

Reconstrucción de datos de series de tiempo: una aplicación a la demanda horaria de la electricidad

Time Series Data Reconstruction: An Application to the Hourly Demand of Electricity

ELKIN CASTAÑO^a

FACULTAD DE CIENCIAS, UNIVERSIDAD NACIONAL DE COLOMBIA, MEDELLÍN, COLOMBIA
FACULTAD DE CIENCIAS ECONÓMICAS, UNIVERSIDAD DE ANTIOQUIA, MEDELLÍN, COLOMBIA

Resumen

Generalmente, la identificación y estimación de modelos ARIMA parten del supuesto de que las series que se van a analizar no contienen datos faltantes, ni observaciones atípicas, ni existen intervenciones en el período de estudio. Sin embargo, en la práctica, estos problemas pueden ocurrir simultáneamente, afectando la identificación del modelo adecuado y por tanto su capacidad de pronóstico. Este artículo presenta un procedimiento que permite estimar el efecto de las intervenciones, de las observaciones atípicas, estimar las observaciones faltantes y simultáneamente identificar el modelo ARIMA. El procedimiento se aplica a una serie de demanda horaria de electricidad en la cual ocurren los tres eventos mencionados.

Palabras clave: observaciones atípicas, observaciones faltantes, intervención, función de transferencia, ARIMA.

Abstract

Usually, in the identification and estimation of ARIMA models it is supposed that the series to analyze contain neither missing data, nor atypical observations, and interventions do not exist under study period. Nevertheless, in the practice, these problems can happen simultaneously, affecting the identification of the suitable model and therefore his forecasting capacity. This article presents a procedure that allows to estimate the effect of the interventions, of the atypical observations, to estimate the missing observations and simultaneously to identify the ARIMA model. The procedure is applied to a series of hourly electricity demand in which the three mentioned events happen.

Key words: Atypical observations, Missing observations, Intervention, Transfer function, ARIMA.

^aProfesor asociado, profesor titular. E-mail: elkincastano@gmail.com

1. Introducción

Generalmente, la identificación y estimación de modelos ARIMA parten del supuesto de que las series que se van a analizar no contienen datos faltantes, ni observaciones atípicas, ni existen intervenciones en el período de estudio. Sin embargo, en la práctica, estos problemas pueden ocurrir simultáneamente, afectando la identificación del modelo y por tanto su capacidad de predicción. En la literatura de series de análisis de series de tiempo se han propuesto métodos para tratar datos faltantes, observaciones atípicas e intervenciones sobre la serie. Algunos de estos procedimientos se encuentran en Box & Tiao (1975), Chow & Lin (1976), Anderson & Moore (1979), Jones (1980), Kohn & Ansley (1983), Harvey & Pierse (1984), Maravall & Peña (1988), Nieto (1989), Peña & Maravall (1990), Chen & Liu (1990), Castañeda (1994). Estos métodos parten del supuesto de que el modelo es conocido o de que hay un subconjunto de observaciones que permiten identificar su estructura. Sin embargo, con frecuencia sucede que la ocurrencia de las observaciones faltantes, de las observaciones atípicas y de las intervenciones en el período de análisis es tal que impide la identificación de un modelo adecuado.

Este artículo presenta un procedimiento, basado en Castaño (1995, 1997), el cual permite estimar el efecto de las intervenciones, de las observaciones atípicas, estimar las observaciones faltantes y simultáneamente identificar el proceso ARIMA que generó el proceso aleatorio. El resultado proporciona una serie reconstruida con estimaciones óptimas para los datos faltantes y para los efectos de observaciones atípicas, la cual puede ser usada efectivamente en predicción. El procedimiento consta de dos etapas básicas: la primera de ellas produce estimaciones preliminares de las intervenciones, de las observaciones atípicas y las observaciones faltantes a partir de una aproximación inicial del proceso ARIMA del ruido por medio de un modelo autorregresivo de alto orden (o de un modelo multiplicativo en caso estacional), del uso del análisis de intervención (Box & Jenkins 1976), del uso del análisis de observaciones atípicas (Chen & Liu 1990) y de la consideración de que una observación faltante puede ser estimada de manera óptima al asignarle un valor tal que pueda ser identificada como una observación atípica aditiva (Chow & Lin 1976); en la segunda etapa se identifica el modelo usando las estimaciones preliminares y se reestiman las intervenciones, las observaciones atípicas y las observaciones faltantes usando de nuevo el análisis de intervención y el análisis de observaciones atípicas sobre el modelo identificado.

El procedimiento se aplica a una serie de demanda de electricidad horaria en la cual se presentan los tres problemas antes mencionados. Los resultados obtenidos muestran que la serie reconstruida puede ser empleada para pronosticar adecuadamente la demanda horaria.

El artículo está compuesto por tres secciones. En la primera se presenta el procedimiento propuesto y sus fundamentos; en la siguiente, la aplicación de la metodología propuesta al caso de la demanda horaria de electricidad; finalmente, en la tercera sección se plantean las conclusiones.

2. Metodología

A continuación se presenta el procedimiento propuesto, mostrando primero que una observación faltante se puede caracterizar como una observación atípica aditiva.

A. Caracterización de una observación faltante como una observación atípica aditiva

Suponga que Z_t es una serie de tiempo que sigue un proceso ARIMA de la forma

$$Z_t = \frac{\theta(B)}{\delta(B)\phi(B)} a_t$$

donde B es el operador de rezagos usual, $\theta(B)$ es el polinomio de medias móviles con todas sus raíces fuera del círculo unidad, $\phi(B)$ es el polinomio autorregresivo con sus raíces fuera del círculo unidad y que no tiene factores comunes con $\theta(B)$, y $\delta(B)$ es el polinomio de diferencias (que induce estacionaridad) con sus raíces sobre el círculo unidad. En primer lugar, consideremos el caso donde hay solamente una observación faltante. Para esto, suponga que la serie se observó durante T períodos y que no se encuentra disponible la observación para el período $t = t^*$, $1 < t^* < T$. Una caracterización natural de un valor faltante es describirlo como una observación atípica aditiva. Esta caracterización ha sido empleada por varios autores entre los que se encuentran Bruce & Martin (1989), Peña & Maravall (1990) y Chen & Liu (1990). La razón es la siguiente: si suponemos que en el período $t = t^*$ ocurre una observación atípica aditiva, podemos representar la serie observada como:

$$NZ_t = \begin{cases} Z_t, & \text{si } t \neq t^*; \\ AF_t^{(t^*)} + Z_t, & \text{si } t = t^*. \end{cases}$$

donde A indica la desviación desde el verdadero valor de Z_{t^*} , y la variable $F_t^{(t^*)}$ toma el valor de uno cuando $t = t^*$ y de cero en otro caso. En este caso, Chen & Liu (1990) mostraron que el valor ajustado de NZ_{t^*} (es decir, después de remover el efecto atípico sobre NZ_{t^*}) es:

$$NZ_{t^*}^* = \frac{\left\{ \sum_{j=1}^{t-1} \left[\sum_{k=j}^{n-t^*+j} \pi_k \pi_{k-j} \right] NZ_{t^*-j} + \sum_{j=1}^{n-t^*} \left[\sum_{k=j}^{n-t^*} \pi_k \pi_{k-j} \right] NZ_{t^*+j} \right\}}{\sum_{k=j}^{n-t^*} \pi_j^2} \quad (1)$$

donde los coeficientes π son obtenidos del polinomio autorregresivo $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots = \phi(B)/\theta(B)$. Si la serie no es estacionaria entonces la parte autorregresiva debe incluir el operador que induce estacionaridad. De acuerdo con (1), el valor interpolado $NZ_{t^*}^*$ está basado en las observaciones de las series anteriores y posteriores de NZ_{t^*} , es decir, en los valores anteriores y posteriores de la serie original Z_t a la observación faltante Z_{t^*} . Por tanto el valor ajustado no tiene nada que ver con la observación atípica NZ_{t^*} .

El resultado anterior sugiere que podemos estimar un valor faltante en una serie de tiempo, tratándola como si fuera una observación atípica aditiva. Por tanto, si conocemos a $\theta(B)$, $\phi(B)$ y $\delta(B)$, el procedimiento para estimar el valor desconocido de Z_{t^*} consiste en asignar un valor atípico cualquiera a la observación faltante y estimar el modelo intervenido:

$$NZ_t = AF_t^{(t^*)} + \frac{\theta(B)}{\delta(B)\phi(B)} a_t$$

La estimación óptima de Z_{t^*} es el valor de NZ_{t^*} menos la estimación de A , (Box & Tiao 1975).

De igual forma se procedería si la serie tuviera m observaciones faltantes no consecutivas en los períodos $t_1^*, t_2^*, \dots, t_m^*$. A cada período donde se desconoce la observación, se asigna un valor atípico y en el modelo

$$NZ_t = \sum_{j=1}^m A_j F_t^{(t_j^*)} + \frac{\theta(B)}{\delta(B)\phi(B)} a_t$$

La estimación óptima de Z_{t_j} se obtiene como el valor de $NZ_{t_j^*}$ menos la estimación de A_j en el modelo anterior, para $j = 1, 2, \dots, m$.

Sin embargo, el problema en la práctica es más complicado pues, en general, no se dispone del conocimiento de $\theta(B)$, $\phi(B)$ y $\delta(B)$, y debemos tratar de identificarlos a partir de la información incompleta inicial. A continuación presentamos el procedimiento que permite estimar el efecto de las intervenciones, de las observaciones atípicas, estimar las observaciones faltantes y simultáneamente identificar el proceso ARIMA que generó el ruido del modelo.

B. El procedimiento de reconstrucción

- i) A cada una de las m observaciones faltantes asigne un valor que sea atípico. Aproxime a Z_t usando las intervenciones y un proceso puro autorregresivo de orden alto. En la elección del orden, se debe tener en cuenta la frecuencia del período de observación, y si el proceso es estacional o no. En otras palabras, Z_t puede aproximarse como el proceso autorregresivo puro:

$$Z_t = \sum_{i=1}^k v_i(B) I_t^{(t_i)} + \sum_{j=1}^m A_j F_t^{(t_j^*)} + \frac{1}{\delta(B)\phi'(B)} a_t$$

donde $v_i(B)$ es la función de transferencia correspondiente a la i -ésima intervención, $i = 1, \dots, k$; $I_t^{(t_i)}$ es una variable indicadora que toma el valor de 1 si $t = t_i$, donde t_i es el período donde ocurre la i -ésima intervención, y de cero en otro caso; $F_t^{(t_j^*)}$ es una variable indicadora que toma el valor de 1 para cuando $t = t_j^*$, el período donde ocurre la j -ésima observación faltante, $j = 1, \dots, m$ y $\phi'(B)$ es el polinomio autorregresivo del orden seleccionado. Para la aproximación de un proceso ARIMA($p, 1, q$), Said & Dickey (1984) obtuvieron el

siguiente resultado: todo proceso ARIMA($p, 1, q$) puede ser adecuadamente aproximado por medio de un proceso ARIMA($n, 1, 0$), donde $n \leq T^{1/3}$.

Estime el modelo anterior y detecte e identifique observaciones atípicas. Verifique si los residuales se comportan como ruido blanco. En caso afirmativo, el orden elegido para el modelo autorregresivo aproxima adecuadamente la estructura dinámica del proceso.

- ii) Especifique un nuevo modelo en el cual se adicionan h observaciones atípicas detectadas. El nuevo modelo por estimar es:

$$Z_t = \sum_{i=1}^k v_i(B)I_t^{(t_i)} + \sum_{j=1}^m A_j F_t^{(t_j^*)} + \sum_{s=1}^h v_s^*(B)O_t^{(t_s)} + \frac{1}{\delta(B)\phi'(B)} a_t$$

donde $v_s^*(B)$ es la función de transferencia para la s -ésima observación atípica detectada y $O_t^{(t_s)}$ es una variable indicadora que toma el valor de 1 si $t = t_s$, el período en el cual ocurrió la s -ésima observación atípica, $s = 1, \dots, h$. Estime el modelo y verifique que no existan nuevas observaciones atípicas y que los residuales son ruido blanco.

- iii) Filtre a Z_t por el modelo estimado sin incluir el modelo del ruido. Esto es lo mismo que estimar el modelo restringido

$$Z_t = \sum_{i=1}^k v_i(B)I_t^{(t_i)} + \sum_{j=1}^m A_j F_t^{(t_j^*)} + \sum_{s=1}^h v_s^*(B)O_t^{(t_s)} + \eta_t$$

donde el valor de los parámetros es restringido a ser igual a los estimados en la etapa anterior. Obtenga los residuales e identifique el modelo ARIMA correspondiente. Sobre esta serie de residuales pueden emplearse las técnicas de identificación de Box-Jenkins para los modelos ARIMA.

- iv) Suponga que el modelo identificado es de la forma

$$\eta_t = \frac{\theta(B)}{\delta(B)\phi(B)} a_t$$

Para obtener una estimación más refinada de las observaciones faltantes ajuste el modelo

$$Z_t = \sum_{i=1}^k v_i(B)I_t^{(t_i)} + \sum_{j=1}^m A_j F_t^{(t_j^*)} + \sum_{s=1}^h v_s^*(B)O_t^{(t_s)} + \frac{\theta(B)}{\delta(B)\phi(B)} a_t$$

El valor de Z_t para el período $t = t_j$ menos la estimación de A_j es la estimación óptima de $Z_{t_j^*}$, $j = 1, 2, \dots, m$, (es decir, $\widehat{Z}_{t_j^*} = Z_t - \widehat{A}_j$, $j = 1, 2, \dots, m$).

En cuanto a la optimalidad de la estimación de valores faltantes usando la caracterización por medio de datos atípicos aditivos, Chen & Liu (1990) señalan que el procedimiento emplea en forma óptima toda la información relevante

para estimar el valor faltante. También señalan que cuando ocurre una secuencia consecutiva de observaciones faltantes, los valores faltantes pueden ser estimados de forma óptima usando los valores observados y los parámetros estimados. Castaño (1995) utilizando series simuladas muestra que el procedimiento de identificación del modelo ARIMA cuando existen diferentes tipos de observaciones atípicas parece funcionar adecuadamente. Para el caso de observaciones faltantes en la serie, Castaño (1997) usando simulaciones, muestra que el procedimiento, además de permitir la identificación del modelo adecuado, genera estimaciones óptimas para las observaciones faltantes aun en muestras relativamente pequeñas ($n = 150$).

3. Un ejemplo

Para planificar el despacho hora a hora de la electricidad requerida por una cierta región, las empresas distribuidoras necesitan pronosticar su demanda futura para cada una de las horas del día. Con este fin se estiman modelos cuya estructura general tiene en cuenta la existencia de eventos calendario que afectan el nivel de la demanda, de otros eventos exógenos que contaminan las observaciones (como apagones o mediciones erróneas), la inexistencia de observaciones en algunos períodos y la existencia de un modelo ARIMA que guía la evolución de la componente estocástica de la serie. Según estos supuestos, el modelo frecuentemente considerado tiene la estructura general:

$$(1 - B)^d(1 - B^s)^D Z_t = (\text{efectos calendario})_t + \sum_{j=1}^m A_j F_t^{(t_j^*)} + \sum_{s=1}^h C_s O_t^{(t_s)} + (\text{modelo ARMA para el ruido})_t$$

donde d y D son los órdenes de diferenciación ordinaria y estacional para estacionarizar el proceso Z_t (el cual debe ser estable en varianza); $(\text{efectos calendario})_t$ indica la posible influencia de un conjunto de 41 eventos calendario sobre el nivel del consumo, y que se modelan como intervenciones aditivas sobre el nivel de la demanda debido a que las series son cronológicamente discontinuas. Un efecto calendario describe el impacto que tiene sobre el nivel de la demanda de electricidad, la ocurrencia de una fecha específica del año, como por ejemplo, el 31 de diciembre.

La expresión $\sum_{j=1}^m A_j F_t^{(t_j^*)}$ indica la existencia de m observaciones faltantes que ocurren en los períodos $t_1^*, t_2^*, \dots, t_m^*$.

La expresión $\sum_{s=1}^h C_s O_t^{(t_s)}$ indica la existencia de h observaciones atípicas de tipo aditivo que ocurren en los períodos t_1, t_2, \dots, t_h , y C_s señala el impacto del evento exógeno sobre la serie.

En general, las funciones de transferencia consideradas en el procedimiento no son exclusivamente aditivas, es decir, funciones que cambian el nivel del proceso solamente para el período en el cual ocurren. Sin embargo, en este caso particular son las más adecuadas debido a que, como se ha mencionado, las series usadas son series cronológicamente discontinuas, en el sentido de que son series conformadas por una sola hora de cada día.

La componente (*modelo ARMA para el ruido*) $_t$ indica la existencia de un modelo ARIMA para el ruido de la serie.

La figura 1 presenta la serie de la demanda de electricidad para la hora 12 de todos los días desde enero 1 de 1996 hasta agosto 31 de 2002, para un total de 2435 observaciones.

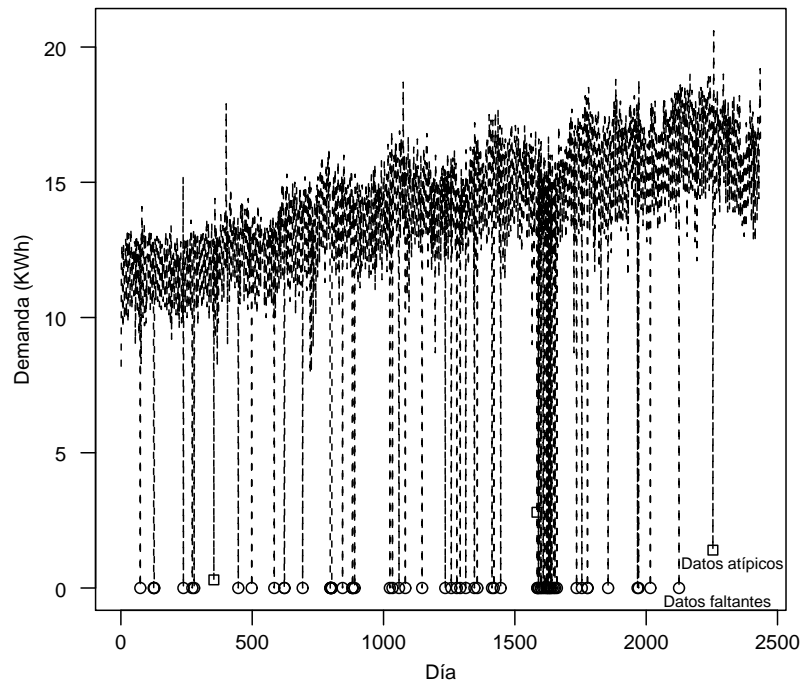


FIGURA 1: Demanda de electricidad hora 12 m.

En esta serie aparecen sistemáticamente datos faltantes (67 en total, que en la figura corresponden a una demanda de 0), y se observan también algunos datos atípicos.

La figura 2, que considera los datos para cuatro semanas, permite observar que la serie posee una componente de estacionalidad de período 7 (estacionalidad semanal).

Observando la figura 1 vemos que este es un caso en el cual la distribución de los datos faltantes es tal que no existe un subconjunto sin observaciones faltantes de longitud suficientemente grande que permita la identificación adecuada del modelo.

El uso del procedimiento de reconstrucción para datos faltantes y atípicos de la serie arrojó los siguientes resultados:

Etapa 1: Estimación preliminar de los efectos de las intervenciones, observaciones atípicas y de las observaciones faltantes.

Sub-etapa 1: Identificación de los órdenes diferenciación.

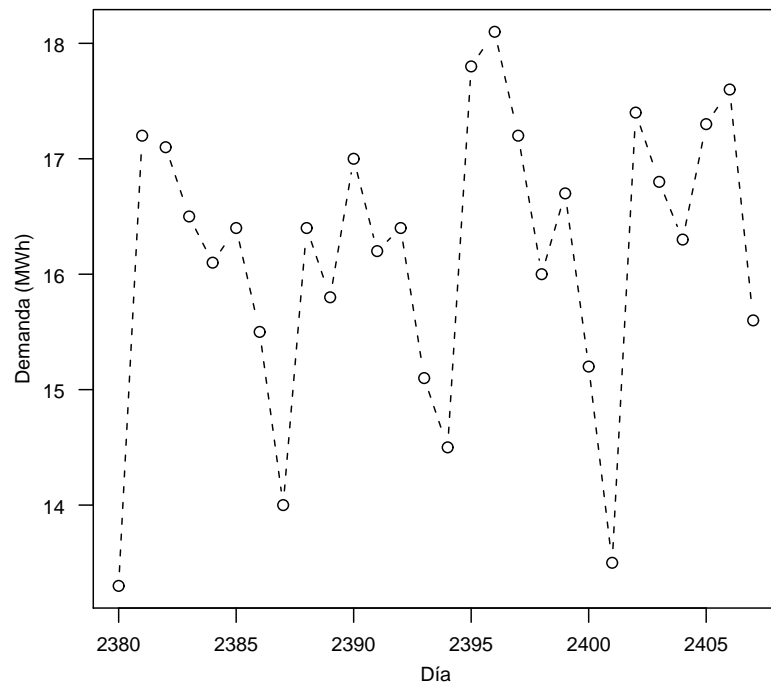


FIGURA 2: Ciclo estacional de la demanda horaria.

La tabla 1 contiene los resultados de la estimación preliminar de los efectos de las intervenciones, observaciones atípicas y de las observaciones faltantes.

Los coeficientes de la parte autorregresiva ordinaria y estacional están cerca de uno, indicando que el proceso parece no ser estacionario tanto en su componente ordinaria como en la estacional. Diferenciar una vez la parte ordinaria y una vez la parte estacional podría estacionarizar la serie.

Sub-etapa 2: Identificación de las observaciones atípicas y faltantes.

Diferenciando la serie mediante los operadores $(1 - B)(1 - B)^7$ y realizando de nuevo la estimación del modelo, se obtiene la tabla 2 con la identificación de las observaciones atípicas y faltantes.

Entre observaciones atípicas y faltantes se detecta un total de 120 observaciones.

Sub-etapa 3: Estimación preliminar de las intervenciones, observaciones atípicas y faltantes.

Las observaciones atípicas detectadas se adicionan al modelo, el cual es estimado de nuevo para obtener una estimación preliminar de las intervenciones, observaciones atípicas y faltantes. Esta etapa es iterativa hasta obtener un resultado en el que no se detecten más observaciones atípicas.

Etapa 2: Identificación del modelo del ruido.

TABLA 1: Resumen de la estimación preliminar.

Parámetro	Variable	Num/Denom	Factor	Orden	Estimac.	Valor t
CONST		CNST	1	0	145.8430	34.33
B1	Ene-01	NUM.	1	0	-44.9850	-4.61
B2	May-01	NUM.	1	0	-10.6130	-1.19
B3	Jul-20	NUM.	1	0	-13.6560	-1.42
B4	Dic-24	NUM.	1	0	-0.3342	-0.34
B5	Dic-25	NUM.	1	0	-38.5210	-3.88
B6	Ene-02	NUM.	1	0	-0.2583	-0.27
B7	Dic-31	NUM.	1	0	-0.8000	-0.81
B8	Ago-07	NUM.	1	0	-0.9646	-1.08
B9	Dic-08	NUM.	1	0	-17.9060	-1.86
B10	DALFENE	NUM.	1	0	-0.2569	-0.27
B11	DOMALF	NUM.	1	0	0.4535	1.32
B12	DOVDIC	NUM.	1	0	0.5876	0.58
B13	DOVENE	NUM.	1	0	0.1704	0.17
B14	DOVMANO	NUM.	1	0	0.2483	0.34
B15	DSS	NUM.	1	0	0.3870	0.43
B16	JSS	NUM.	1	0	-1.6540	-1.81
B17	JUVDIC	NUM.	1	0	0.6001	0.63
B18	JUVENE	NUM.	1	0	0.3811	0.46
B19	JUVMANO	NUM.	1	0	-0.5452	-0.90
B20	LF	NUM.	1	0	-18.2720	-5.31
B21	LFENE	NUM.	1	0	-27.2390	-2.95
B22	LUVDIC	NUM.	1	0	0.8273	0.92
B23	LUVENE	NUM.	1	0	0.1739	0.13
B24	LUVMANO	NUM.	1	0	-0.8888	-1.22
B25	MAVDIC	NUM.	1	0	0.8519	0.89
B26	MAVENE	NUM.	1	0	-0.0674	-0.07
B27	MAVMANO	NUM.	1	0	0.0308	0.05
B28	MIVDIC	NUM.	1	0	-0.0605	-0.06
B29	MIVENE	NUM.	1	0	0.1140	0.13
B30	MIVMANO	NUM.	1	0	-11.2530	-1.86
B31	MSS	NUM.	1	0	0.6388	0.71
B32	SAALF	NUM.	1	0	-0.0961	-0.28
B33	SALFENE	NUM.	1	0	0.1091	0.12
B34	SAVDIC	NUM.	1	0	0.3588	0.38
B35	SAVENE	NUM.	1	0	-0.0616	-0.06
B36	SAVMANO	NUM.	1	0	-0.2871	-0.39
B37	SSS	NUM.	1	0	-0.2170	-0.24
B38	VIVDIC	NUM.	1	0	0.5848	0.57
B39	VIVENE	NUM.	1	0	0.1588	0.20
B40	VIVMANO	NUM.	1	0	-0.6835	-1.09
B41	VSS	NUM.	1	0	-32.1610	-3.52
	DEMHOR	MA	1	1	0.5227	6.98
	DEMHOR	MA	2	7	0.8227	48.26
	DEMHOR	D-AR	1	1	0.8890	10.45
	DEMHOR	D-AR	2	7	0.9692	130.20

Sub-etapa 1: Se estima el modelo anterior sin modelo en el ruido y restringiendo los parámetros a los estimados en la etapa 1. Los residuales del modelo son usados para la identificación del modelo del ruido. La figura 3 presenta las ACF y PACF muestrales.

TABLA 2: Resumen de la detección de observaciones atípicas: Tipo AO.

Obs.	Estimac.	Valor t	Obs.	Estimac.	Valor t	Obs.	Estimac.	Valor t
55	-2.405	-4.55	838	-2.068	-3.91	1584	-13.195	-23.84
69	-2.172	-4.11	844	-14.985	-28.36	1585	-15.900	-27.51
74	-12.556	-23.76	881	-13.604	-24.58	1586	-16.034	-28.96
77	-2.283	-4.32	882	-11.889	-20.62	1590	-16.088	-30.43
90	-2.168	-4.10	883	-14.562	-26.27	1592	-16.091	-30.32
125	-11.664	-21.07	890	-14.396	-27.20	1594	-4.878	-9.21
126	-10.315	-17.89	914	-2.763	-5.23	1599	-16.011	-30.22
127	-12.651	-22.86	957	-2.180	-4.12	1601	-16.154	-30.44
193	-3.358	-6.36	968	-3.943	-7.46	1608	-15.650	-29.49
202	2.583	4.89	1013	-3.518	-6.66	1613	-15.392	-29.05
222	-2.564	-4.85	1024	-15.461	-29.26	1615	-3.287	-6.19
236	-2.347	-4.24	1030	-4.170	-7.89	1620	-15.175	-28.59
237	3.884	6.74	1034	-15.463	-29.26	1622	-15.671	-29.53
238	-9.987	-18.05	1058	-8.259	-15.00	1627	-15.747	-29.67
272	-11.658	-22.02	1059	-15.105	-27.44	1629	-15.982	-30.12
279	-11.857	-22.40	1075	3.015	5.71	1634	-15.707	-29.59
293	-2.097	-3.97	1083	-15.503	-29.34	1636	-15.427	-29.12
328	-1.859	-3.52	1147	-14.386	-27.23	1641	-15.421	-29.06
330	-2.124	-4.02	1152	-2.440	-4.62	1648	-15.155	-28.56
351	-2.270	-4.30	1180	-2.676	-5.06	1653	-15.315	-28.93
354	-12.596	-23.84	1192	-2.070	-3.92	1655	-4.763	-8.99
401	4.448	8.42	1196	-3.557	-6.46	1660	-15.377	-29.05
406	-1.931	-3.65	1197	-3.886	-7.06	1725	-7.545	-14.28
442	-1.896	-3.59	1221	-1.988	-3.76	1735	-15.423	-29.19
447	-12.706	-24.05	1235	-15.253	-28.86	1755	-16.657	-31.52
498	-11.040	-20.89	1243	-2.144	-4.06	1776	-17.348	-31.34
584	-13.386	-25.34	1256	-2.109	-3.99	1777	-16.202	-28.10
622	-12.703	-23.08	1258	-15.595	-29.51	1778	2.349	4.25
623	-10.986	-19.96	1279	-14.743	-27.90	1800	-4.301	-8.14
634	-1.948	-3.69	1292	-14.775	-27.96	1804	-3.637	-6.88
692	-12.976	-24.56	1298	-2.179	-4.12	1855	-13.745	-26.01
721	-3.797	-7.19	1313	-15.413	-29.17	1867	-4.091	-7.74
797	-13.943	-25.12	1344	-2.624	-4.97	1945	-2.006	-3.80
798	-12.309	-21.13	1346	-16.313	-30.87	1967	-14.021	-26.53
799	-14.988	-25.55	1357	-14.816	-28.04	1971	-17.833	-33.74
800	-14.826	-25.25	1413	-15.163	-28.65	2016	-13.671	-25.87
801	-14.826	-25.28	1420	-15.381	-29.06	2125	-18.028	-34.10
802	-14.802	-25.41	1441	-3.030	-5.73	2156	-1.984	-3.75
803	-14.821	-26.70	1446	-16.535	-31.29	2254	-13.098	-24.75
817	-2.145	-4.06	1566	-7.295	-13.80	2257	3.507	6.63

Número total de observaciones: 2435

Número efectivo de observaciones: 2419

Error estándar residual (sin ajuste por observaciones atípicas): 0.257497E+01

Error estándar residual (con ajuste por observaciones atípicas): 0.569947E+00

Los resultados parecen ser consistentes con un modelo ARMA de media móviles estacional multiplicativo de la forma $ARMA(0, 1) \times (0, 1)_{S=7}$. El análisis posterior indica que es más adecuado un modelo $ARMA(1, 1) \times (0, 1)_{S=7}$.

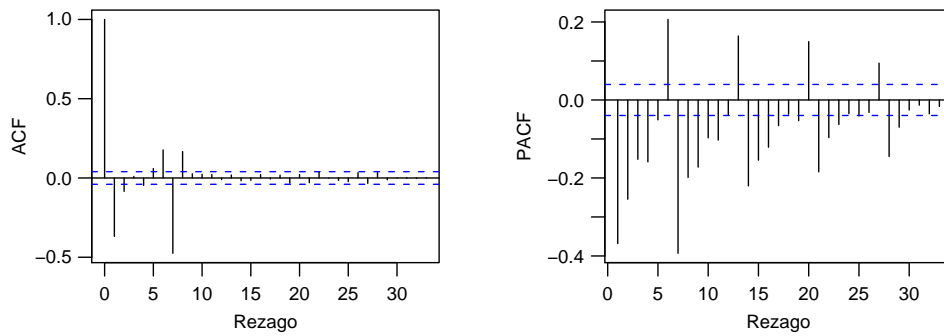


FIGURA 3: Funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) de los residuales del modelo (Sub-etapa 1).

Sub-etapa 2: Estimación con el modelo identificado para el ruido.

Introduciendo el modelo $ARMA(1, 1) \times (0, 1)_{S=7}$ para el ruido, se estima de nuevo el modelo.

Finalmente, la imposición de restricciones nulidad para las estimaciones estadísticamente no significativas (usando como criterio que el valor absoluto de su respectivo estadístico t sea menor que 1) proporciona el modelo final, presentado en la tabla 3:

TABLA 3: Resumen de la estimación del modelo final.

Parámetro	Variable	Factor	Orden	Estimación	Valor t	Parámetro	Variable	Factor	Orden	Estimación	Valor t
B1	Ene-01	1	0	-51.675	-23.05	OAO1024	AO1024	1	0	-154.292	-31.24
B2	May-01	1	0	-17.738	-9.44	OAO1030	AO1030	1	0	-41.851	-8.47
B3	Jul-20	1	0	-19.299	-8.66	OAO1034	AO1034	1	0	-155.111	-31.38
B4	Dic-24	1	0	-0.654	-3.01	OAO1052	AO1052	1	0	-18.364	-3.71
B5	Dic-25	1	0	-45.659	-21.08	OAO1058	AO1058	1	0	-85.257	-16.41
B6	Ene-02	1	0	-0.576	-2.88	OAO1059	AO1059	1	0	-159.048	-29.28
B7	Dic-31	1	0	-10.462	-4.82	OAO1060	AO1060	1	0	-21.011	-4.05
B8	Ago-07	1	0	-15.485	-8.25	OAO1075	AO1075	1	0	30.157	6.11
B9	Dic-08	1	0	-15.342	-6.94	OAO1083	AO1083	1	0	-155.247	-31.43
B10	DALFENE	1	0	-0.609	-2.90	OAO1147	AO1147	1	0	-143.333	-29.02
B16	JSS	1	0	-24.032	-11.56	OAO1152	AO1152	1	0	-24.496	-4.96
B17	JUVDIC	1	0	-0.215	-1.08	OAO1180	AO1180	1	0	-26.812	-5.43
B20	LF	1	0	22.063	-30.65	OAO1192	AO1192	1	0	-20.958	-4.24
B21	LFENE	1	0	-29.493	-14.55	OAO1196	AO1196	1	0	-35.404	-6.86
B22	LUVDIC	1	0	0.232	1.23	OAO1197	AO1197	1	0	-39.021	-7.56
B24	LUVMANO	1	0	-0.256	-1.73	OAO1221	AO1221	1	0	-20.037	-4.06
B25	MAVDIC	1	0	0.310	1.50	OAO1235	AO1235	1	0	-152.642	-30.91
B26	MAVENE	1	0	-0.240	-1.25	OAO1243	AO1243	1	0	-21.489	-4.35
B28	MIVDIC	1	0	-0.282	-1.37	OAO1256	AO1256	1	0	-21.495	-4.35
B31	MSS	1	0	-0.216	-1.09	OAO1258	AO1258	1	0	-15.608	-31.60
B33	SALFENE	1	0	-0.482	-2.33	OAO1279	AO1279	1	0	-147.374	-29.11
B34	SAVDIC	1	0	-0.321	-1.68	OAO1292	AO1292	1	0	-147.007	-29.77
B35	SAVENE	1	0	-0.513	-2.26	OAO1298	AO1298	1	0	-21.872	-4.39
B37	SSS	1	0	-0.713	-3.61	OAO1313	AO1313	1	0	-154.471	-31.28
B39	VIVENE	1	0	-0.324	-2.01	OAO1325	AO1325	1	0	-18.567	-3.76
B40	VIVMANO	1	0	-0.171	-1.39	OAO1344	AO1344	1	0	-26.811	-5.43
B41	VSS	1	0	-39.058	-18.79	OAO1346	AO1346	1	0	-163.394	-33.09

TABLA 3: Resumen de la estimación del modelo final. (continuación)

Parámetro	Variable	Factor	Orden	Estimación	Valor t	Parámetro	Variable	Factor	Orden	Estimación	Valor t
OA055	AO55	1	0	-25.754	-5.19	OA01357	AO1357	1	0	-148.463	-30.07
OA069	AO69	1	0	-23.178	-4.67	OA01413	AO1413	1	0	-151.926	-30.73
OA074	AO74	1	0	-126.841	-25.59	OA01420	AO1420	1	0	-154.105	-31.21
OA077	AO77	1	0	-23.561	-4.75	OA01441	AO1441	1	0	-30.612	-6.20
OA090	AO90	1	0	-2.359	-4.76	OA01446	AO1446	1	0	-165.237	-33.46
OA0125	AO125	1	0	-118.488	-22.75	OA01543	AO1543	1	0	-18.577	-3.76
OA0126	AO126	1	0	-104.022	-19.13	OA01566	AO1566	1	0	-73.158	-14.81
OA0127	AO127	1	0	-12.697	-24.39	OA01584	AO1584	1	0	-131.991	-25.23
OA0193	AO193	1	0	-33.435	-6.75	OA01585	AO1585	1	0	-159.283	-29.27
OA0202	AO202	1	0	24.749	4.56	OA01586	AO1586	1	0	-160.157	-30.83
OA0222	AO222	1	0	-24.828	-5.02	OA01590	AO1590	1	0	-160.924	-32.56
OA0236	AO236	1	0	-23.355	-4.49	OA01592	AO1592	1	0	-161.335	-32.61
OA0237	AO237	1	0	37.674	6.93	OA01594	AO1594	1	0	-48.979	-9.90
OA0238	AO238	1	0	-100.347	-19.31	OA01599	AO1599	1	0	-160.856	-32.52
OA0272	AO272	1	0	-117.681	-23.80	OA01601	AO1601	1	0	-161.657	-32.66
OA0279	AO279	1	0	-119.478	-24.16	OA01608	AO1608	1	0	-156.446	-31.61
OA0293	AO293	1	0	-21.702	-4.39	OA01613	AO1613	1	0	-15.452	-31.23
OA0328	AO328	1	0	-19.358	-3.92	OA01615	AO1615	1	0	-3.287	-6.64
OA0330	AO330	1	0	-21.324	-4.31	OA01620	AO1620	1	0	-151.938	-30.68
OA0351	AO351	1	0	-22.568	-4.56	OA01622	AO1622	1	0	-156.703	-31.66
OA0354	AO354	1	0	-126.684	-25.65	OA01627	AO1627	1	0	-157.893	-31.91
OA0401	AO401	1	0	44.794	9.07	OA01629	AO1629	1	0	-159.981	-32.31
OA0406	AO406	1	0	-19.251	-3.90	OA01634	AO1634	1	0	-157.067	-31.74
OA0442	AO442	1	0	-19.016	-3.85	OA01636	AO1636	1	0	-154.358	-30.44
OA0447	AO447	1	0	-127.768	-25.87	OA01641	AO1641	1	0	-154.092	-31.15
OA0498	AO498	1	0	-110.288	-22.18	OA01648	AO1648	1	0	-151.274	-30.58
OA0584	AO584	1	0	-133.939	-26.97	OA01653	AO1653	1	0	-152.951	-29.75
OA0622	AO622	1	0	-126.928	-24.58	OA01655	AO1655	1	0	-47.192	-9.54
OA0623	AO623	1	0	-109.498	-21.22	OA01660	AO1660	1	0	-152.856	-30.93
OA0634	AO634	1	0	-20.197	-4.09	OA01667	AO1667	1	0	18.452	3.73
OA0692	AO692	1	0	-129.741	-26.28	OA01725	AO1725	1	0	-75.739	-15.33
OA0718	AO718	1	0	-17.884	-3.62	OA01735	AO1735	1	0	-154.113	-31.20
OA0721	AO721	1	0	-38.041	-7.65	OA01755	AO1755	1	0	-166.659	-33.74
OA0797	AO797	1	0	-13.917	-26.68	OA01776	AO1776	1	0	-17.374	-33.45
OA0798	AO798	1	0	-122.845	-22.33	OA01777	AO1777	1	0	-162.293	-29.89
OA0799	AO799	1	0	-149.537	-26.96	OA01778	AO1778	1	0	23.256	4.48
OA0800	AO800	1	0	-148.857	-26.81	OA01800	AO1800	1	0	-42.206	-8.55
OA0801	AO801	1	0	-148.626	-26.81	OA01804	AO1804	1	0	-36.195	-6.69
OA0802	AO802	1	0	-148.341	-26.97	OA01855	AO1855	1	0	-137.596	-27.87
OA0803	AO803	1	0	-148.085	-28.39	OA01867	AO1867	1	0	-40.525	-8.20
OA0817	AO817	1	0	-2.156	-4.36	OA01884	AO1884	1	0	17.848	3.61
OA0828	AO828	1	0	-19.007	-3.82	OA01945	AO1945	1	0	-19.858	-4.02
OA0838	AO838	1	0	-21.136	-4.28	OA01951	AO1951	1	0	-18.146	-3.68
OA0844	AO844	1	0	-150.019	-30.39	OA01967	AO1967	1	0	-140.398	-28.44
OA0881	AO881	1	0	-136.002	-26.19	OA01971	AO1971	1	0	-178.681	-36.19
OA0882	AO882	1	0	-118.914	-21.91	OA02016	AO2016	1	0	-136.688	-27.68
OA0883	AO883	1	0	-145.518	-28.00	OA02125	AO2125	1	0	-179.962	-36.45
OA0890	AO890	1	0	-143.778	-29.09	OA02156	AO2156	1	0	-20.176	-4.08
OA0914	AO914	1	0	-26.604	-5.38	OA02254	AO2254	1	0	-131.168	-26.52
OA0957	AO957	1	0	-21.759	-4.40	OA02257	AO2257	1	0	35.468	7.17
OA0968	AO968	1	0	-40.175	-8.13	OA02434	AO2434	1	0	17.883	3.57
OA01006	AO1006	1	0	-18.174	-3.68		DEMHOR	1	1	0.909	93.06
OA01013	AO1013	1	0	-36.378	-7.36		DEMHOR	2	7	0.974	206.63

Número total de observaciones: 2435

Número efectivo de observaciones: 2426

Error estándar residual (sin ajuste por observaciones atípicas): 0.535876E+00

Los diagnósticos básicos, presentados en las figuras 4 y 5, indican que los errores de modelo son ruido blanco, que no se rechaza el supuesto de normalidad y que las raíces de los polinomios AR y MA tienen raíces fuera del círculo unidad.

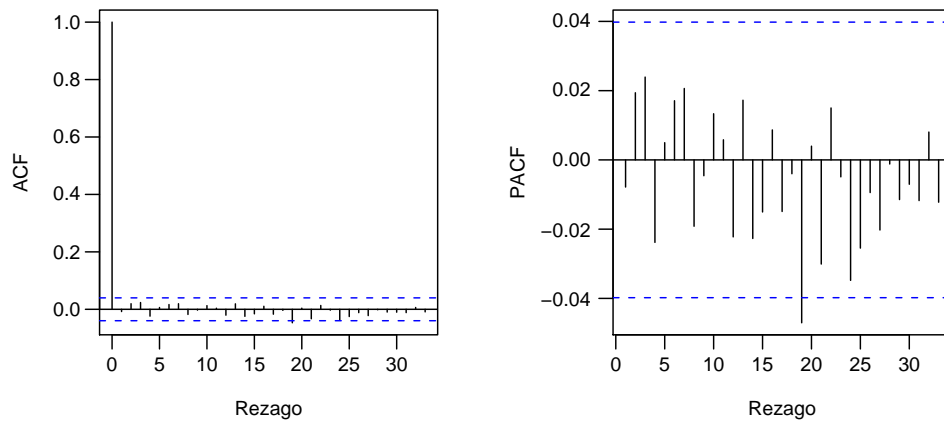


FIGURA 4: Funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) de los residuales del modelo (modelo final).

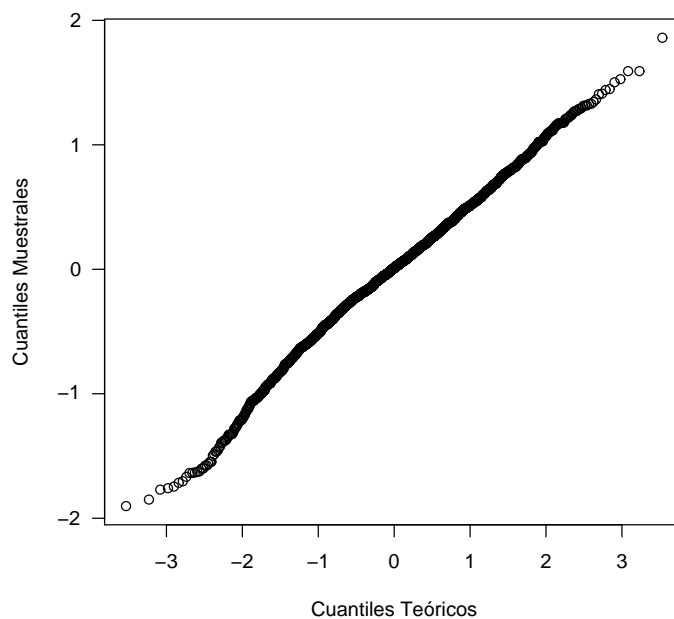


FIGURA 5: Gráfico de Q-Q normal.

La reconstrucción de la serie por observaciones faltantes y atípicas se basa en filtrar la serie original usando las estimaciones obtenidas en el modelo final.

La tabla 4 muestra parcialmente algunos resultados, donde DEMHOR es la demanda original, NDEMHOR es la DEMANDA reconstruida, INDIC es una variable que toma el valor de 1 si la observación es atípica o faltante y EFECT es la cantidad que hay que restar a la demanda para obtener la demanda “normal”.

TABLA 4: Reconstrucción de la demanda.

Variable	Demhor	Ndemhor	Indic	Efect
68	12.7	12.700	0	0.000
69	9.3	11.618	1	-2.318
70	9.9	9.900	0	0.000
71	12.4	12.400	0	0.000
72	12.7	12.700	0	0.000
73	12.5	12.500	0	0.000
74	0.0	12.684	1	-12.684
75	12.7	12.700	0	0.000
76	12.2	12.200	0	0.000
77	8.3	10.656	1	-2.356
78	13.2	13.200	0	0.000
79	12.3	12.300	0	0.000
80	14.1	14.100	0	0.000
81	12.7	12.700	0	0.000
82	13.4	13.400	0	0.000
83	11.5	11.500	0	0.000
84	10.0	10.000	0	0.000
85	10.9	10.900	0	0.000
86	12.3	12.300	0	0.000
87	12.6	12.600	0	0.000
88	13.0	13.000	0	0.000
89	13.4	13.400	0	0.000
90	9.7	12.059	1	-2.359
91	10.5	10.500	0	0.000
92	12.7	12.700	0	0.000
93	12.5	12.500	0	0.000
94	13.1	13.100	0	0.000
95	10.0	10.000	0	0.000

En figura 6 aparece la serie original y la serie reconstruida.

Finalmente, la figura 7 presenta los pronósticos para la demanda de las 168 horas de la siguiente semana, después de elaborar los modelos para cada una de las 24 horas, siguiendo la metodología anterior.

Los resultados muestran una gran capacidad de la serie reconstruida para pronosticar la demanda desconocida.

4. Conclusiones

- i) Los métodos de imputación de valores faltantes suponen conocimiento del modelo o que los datos sean tales que un subconjunto de ellos permitan identificar el modelo. Otros métodos que hacen imputaciones ad hoc basados en

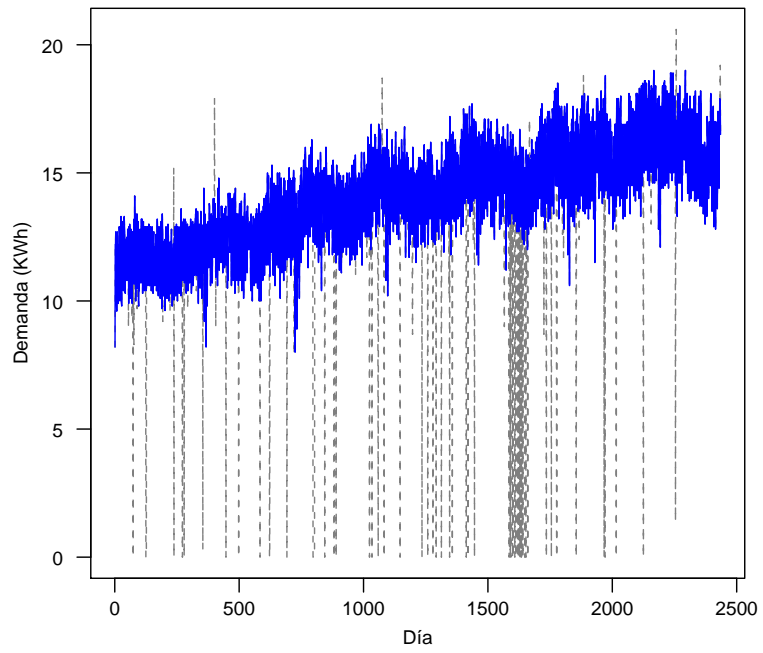


FIGURA 6: Demanda horaria reconstruida.

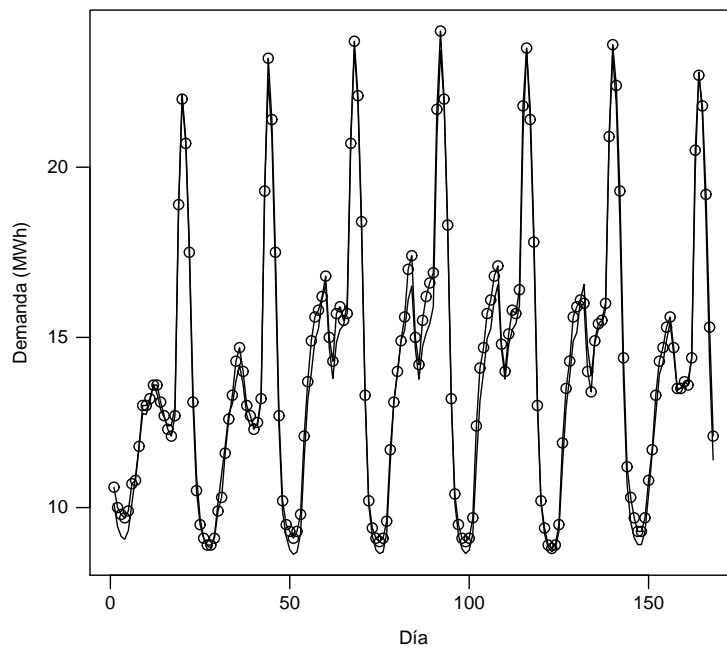


FIGURA 7: Demanda horaria real y pronosticada (la serie de la demanda real corresponde a la línea con puntos).

promedio de observaciones vecinas o estacionalmente vecinas, pueden generar problemas en la identificación si existen observaciones atípicas. El método propuesto permite la identificación del modelo y estimación conjunta de datos faltantes, efectos de intervenciones y observaciones atípicas y de los parámetros del modelo.

- ii) El procedimiento puede ser extendido al caso de un vector de series de tiempo con datos faltantes y observaciones atípicas.
- iii) La aplicación del procedimiento a los datos de una serie real de demanda horaria de electricidad, muestra que el procedimiento parece comportarse bien.

Recibido: mayo de 2007
Aceptado: octubre de 2007

Referencias

- Anderson, B. D. O. & Moore, B. J. (1979), *Optimal Filtering*, Prentice-Hall, Englewood, Cliffs, NJ.
- Box, G. E. P. & Jenkins, G. M. (1976), *Time Series Analysis, Forecasting and Control*, second edn, Holden-Day.
- Box, G. E. P. & Tiao, G. C. (1975), 'Intervention Analysis with Applications to Economic and Environmental Problems', *Journal of the American Statistical Association* **70**, 335–365.
- Bruce, A. & Martin, R. (1989), 'Leave-k-Out Diagnostics for Time Series', *Journal of the Royal Statistical Society* **51**(3), 363–424.
- Castañeda, M. (1994), Reconstrucción de series de tiempo univariadas mediante el enfoque de pronósticos con restricciones, Tesis de Maestría, Estadística, Universidad Nacional de Colombia, Facultad de Ciencias, Departamento de Estadística, Bogotá.
- Castaño, E. (1995), 'Identificación de un modelo ARIMA contaminado', *Lecturas de Economía* **42**, 49–70.
- Castaño, E. (1997), 'Identificación de un modelo ARIMA cuando existen observaciones faltantes', *Lecturas de Economía* **47**, 25–45.
- Chen, C. & Liu, L. (1990), 'Joint Estimation of Models Parameters and Outliers Effects in Time Series', *Working Papers Series, Scientific Computing Associates* pp. P.O.Box 625, DeKalb, Illinois 60115.
- Chow, G. & Lin, A. (1976), 'Best Linear Unbiased Estimation of Missing Observation in a Economic Time Series', *Journal of the American Statistical Association* **71**, 719–721.

- Harvey, A. C. & Pierse, R. G. (1984), 'Estimating Missing Observations in Economic Time Series', *Journal of the American Statistical Association* **79**(385), 15–131.
- Jones, R. H. (1980), 'Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations', *Technometrics* **22**, 389–395.
- Kohn, R. & Ansley, F. (1983), 'Fixed Interval Estimation in State Space Models when Some of the Data are Missing or Aggregated', *Biometrika* **70**, 683–688.
- Maravall, A. & Peña, D. (1988), Missing Observations of Time Series and the Dual Autocorrelation Function, *Servicio de Estudios*, Banco de España, Madrid. Documento de trabajo No. 8830.
- Nieto, F. (1989), Reconstrucción de una serie de tiempo censurada usando filtros de Kalman, Tesis de Maestría, Estadística, Universidad Nacional de Colombia, Facultad de Ciencias, Departamento de Estadística, Bogotá.
- Peña, D. & Maravall, A. (1990), 'Interpolation, Outliers and the Inverse Autocorrelations', *Communications in Statistics* **A20**(10), 3175–3186.
- Said, S. & Dickey, D. (1984), 'Testing Unit Roots in Autoregressive-Moving Average Models with Unknown Order', *Biometrika* **71**, 599–601.