



Facultad de Ciencias Económicas y de Administración  
Universidad de la República

**UNIVERSIDAD DE LA REPUBLICA  
FACULTAD DE CIENCIAS ECONOMICAS Y DE ADMINISTRACION**

**Pasantía para el título de Licenciado en Estadística**

**APLICACIÓN DE LA MODELIZACIÓN AUTOMÁTICA DE SERIES DIARIAS  
CON DETECCIÓN Y ESTIMACIÓN CONJUNTA DE OUTLIERS.**

**por**

**CONSTANCIA URRUTIA.**

**Montevideo  
URUGUAY  
2010**

# Aplicación de la modelización automática de series diarias con detección y estimación conjunta de outliers

Constancia Urrutia  
Pasantía Licenciatura en Estadística  
Facultad de Ciencias Económicas y Administración  
*Universidad de la República*

21 de febrero de 2010

## 1. Introducción

Siguiendo la metodología de Cilintano y Urrutia[8] y la de Chen y Liu[7], en este artículo se presenta una aplicación de la modelización automática para series diarias de actividad económica con detección y estimación conjunta de outliers, mediante un software de libre uso llamado R ([www.R-project.org](http://www.R-project.org)).

El objetivo que se persigue es mejorar el programa previamente realizado en Cilintano y Urrutia[8] para poder identificar un modelo que realice predicciones satisfactoriamente sobre arribos de ómnibus a una determinada terminal, partiendo de una serie diaria que comienza el 1° de enero de 1997 y finaliza el 31 de diciembre de 2004.

Para ello se diseña un programa en R que sigue los siguientes pasos:

1. Ingreso de la serie y fechado
2. Especificación de la transformación logarítmica
3. Modelización del Efecto Calendario
4. Identificación de la estacionalidad dominante y cálculo de diferencias
5. Identificación y Estimación de un modelo siguiendo el criterio de BIC
6. Detección y Estimación Conjunta de outliers
7. Diagnóstico de los Residuos

## 8. Predicción del modelo seleccionado

A continuación se describen cada uno de ellos.

### 1.1. Ingreso y fechado de la Serie

En esta etapa el usuario debe ingresar la serie diaria a ser modelizada ( $Y_t$ ) y la fecha de la primera observación. Luego el programa en forma automática asigna las fechas de las observaciones siguientes<sup>1</sup> y calcula algunos estadísticos descriptivos de la serie como la media, la mediana y la varianza. También imprime en pantalla el valor máximo, mínimo, el primer y segundo cuartil, el número de observaciones y el número de datos faltantes existentes.

### 1.2. Test para la Transformación Logarítmica

Cuando la variabilidad de la serie aumenta linealmente con el nivel de la serie, si se toma logaritmos naturales de la serie se consigue una variabilidad constante.

Box-Cox planteó una transformación<sup>2</sup> a realizar para estabilizar la varianza cuando existe una relación entre las medias ( $\mu_i$ ) de un conjunto de variables aleatorias  $y_i$  y sus desviaciones típicas ( $\sigma_i$ ) del tipo

$$\sigma_i = k\mu_i^\alpha$$

Entonces como primer paso de esta etapa se procede a realizar una prueba para ver si es necesario hacer una transformación a la serie. La prueba se realiza particionando la serie en grupos de siete observaciones consecutivas, se calcula la desviación típica  $s_i$  y la media  $\bar{y}_i$  de estos grupos. Se toma logaritmos en la ecuación anterior y se realiza la regresión:

$$\log(s_i) = c + \alpha \log(\bar{y}_i)$$

La pendiente estimada es el valor de  $\alpha$  que se aplica a la transformación de Box-Cox. Entonces según el coeficiente que se estime en la regresión se realiza la transformación de Box-Cox, con parámetro  $\lambda = 1 - \hat{\alpha}$ , si  $\hat{\alpha}$  es mayor o igual a uno se procede a realizar el logaritmo natural de la serie, sino se continúa con los datos originales.

---

<sup>1</sup>La serie debe ser de 7 días a la semana.

<sup>2</sup>Transformación de Box-Cox: sea  $\lambda = 1 - \hat{\alpha}$   $\left\{ \begin{array}{l} z_i = \frac{y_i^\lambda - 1}{\lambda} \text{ si } \lambda \neq 0 \\ z_i = \log(y_i) \text{ si } \lambda = 0 \end{array} \right.$

### 1.3. Efecto Calendario

En esta etapa se modeliza el Efecto Calendario (EC) mediante variables indicatrices con los feriados, vacaciones y algunos eventos especiales presentes en el calendario uruguayo y luego se ponen a prueba todos los coeficientes estimados y se eliminan aquellos que no sean significativos.

A la serie filtrada por el EC se le llamará  $Y_{EC_0}$ .

Cuando se filtra a la serie  $Y_t$  se quiere aislar el efecto que produce el EC, para luego modelizar a  $Y_{EC_0}$  con modelos SARIMA.

Para filtrar a la serie el programa sigue los siguientes pasos:

1. Se utilizan variables que identifican a los días feriados del Uruguay, a sus días contiguos, vacaciones y eventos especiales. Todos estos días se captan a través de un esquema dinámico estándar con variables indicatrices para el/los días que se quiere/n captar.
2. Se realiza una regresión lineal múltiple para estimar los coeficientes asociados a las variables y así poder marcar el efecto de estos días particulares.
3. Se obtiene  $Y_{EC_0} = Y_t - X\hat{\beta}_0$  la serie filtrada por el efecto calendario. Donde  $X$  es la matriz que contiene a cada una de las variables del paso uno, y  $\hat{\beta}_0$  es el vector de los coeficientes estimados para cada variable.

### 1.4. Identificación de la estacionalidad dominante y transformación estacionaria

#### *Criterio de la Varianza Residual*

“La Varianza Residual (VR) es un criterio utilizado para determinar el número de diferencias a aplicar para convertir a una serie en estacionaria. Consiste en aplicar diferencias en tanto disminuya la varianza de la serie resultante, y parar de diferenciar cuando la varianza aumente al tomar una nueva diferencia.”[23]

$$VR = \frac{1}{T - PM} \sum_{t=1}^T (Y_{EC_0,t} - \bar{Y}_{EC_0})^2$$

donde  $PM$  es el número de parámetros del modelo y  $T$  el número de observaciones.

También se utilizan otros estadísticos complementarios, para comparar diferentes modelos. Estos son la desviación residual ( $DR = \sqrt{VR}$ ) y el criterio de Amemiya (PC):

$$PC = VR \left( 1 + \frac{PM}{T} \right)$$

Se calcula la VR de las siguientes diferencias  $\Delta$ ,  $\Delta^2$ ,  $\Delta_7$ ,  $\Delta_{365}$ ,  $\Delta\Delta_7$ ,  $\Delta\Delta_{365}$  sobre la

serie  $Y_{EC_0}$ . Luego el programa compara la VR de las diferentes transformaciones estacionarias con la de la serie sin diferencias y selecciona el esquema que tenga menor VR para seguir con la metodología.

Con ello se escogerá un operador  $\Delta^d U_{s1-1} = \Delta^d(1 + L + L^2 + \dots + L^{s1-1})$  considerando a  $s1$  como la estacionalidad dominante. Si el esquema elegido fuera  $\Delta^d$  se escoge como estacionalidad principal,  $sj$ , la que de lugar a una varianza residual mínima sobre el esquema  $\Delta^d U_{sj-1}$ .

Luego de elegir la estacionalidad dominante mediante esquemas estocásticos, se prueban dos esquemas determinísticos para modelizar la estacionalidad mensual y semanal. Al primer esquema lo constituyen 12 variables indicatrices de los meses, con unos indicando un mes en particular y ceros para el resto del año. El otro esquema determinístico a utilizar es el semanal, donde las variables presentan unos para un día de la semana en particular y ceros para el resto de los días, son 7 variables en total. Se prueban estos dos esquemas siguiendo el criterio de VR mínima.

Y se obtiene  $Y_{EC_1} = Y_t - X_1 \hat{\beta}_1$  donde  $X_1$  es una matriz formada por las variables del efecto calendario y las variables de las dummy semanales y/o mensuales.

Una vez que se tiene el modelo estimado con todos sus coeficientes se procede a estudiar la significación de los mismos, y en caso de que se detecte a través de *pruebas t* que los coeficientes estimados no son distintos de cero, entonces se procede a eliminar esos coeficientes junto con su variable asociada.

También se utilizan una serie de *pruebas F* para detectar igualdad de coeficientes para un mismo feriado en distintos días de la semana, en días de entre la semana, y para los fines de semana, y en el caso de presentar igualdad asociar dichas variables.

La etapa de reducción se realiza no sólo para que el modelo sea mas parsimonioso, sino porque también el R presenta ciertas limitaciones de velocidad cuando se trabaja con una matriz de regresión de gran dimensión a la hora de estimar un modelo SARIMA.

## 1.5. Identificación y Estimación

En esta etapa se identifican los ordenes (p,d,q)(P,D,Q) para la modelización SARI-MA de la serie  $Y_{EC_1}$  y se estima el modelo que mejor ajusta según el criterio BIC.

Las diferencias d y D vienen determinadas de la etapa anterior.

Para identificar los ordenes asociados al modelo que se ajusta a  $Y_{EC_1}$  primero se estiman todos los modelos posibles dada la combinación de órdenes máximos:  $p=3$ ,  $q=3$ ,  $P=2$  y  $Q=2$ .

Una vez estimados todos los modelos por máxima verosimilitud se escoge el 'mejor' de ellos según el menor valor del estadístico BIC.<sup>3</sup>

Luego el programa utiliza los ordenes de los polinomios autorregresivos y de medias

---

<sup>3</sup>En esta etapa el analista puede elegir el criterio de selección entre el AIC, AICC y BIC, por defecto se utiliza el BIC.

móviles obtenidos anteriormente para modelizar la serie. Si  $P$  y/o  $Q$  es cercano a uno, el programa realiza una diferencia estacional y se vuelve a re-calcular todo nuevamente.

## 1.6. Detección y Estimación Conjunta de Outliers

En esta etapa se realiza una mejora al programa de modelización automática, realizado por Cilintano y Urrutia[8], ya que el mismo anteriormente calculaba los outliers que quedaban fuera del intervalo de  $-/+ 3$  desvíos de la media de los  $\hat{\epsilon}_t$ .

Citando a Gomez y Taguas[13], el enfoque tradicional para tratar el problema de los outliers, suponiendo que se ha especificado un modelo ARMA correctamente, consiste en identificar primero la posición y el tipo de la anomalía y utilizar, a continuación, el análisis de intervención propuesto por Box y Tiao[1]. Esta aproximación requiere iteraciones entre diferentes etapas de detección y estimación de outliers dentro de un modelo de intervención. Entre los primeros intentos en esta dirección cabe citar los de Chang, Tiao y Chen[5], Hillmer, Bell y Tiao[17], y Tsay[26]. Estos procedimientos funcionan bien cuando se trata de detectar y estimar los efectos de outliers de gran relevancia y aislados, pero no resuelven el problema debido a que:

- La presencia de varios outliers puede hacer que el modelo no se especifique adecuadamente.
- Incluso si el modelo está adecuadamente especificado, los outliers pueden producir sesgos importantes en los estimadores de los parámetros.
- Puede que no se identifiquen algunos outliers debido a un efecto de enmascaramiento.

Chung Chen y Lon-Mu Liu[7], proponen un método de tratamiento de outliers para resolver algunos de estos problemas, el cual es desarrollado a lo largo del capítulo. Siguiendo su metodología, se trata de automatizar este proceso de detección y corrección de outliers siguiendo las tres etapas propuestas por los autores:

1. Estimación inicial de los parámetros y detección de los outliers.
2. Estimación conjunta de los efectos de los outliers y los parámetros del modelo.
3. Detección de outliers basados en las estimaciones finales de los parámetros.

### Definición de los cuatro tipos de Outliers

Los tipos de outliers que se han considerado son:

- El outlier innovativo (IO) es un suceso cuyo efecto se propaga en conformidad con el modelo ARMA del proceso, afectando a todos los valores observados después de su ocurrencia. El efecto de este outlier depende de la estructura ARMA de la serie.
- El outlier aditivo (A0) es un suceso (efecto externo) que afecta a una serie en un solo instante temporal ( $t = t_j$ ), causa un efecto inmediato y único en la serie observada en el período  $t_j$ , de magnitud  $w$ .
- El cambio temporal (TC) es un suceso que tiene un efecto inicial  $w$  en el período  $t_j$  y luego decae exponencialmente en conformidad con un factor de amortiguación, que denotaremos mediante el parámetro  $\delta$ .
- El cambio de nivel (LS) es un suceso que afecta a una serie en un período dado, y cuyo efecto es permanente de tipo escalón.

### 1.6.1. Detección

Realizada la identificación y estimación del modelo ajustado al que denotaremos como  $Z_t^*$ , se puede considerar a los residuos del mismo ( $\hat{e}_t$ ) como base para detectar a los outliers presentes en una serie temporal, ya que la mayoría de los diagnósticos de identificación de modelos se basan en ellos. Los outliers pueden afectar al modelo identificado como subyacente a la serie, así como a los parámetros estimados del modelo. Tal vez no es evidente el modo de como los residuos pueden ser útiles en la detección de outliers, así que para una mejor comprensión de como un outlier afecta a la serie residual, se considera la serie que se obtiene de la aplicación de un filtro:  $\hat{e}_t = \hat{\pi}(B)Z_t^*$ , donde  $\hat{\pi}(B)$  es un operador de polinomios del modelo ARMA:

$$\hat{\pi}(B) = \hat{\phi}(B)\alpha(B)/\hat{\theta}(B) = 1 - \hat{\pi}_1B - \hat{\pi}_2B^2 - \dots$$

donde los pesos  $\hat{\pi}_j$  para  $j$  más allá de un largo  $J$  son iguales a 0, porque las raíces de  $\theta(B)$  están fuera del círculo unitario.

Una vez que se obtiene  $\hat{\pi}(B)$ , los residuos estimados  $\hat{e}_t$  (los cuales pueden estar contaminados por outliers) en el modelo ajustado  $Z_t^*$  pueden ser expresados:

$$\hat{e}_t = \sum_{j=1}^m w_j \hat{\pi}(B) L_j(B) I_t(t_j) + a_t$$

donde  $w_j$  denota la magnitud del outlier  $j$  y  $L_j(B) = \theta(B) / \alpha(B)\phi(B)$  para un IO,  $L_j(B) = 1$  para un AO,  $L_j(B) = 1/(1 - B)$  para un LS y  $L_j(B) = 1/(1 - \delta^4)$  y para un TC en  $t = t_j$ .  $I_t(t_j)$  es una función indicatriz donde  $I_t(t_j) = 1$  cuando ocurre el impacto del outlier en  $t = t_j$  y  $I_t(t_j) = 0$  en otro caso, con  $t_j$  la posible ubicación del outlier.

En otras palabras, la información de un IO está contenida en los residuos  $\hat{e}_{t_j}$  en un momento particular  $t_j$ , mientras que en un AO, TC o LS la información está dispersa sobre una serie de residuos  $\hat{e}_{t_j}, \hat{e}_{t_{j+1}}, \dots$  que siguen el modelo.

De la teoría de mínimos cuadrados, los estimadores de los impactos de los outliers en  $t = t_j$  son dados por dos modelos, ya que para AO, TC y LS es igual:

$$\hat{w}_{IO}(t_j) = \hat{e}_{t_j}$$

---

<sup>4</sup>En Chen y Liu[7] recomiendan fijar el valor de  $\delta$  para un outlier del tipo TC en 0.7

$$\hat{w}_{AO}(t_j) = \left[ \sum_{t=t_j}^n \hat{e}_t x_{2t} \right] / \left[ \sum_{t=t_j}^n x_{2t}^2 \right]$$

$$\hat{w}_{TC}(t_j) = \left[ \sum_{t=t_j}^n \hat{e}_t x_{4t} \right] / \left[ \sum_{t=t_j}^n x_{4t}^2 \right]$$

$$\hat{w}_{LS}(t_j) = \left[ \sum_{t=t_j}^n \hat{e}_t x_{3t} \right] / \left[ \sum_{t=t_j}^n x_{3t}^2 \right]$$

donde  $x_{it} = 0 \forall i$  y  $t < t_j$ ,  $x_{it_j} = 1 \forall i$  y  $k \geq 1$ ,  $x_{2(t_j+k)} = -\hat{\pi}_k$ ,  $x_{3(t_j+k)} = 1 - [\sum_{j=1}^k \hat{\pi}_j]$  y  $x_{4(t_j+k)} = \delta^k - [\sum_{j=1}^{k-1} \delta^{k-j} \hat{\pi}_j - \hat{\pi}_k]$ .

Las tres últimas fórmulas también se pueden resumir como  $\hat{w}(t_j) = \hat{\rho}^2 \hat{\pi}(F) \hat{e}_{t_j}$ , donde  $\hat{\rho}^2 = (1 + \hat{\pi}_1^2 + \hat{\pi}_2^2 + \dots + \hat{\pi}_{n-t_j}^2)^{-1}$  y F es el operador tal que  $F(\hat{e}_t) = \hat{e}_{t+1}$ .

El efecto de un IO en el momento  $t_j$  puede ser estimado por el residuo de ese período  $\hat{e}_{t_j}$ , pero el mejor estimador del efecto de un AO, TC o LS es una combinación lineal de  $\hat{e}_{t_j}, \hat{e}_{t_j+1}, \dots$  con ponderaciones que dependen del modelo por el cual el proceso es generado.

Entonces, una posible aproximación para detectar outliers es examinar el valor máximo de los estadísticos estandarizados de los efectos de los outliers, siendo:

$$\hat{\tau}_{IO}(t_j) = \hat{w}_{IO}(t_j) / \hat{\sigma}_a$$

$$\hat{\tau}_{AO}(t_j) = [\hat{w}_{AO}(t_j) / \hat{\sigma}_a] \left[ \sum_{t=t_j}^n x_{2t}^2 \right]^{1/2}$$

$$\hat{\tau}_{TC}(t_j) = [\hat{w}_{TC}(t_j) / \hat{\sigma}_a] \left[ \sum_{t=t_j}^n x_{4t}^2 \right]^{1/2}$$

$$\hat{\tau}_{LS}(t_j) = [\hat{w}_{LS}(t_j) / \hat{\sigma}_a] \left[ \sum_{t=t_j}^n x_{3t}^2 \right]^{1/2}$$

Para  $t = 1, \dots, n$  se calculan los  $\hat{\tau}_{IO}$ ,  $\hat{\tau}_{AO}$ ,  $\hat{\tau}_{TC}$  y  $\hat{\tau}_{LS}$  utilizando los residuos y como estimador de la desviación estándar ( $\hat{\sigma}_a$ ) se utiliza la mediana de la desviación absoluta (MAD)<sup>5</sup> (ya que cuando hay outliers presentes no es recomendable usar el desvío estándar porque éste puede conducir a errores de estimación).

Para una ubicación determinada, los  $\hat{\tau}$  siguen una distribución aproximadamente normal. Y conociendo el tipo y ubicación de los outliers se pueden ajustar los efectos de los mismos en las observaciones y en los residuos utilizando las ecuaciones mencionadas anteriormente.

Sea  $\lambda_t = \max_t \lambda_t = \hat{\tau}_{IO}, \hat{\tau}_{AO}, \hat{\tau}_{TC}, \hat{\tau}_{LS}$ , si el  $\max_t \lambda_t > C$ , donde  $C$  es un valor crítico predeterminado<sup>6</sup>. Entonces, cuando ocurre esto hay una posibilidad de que haya un outlier del tipo  $tp$  en el tiempo  $t_j$ , donde  $tp$  puede ser IO, AO, TC o LS.

Si no se encuentra ningún outlier, se pasa directo a la etapa de diagnóstico utilizando el modelo ajustado en la etapa de identificación si no, se crean variables indicatrices para cada outlier encontrado.

### 1.6.2. Estimación conjunta

Para describir la serie de tiempo sujeta a  $m$  intervenciones, en  $t_1, t_2, \dots, t_m$  resultado de varios tipos de outliers detectados en la etapa anterior, se considera el siguiente modelo:

$$Z_t^* = \sum_{j=1}^m w_j L_j(B) I_t(t_j) + Z_t$$

donde  $Z_t = [\theta(B)/\alpha(B)\phi(B)]a_t$ ,  $t = 1, \dots, n$ , sigue un proceso ARIMA,  $n$  es el número de observaciones de la serie;  $\theta(B)$ ,  $\phi(B)$ , y  $\alpha(B)$  son polinomios en  $B$ ; donde todas las raíces de  $\theta(B)$  y  $\phi(B)$  están fuera del círculo unitario; y todas las raíces de  $\alpha(B)$  están sobre el círculo unitario.

Entonces, para los outliers potenciales ( $\lambda_t > C$ ), se estiman conjuntamente los efectos de los outliers mediante una regresión múltiple entre  $e_t$  y  $L_j(B)I_t(t_j)$ .

Luego se realizan pruebas donde se eliminan las variables de los outliers para los que cumplan  $(\hat{w}_t/\hat{\sigma}_t) < C$ . Una vez que se obtienen los outliers que cumplen la condición anterior, se realiza otra regresión. Pero esta vez entre la serie  $Z_t^*$  y el resto de las variables, del efecto calendario y variables determinísticas mensuales significativas junto con las variables de los outliers que no fueron eliminados en la etapa anterior.

Una vez estimadas conjuntamente se realizan nuevamente pruebas  $t$  correspondientes

<sup>5</sup> $\hat{\sigma}_a = 1.483 * \text{mediana}(\hat{e}(t) - \tilde{e})$ , donde  $\tilde{e}$  es la mediana de los residuos estimados.

<sup>6</sup>Al valor crítico  $C$  lo puede especificar cada analista dependiendo de la potencia que se quiera obtener en las pruebas.

para testear la significación de los parámetros uno a uno, con un nivel de confianza del 95 %, y si no se rechaza la  $H_0$  entonces se elimina ese parámetro y su variable asociada, y se crea una nueva matriz con las variables significativas para identificar el nuevo modelo que se le ajustará a la serie.

## 1.7. Diagnóstico

En esta etapa se procede a realizar pruebas y gráficos de diagnóstico a los residuos del modelo ajustado de forma automática por el programa.

Se calculan los p-valores del estadístico de Ljung-Box para testear incorrelación de los residuos, se realizan los gráficos de los residuos estandarizados y su autocorrelograma.

## 1.8. Predicción

Para realizar la predicción, el R utiliza un algoritmo basado en el Filtro de Kalman. Este algoritmo permite estimar el valor que se quiere predecir dada la información disponible hasta ese momento en tiempo real y de forma recursiva.

En este paso se calcula la predicción junto con los intervalos de predicción que se le designe (a un 80 % y 95 % por defecto), también se grafica la predicción junto con su intervalo y la serie ajustada.

Por defecto se predice a 7 pasos, ya que se quieren captar las predicciones en los distintos días de la semana.

Si el programa procedió a realizar las predicciones con la serie transformada ( $Z_t$ ), los resultados de los mismos son devueltos en la unidad de medida de la serie original ( $Y_t$ ) con sus respectivas bandas de confianza.

También se calculan en forma automática algunos estadísticos con el fin de comparar la bondad de las predicciones con otros modelos estimados para la serie. Se calcula el error medio (EM), la raíz del error cuadrático medio (RECM), y el error absoluto (EA) de las predicciones.

Sea  $e_t = y_t - \hat{y}_t$  el error de predicción. Entonces:

$$EM = E[e_t]$$

$$RECM = \sqrt{E[e_t^2]}$$

$$EA = E[|e_t|]$$

## 2. Resultados

Se comenzó realizando un análisis descriptivo de la serie de arribos diarios de ómnibus ( $Y_t$ ). El gráfico de la serie completa se presenta en la Figura 1, donde se puede observar como a principio de cada año la serie aumenta su nivel, este aumento se puede asociar a las vacaciones de verano donde las frecuencias de ómnibus seguramente se incrementen, con respecto al resto del año. Por otro lado se observa un cambio de nivel de carácter permanente a partir de mediados del año 2002.

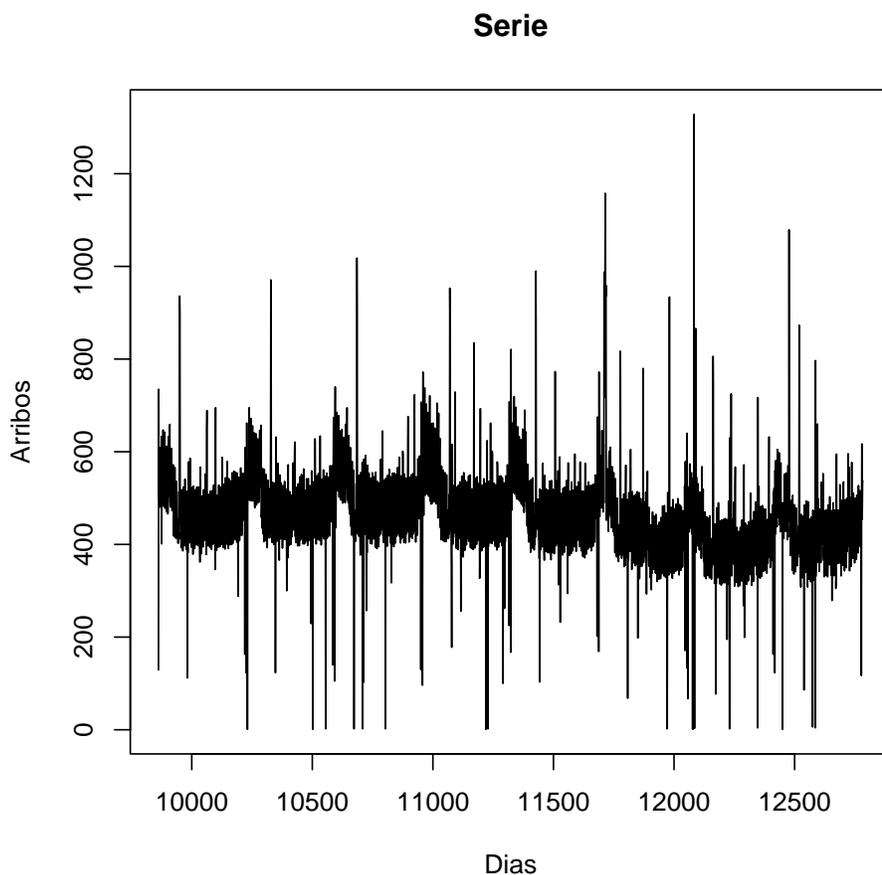


Figura 1: Serie de arribos diarios desde el 1 de enero de 1997 al 31 de diciembre de 2004.

Se calculan los estadísticos descriptivos básicos como la media, la mediana, el desvío, valor máximo y mínimo (ver Cuadro 1).

La media para la serie es de 442 arribos, mientras que el mínimo es de 1 arribo, y este valor se da para las siguientes fechas: 05/01/1998, 03/10/1998, 19/09/2000, 22/01/2003 y 30/01/2004.

El máximo de la serie se da para el domingo 28/01/2003 con 1328 arribos. Todos estos valores junto con otros atípicos serán tratados más adelante.

	<b>Arribos</b>
Media	442
Mediana	424
Mínimo	1
Máximo	1328
Desviación standar	100.6514

Cuadro 1: Estadísticos Descriptivos de la serie de arribos  $Y_t$

Dada la marcada estacionalidad existente en la serie, a efectos descriptivos se agrupó a la serie en submuestras, por un lado según el día de la semana y por otro según el mes del año. Y se calcularon las medias y desvíos para esos grupos, para observar si existen diferencias para esos períodos. Estos estadísticos se presentan en los Cuadros 2 y 3.

<b>Día</b>	<b>Mediana</b>	<b>Desvío</b>
Lunes	435	98
Martes	421	86
Miércoles	414	87
Jueves	415	80
Viernes	441	83
Sábados	400	87
Domingos	415	114

Cuadro 2: Medianas y desvíos de los agrupamientos de  $Y_t$  según el día de la semana.

En el Cuadro 2 se puede ver que el comportamiento de los arribos es bastante estable a lo largo de la semana, exceptuando los lunes y viernes. Este comportamiento se podría asociar a que las personas salen de Montevideo los fines de semana, para retornar a la misma los días lunes y las personas que viven en el interior viajen a Montevideo los días viernes.

Los meses de verano (Enero y Febrero) tienen una mediana mayor que el resto de los meses. Esto se puede asociar al período vacacional de verano, donde el flujo de ómnibus aumenta (ver Cuadro 3).

La Figura 2 respresenta el gráfico de la serie de arribos para los meses de enero y febrero de 1997, y aquí se podría pensar en una estacionalidad semanal presente en la serie.

Mes	Mediana	Desvío
Enero	514	155
Febrero	497	91
Marzo	434	88
Abril	423	122
Mayo	414	85
Junio	411	81
Julio	415	78
Agosto	409	70
Septiembre	411	74
Octubre	414	87
Noviembre	417	76
Diciembre	450	87

Cuadro 3: Medianas y desvíos de los agrupamientos de  $Y_t$  según el mes del año.

## 2.1. Efecto Calendario

Para este trabajo se utilizaron 175 variables auxiliares que señalan no sólo a los feriados, eventos especiales y vacaciones significativos para el Uruguay, sino que también a los días contiguos posteriores y anteriores a ellos.

Se analizaron los feriados no laborables por separado, ya que después de varias pruebas y agrupaciones de los mismos se decidió que este tipo de feriados afectan de manera importante a la serie y por eso no se agruparon en una misma variable.

El Cuadro 4 muestra los estadísticos de la varianza residual (VR), desvío estándar de los residuos (DR) y el criterio de información de Amemiya<sup>7</sup> (PC) luego de filtrar al logaritmo de la serie por el efecto calendario, a esta serie se le llama  $Y_{EC_0} = Y_t - \beta_0 X_0$ . Donde  $X_0$  es la matriz que contiene las variables del efecto calendario y  $\beta_0$  el vector de sus coeficientes estimados asociados.

$Y_{EC_0}(t)$	Valor
N° de Variables	175
VR	5442.07
DR	73.77
PC	5769.86

Cuadro 4: Dimensión de  $\beta_0$ , Varianza Residual, Desvío Residual y Criterio de Amemiya

<sup>7</sup>Es un criterio utilizado para selección de modelos:  $PC = VR(1 + (PM/T))$ , siendo PM el número de parámetros del modelo y T el número de observaciones de  $Y_{EC_0}$

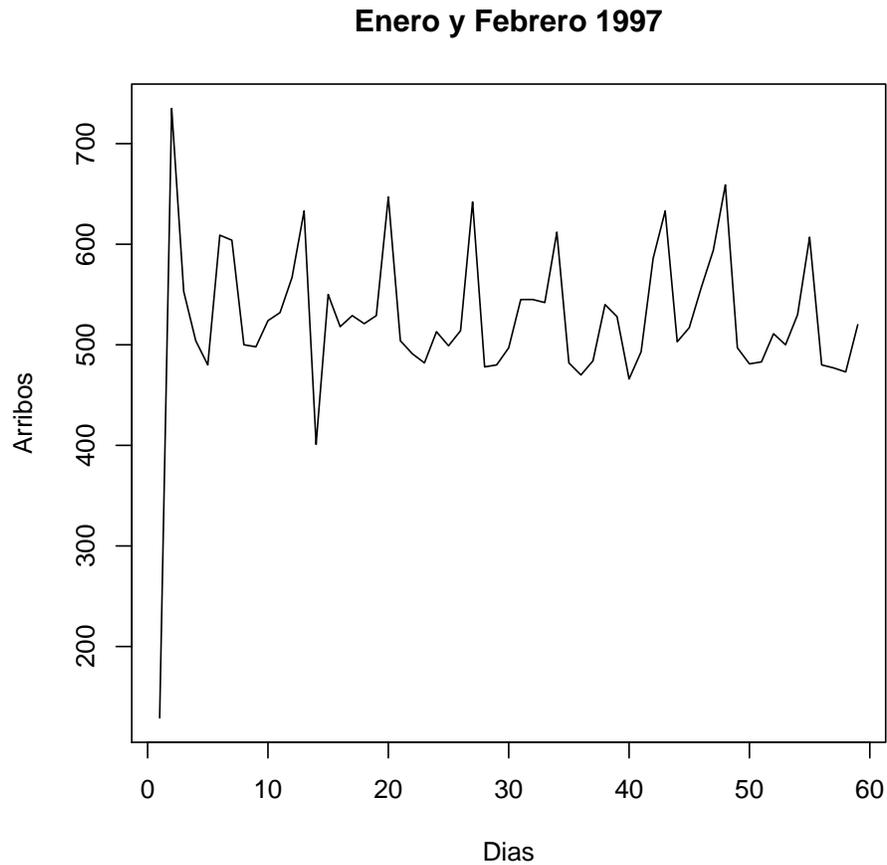


Figura 2: Serie de arribos diarios para los meses de enero y febrero de 1997.

## 2.2. Esquemas estocásticos

Se realizan una serie de transformaciones estacionarias para aplicar a la serie  $Y_{EC_0}$ . Los esquemas estocásticos que se proponen inicialmente son:

- $Y_{EC_0}$  (no aplicarle ningún esquema)
- $\Delta Y_{EC_0}$
- $\Delta^2 Y_{EC_0}$
- $\Delta_7 Y_{EC_0}$
- $\Delta_{365} Y_{EC_0}$
- $\Delta\Delta_7 Y_{EC_0}$
- $\Delta\Delta_{365} Y_{EC_0}$

Transformación	VR
$Y_{EC_0}$	5442.07
$\Delta Y_{EC_0}$	9497.50
$\Delta^2 Y_{EC_0}$	27165.32
$\Delta_7 Y_{EC_0}$	11050.16
$\Delta_{365} Y_{EC_0}$	11928.27
$\Delta \Delta_7 Y_{EC_0}$	21410.04
$\Delta \Delta_{365} Y_{EC_0}$	21410.05

Cuadro 5: Varianza Residual para distintas transformaciones de  $Y_{EC_0}$ 

En el Cuadro 5 se puede observar como el esquema sin diferencias es el que tiene menor VR, por lo tanto el programa automático continúa trabajando con éste.

### 2.3. Esquemas Determinísticos para modelizar la estacionalidad

En esta etapa se agregan esquemas determinísticos para captar otras estacionalidades secundarias de otros períodos que aún pueden permanecer presentes en la serie (como la estacionalidad anual o la mensual).

El esquema corresponde a variables que indican los meses del año, son 12 variables con unos para todos los días de determinado mes y ceros para el resto (el programa cuando realiza la regresión deja un mes afuera de forma automática).

El procedimiento consiste en agregar las variables 'mensuales' a la matriz  $X_0$  (formada por las variables del EC) y de esta forma obtener la matriz  $X_1$ . Luego filtrar a  $Y_t$  por las variables se obtiene:  $Y_{EC_1} = Y_t - \beta_1 X_1$ .

A efectos de continuar con el análisis se sigue con el modelo que incluye las variables mensuales, ya que luego de que se agregaron dichas variables, la VR disminuyó un 4.70% como muestra el Cuadro 6.

$Y_{EC_1}$	Valor	Reducción (%)
N° var	186	-0.63
VR	5186.14	4.70
DR	72.01	2.38
PC	5518.04	4.36

Cuadro 6: Dimensión de  $\beta_1$ , Varianza Residual, Desvío Residual y Criterio de Amemiya de  $Y_{EC_1}$  y reducción en porcentaje de ellos en comparación con los estadísticos de  $Y_{EC_0}$ 

Teniendo el modelo estimado se procede a realizar las pruebas t correspondientes

para poner a prueba la significación de los parámetros uno a uno, con un nivel de confianza del 95 % ( $H_0$ )  $\beta_{2,i} = 0$ , y si no se rechaza la  $H_0$  entonces se elimina ese parámetro y su variable asociada, y se crea la matriz  $X_2$  que contiene solo a las variables significativas. Se filtra a la serie  $Y_t$  por las variables  $X_2$  y se obtiene  $Y_{EC_2} = Y_t - \beta_2 X_2$ . Los parámetros que tenía el modelo hasta el momento eran 186, y luego de la reducción quedaron 67 parámetros significativos. (Ver Cuadro 7).

$Y_{EC_2}$	Valor	Reducción (%)
N° var	67	63.98
VR	5176.768	0.18
DR	71.94976	0.09
PC	5297.241	4.00

Cuadro 7: Dimensión de  $\beta_2$ , Varianza Residual, Desvío Residual y Criterio de Amemiya de  $Y_{EC_2}$  y reducción en porcentaje de ellos en comparación con los estadísticos de  $Y_{EC_1}$

Hasta aquí se vino utilizando la VR como criterio de inclusión. En los Cuadros anteriores también figuran otros estadísticos como el desvío residual (DR) y el criterio de Amemiya (PC) que también se reducen cuando se reduce la VR.

## 2.4. Identificación y Estimación del Modelo

### Órdenes $p$ , $q$ , $P$ y $Q$

Luego de filtrar a la serie por el Efecto Calendario, las variables determinísticas mensuales y reducirlo mediante pruebas t, se calculan los órdenes  $p$ ,  $q$ ,  $P$  y  $Q$  del modelo SARIMA a ajustar. Para ello existe una función en R llamada “Best Arima” la cual utiliza los residuos de la serie con la que se viene trabajando, por lo que solamente hay que especificarle el valor máximo de los órdenes de cada proceso ( $p=3$ ,  $q=3$ ,  $P=2$  y  $Q=2$ ), ya que el orden de la diferencia regular y el orden de la diferencia estacional ( $d=0$  y  $D=0$ ) vienen determinados por la etapa 2.2, así como también el período también viene determinado desde que se identificó una estacionalidad como principal (período = 7).

En esta función el analista puede especificar el estadístico para la selección del modelo (BIC, AIC o AICC), por defecto el programa utiliza el BIC, ya que es el que tiende a elegir modelos más parsimoniosos.

Luego de que el programa corre el Best Arima, si los resultados de  $P$  y/o  $Q$  son cercanos a 1, se le realiza una diferencia estacional a la serie y se vuelve a correr el programa.

El modelo intermedio identificado que presentó menor BIC para la serie filtrada y diferenciada fue un  $SARIMA(1, 0, 1)(0, 1, 1)_7$ , con una varianza estimada igual a 3334.

## 2.5. Estimación Conjunta de los parámetros del modelo y los efectos de los outliers en la serie

Para diferenciar las etapas en las que se trabaja, desde aquí en adelante llamaremos a la Serie  $Z_t$  en lugar de  $Y_t$  simplemente por un tema de notación.

Para comenzar, se definen las siguientes variables para el programa:

$n$  = largo de la serie = 2922 días.

object = modelo intermedio =  $SARIMA(1, 0, 1)(0, 1, 1)_7$

resid = residuos del modelo identificado en la etapa anterior  $\hat{e}_{EC2}$  que desde ahora llamaremos simplemente  $\hat{e}_t$

C = valor critico = 4.3

Una vez definidas estas variables, el programa calcula los efectos  $\hat{w}_t$  para los residuos  $\hat{e}_t$ , luego los estandariza (donde se obtienen los  $\hat{\tau}_t$ ) utilizando una estimación robusta de  $\sigma_a$  (MAD), y calcula el máximo para cada tipo de outlier ( $\hat{\lambda}_t$ ). Si  $\hat{\lambda}_t$  resulta ser mayor que  $C$  se los clasifica como outlier y mantienen el tipo (IO, AO, TC o LS) dentro de los cuales fueron clasificados anteriormente.

Al encontrar outliers de los cuatro tipos se pasa a la siguiente etapa donde se realiza una primera regresión múltiple para estimar el efecto de los outliers en su conjunto,  $e_o = \hat{e}_t - \beta_o X_o$ , siendo  $X_o$  una matriz formada por 565 variables de potenciales outliers y  $\beta_o$  el tamaño del efecto a estimar. Una vez obtenidas las estimaciones se eliminan las variables de los outliers que no cumplen:  $[\hat{w}_j / \sigma_{\hat{w}_j}] < C$ , y se procede a realizar una nueva regresión ( $e_{o1} = \hat{e}_t - \beta_{o1} X_{o1}$ ) para estimar los coeficientes asociados a las variables de los outliers que si cumplían la condición, donde se obtuvieron 44 de los cuales, 35 son del tipo IO y 4 son del tipo AO y 5 del tipo LS.

Las variables de la matriz  $X_{o1}$  obtenidas anteriormente se agregaron a la matriz  $X_2$  (formada por las variables significativas del EC y las variables determinísticas mensuales obtenidas en 2.3) y de esta forma obtener la matriz  $X_3$  con 111 variables. Luego de filtrar a  $Z_t$  por  $X_3$  se obtiene  $Z_3 = Z_t - \beta_3 X_3$ . Teniendo el modelo estimado se procede a realizar las pruebas  $t$  correspondientes para someter a prueba la significación de los parámetros uno a uno, con un nivel de confianza del 95 %, con  $((H_o): \beta_{3,i}=0)$  y eliminando el parámetro y su variable asociada cuando no se rechaza  $(H_o)$ . A partir de aquí se crea la matriz  $X_4$  formada solo por las 91 variables significativas obteniendo  $Z_4 = Z_t - \beta_4 X_4$ .

El Cuadro 8 muestra los estadísticos VR, DR y PC luego de filtrar la serie por el Efecto Calendario y los Outliers. Dicho Cuadro también muestra el porcentaje de la reducción de los estadísticos VR, DR y PC en comparación con los de  $Y_2$ .

Una vez que están los coeficientes significativos se utiliza la matriz  $X_4$  en la función “Best arima” para identificar el modelo que se le ajustará luego a la serie.

El modelo identificado es un  $SARIMA(1, 0, 1)(0, 1, 2)_7$  con una varianza estimada igual a 2532. Los resultados de la estimación se presentan en el Cuadro 9 y el resto de

$Z_4$	Valor	Reducción (%)
N° var	91	-35.82
VR	4400.35	14.99
DR	66.33	7.81
PC	4538.90	14.31

Cuadro 8: Dimensión de  $\beta_4$ , Varianza Residual, Desvío Residual y Criterio de Amemiya de  $Z_4$  y reducción en porcentaje de ellos en comparación con los estadísticos de  $Y_{EC_2}$

la salida se puede encontrar en el Apéndice.

	AR1	MA1	SMA1	SMA2
Coef.	0.749	-0.590	-0.785	-0.172
S.e.	0.056	0.073	0.020	0.020

Cuadro 9: Estimación de los parámetros del modelo

## 2.6. Diagnóstico del Modelo

A continuación se presentan las pruebas realizadas a los residuos del modelo  $\Delta_7 Z_4^*$ . Primero se hizo una prueba t para testear el valor de la media de los residuos:

t-test

datos:

-  $\Delta_7(Z_4^*)$

- t = -0.4313

- df = 2921

- p-value = 0.6663

- Hipótesis alternativa: La media no es igual a 0.

- Intervalo de confianza al 95 %:

(-2.223980 , 1.422056 ).

- Estimaciones muestrales: media de x: -0.400962

Y luego se realizó el contraste de Ljung Box para observar la incorrelación de los residuos y estos presentan correlaciones en sus retardos.

## 2.7. Predicción del Modelo

Siguiendo con la modelización, en este paso se procede a probar al modelo estimado  $SARIMA(1, 0, 1)(0, 1, 2)_7$  en predicción.

Se realiza la predicción a 7 pasos para ver la evolución de los valores predichos en la semana.

En el Cuadro 10, se presentan los resultados de las predicciones del modelo, su intervalo de predicción al 95 % de confianza y los valores reales de la serie de arribos para los primeros 7 días de enero.

Día	O.Real	Estimación	inf 95 %	sup 95 %	Error Relativo
1/1/05	2	122	23	221	-60
2/1/05	576	726	626	826	-0.26
3/1/05	695	641	541	742	0.08
4/1/05	500	516	415	617	-0.03
5/1/05	467	426	325	528	0.09
6/1/05	414	528	427	630	-0.27
7/1/05	492	561	459	662	-0.14

Cuadro 10: Valor real de la serie de arribos y predicciones del modelo para los primeros 7 días de enero de 2005 junto con su intervalo de predicción al 95 % y el error relativo de predicción.

Como se puede observar en el Cuadro 10 el valor real del 1o de Enero del 2005 es un outlier por lo que no se debería considerar el error relativo entre la observación y la predicción. Para el resto de las predicciones puede observarse que los errores relativos son cercanos a cero. Los valores de los estadísticos del Error medio (EM), la raíz del error cuadrático medio (RECM) y el error absoluto medio (EA) para las predicciones del modelo se presentan en el Cuadro 11.

Modelo	Error Medio	RECM	Error Abs. Medio
M.c/det out	-0.40	50.25	24.54

Cuadro 11: Error medio, Raíz del error cuadrático medio y error absoluto medio de la predicción del modelo.

### 3. Conclusión

La metodología que se describe en este trabajo fue realizada mayormente en Cilintano y Urrutia[8] a la cual se le realiza una mejora en la etapa de detección y estimación de outliers. Previamente se detectaban los outliers que se encontraban por debajo y por encima de un intervalo de 3 desvíos de la media de los residuos.

En el presente trabajo se diseña un programa automático que detecta a los outliers a partir de un modelo intermedio, los clasifica en IO, AO, TC y LS y los estima conjuntamente.

Esta metodología fue aplicada a una serie de arribos de ómnibus de ocho años, la cual se filtró por el efecto calendario (EC) intentando modelizar sus estacionalidades mediante una estructura mixta(estocástica y determinística). Luego se identificó un modelo intermedio que fue utilizado como base para detectar los outliers, los cuales fueron satisfactoriamente clasificados y estimados. Una vez concluida la etapa anterior se identificó el modelo final, el cual se utilizó para realizar predicciones del 1o de Enero al 7 de Enero del 2005 obteniendo resultados aceptables como se puede observar en el Cuadro 10.

La mayor ventaja de trabajar con el software R utilizado para llevar a cabo la programación de dicha metodología es que el mismo es de libre uso y además está al alcance de cualquier analista que desee trabajar con series de cualquier tipo de agregación temporal.

Una de las desventajas que se encontró al trabajar con R es el tiempo que lleva correr algunos programas cuando se trabaja con muchos datos (8 años en este caso) y la otra es cuando se utiliza la función “Best Arima”, a la cual no se le puede especificar el máximo de las diferencias(regular o estacional) por lo que deben ser identificadas anterior o posteriormente.

## Referencias

- [1] **Box, G.E. Tiao, G.C.(1975)** '*Intervention Analysis with Applications to Economic and Environmental Problems.*' Journal of The American Statistical Association. Vol 70 N°70-79.
- [2] **Cancelo, J.R., Espasa, A.(1991).** '*Forecasting Daily Demand for Electricity with Multiple-Input NonLinear Transfer Function Models: A case Study.*'. Working Papers 91-05 Universidad Carlos III de Madrid.
- [3] **Cancelo, J.R., Espasa, A.(1991).** '*Threshold Modelling of Nonlinear Dynamic Relationships: An Application to a Daily Series of Economic Activity.*'. Working Papers 91-05 Universidad Carlos III de Madrid.
- [4] **Cancelo, J.R. y Espasa, A. (1996)** '*Modelling and Forecasting Daily Series of Electricity Demand.*' Investigaciones Económicas. Volúmen XX(3), pp.359-376.
- [5] **Chang, I., Tiao, G. C., y Chen, C. (1988)** '*Estimation of time series parameters in the presence of outliers.*' Technometrics, 30, 193204.
- [6] **Chen, R. y Liu, L.(1993)** '*Forecasting time series with outliers.*' Journal of Forecast, 12, 1335.
- [7] **Chen, R. y Liu, L.(1993)** '*Joint Estimation of Model Parameters and Outlier Effects in Time Series.*' Journal of The American Statistical Association. Vol.88 N°421.
- [8] **Cilintano, S. y Urrutia, C.(2007)** '*Modelización automática de series diarias de actividad económica.*' Pasantía Licenciatura en Estadística.
- [9] **Espasa, A. Revuelta, J.M. Cancelo, J. (1996)** '*Modelización Automática de Series Diarias de Actividad Económica.*' Working Papers 96-07. Universidad Carlos III de Madrid.
- [10] **Espasa, A. (1993)** '*Modelling Daily Series of Economic Activity.*' Working Papers 93-32 Universidad Carlos III de Madrid.
- [11] **Espasa, A. y Cancelo, J. (1993)** '*Métodos Cuantitativos para el Análisis de la Coyuntura Económica.*', Alianza Editorial, Madrid.
- [12] **Gómez, V., Maravall, A. y Peña, D. (1999)** '*Missing observations in ARIMA models: Skipping approach versus additive outlier approach.*' Journal of Econometrics 88(1999) 341-363.
- [13] **Gómez y Taguas (1995)** '*Detección y Corrección Automática de Outliers con Tramo: Una aplicación al IPC de Bienes Industriales No Energéticos.*' D-95006 Noviembre 1995.

- 
- [14] **Gómez, V. y Maravall, A. (1994)** '*Estimation, Prediction, and Interpolation for Nonstationary Series with the Kalman Filter*'. Journal of The American Statistical Association. Vol.89 611-624.
- [15] **Gómez, V. y Maravall, A. (1996)** '*Programs TRAMO and SEATS. Instruction for the User.*' Beta Version: October 1996.
- [16] **Gómez y Maravall (1998)** '*Automatic Modeling Methods for Univariate Series.*' Banco de España-Servicio de Estudios. Documento de Trabajo n°9808.
- [17] **Hillmer, S.C., Bell, W.R. and Tiao, G.C. (1983)** '*Modeling Considerations in the Seasonal Adjustment of Economic Time Series.*' Washington, D.C.: US Bureau of the Census, 74-100.
- [18] **Lon-Mu Liu (1980)** '*Analysis of Time Series with Calendar Effects*'. Management Science. Vol 26, N°1.
- [19] **Maravall, A. (1981)** '*Desestacionalización y Política Monetaria.*' Banco de España-Servicio de Estudios. Estudios Económicos N°19.
- [20] **Maravall, A. y Kaiser, R.(2000)** '*Notes on time series analysis ARIMA models and signal extraction.*' Banco de España-Servicio de Estudios. Documento de trabajo N° 0012.
- [21] **Peña, D. (1989)** '*Modelos y Métodos 2. Modelos Lineales y Series Temporales.*' Segunda Edición, Alianza Editorial.
- [22] **Pollock, D.(1978)** '*The methods of Time Series Analysis.*' Interdisciplinary Science Reviews 12, N°2.
- [23] **Revuelta, J. M.(2000)** '*Desarrollo de una Metodología Automática de Modelización de Series Diarias de Actividad Económica.*' Tesis Doctoral Universidad Carlos III de Madrid.
- [24] **Tiao, G. y Tsay, R.(1983)** '*Consistency Properties of Least Squares Estimates of Autoregressive Parameters in ARMA Models.*' The Annals of Statistics, Vol 11 N°3.
- [25] **Tsay, R.(1984)** '*Regression Models with Time Serie Errors.*' Journal of the American Statistical Association, Vol 79 N°385.
- [26] **Tsay, R.S. (1988)** '*Outliers, Level Shifts, and Variance Changes in Time Series.*' Journal of Forecasting 7: 1-20.

# Apéndice

Constancia Urrutia

Pasantía Licenciatura en Estadística

Facultad de Ciencias Económicas y Administración

*Universidad de la República*

21 de febrero de 2010

## 0.1. Gráficos de las diferentes estructuras utilizadas para modelizar la estacionariedad

A continuación se presentan los gráficos de los autocorrelogramas de las diferencias aplicadas a la serie  $Y_{EC_0}$  de los arribos filtrada por el efecto calendario.

Dichas transformaciones son:

- $Y_{EC_0}$  (no aplicarle ningún esquema)
- $\Delta Y_{EC_0}$
- $\Delta^2 Y_{EC_0}$
- $\Delta_7 Y_{EC_0}$
- $\Delta_{365} Y_{EC_0}$
- $\Delta\Delta_7 Y_{EC_0}$
- $\Delta\Delta_{365} Y_{EC_0}$

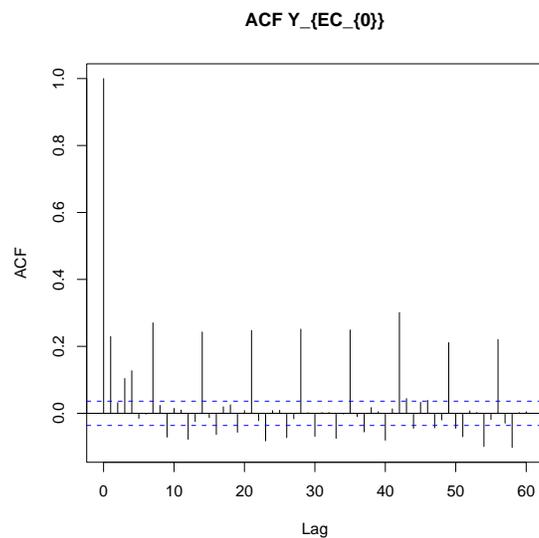


Figura 1: ACF de la serie  $Y_{EC_0}$

En la figura 8 se puede observar el ACF de la Serie filtrada por el efecto calendario y las variables mensuales, la cual es la seleccionada por tener menor VR.

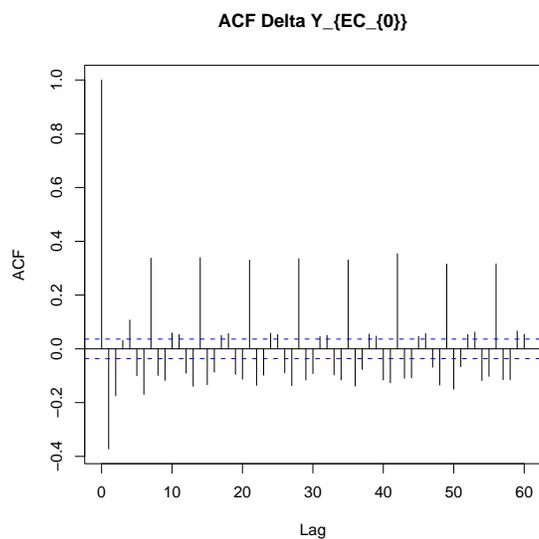


Figura 2: ACF de la serie  $\Delta Y_{EC_0}$

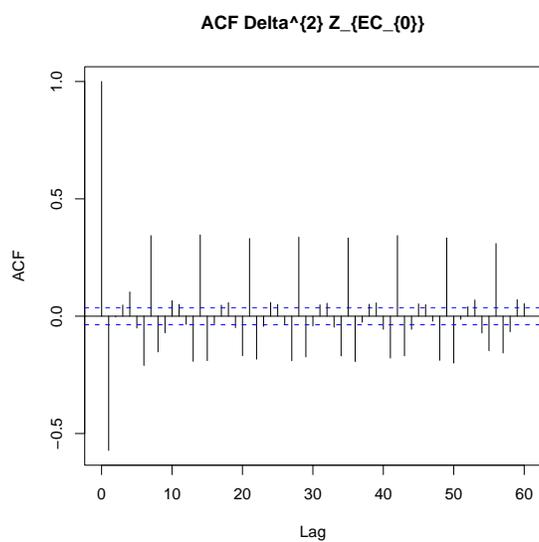


Figura 3: ACF de la serie  $\Delta^2 Y_{EC_0}$

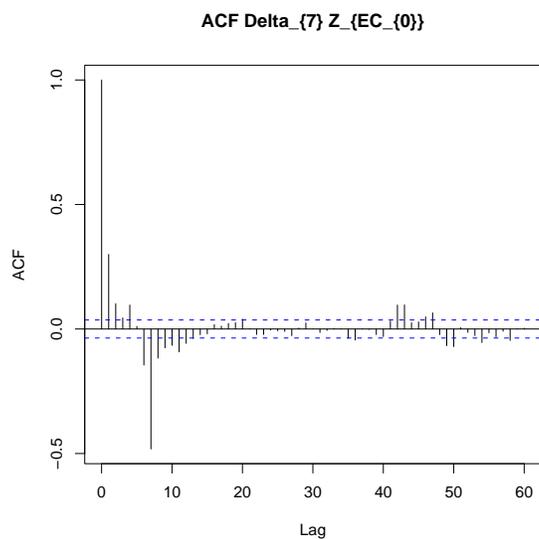


Figura 4: ACF de la serie  $\Delta_7 Y_{EC_0}$

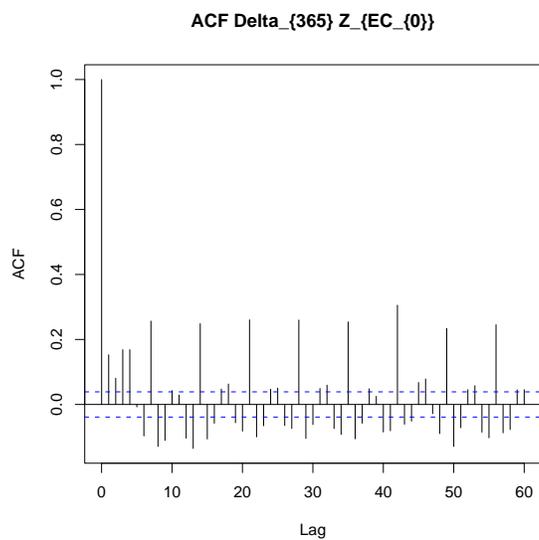


Figura 5: ACF de la serie  $\Delta_{365} Y_{EC_0}$

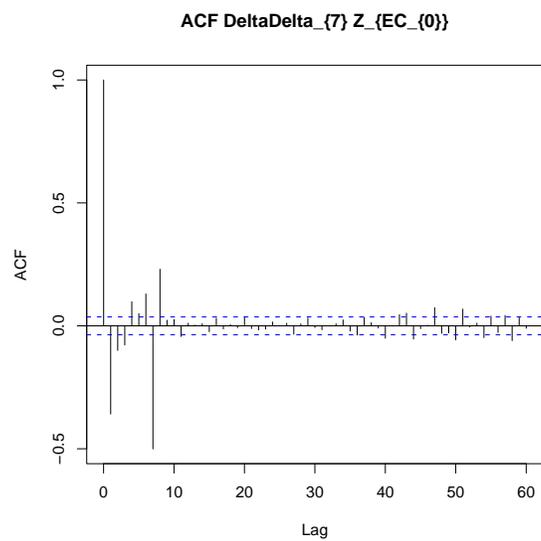


Figura 6: ACF de la serie  $\Delta\Delta_7 Y_{EC_0}$

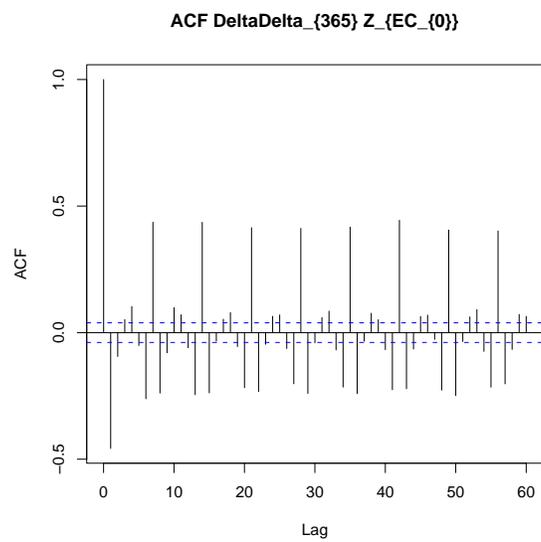


Figura 7: ACF de la serie  $\Delta\Delta_{365} Y_{EC_0}$

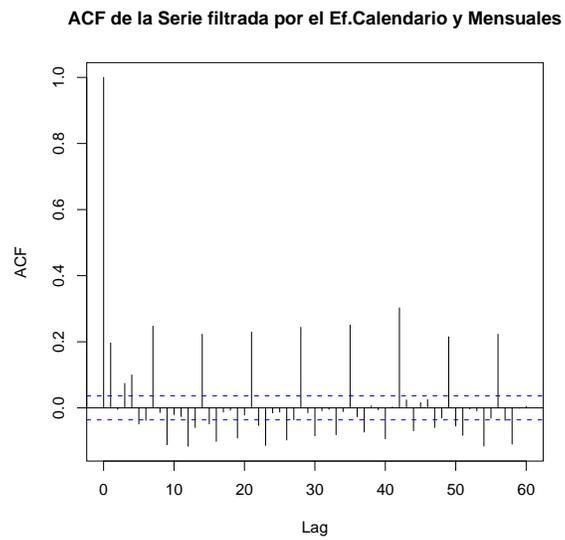


Figura 8:  $Y_{EC1}$ , Serie filtrada por EC y variables determinísticas mensuales

## 0.2. Gráficos de los Outliers Significativos en la Quinta Regresión

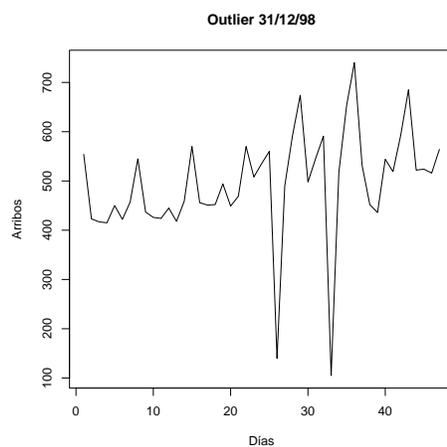


Figura 9: Outlier IO.20 en la serie desde 1/12/1998-15/1/1999

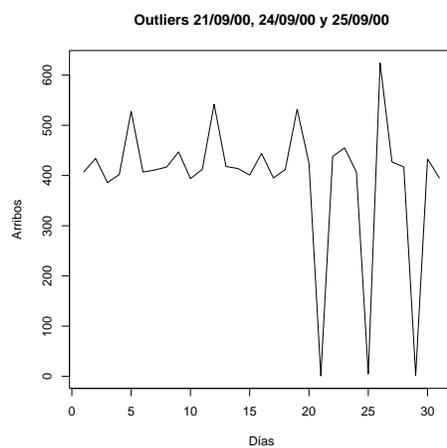


Figura 10: Outliers IO.45,46 y 47 en la serie desde 1/9/2000 al 1/10/2000

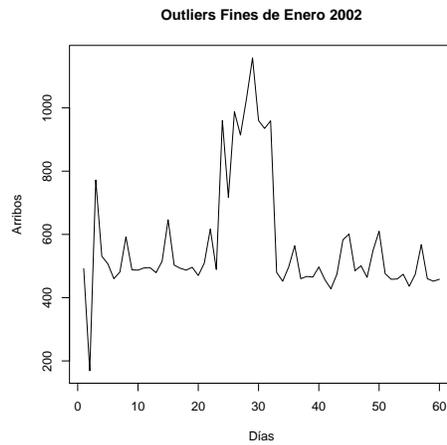


Figura 11: Outliers IO.67,69,70,71,72 y 74 en la serie desde 1/1/2002-1/3/2002

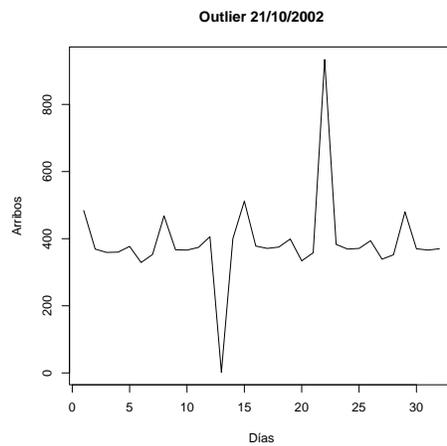


Figura 12: Outlier IO.83 en la serie desde 1/10/2002-1/11/2002

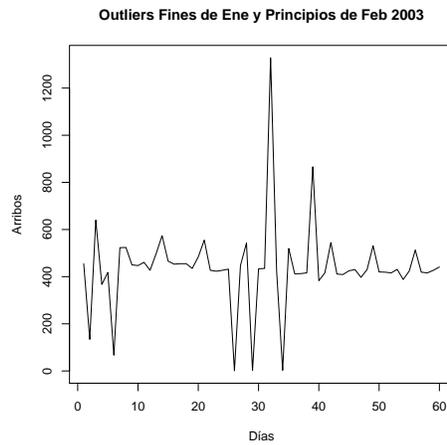


Figura 13: Outliers IO.87,89,90 y 91 en la serie desde 1/1/2003-1/3/2003

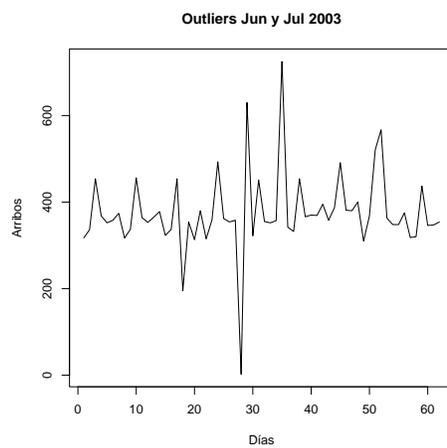


Figura 14: Outliers IO.95 y 96 en la serie desde 1/6/2003-1/8/2003

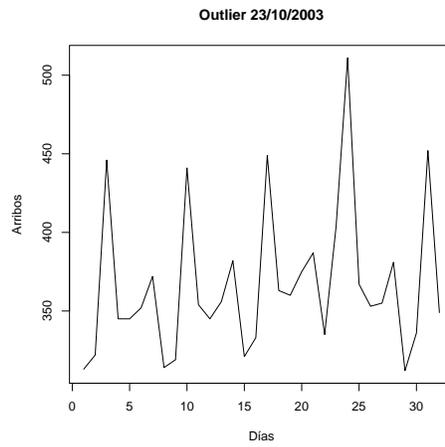


Figura 15: Outliers IO.87,89,90 y 91 en la serie desde 1/10/2003-1/11/2003

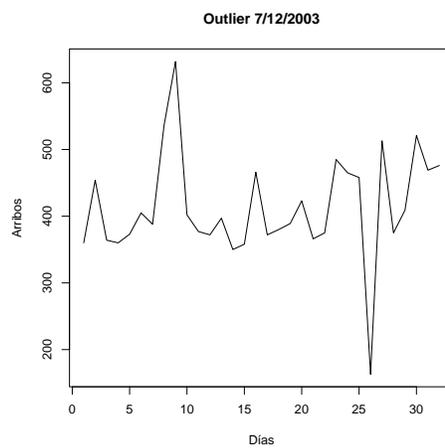


Figura 16: Outliers IO.100 en la serie desde 1/12/2003-1/1/2004

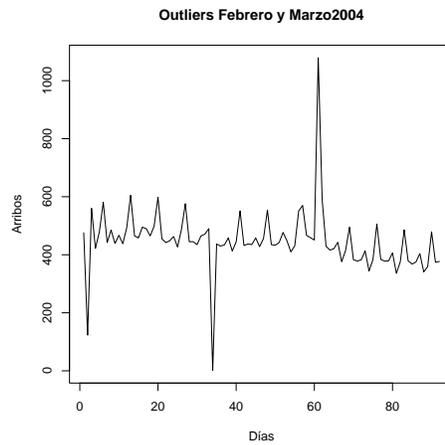


Figura 17: Outliers IO.102, IO.103 y LS.589 en la serie desde 1/2/2004-1/4/2004

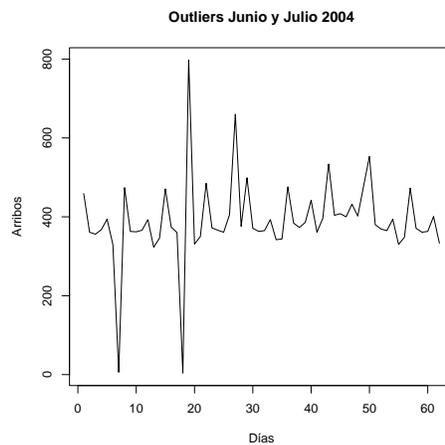


Figura 18: Outliers IO.106, IO.107, LS.674 y LS.679 en la serie desde 1/6/2004-1/8/2004

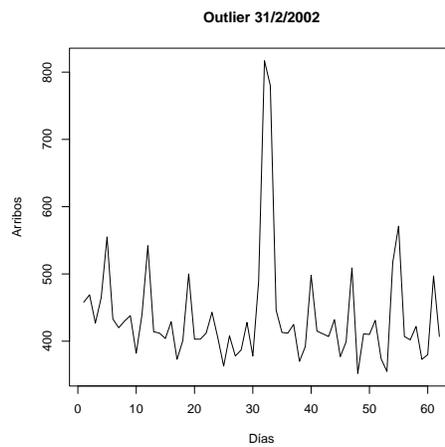


Figura 19: Outliers AO.146 en la serie desde 1/3/2002-1/5/2002

### 0.3. Salida del Modelo Seleccionado $SARIMA(1, 0, 1)(0, 1, 2)_7$

Coefficientes	estimación	S.E.
ar1	0.7488	0.0564
ma1	-0.5900	0.0726
sma1	-0.7851	0.0203
sma2	-0.1716	0.0200
crisis	-51.698	9.700
ene1.fn1	-500.6248	49.5276
ene1.fn1m	273.5139	49.4772
ene1.fn1m	-343.1277	49.5305
ene1.fn1mx	255.1270	49.4883
ene1.fn1x	-392.3875	22.1102
ene1.fn1xj	200.8287	19.5213
ene1.fn1j	-401.3881	35.0549
ene1.fn1v	-488.0594	49.1444
ene1.fn1vl	52.7810	49.1575
ene1.fn1s	-466.2981	48.8108
may1.fn1	-360.1811	48.5078
may1.fn1m	178.0352	48.3875
may1.fn1m	-304.5656	48.0524
may1.fn1x	-334.0550	47.9562
may1.fn1j	-290.3247	33.9349
may1.fn1v	-325.1643	47.9894
may1.fn1vl	76.5322	48.0311
may1.fn1s	-280.9318	33.9212
jul18.fn1j	-106.9155	48.1234
jul18.fn1v	187.6023	34.2107
jul18.fn1vl	130.7927	34.1952
jul18.fn1sd	120.8051	47.9956
jul18.fn1d	101.8983	34.3070
jul18.fn1dl	76.7520	34.3258

Cuadro 1: Estimación y Error estándar de los coeficientes asociados a las variables del modelo Maut, Primera Parte

Coefficientes	estimación	S.E.
dic25.fnl	-386.7548	49.4794
dic25.fnlm	206.7282	49.4703
dic25.fnm	-268.4935	49.4842
dic25.fnmx	197.5867	49.4710
dic25.fnx	-286.7514	48.4116
dic25.fnj	-308.3483	34.7678
dic25.fnv	-399.5747	49.0892
dic25.fnvl	45.3020	49.1335
dic25.fns	-362.0648	35.4144
ago25.fnl	29.3642	34.2219
ago25.fnlm	261.0914	34.2157
ago25.fnx	-100.9640	33.9709
ago25.fnvl	142.6628	48.0727
flcont.fld	-36.4910	10.4241
flcont.fl	-4.7610	10.4813
flcont.flm	141.7628	10.4011
flscont.fls	-81.4425	21.7477
fldcont.fdl	8.3265	19.7682
fnclcont.fncl	-91.7014	15.7909
fnclcont.fnclm	-326.8638	33.9922
fnclcont.fnclx	-66.5028	34.1531
santa.m1ss	-29.1437	17.1781
santa.j1ss	73.9726	17.1258
santa.s2ss	169.9236	17.4744
santa.d2ss	518.2578	18.5641
santa.l2ss	343.5246	17.4992
carna.cm	131.2791	17.8635
carna.cx	165.7663	17.3244
eleccion.d281199	173.3244	48.6623
eleccion.l281199	150.5557	48.6028
eleccion.d311004	178.2872	48.8322
eleccion.l311004	346.4430	59.5189
eleccion.d140500	152.6127	48.4629
eleccion.l140500	184.0431	48.4649

Cuadro 2: Estimación y Error estándar de los coeficientes asociados a las variables del modelo Maut, Segunda Parte

Coefficientes	estimación	S.E.
verano	32.1566	6.3556
enero	90.4301	7.8740
febrero	47.6064	8.0380
marzo	17.7826	6.0095
noviembre	9.4298	5.9399
diciembre	34.6590	6.4762
IO(31/12/1998)	92.0852	32.7263
IO(21/9/2000)	251.3011	33.1147
IO(24/9/2000)	-223.7482	32.6170
IO(24/1/2002)	30.6637	42.6308
IO(26/1/2002)	216.5843	33.7612
IO(27/1/2002)	213.2902	32.8106
IO(28/1/2002)	213.8462	32.6722
IO(29/1/2002)	152.863	32.682
IO(31/1/2002)	148.8565	40.7803
IO(21/10/2002)	232.6136	32.4336
IO(25/1/2003)	-184.1526	32.6495
IO(31/1/2003)	503.9439	40.2961
IO(2/2/2003)	-238.4774	32.4861
IO(7/2/2003)	441.9296	40.2526
IO(28/6/2003)	210.9191	32.9935
IO(4/7/2003)	200.5967	32.5635
IO(23/10/2003)	219.3258	32.8245
IO(7/12/2003)	87.4340	32.5537
IO(2/2/2004)	-286.2615	32.5335
IO(29/2/2004)	337.1837	32.5327
IO(18/6/2004)	248.5487	33.7248
IO(26/6/2004)	156.7909	32.8436
AO(31/3/2002)	-86.4263	51.4521
LS(2/3/2004)	0.3335	0.1207
LS(25/6/2004)	55.9880	20.2577
LS(2/7/2004)	93.3531	33.7863

Cuadro 3: Estimación y Error estándar de los coeficientes asociados a las variables del modelo Maut, Tercera Parte

## 0.4. Gráfico de la predicción

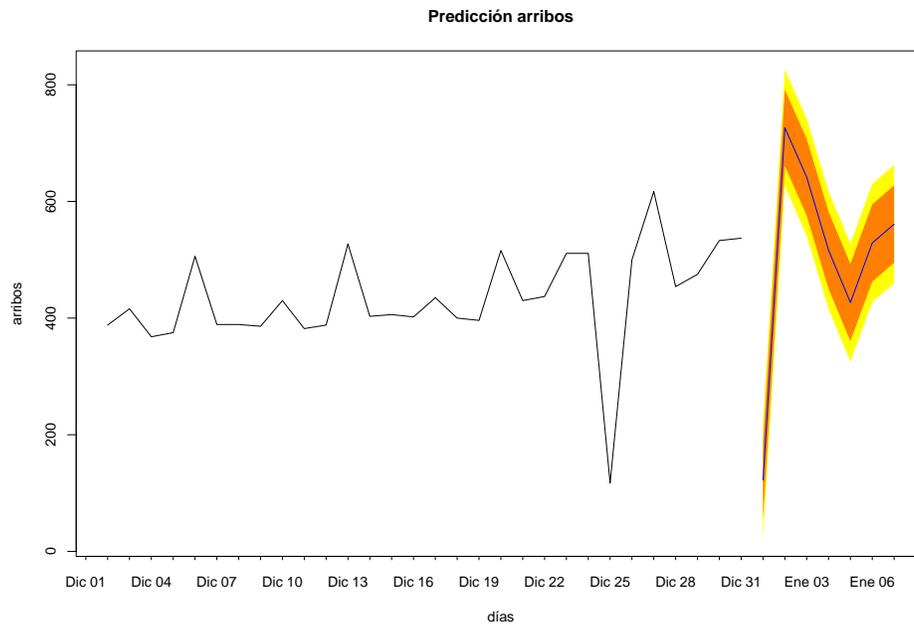


Figura 20: Serie de arribos de ómnibus desde el 1° al 31 de diciembre de 2004 y predicción del modelo Maut con detección de outliers para los primeros 7 días de enero de 2005 con sus intervalos de predicción a un 80 % y 95 % de confianza.