

# **El "efecto calendario" y la especificación de un modelo de series de tiempo para la demanda de electricidad**

## **Elkin Castaño Vélez**

*Lecturas de Economía*. No. 35

-Introducción, 153. -I. Procedimiento para la construcción del modelo, 154. -II. Una aplicación, 161. -III. Conclusiones, 169. -Bibliografía, 170.

## **Introducción**

**C**on frecuencia las series de tiempo son medidas en períodos de tiempo cuya longitud es cambiante (por ejemplo las series económicas e hidrológicas medidas mensualmente) y algunas están afectadas también por otros eventos tales como Semana Santa y días de fiesta.

Bell y Hillmer (1983) proponen una metodología basada en el análisis de funciones de transferencia (Box-Jenkins, 1976) para especificar mejor el proceso generador de la serie y producir pronósticos más precisos.

Suponga que  $Z_t$  es una serie de tiempo tal que

$$Z_t = v(X_t, \beta) + N_t \quad (1)$$

donde  $v$  es una función del vector de parámetros,  $\beta$  y del vector de variables independientes y fijas,  $X_t$  observadas en el período  $t$ , y  $N_t$  es una serie de ruido. Si  $N_t$  es ruido blanco, el modelo (1) corresponde al modelo de regresión clásico lineal (o no lineal). Sin embargo, en el contexto de series de tiempo,  $N_t$  es generalmente autocorrelacionado y no estacionario, y (1) en ese caso es conocido como una función de transferencia y ruido (Box-Jenkins, 1976). El término  $v(X_t, \beta)$ , denominado función de transferencia del modelo, se empleará para especificar los efectos de variación de los días comerciales y de Semana Santa.

Asumiremos que  $N_t$  sigue un proceso ARIMA de la forma

$$\phi(B) \delta(B) N_t = \theta(B) a_t \quad (2)$$

donde  $B$  es el operador de rezagos  $BN_t = N_{t-1}$ .

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

y donde  $\phi(B)$  y  $\theta(B)$  tienen todos sus ceros fuera del círculo unidad y no tienen ceros comunes.

$\delta(B)$  es un operador diferencias (con todos sus ceros sobre el círculo unidad) que estacionariza el proceso  $N_t$ .

$(a_t)$  es un proceso de ruido blanco de varianza  $\sigma^2$ .

Algunos  $\phi$  y/o  $\theta$  pueden ser cero o estar restringidos de modo que el modelo puede ser estacional multiplicativo.

## I. Procedimiento para la construcción del modelo

Para la elaboración de modelos de la forma (1), se empleará el proceso de construcción en tres etapas sugerido por Box y Jenkins

(1976): identificación, estimación y validación. A continuación haremos una breve descripción de cada paso.

## A. Identificación

El proceso de identificación debe efectuarse tanto para la función de transferencia  $v(X_t, \beta)$  como para el ruido  $N_t$ .

### 1. Identificación de la función de transferencia

La función de transferencia  $v(X_t, \beta)$  puede ser especificada considerando la naturaleza del efecto calendario que trata de describir, ya sea de variación en los días comerciales o de Semana Santa.

#### a. Modelación de la variación de los días comerciales

Suponga que  $E_i$ ,  $i = 1, 2, \dots, 7$ , representa la tasa promedio de actividad en los días lunes, martes, miércoles, jueves, viernes, sábado y domingo para la serie que va a ser modelada. Entonces el efecto que se puede atribuir al número de veces que cada día de la semana ocurre en el mes  $t$  es

$$TD_t = \sum_{i=1}^7 E_i X_{it} \quad (3)$$

donde las  $X_{it}$  son, respectivamente, el número de lunes, martes, etc., en el mes  $t$ . El modelo anterior permite incorporar variaciones en el nivel debido a las diferentes longitudes de los meses y variaciones en el nivel, debido a la diferente composición de los días de la semana de los meses con la misma longitud.

Un modelo que incorpora el efecto de la variación en los días comerciales es

$$Z_t = TD_t + N_t \quad (4)$$

Debido a que las variables anteriores están altamente correlacionadas (y por tanto también los estimadores de los coeficientes  $E_i$ ), es útil la siguiente reparametrización sugerida en Bell Y Hillmer:

Sea  $T_{it} = X_{it} - X_{7t}$ ,  $i = 1, 2, \dots, 6$  y sea  $T_{7t}$  la suma de las  $X_{it}$  para  $i = 1, 2, \dots, 7$ , la cual denota la longitud del mes  $t$ .

Entonces el modelo anterior puede ser escrito como:

$$TD_t = \sum_{i=1}^7 (E_i - M(E_i)) (X_{it} - X_{7t}) + \sum_{i=1}^7 (E_i - M(E_i)) X_{7t} + M(E_i) \sum_{i=1}^7 X_{it}$$

$$TD_t = \sum_{i=1}^7 \beta_i T_{it}$$

donde  $M(E_i) = \frac{1}{7} \sum_{i=1}^7 E_i$

$$\beta_i = E_i - M(E_i), i = 1, 2, \dots, 6$$

y  $\beta_7 = M(E_i)$ .

El modelo queda de la forma

$$Z_t = \sum_{i=1}^7 \beta_i T_{it} + \frac{\theta(B)}{\phi(B)\delta(B)} a_t$$

Los parámetros  $\beta_i = E_i - M(E_i)$ ,  $i=1, 2, \dots, 6$ , miden las diferencias entre los efectos del lunes, martes, etc., y el promedio de los efectos diarios  $\beta_7 = M(E_i)$ .

La diferencia entre el efecto del domingo y el promedio de los efectos diarios está dado por

$$E_7 - M(E_1) = - \sum_{i=1}^6 \beta_i$$

y el efecto del domingo es

$$E_7 = \beta_7 - \sum_{i=1}^6 \beta_i$$

El paquete estadístico SCA (*Scientific Computing Associates*, Liu y Hudak, 1986) permite construir las variables  $X_{it}$  y  $T_{it}$  para considerar el efecto de la variación de los días comerciales en modelación de series de tiempo por medio del parágrafo *DAYS*.

#### b. Modelación del efecto Semana Santa

Con frecuencia la Semana Santa tiene algún efecto sobre las actividades comerciales en el período inmediato que la antecede. Este efecto puede influenciar cualquier modelo de series de tiempo si la serie se relaciona con dichas actividades, y entonces puede ser necesaria la inclusión de la información sobre la Semana Santa. En el caso de la demanda mensual de electricidad podemos esperar que las demandas mensuales de algunos sectores tales como el residencial, industrial y el comercial principalmente, se vean afectadas en los meses en los que ocurre la Semana Santa.

Ahora bien, las fechas extremas entre las que puede caer la Semana Santa (la más temprana y la última) son marzo 22 y abril 25. De ahí que la especificación de la forma funcional para el efecto de la Semana Santa no es tan simple como en el caso de la variación en los días comerciales.

Suponga que  $\alpha_i$  es el efecto sobre la serie modelado sobre el  $i$ -ésimo día antes de Semana Santa y sea  $h(i,t)$  igual a uno cuando el  $i$ -ésimo

día antes de Semana Santa cae en el mes correspondiente al período  $t$ , e igual a cero en otro caso. Entonces el efecto Semana Santa en  $t$ ,  $S_t$  es

$$S_t = \sum_{i=1}^k \alpha_i h(i, t) \quad (5)$$

Si consideramos que la serie de tiempo posee también efecto por variación en los días comerciales, el modelo que debemos considerar es de la forma

$$Z_t = TD_t + S_t + N_t \quad (6)$$

donde  $TD_t$  y  $N_t$  fueron definidas anteriormente.

En la relación dada por (6) se consideró el impacto diario de la Semana Santa sobre el nivel de la serie. Desafortunadamente, estos efectos tan generales no pueden ser estimados en la mayoría de los casos prácticos, pues sólo se dispone de información mensual. Sin embargo podemos considerar ciertos patrones especiales para los  $\alpha$ . Por ejemplo, suponga que  $\alpha_1 = \dots = \alpha_m = \alpha$  y  $\alpha_{m+1} = \dots = \alpha_k = 0$  para algún valor de la  $m$ . Esto implica que

$$S_t = \alpha^* \cdot H(m, t) \quad (7)$$

donde  $\alpha^* = \alpha_m$  y

$$H(m, t) = \frac{1}{7} \sum_{i=1}^m h(i, t)$$

Dado  $m$ ,  $H(m, t)$  puede ser definida como la proporción del período de  $m$  días antes de Semana Santa que caen en el mes  $t$ . Con esta definición  $H(m, t)$  puede ser obtenida para todo  $m > 0$ . Bell y Hillmer

discuten un *test* para chequear la validez de (7). Otros patrones diferentes de (7) son posibles de definir. Sin embargo, es poco probable que puedan ser detectados con los datos. La posibilidad de detectarlos aumenta a medida que se agregan más datos que cubran diferentes fechas en la Semana Santa. El paquete estadístico SCA permite construir la variable  $H(m,t)$  usando el parágrafo *EASTER*.

## 2. Identificación del modelo del ruido

Para la identificación de la forma de  $N_t$ , Bell y Hillmer recomiendan:

a) emplear la función de autocorrelación muestral (FACM) de  $Z_t$  para identificar el grado de diferenciación adecuada,  $\delta(b)$ , en  $N_t$ .

b) para la identificación de  $\phi(B)$  y  $\theta(B)$  no se deberían emplear la FACM, ni la función de autocorrelación parcial muestral (FACPM), ni la función de autocorrelación extendida muestral (FACEM) de  $Z_t$ , pues los efectos de calendario tienden a oscurecer su información.

Se deben remover primero los efectos aproximados de  $v(X_t, \beta)$  ajustando el modelo

$$\delta(B)Z_t = v(X_t, \beta) + \varepsilon_t \quad (8)$$

por mínimos cuadrados (lineales o no lineales) y examinar la FACM, la FACPM y la FACEM de los residuales obtenidos de esa regresión para identificar tentativamente el modelo del ruido.

## B. Estimación del modelo

Usando (1) y (2) el modelo puede ser escrito como

$$\delta(B)Z_t = \delta(B)v(X_t, \beta) + \theta(B)/\phi(B)a_t \quad (9)$$

Asumiendo normalidad en  $a_t$  podemos estimar los parámetros en  $\beta$ , los  $\phi$  y los  $\theta$ . El estimador de  $\sigma_a^2$  es  $(n-r)^{-1}SCR$ , donde SCR es la suma de cuadrados de los residuales del modelo ajustado,  $n$  es número de observaciones menos el grado de  $\delta(B)\phi(B)$  y  $r$  es el número de parámetros en (9).

Pierce (1971) discute la inferencia de (9) cuando  $v(X_t, \beta)$  es lineal en  $\beta$ , que es nuestro caso.

### C. Diagnósticos

En general, deben validarse las formulaciones hechas para  $v(X_t, \beta)$  y para el modelo del ruido (2).

Para la forma de  $v(X_t, \beta)$  se pueden graficar los residuales contra cada una de las  $X_{jt}$ , y contra cualquier otra variable independiente posible.

Los residuales pueden ser graficados contra el tiempo para verificar varianza constante y la existencia de tendencias. Se debe efectuar un análisis de observaciones atípicas (*outliers*) sobre los residuales, para tratar de explicar efectos de factores exógenos sobre las series (Chen y Liu, 1990).

La FACM de los residuales debería ser examinada para determinar si aún existe algún grado de autocorrelación.

Ljung y Box (1978) muestran que bajo la hipótesis de que el modelo sea correcto y  $n$  sea grande el estadístico

$$Q = n(n+2) \sum_{k=1}^L \frac{\gamma_k^2}{(n-k)}$$

tiene una distribución chi-cuadrado con  $L$ -s grados de libertad, aproximadamente, donde  $r_k$  es el coeficiente de autocorrelación muestral de



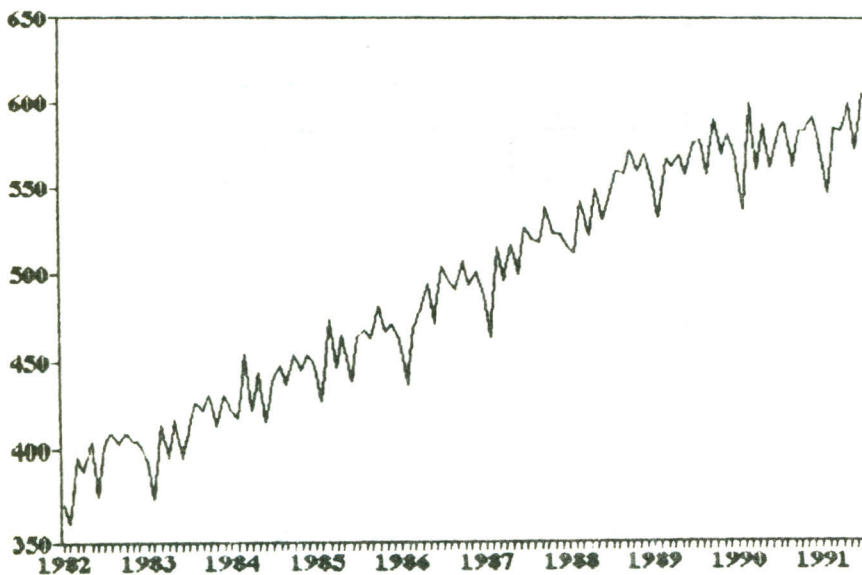
los residuos de orden  $k$  y  $s$  es el número de parámetros del modelo del ruido (2). El modelo del ruido no es adecuado si  $Q$  es mayor que el percentil  $\alpha$ -superior de la distribución chi-cuadrado con  $L-s$  grados de libertad.

## II. Una aplicación

Para ilustrar la metodología anterior consideremos la gráfica No. 1 que representa la serie de demandas mensuales de energía eléctrica atendida por las Empresas Públicas de Medellín -EEPPM- desde enero de 1982 hasta julio de 1991. Contamos con un total de 115 observaciones.



Gráfica No. 1 Demandas mensuales de electricidad EPM

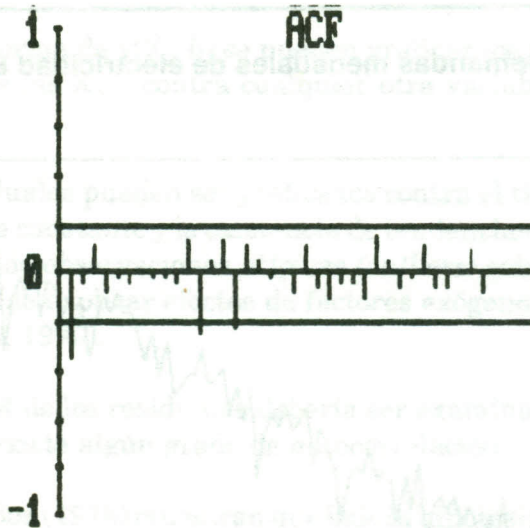


**A. Identificación**

La gráfica No. 2 muestra la función de autocorrelación de los residuales de la estimación de la función de transferencia donde se ha elegido  $m=4$

$$Z_t = \sum \beta_i T_{it} + \alpha H(4, t) + \frac{N_t}{(1-B)(1-B^{12})}$$

Gráfica No. 2



Por tanto, para el ruido, un modelo MA de orden 1, tanto en la parte ordinaria como en la estacional, podría ser el indicado.

El modelo tentativo completo para la serie es de la forma:

$$Z_t = \omega_1 I_t(80) + \sum_{i=1}^7 \beta_i T_{it} + \alpha H(m,t) + \frac{(1-\theta_1 B)(1-\theta_{12} B^{12})}{(1-B)(1-B^{12})} a_t$$

donde  $I_t(80)$  es una variable dummy que indica la ocurrencia de un (*outlier de*) cambio de nivel a partir de la observación 80.

### B. Estimación

Podemos obtener los estimadores máximo verosímiles para los parámetros del modelo identificado tentativamente, incluyendo  $m$ , definiendo el log-verosimilitud asintótica.

$$L_{\max}(m) = \max_{\beta, \alpha, \phi, \theta, \sigma^2} L(\beta, \alpha, \phi, \theta, \sigma^2)$$

$$L_{\max}(m) = \frac{n}{2} \ln \sigma^{*2}(m) + \text{constante}$$

donde  $\sigma^{*2}(m)$  es el estimador de  $\sigma^2$  para  $m$  fijo, y maximizándola con respecto a  $m$ . La siguiente tabla da los valores de  $\sigma^{*2}$  para  $m=1, \dots, 25$

---

Estimación de m							
m	1	2	3	4	5	6	7
$\sigma^{*2}(m)$	30.55	30.55	29.15	29.86	30.59	31.15	32.0
$\alpha^*(m)$	-16.13	-16.13	-17.07	-16.38	-15.68	-15.11	-14.9
desv.	1.75	1.75	1.79	1.88	1.91	1.93	2.02
m	8	9	10	11	12	13	14
$\sigma^{*2}(m)$	32.72	33.38	33.93	33.77	33.70	33.69	33.7
$\alpha^*(m)$	-14.52	-14.12	-13.73	-13.97	-14.10	-14.16	14.2
desv.	2.09	2.14	2.18	2.21	2.24	2.26	2.28

---

Continuación. Estimación de m

m	15	16	17	18	19	20	21
$\sigma^2(m)$	34.05	34.39	34.74	35.09	35.15	35.23	35.33
$\alpha^*(m)$	-14.32	-14.38	-14.38	-14.31	-14.75	-15.12	-15.4
desv.	2.36	2.43	2.49	2.54	2.64	2.73	2.81
m	22	23	24	25			
$\sigma^2(m)$	35.33	35.33	35.33	35.33			
$\alpha^*(m)$	-16.17	-16.90	-17.64	-18.4			
desv.	2.95	3.08	3.22	3.35			

Donde desv. es el estimador del error estándar de  $\alpha^*(m)$ . De aquí el estimador máximo verosímil de m es aproximadamente 3.

Para este valor de m, los estimadores para los demás parámetros se encuentran a continuación.

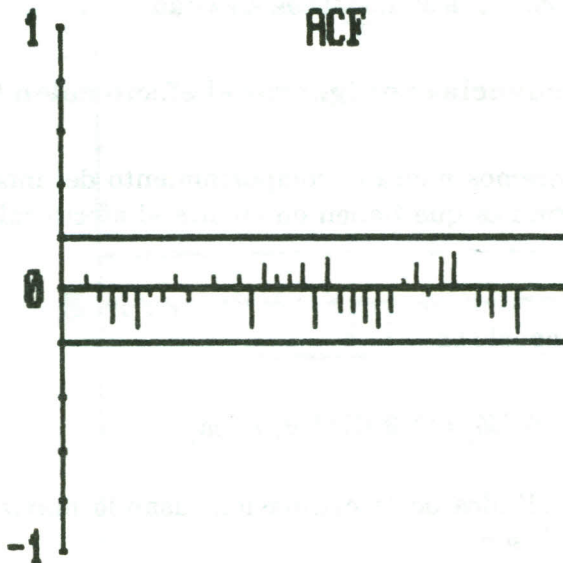
Parámetro label	Variable nombre	Factor	orden	Valor	STD error	T valor
1 $W_{1a}$	$I_1(80)$	1	0	13.5902	4.0711	3.34
2 $B_1$	$T_1$	1	0	1.2602	.8661	1.46
3 $B_2$	$T_2$	1	0	-.3511	.9164	-.38
4 $B_3$	$T_3$	1	0	-2.7332	.8622	-3.17
5 $B_4$	$T_4$	1	0	4.6051	.8419	5.47
6 $B_5$	$T_5$	1	0	-.2568	.8439	-.30
7 $B_6$	$T_6$	1	0	-1.8275	.8612	-2.12
8 $B_7$	$T_7$	1	0	17.9603	2.8698	6.26
9 $\alpha$	$H(3.1)$	1	0	-17.0696	1.7856	-9.56
10 $\theta_1$	EPM	1	1	.4648	.0901	5.16
11 $\theta_{12}$	EPM	2	12	.3878	.0909	4.27

Continuación

Total sum of squares	.527893E+06
Total number of observations	115
Residual sum of squares	.297308E+04
R-Square	.994
Effective number of observations	102
Residual variance estimate	.291478E+02
Residual standard error	.539887E+01
AIC	745.749

La gráfica No. 3 muestra la función de autocorrelación muestral para los residuales de este modelo y a continuación una tabla con las autocorrelaciones, su desviación estándar y el estadístico Q de Ljung y Box (1978).

Gráfica No. 3



Autocorrelaciones

1-12	-0.00	.05	-.02	-.11	-.03	-.15	-.01	-.02	.05	-.02	.01	.06
ST.E.	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10
Q	.0	.2	.3	1.6	1.7	4.0	4.0	4.1	4.3	4.4	4.4	4.8
13-24	.01	.05	-.13	.09	.06	.06	.08	-.14	.11	-.08	-.07	-.11
ST.E.	.10	.10	.10	.11	.11	.11	.11	.11	.11	.11	.11	.11
Q	4.8	5.1	7.3	8.2	8.6	9.1	9.9	12.5	13.9	14.7	15.4	17.1
25-36	-.15	-.07	.03	.09	-.01	.11	.14	.01	-.04	-.09	-.05	-.17
ST.E.	.11	.11	.11	.11	.11	.11	.12	.12	.12	.12	.12	.12
Q	20.0	20.7	20.9	22.0	22.0	23.6	26.7	26.7	27.0	28.4	28.8	33.4

De estos resultados podemos concluir que los residuales parecen ruido blanco y que el modelo es adecuado.

**C. Consecuencias por ignorar el efecto calendario**

Consideremos ahora el comportamiento del modelo identificado sin los términos que tienen en cuenta el efecto calendario sobre la serie.

Dicho modelo es

$$(1-B)(1-B^{12})Z_t = (1-\theta_1 B)(1-\theta_{12} B^{12})a_t$$

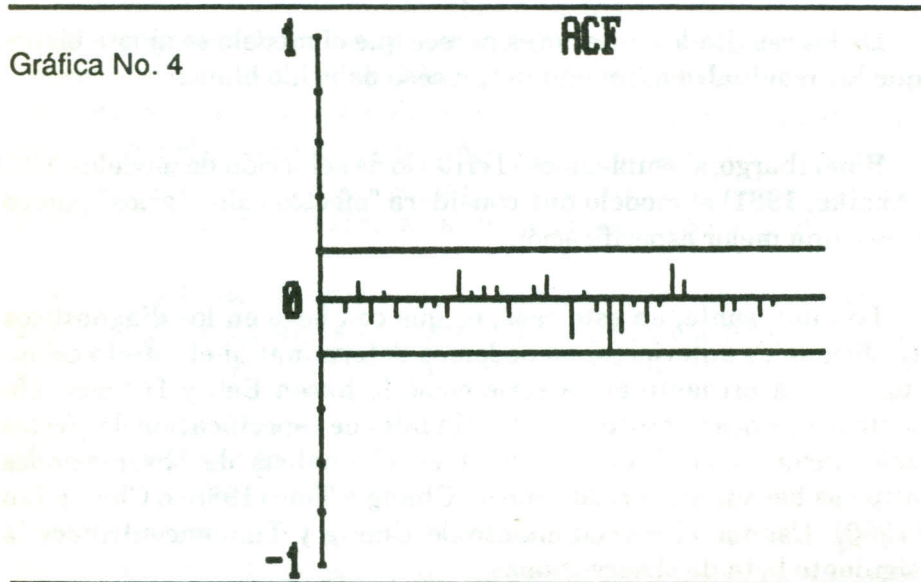
Los resultados de la estimación, usando máxima verosimilitud condicional, son:

---

Parámetro <i>label</i>	Variable nombre	Factor	Orden	Valor	STD error	T valor
1	$\theta_1$	EPM	1	.6626	.0738	8.98
2	$\theta_{12}$	EPM	2	.7980	.0749	10.65
<i>Total sum of squares</i>				.527893E+06		
<i>Total number of observations</i>				115		
<i>Residual sum of squares</i>				.679698E+04		
<i>R-square</i>				.985		
<i>Effective number of observations</i>				102		
<i>Residual variance estimate</i>				.666371E+02		
<i>Residual standard error</i>				.816315E+01		
<i>AIC</i>				813.781		

---

La gráfica No. 4 muestra la función de autocorrelación de los residuales de este modelo



La siguiente tabla muestra las autocorrelaciones, sus errores estándar y el valor del estadístico Q de Ljung-Box

---

### Autocorrelaciones

1-12	.01	-.02	.07	-.09	.03	-.05	.02	-.01	-.01	-.05	.11	.02
ST.E	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10
Q	.0	.0	.6	1.4	1.5	1.8	1.8	1.8	2.1	3.5	3.5	3.5
13-24	.04	.05	-.05	.01	.06	.08	.01	-.13	.04	-.04	-.19	-.10
ST.E	.10	.10	.10	.10	.10	.10	.10	.10	.11	.11	.11	.11
Q	3.7	4.0	4.4	4.4	4.8	5.6	5.6	7.8	8.0	8.2	12.9	14.3
25-36	.01	-.09	-.01	.14	.07	.00	.02	-.04	-.02	.01	-.06	-.01
ST.E	.11	.11	.11	.11	.11	.11	.11	.11	.11	.11	.11	.11
Q	4.3	15.4	15.4	18.2	18.9	18.9	19.0	19.2	19.3	19.3	19.8	19.9

---

De los resultados anteriores parece que el modelo se ajusta bien y que los residuales asemejan un proceso de ruido blanco.

Sin embargo, si empleamos el criterio de selección de modelos AIC, (Akaike, 1981) el modelo que considera "efectos calendarios" parece tener una mejor especificación.

Lo interesante, en este caso, es que con base en los diagnósticos tradicionales anteriores, no podemos determinar si el "efecto calendario" está presente en la serie como lo hacen Bell y Hillmer. Un método que nos permite descubrir la falta de especificación de efectos calendario en el modelo se basa en el análisis de observaciones atípicas basado en el resultado de Chang y Tiao (1988) o Chen y Liu (1990). Usando el procedimiento de Chang y Tiao encontramos la siguiente lista de observaciones:



---

Observaciones	Estimación	Valor-T	Tipo	Mes
52	17.63	3.32	LS	Abril
99	21.18	3.32	AO	Marzo
26	16.36	2.81	AO	Febrero
27	15.95	2.83	AO	Marzo
74	13.36	2.98	LS	Febrero
93	-17.63	-2.99	IO	Septiembre
105	-17.25	-3.05	IO	Septiembre
39	12.72	2.63	AO	Marzo
63	10.23	2.67	LS	Marzo
55	11.61	2.60	AO	Julio
76	-11.25	-2.56	AO	Abril

*Adjusted residual standard error = 4.9069*

---

Esta tabla nos revela que muchas de estas observaciones atípicas están señalando falta de especificación en el modelo por "efectos calendario" tales como los debidos a meses de diferentes longitud (febrero) y a la ocurrencia de Semana Santa (marzo, abril). Estos efectos introducen una variabilidad adicional que el modelo no alcanza explicar. La varianza estimada del error en el modelo simple es de 66.6371 contra 29.1478, la varianza del modelo completo.

Es importante observar que la existencia de las observaciones atípicas, que señalan la existencia de "efectos calendario", tienen posiblemente el efecto de enmascarar la función de autocorrelación.

### III. Conclusiones

Se pueden introducir mejoras sustanciales en la capacidad predictiva de un modelo de series de tiempo si se introducen variables

exógenas que expliquen los posibles efectos que producen las variaciones en los días comerciales y la Semana Santa.

Estos efectos pueden ser introducidos a través de funciones de transferencia.

Si estos efectos están presentes sobre la serie de tiempo, en algunas ocasiones, ellos pueden enmascarar la función de autocorrelación impidiendo su identificación desde la especificación inicial del modelo como lo presentan Bell y Hillmer. En estos casos es importante usar el procedimiento de detección de observaciones atípicas el cual puede sugerir la presencia o no de efectos calendario sobre la serie.

### Bibliografía

Akaike, H. (1981), "Likelihood of a Model and Information Criteria", *Journal of Econometrics*, 16, 3-14.

Bell, W.R. and Hillmer, S.C. (1983), "Modeling Time Series with Calendar Variation", *Journal of the American Statistical Association*, vol. 78, No. 383, 526-534.

Box, G.E.P. and Jenkins, G.M. (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-day.

Chang, I. and Tiao, G.C. (1988), "Estimation of time Series Parameters in the Presence of Outliers", *Technometrics*, 30, 193-204.

Chen, C. and Liu, L.M. (1990), "Joint Estimation of the Model Parameters and Outliers Effects in Time Series", Working Papers Series, Scientific Computing Associates, Illinois.

Liu, L.M. and Hudak, G.B. (1986), *The SCA Statistical System Reference Manual*, Scientific Computing Associates, Dekalb, Illinois.

Ljung, G.M. and Box, G.E.P. (1978), "On a Measure of Lack of Fit in Time Series Models", *Biometrika*, 65, 297-303.

Pierce, D.A. (1971), "Least squares Estimation in the Regression Model with Autorregresive-moving Average Errors", *Biometrika*, 58, 299-312.