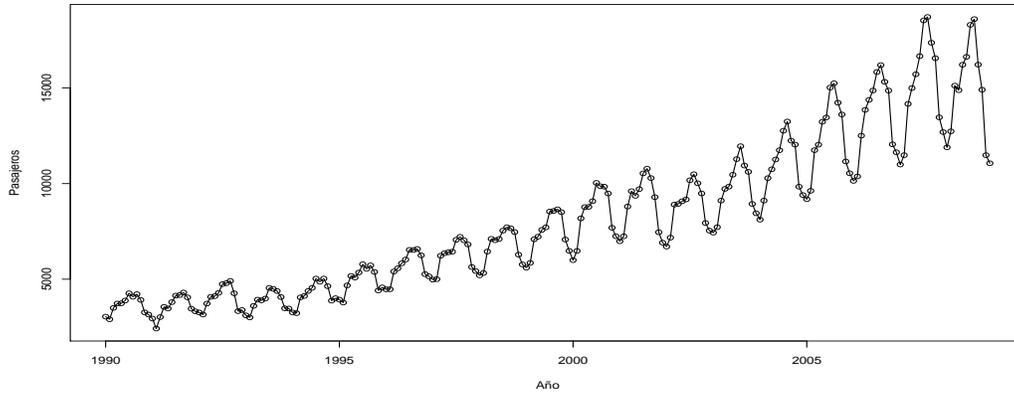


Figura 2: Gráfico de secuencia de la serie pasajeros1

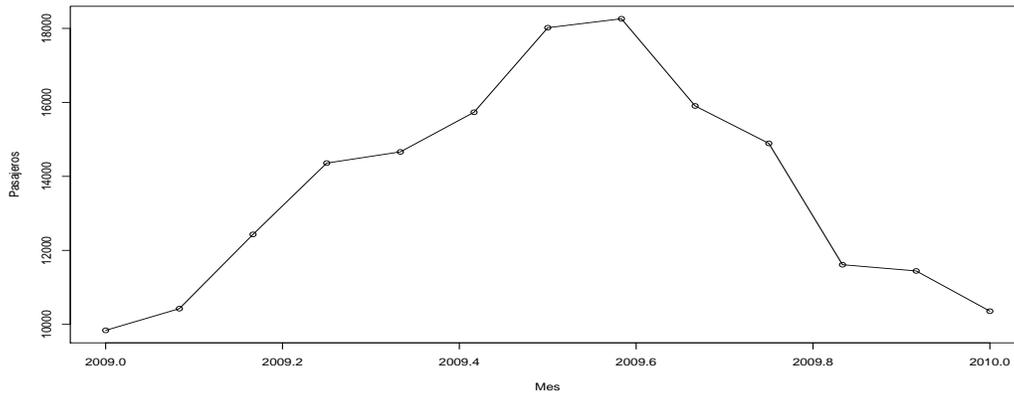


Figura 3: Gráfico de secuencia de la serie pasajeros2

2. Modelización de la serie: Construcción, identificación y diagnosis del modelo

2.1. Corregir heterocedasticidad

En primer lugar, observamos que la serie pasajeros1 (Figura 2) presenta heterocedasticidad; pues la variabilidad de la serie no es constante, ésta aumenta al hacerlo el nivel de la serie. Luego, realizaremos una transformación de Box-Cox (1) para estabilizar la varianza.

$$\begin{cases} \frac{x_t^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0, \\ \ln(x_t) & \text{si } \lambda = 0, \end{cases} \quad (1)$$

Si la desviación típica es función potencial de la media, $\sigma_t = k\mu_t^{1-\lambda}$, entonces la transformación de Box-Cox con parámetro λ consigue estabilizar la varianza. La situación más usual y la más fácil de interpretar es en aquella en que $\sigma_t = k\mu_t$. Pues, en este caso, $\lambda = 0$ y la aplicación de logaritmo neperiano estabiliza la varianza. Por ello, siempre que se pueda asumiremos que $\lambda = 0$. A continuación obtendremos un valor apropiado para λ . Para ello, se parte de la ecuación

$$\sigma_t = k\mu_t^{1-\lambda}$$

al aplicarle la función logaritmo se obtiene que

$$\log(\sigma_t) = a + b \log(\mu_t),$$

dónde $a = \log(k)$ y $b = 1 - \lambda$. Por tanto, si disponemos de estimaciones $\hat{\mu}_t$ y $\hat{\sigma}_t$ y ajustamos un modelo lineal a $(\log(\hat{\mu}_t), \log(\hat{\sigma}_t))$, obtendremos una estimación para el valor de λ :

$$\hat{\lambda} = 1 - \hat{b}.$$

Al aplicar esto a nuestra serie de datos obtenemos que el mejor valor de λ es $-0,11079$, por tanto, podemos asumir que $\lambda = 0$. Entonces, transformando la serie mediante el logaritmo neperiano (Figura 4) conseguimos estabilizar la varianza.

Además, en el gráfico de secuencia de la serie pasajeros1 (Figura 2) se observa una clara tendencia creciente y una componente estacional, debido a que en los periodos estivales hay más pasajeros que en los periodos no estivales. Notemos que lo dicho para esta partición de la serie es válido para la serie original (Figura 2); aunque en esta también se observa el efecto de la crisis económica de 2009 pues si comparamos el número de pasajeros en el periodo estival de 2009 con los años anteriores se observa que este es más bajo que los anteriores y lo mismo ocurre al comparar los periodos no estivales.

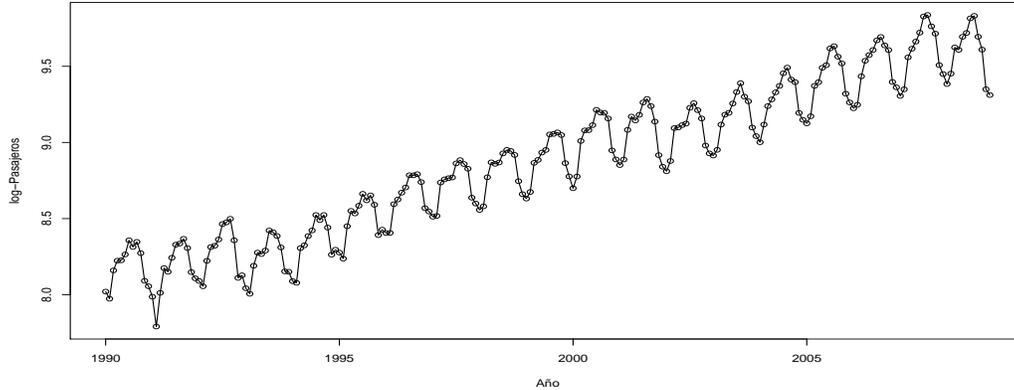


Figura 4: Gráfico de secuencia de la serie pasajeros1 transformada mediante la aplicación logaritmo

2.2. Identificación de los ordenes del modelo ARIMA (p,d,q) y (P,D,Q) y del periodo estacional s

En primer lugar, notemos que el gráfico de secuencia de la serie pasajeros1 (Figura 2) sugiere presencia de tendencia y componente estacional. Para verificarlo, representamos las correlaciones totales y parciales (fas, fap) (Figura 5) de la serie pasajero1 transformada mediante la aplicación logaritmo neperiano (Figura4), dónde se observa la presencia de tendencia. Por tanto, diferenciamos regularmente la serie (d=1) con el objetivo de eliminar la tendencia.

Al representar la serie diferenciada regularmente (Figura 6), se observa que la tendencia ha sido eliminada. Aunque tanto en el gráfico de secuencia de la serie diferenciada regularmente como en la representación gráfica de sus correlaciones parciales y totales (Figura 7) se observa la presencia de una componente estacional con periodo estacional anual, s=12. Tras diferenciarla de nuevo, mediante una diferenciación estacional (D=1), se observa que se ha eliminado la componente estacional; tanto en el grafico de secuencia (Figura 8) de la serie nuevamente diferenciada como en la representación de las correlaciones totales y parciales (Figura 9).

De todo esto se deduce que la serie transformada diferenciada, tanto regular como estacionariamente, proviene de un proceso ARMA estacionario multiplicativo. A la vista de las fas y fap (Figura 9), se deduce que la serie pasjeros1 transformada mediante el logaritmo neperiano puede provenir de un $ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$ o $ARIMA(1, 1, 0) \times (0, 1, 1)_{12}$.

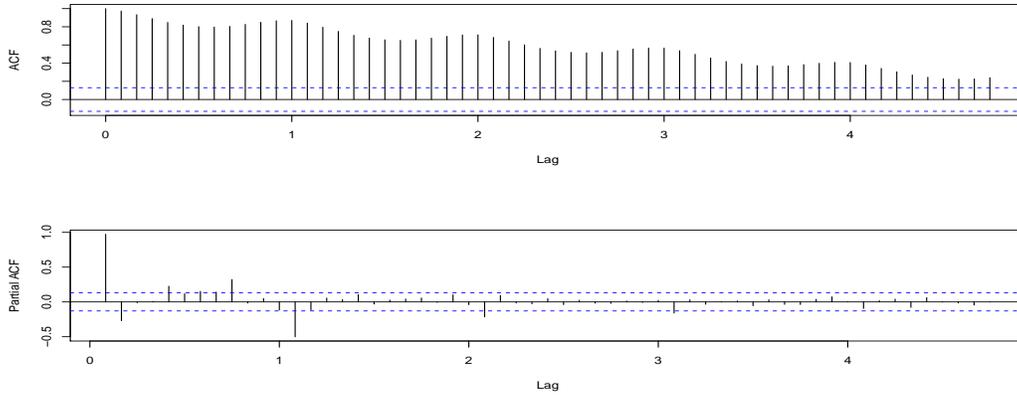


Figura 5: Correlaciones totales y parciales de la serie transformada pasajeros1

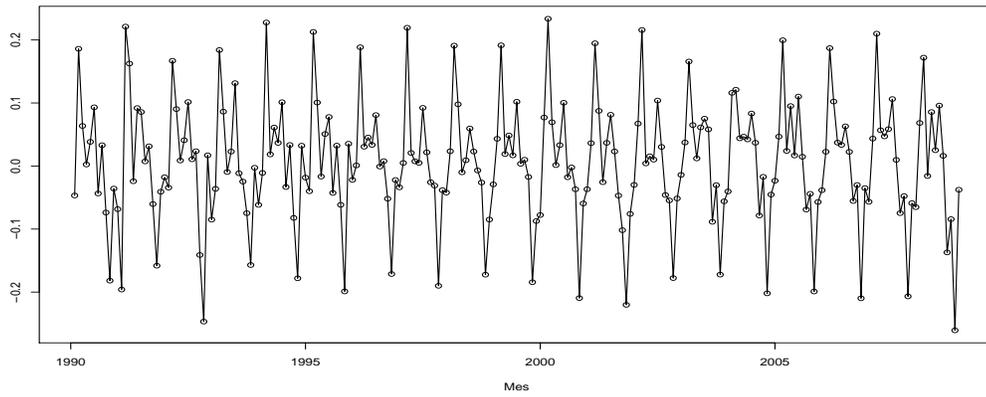


Figura 6: Gráfico de secuencia de la serie transformada pasajeros1 diferenciada regularmente

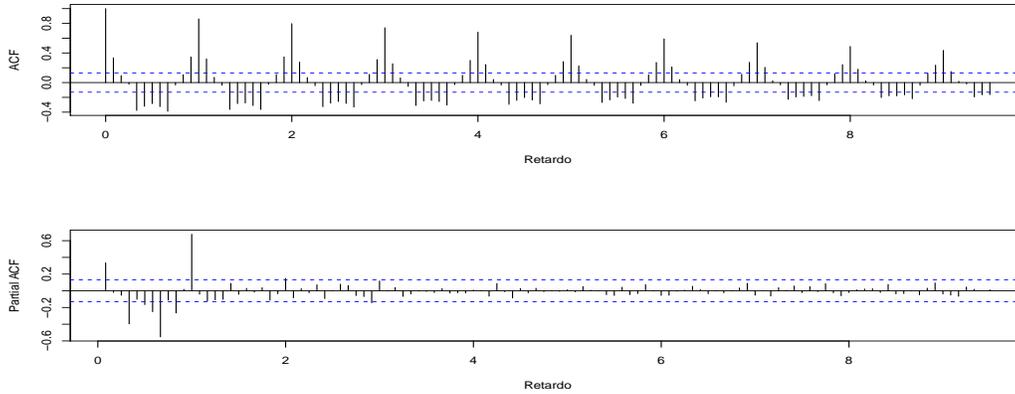


Figura 7: Correlaciones totales y parciales de la serie transformada pasajeros1 diferenciada regularmente

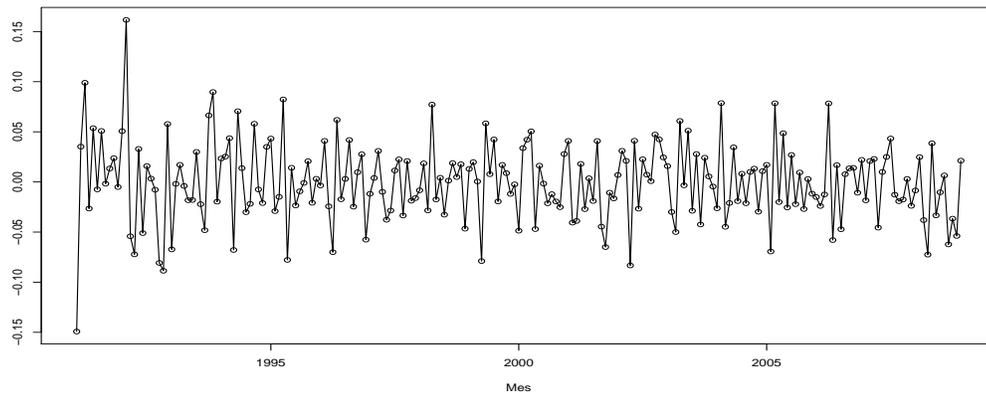


Figura 8: Gráfico de secuencia de la serie transformada pasajeros1 diferenciada regular y estacionariamente

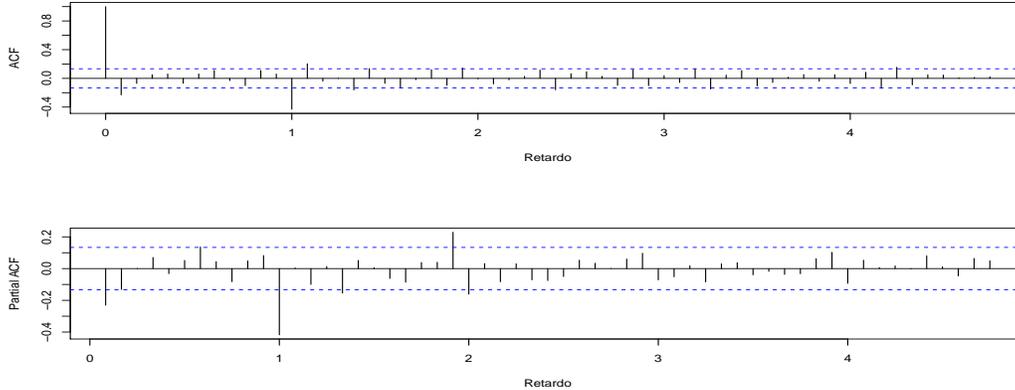


Figura 9: Correlaciones totales y parciales de la serie transformada pasajeros1 diferenciada regular y estacionariamente

A continuación, estudiaremos la validez de los modelos propuestos: $ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$ o $ARIMA(1, 1, 0) \times (0, 1, 1)_{12}$

2.3. Estimación de los modelos identificados

En este apartado, se obtendrán los coeficientes de los dos modelos propuestos, esta estimación se realiza por máxima verosimilitud.

Modelo $ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$

- $\hat{\theta}_1 = -0,2012$; (0.0733)
- $\hat{\Theta}_1 = -0,5969$; (0.0573)
- $\hat{\mu} = -1e - 04$; (8e-04)
- $\hat{\sigma}^2 = 0,001062$

Todos los coeficientes son significativos excepto la media $\hat{\mu}$ del proceso que no es significativamente distinta de cero. Entonces, se ajusta un $ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$ con $\hat{\mu} = 0$, sin constante a la serie sin diferenciar. Obteniéndose:

- $\hat{\theta}_1 = -0,2011$; (0.0733)
- $\hat{\Theta}_1 = -0,5970$; (0.0573)
- $\hat{\sigma}^2 = 0,001062$

Modelo $ARIMA(1, 1, 0) \times (0, 1, 1)_{12}$

- $\hat{\phi}_1 = -0,1618$; (0.0679)

- $\hat{\Theta}_1 = -0,5991; (0.0565)$
- $\hat{\mu} = -1e - 04; (8e-04)$
- $\hat{\sigma}^2 = 0,001069$

Todos los coeficientes son significativos excepto la media $\hat{\mu}$ del proceso que no es significativamente distinta de cero. Entonces, se ajusta un $ARIMA(1, 1, 0) \times (0, 1, 1)_{12}$ con $\hat{\mu} = 0$, sin constante a la serie sin diferenciar. Obteniéndose:

- $\hat{\phi}_1 = -0,1617 ; (0.0679)$
- $\hat{\Theta}_1 = -0,5992; (0.0565)$
- $\hat{\sigma}^2 = 0,001069$

El modelo $ARIMA(1, 1, 0) \times (0, 1, 1)_{12}$, sin constante, es el que proporciona menor BIC (Tabla 1). Por tanto, si este modelo pasa el análisis de los residuos, podremos admitir que este modelo puede ser utilizado como generador de la serie de pasajeros1 transformada mediante el logaritmo neperiano.

Modelo tentativo	Grados de libertad	Criterio BIC
$ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$	3	-840.6500
$ARIMA(1, 1, 0) \times (0, 1, 1)_{12}$	3	-839.2214

Cuadro 1: Resultados del criterio BIC

2.4. Diagn osis del modelo: An alisis de residuos

El modelo ajustado $ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$ sin constante (constante nula) puede ser utilizado como generador de la serie de pasajeros1 transformada mediante la funci n logaritmo neperiano, pues las innovaciones son independientes (Figura 12), con media cero (Figura 10) y no son gaussianas (Figura 11). Mediante el test de *Jarque Bera* o el de *Shapiro-Wilks* verificamos que las innovaciones no son gaussianas tal como apuntaba la (Figura 11) y mediante el *t-test* verificamos, que efectivamente, las innovaciones s  tienen media cero (Tabla 2).

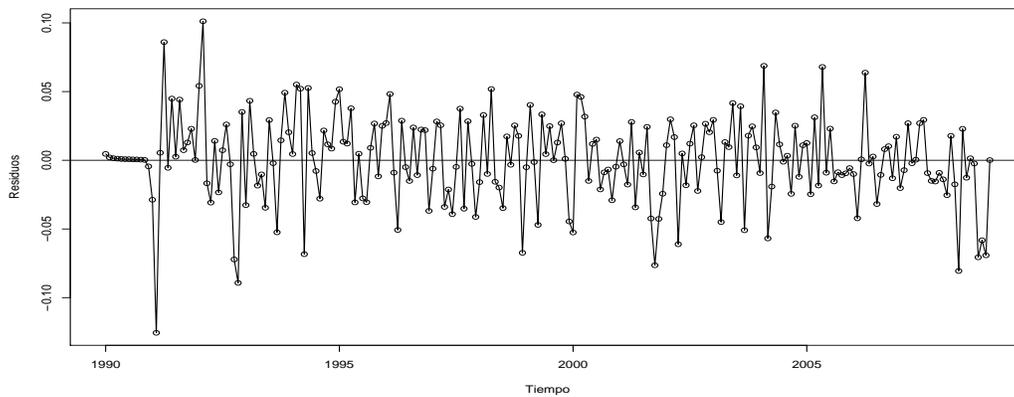


Figura 10: Gr fico de secuencia de los residuos

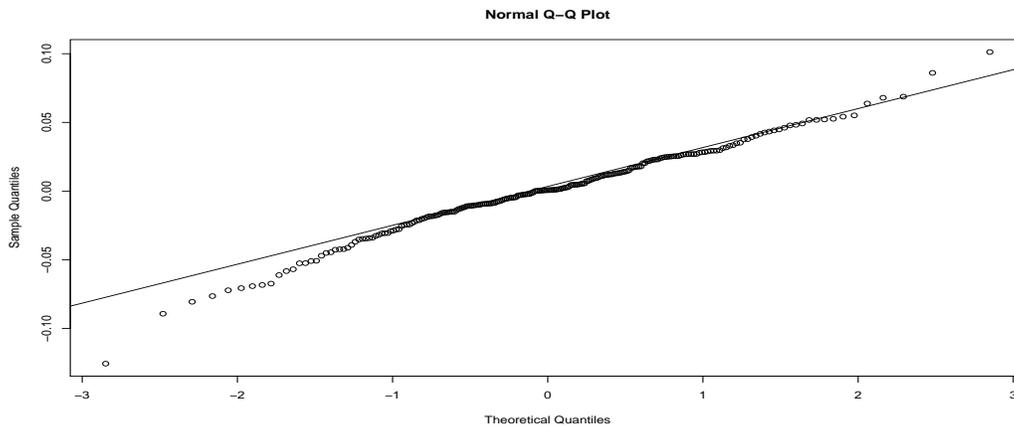


Figura 11: Q-Q Normal

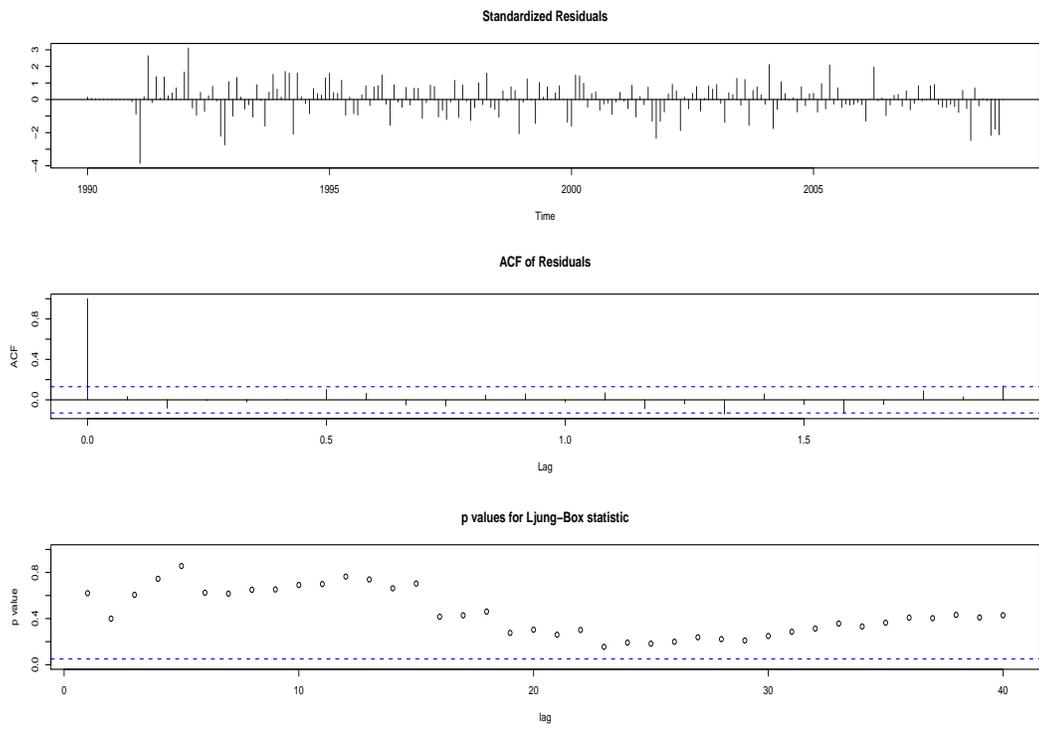


Figura 12: Fas estimadas y Ljung-Box

Test	p-valor	Hipótesis nula	Conclusión
t-test	0.9466 (>0.005)	La media es igual a cero	ACEPTAMOS
Jarque Bera Test	0.0001526 (<0.005)	La serie de residuos es normal	RECHAZAMOS
Shapiro-Wilk normality test	0.01204 (<0.005)	La serie de residuos es normal	RECHAZAMOS

Cuadro 2: Resultados del t-test, Jarque Bera Test y Shapiro-Wilk normality test

3. Predicción en base al modelo $ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$

Para finalizar el estudio, el modelo $ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$ que hemos seleccionado y estimado será el utilizado para realizar predicciones con origen $T = 228$ y horizontes de predicción $k = 1, \dots, 13$ (Figura 13). Cabe notar que no podemos realizar los intervalos de confianza de predicción puesto que las innovaciones no son gaussianas.

En (Tabla 3) podemos observar cuál fue el número de pasajeros en el transporte aéreo español en el periodo comprendido entre Enero de 2009 y Enero de 2010; y comparar los valores reales con las predicciones obtenidas.

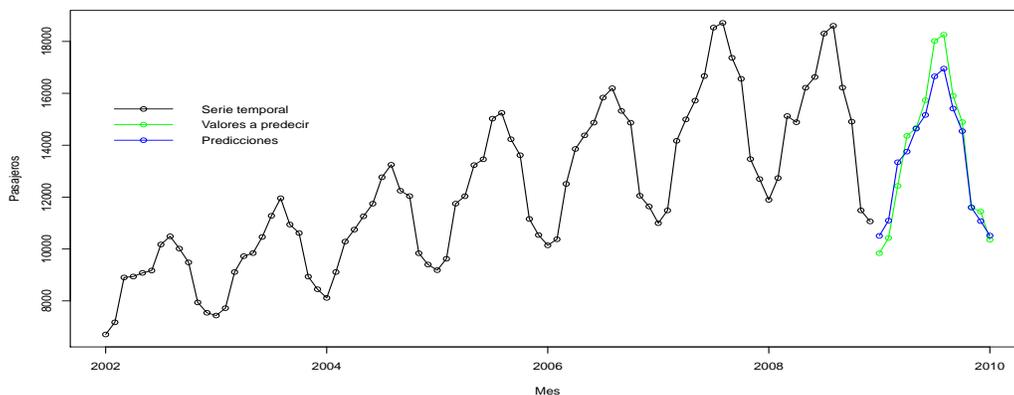


Figura 13: Representación de los valores observados junto con los valores reales de la serie original

Mes	Número de pasajeros	Predicción
Enero	9838.0	10505.78
Febrero	10424.3	11095.83
Marzo	12431.9	13344.60
Abril	14359.3	13753.43
Mayo	14662.6	14637.73
Junio	15734.2	15165.70
Julio	18018.5	16653.14
Agosto	18260.5	16952.22
Septiembre	15904.0	15413.31
Octubre	14892.2	14544.60
Noviembre	11612.4	11597.86
Diciembre	11444.2	11076.33
Enero (2010)	10357.0	10517.76

Cuadro 3: Valores numéricos de predicción