

ESTADISTICA ESPAÑOLA
Vol. 36, Núm. 135, 1994, págs. 5 a 20

**Un Estudio de Simulación Comparativo
de Técnicas No Paramétricas,
Semiparamétricas y Box-Jenkins para la
Predicción con Datos Dependientes***

por

RICARDO CAO ABAD

Departamento de Matemáticas

Facultad de Informática

Universidad de La Coruña

MANUEL FEBRERO BANDE, IGNACIO GARCIA JURADO,
WENCESLAO GONZALEZ MANTEIGA y JOSE MANUEL PRADA SANCHEZ

Departamento de Estadística e I.O.

Facultad de Matemáticas

Universidad de Santiago de Compostela

Un Estudio de Simulación Comparativo de Técnicas No Paramétricas, Semiparamétricas y Box-Jenkins para la Predicción con Datos Dependientes*

por
RICARDO CAO ABAD
Departamento de Matemáticas
Facultad de Informática
Universidad de La Coruña

MANUEL FEBRERO BANDE, IGNACIO GARCIA JURADO,
WENCESLAO GONZALEZ MANTEIGA y JOSE MANUEL PRADA SANCHEZ
Departamento de Estadística e I.O.
Facultad de Matemáticas
Universidad de Santiago de Compostela

RESUMEN

Dada una muestra $\{X_{t-n+1}, \dots, X_{t-1}, X_t\}$ correspondiente a una serie de tiempo estacionaria $\{X_t\}_{t \in \mathbb{Z}}$, se trata, en este trabajo, el problema de la predicción a k retardos, \hat{X}_{t+k} , de dicha serie. Para ello, se presenta un estudio de simulación comparativo de técnicas Box-Jenkins, técnicas no paramétricas basadas en la estimación no paramétrica tipo núcleo de la función de autorregresión y técnicas semiparamétricas diseñadas como la suma de la predicción no paramétrica anterior y de la predicción Box-Jenkins construida con la serie residual.

* Este trabajo ha sido financiado parcialmente con los proyectos PB91-0794 (DGICYT) y XUGA20701B92 (Xunta de Galicia). Los autores agradecen las sugerencias de dos *referees* anónimos, gracias a los cuales la versión final de este artículo ha mejorado notablemente.

Palabras Clave: Series de tiempo, predicción, estimación no paramétrica, estimación semiparamétrica, metodología Box-Jenkins, simulación.

Clasificación AMS: 62M20, 62G99.

1. INTRODUCCIÓN

Uno de los problemas de mayor importancia en Estadística, especialmente en el contexto de datos dependientes, es el de la predicción. Un planteamiento general para este problema puede ser el que se describe a continuación. Si (Z_t, Y_t) es una serie estacionaria $(p+1)$ -dimensional, $t = 0, \pm 1, \dots$, se trata de estimar $\phi(z_t^0) = \phi[F(\cdot | Z_t = z_t^0)]$, donde $F(\cdot | Z_t = z_t^0)$ es la distribución de Y_t condicionada a $Z_t = z_t^0$, a partir de una muestra $\{(Z_i, Y_i)\}_{i=1}^n$ de dicha serie (frecuentemente ϕ será el funcional media o mediana). Un ejemplo de especial importancia es aquél en el que $Y_t = X_{t+k}$, $k \geq 1$, y $Z_t = (X_t, \dots, X_{t-r+1})$, donde X_t es una serie estacionaria, y se quiere estimar la función $\phi(x_1^0, \dots, x_r^0) = E[X_{t+k} | (X_t, \dots, X_{t-r+1})] = (x_1^0, \dots, x_r^0)$, denominada función de autorregresión, a partir de una muestra $\{X_i\}_{i=1}^m$, donde m sería tal que $n = m - (r + k) + 1$ en el contexto general que acabamos de describir.

En la literatura relativa a la estimación de la función de autorregresión para la predicción, destacamos tres diferentes aproximaciones al problema. La primera, que se basa en la metodología Box-Jenkins, consiste en estimar la función de autorregresión por la mejor combinación lineal de la serie mediante el criterio de minimización del error cuadrático medio. Así, para un modelo $AR(p): X_t = \varphi_0 + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + a_t$ o $\Phi(B)X_t = a_t$ (en notación Box-Jenkins), donde $\{a_t\}$ es ruido blanco, es decir, variables aleatorias independientes e idénticamente distribuidas de media cero y varianza σ^2 , r toma ahora el valor p , la función de autorregresión viene dada por $\phi(x_1^0, \dots, x_p^0) = \varphi_0^k + \varphi_1^k x_1^0 + \dots + \varphi_p^k x_p^0$ y el predictor óptimo mínimo cuadrático es $\hat{\phi}(X_t, \dots, X_{t-p+1}) = \hat{X}_{t+k} = \varphi_0^k + \varphi_1^k X_t + \dots + \varphi_p^k X_{t-p+1}$ (los coeficientes φ_i^k se estimarían con la muestra observada $\{X_i\}_{i=1}^m$). Para un modelo más general $ARMA(q, s)$ dado por $\Phi(B)X_t = \Theta(B)a_t$, donde $\Phi(B) = (1 - \varphi_1 B - \dots - \varphi_q B^q)$ y $\Theta(B) = (1 - \theta_1 B - \dots - \theta_s B^s)$, el predictor lineal óptimo no viene caracterizado por una función de autorregresión sino que, en general, será de la forma $\hat{\phi}(X_t, \dots, X_{t-m+1}) = \hat{X}_{t+k} = \varphi_{0m}^k + \varphi_{1m}^k X_t + \dots + \varphi_{mm}^k X_{t-m+1}$, que se puede obtener utilizando el algoritmo de Durbin-Levinson (Proposición 5.2.1., pág. 169, Brockwell-Davis, 1991).

La segunda aproximación es la denominada «no paramétrica», que consiste en estimar directamente la función sin hacer ninguna suposición finito dimensional sobre la misma. En general, a partir de una muestra $\{(Z_i, Y_i)\}_{i=1}^n$, la estimación viene dada por:

$$\hat{\phi}(z_t^0) = \sum_{i=1}^n \omega_{ni}(z_t^0, (Z_1, Y_1), \dots, (Z_n, Y_n)) Y_i \quad [1]$$

donde $\{\omega_{ni}\}$ puede ser una sucesión de pesos tipo núcleo, k_n -puntos más próximos, etc.

En este contexto, la estimación de la función de autorregresión está totalmente ligada a la predicción, sobre todo cuando el predictor viene caracterizado por una función de autorregresión. El estudio desde esta perspectiva es más reciente, destacando, entre otros trabajos:

a) Yakowitz (1985), donde se predice $Y_t = X_{t+1}$ a partir de $Z_t = X_t$ con la muestra $\{X_1, \dots, X_n\}$ en un modelo markoviano estacionario con densidad de $X_{t+1}/X_t = x$, $f(y/x)$, y densidad estacionaria de X_t , $\pi(x)$. Se estima $\varphi(x) = E(X_{t+1}/X_t = x)$ por el método núcleo, es decir, para cada $i=1, \dots, n-1$,

$$\omega_{ni}(x, (X_1, X_2), \dots, (X_{n-1}, X_n)) = \frac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{h_n}\right)} \quad [2]$$

en la sucesión de pesos de [1], con K la función núcleo y h_n el parámetro ventana. Yakowitz presenta en este trabajo el predictor $\hat{\varphi}(X_t)$ como una buena alternativa a los modelos ARMA en la aplicación a la predicción de inundaciones;

b) Yakowitz (1987), donde se predice $Y_t = X_{t+1}$ a partir de

$$Z_t = (X_t, X_{t-1}, \dots, X_{t-p_1+1}, U_t, \dots, U_{t-p_2+1})$$

con $p_1+p_2 = p$. X_t es una serie estacionaria, en general markoviana, U_t una serie exógena estacionaria y la sucesión de pesos es del tipo k_n -puntos más próximos:

$$\omega_{ni}(z^0, (Z_1, Y_1), \dots, (Z_n, Y_n)) = \frac{(1_{\{||z^0 - (Z_p, Y_p)|| \leq R(n)\}})}{k_n} \quad [3]$$

donde $R(n)$ es la distancia entre z y el k_n dato más próximo. Yakowitz prueba un funcionamiento competitivo en comparación con los modelos ARMAX, $\varphi(B)X_t = \varphi^*(B)U_t + \theta(B)a_t$, en contextos de hidrología;

c) Härdle-Vieu (1992), donde se predice Y_t a partir de Z_t en un modelo $Y_t = \varphi(Z_t) + \varepsilon_t$ con ε_t de media cero y varianza σ^2 independiente de Z_t . Considerando una sucesión de pesos en [1] de tipo núcleo como en [2], Härdle-Vieu estudian elecciones óptimas para el parámetro ventana. El caso $Y_t = \frac{Z_{t+1}-Z_t}{Z_t}$ (relativo a la denominada elasticidad), es tratado como ejemplo de aplicación cuando Z_t es el precio del oro en el instante t .

Obsérvese que estos tres artículos no presentan enfoques diferentes, sino más bien complementarios. De hecho, en los dos primeros, se parte de hipótesis distintas y no se trata la elección del parámetro ventana (h_n o k_n), aspecto éste analizado en el tercer artículo.

Finalmente, la tercera aproximación es la de tipo semiparamétrico, en la que se supone que

$$Y_t = \varphi(Z_t) + \varepsilon_t \quad [4]$$

donde ε_t sigue un modelo ARMA (q,s) independiente de Z_t . El predictor \hat{Y}_t viene dado por $\hat{\varphi}(Z_t) + \hat{\varepsilon}_t$, donde $\hat{\varphi}$ es la predicción no paramétrica construida a partir de la muestra con pesos del tipo [2] o [3] y $\hat{\varepsilon}_t$ es la predicción Box-Jenkins construida a partir de la serie residual $\hat{\varepsilon}_t = Y_t - \hat{\varphi}(Z_t)$. Una aplicación de esta técnica se utilizó para el análisis de datos de inmisión obtenidos en las proximidades de la central térmica de As Pontes con $Y_t = X_{t+6}$ y $Z_t = (X_t, X_{t-1})$, siendo X_t el nivel de inmisión en el instante t (ver Cao y otros, 1992; González-Manteiga y otros, 1993).

En este trabajo presentamos un amplio estudio de simulación comparativo de las estimaciones de la predicción puramente no paramétricas, Box-Jenkins y las de tipo semiparamétrico. Por último, incluimos un apartado de conclusiones relativas a los resultados obtenidos.

2. ESTUDIO DE SIMULACION

A continuación describimos el estudio de simulación llevado a cabo. Se obtienen 100 series con 200 observaciones cada una (X_1^i, \dots, X_{200}^i con $1 \leq i \leq 100$) y, para cada una de ellas, se obtienen 1.000 posibles realizaciones para k períodos por delante ($X_{200+k}^i (1), \dots, X_{200+k}^i (1.000)$), que se comparan con la predicción que se realizó a partir de la muestra X_1^i, \dots, X_{200}^i .

Las series simuladas son las siguientes:

- Serie 1. Modelo AR (1): $X_t = 5 + 0.75X_{t-1} + a_t$, $a_t \sim N(0,1)$.
- Serie 2. Modelo ergódico: $(X_t/X_{t-1} = x) \sim N(\text{sign}(x)|x|^{0.8}, \frac{0.25}{1 + \exp(x)})$ (Yakowitz, 1985).
- Serie 3. Modelo pseudoestacionario: $X_t = S_t + P_t$ con $S_t = 3\sin \omega t$, $\omega = 2\pi/30$ y P_t una serie ARMA (1,1) con parámetros 0.75 y 0.4.
- Serie 4. AR no lineal: $X_t = R(X_{t-1}) + \varepsilon_t$ con $R(x) = \frac{x}{1+x^2}$, $x \in (-1,1)$ y $\varepsilon_t \sim U_{[-\frac{1}{2}, \frac{1}{2}]}$.
- Serie 5. AR no lineal: $X_t = R(X_{t-1}) + \varepsilon_t$ con $R(x) = \frac{19\pi}{20} \sin x$, $x \in (0,\pi)$ y $\varepsilon_t \sim U_{[-\frac{\pi}{20}, \frac{\pi}{20}]}$.
- Serie 6. ARMA (1,1): $X_t = 0.7X_{t-1} + a_t - 0.4a_{t-1}$ con $a_t \sim N(0,1)$.
- Serie 7. MA (1): $X_t = 5 + a_t - 0.75a_{t-1}$ con $a_t \sim N(0,1)$.

Para cada una de estas series se comparan los tres predictores que se detallan a continuación.

- a) El predictor Box-Jenkins óptimo estimado correspondiente al modelo seleccionado entre posibles $ARIMA(q, l, s)$ (con $q \leq 6, l \leq 1, s \leq 6$) utilizando la rutina FTCP de la librería IMSL (versión 1982), que procede de modo iterativo utilizando como regla de parada el test de Ljung-Box (para más detalles, ver manuales de la IMSL). Esta predicción se realiza en cada uno de los primeros seis retardos.
- b) El predictor puramente no paramétrico correspondiente a la estimación por el método núcleo de $E(X_{t+k}/X_t = x)$, $k = 1, \dots, 6$. El núcleo considerado es el gaussiano (peso del tipo [2]) y la ventana h_n es elegida por un método de validación cruzada adaptado a cada punto x , $h_n = h_n(x)$ (Vieu, 1991). La función ponderación en este método de validación cruzada es elegida siguiendo las sugerencias de dicho artículo.
- c) El predictor semiparamétrico obtenido según se indica en Cao y otros (1992). Como estimador no paramétrico se considera el elegido en b), con el mismo parámetro ventana, y como estimador de la parte Box-Jenkins el óptimo seleccionado como en a), pero para la serie residual.

De este modo, como ya indicamos anteriormente, llamando X_1^i, \dots, X_{200}^i , $i = 1, \dots, N = 100$, a cada una de las series simuladas $(1, 2, \dots, 7)$ y, considerando \hat{X}_{200+k}^{ia} , \hat{X}_{200+k}^{ib} , y \hat{X}_{200+k}^{ic} , $k = 1, \dots, 6$ como cada uno de los predictores a k retardos según los métodos a), b) y c) respectivamente, se comparan los métodos mediante los criterios cuadrático y en valor absoluto

$$\frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{j=1}^M (X_{200+k}^{ij} - \hat{X}_{200+k}^{ij})^2 \quad [5]$$

$$\frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{j=1}^M |X_{200+k}^{ij} - \hat{X}_{200+k}^{ij}| \quad [6]$$

donde X_{200+k}^{ij} representa el valor observado en la j -ésima prolongación de la serie $\{X_m^i\}_{m=1}^{200}$, $j = 1, \dots, M = 1000$, $i = a, b$ o c y $k = 1, \dots, 6$. Estos criterios proporcionan estimaciones del error cuadrático $E(\hat{X}_{t+k} - X_{t+k})^2$ y del error absoluto $E(|\hat{X}_{t+k} - X_{t+k}|)$ respectivamente.

Los resultados obtenidos se presentan en las Tablas 1 a 7. En cada una de ellas se incluyen los resultados obtenidos para cada ventana utilizada (nótese que se utilizaron diferentes ventanas obtenidas multiplicando por diversos factores la óptima de validación cruzada local). Por «Cuad(B-J,SP,NP)» entendemos el error cuadrático [5] para el método Box-Jenkins, semiparamétrico y no paramétrico respectivamente. De forma análoga, utilizamos «Abs(B-J,SP,NP)» para referirnos al error absoluto [6]. A su vez, también se incluyen en las tablas las frecuencias de veces que modelizaron las metodologías Box-Jenkins y semiparamétrica las cien series simuladas.

Tabla 1
MODELO AR(1)

Ventana X factor	Cuad. B-J	Cuad. SP			Cuad. NP		
		—	0.1	1	10	0.1	1
Retardo 1	1.0067	1.1112	1.0385	1.0204	1.1114	1.1397	1.3731
Retardo 2	1.2467	1.3649	1.2936	1.2770	1.3628	1.3668	1.4853
Retardo 3	1.3836	1.4414	1.4110	1.4023	1.4397	1.4449	1.5080
Retardo 4	1.4498	1.5346	1.4716	1.4701	1.5152	1.4822	1.5316
Retardo 5	1.4860	1.5763	1.5009	1.5001	1.5799	1.5089	1.5333
Retardo 6	1.5085	1.5523	1.5198	1.5198	1.5523	1.5243	1.5330
Ventana X factor	Abs. B-J	Abs. SP			Abs. NP		
		—	0.1	1	10	0.1	1
Retardo 1	0.8034	0.8981	0.8331	0.8147	0.8989	0.9249	1.1469
Retardo 2	1.0056	1.1005	1.0362	1.0204	1.0988	1.0986	1.2083
Retardo 3	1.1047	1.1534	1.1279	1.1200	1.1519	1.1568	1.2127
Retardo 4	1.1564	1.2322	1.1754	1.1734	1.2147	1.1843	1.2270
Retardo 5	1.1856	1.2661	1.1985	1.1978	1.2698	1.2051	1.2258
Retardo 6	1.2032	1.2403	1.2125	1.2129	1.2403	1.2161	1.2237

Nº de series sin fallo: 100 100 100

Tabla 2
MODELO ERGODICO

Ventana X factor	Cuad. B-J	Cuad. SP			Cuad. NP		
		—	0.1	1	10	0.1	1
Retardo 1	0.3210	0.3834	0.3238	0.3266	0.4036	0.3548	0.6637
Retardo 2	0.4301	0.4954	0.4324	0.4415	0.5093	0.4506	0.7517
Retardo 3	0.5142	0.5924	0.5212	0.5181	0.5996	0.5412	0.7824
Retardo 4	0.5757	0.6633	0.6001	0.5816	0.6696	0.6131	0.8073
Retardo 5	0.6271	0.6802	0.6322	0.6227	0.6828	0.6433	0.8570
Retardo 6	0.6683	0.7122	0.6812	0.6624	0.7178	0.6875	0.8615
Ventana X factor	Abs. B-J	Abs. SP			Abs. NP		
		—	0.1	1	10	0.1	1
Retardo 1	0.2567	0.3178	0.2589	0.2617	0.3386	0.2902	0.6009
Retardo 2	0.3487	0.4068	0.3454	0.3546	0.4213	0.3628	0.6644
Retardo 3	0.4119	0.4874	0.4175	0.4166	0.4947	0.4367	0.6795
Retardo 4	0.4624	0.5437	0.4836	0.4691	0.5505	0.4951	0.6926
Retardo 5	0.5028	0.5492	0.5048	0.4978	0.5520	0.5132	0.7328
Retardo 6	0.5355	0.5702	0.5438	0.5296	0.5760	0.5474	0.7283

Nº de series sin fallo: 99 100 100

Tabla 3
MODELO PSEUDO-ESTACIONARIO

Ventana X factor	Cuad. B-J	Cuad. SP			Cuad. NP		
	—	0.5	1	10	0.5	1	10
Retardo 1	1.2220	1.1695	1.1761	1.2501	1.2015	1.2239	2.6709
Retardo 2	1.4391	1.3860	1.3580	1.4606	1.4806	1.5328	2.8774
Retardo 3	1.5894	1.4597	1.4534	1.5850	1.6831	1.8259	2.9907
Retardo 4	1.6541	1.7297	1.7785	1.7627	1.7881	1.9082	2.8900
Retardo 5	1.7268	2.1709	1.8148	1.6531	1.9341	2.1106	2.7117
Retardo 6	1.6913	1.9246	1.4585	1.6490	1.9267	2.0259	2.4505

Ventana X factor	Abs. B-J	Abs. SP			Abs. NP		
	—	0.5	1	10	0.5	1	10
Retardo 1	0.9967	0.9464	0.9515	1.0273	0.9752	0.9946	2.4791
Retardo 2	1.1981	1.1473	1.1206	1.2258	1.2297	1.2814	2.6852
Retardo 3	1.3497	1.2212	1.2094	1.3473	1.4282	1.5740	2.7964
Retardo 4	1.4092	1.4859	1.5332	1.5203	1.5297	1.6518	2.6864
Retardo 5	1.4819	1.9400	1.5720	1.4086	1.6780	1.8647	2.4943
Retardo 6	1.4457	1.6947	1.2156	1.4057	1.6647	1.7687	2.1270

Nº de series sin fallo: 34 100 100

Tabla 4
MODELO AR NO LINEAL

Ventana X factor	Cuad. B-J	Cuad. SP			Cuad. NP		
	—	0.1	1	10	0.1	1	10
Retardo 1	0.2927	0.3114	0.2967	0.2959	0.3107	0.2987	0.3846
Retardo 2	0.3587	0.3826	0.3660	0.3664	0.3792	0.3731	0.4188
Retardo 3	0.3926	0.4123	0.3959	0.3982	0.4095	0.3997	0.4207
Retardo 4	0.4100	0.4294	0.4173	0.4159	0.4269	0.4190	0.4268
Retardo 5	0.4207	0.4390	0.4261	0.4245	0.4365	0.4272	0.4291
Retardo 6	0.4246	0.4375	0.4313	0.4277	0.4370	0.4314	0.4291
Ventana X factor	Abs. B-J	Abs. SP			Abs. NP		
	—	0.1	1	10	0.1	1	10
Retardo 1	0.2522	0.2650	0.2548	0.2542	0.2646	0.2560	0.3222
Retardo 2	0.2981	0.3143	0.3011	0.3013	0.3116	0.3065	0.3445
Retardo 3	0.3208	0.3365	0.3234	0.3253	0.3343	0.3266	0.3441
Retardo 4	0.3348	0.3504	0.3407	0.3396	0.3482	0.3422	0.3489
Retardo 5	0.3437	0.3588	0.3481	0.3469	0.3569	0.3492	0.3512
Retardo 6	0.3466	0.3571	0.3523	0.3493	0.3567	0.3525	0.3507

Nº de series sin fallo: 100 100 100

Tabla 5
MODELO AR NO LINEAL II

Ventana X factor	Cuad. B-J	Cuad. SP			Cuad. NP		
		—	0.1	1	10	0.1	1
Retardo 1	0.8032	0.2985	0.2869	0.6042	0.3575	0.2596	0.6440
Retardo 2	0.8356	0.4313	0.4448	0.7674	0.4142	0.4582	0.8508
Retardo 3	0.9236	0.6362	0.7308	0.9393	0.6398	0.7476	1.1849
Retardo 4	0.9622	0.8906	0.9040	0.9850	0.8916	0.9063	1.2626
Retardo 5	0.9900	0.9595	0.9887	1.0074	0.9602	0.9854	1.3004
Retardo 6	1.0211	0.9934	1.0042	1.0369	0.9941	1.0042	1.3475
Ventana X factor	Abs. B-J	Abs. SP			Abs. NP		
		—	0.1	1	10	0.1	1
Retardo 1	0.7943	0.2818	0.2731	0.5930	0.3010	0.2453	0.6324
Retardo 2	0.8043	0.3903	0.4064	0.7299	0.3725	0.4194	0.8143
Retardo 3	0.8495	0.5564	0.6538	0.8655	0.5600	0.6705	1.1154
Retardo 4	0.8480	0.7582	0.7788	0.8704	0.7591	0.7822	1.1510
Retardo 5	0.8614	0.8158	0.8589	0.8765	0.8165	0.8571	1.1634
Retardo 6	0.8799	0.8425	0.8584	0.8941	0.8429	0.8628	1.1972

Nº de series sin fallo: 100 100 100

Tabla 6
MODELO ARMA (1,1)

Ventana X factor	Cuad. B-J	Cuad. SP			Cuad. NP		
		—	0.1	1	10	0.1	1
Retardo 1	1.0311	1.1113	1.0330	1.0295	1.1115	1.0555	1.1081
Retardo 2	1.0715	1.0975	1.0743	1.0813	1.0986	1.0784	1.1031
Retardo 3	1.0982	1.1217	1.0881	1.0970	1.1220	1.0900	1.1007
Retardo 4	1.0932	1.1238	1.0921	1.0936	1.1242	1.0935	1.0946
Retardo 5	1.0962	1.1143	1.1013	1.0972	1.1146	1.1025	1.0976
Retardo 6	1.0919	1.1351	1.0988	1.0922	1.1349	1.0993	1.0928
Ventana X factor	Abs. B-J	Abs. SP			Abs. NP		
		—	0.1	1	10	0.1	1
Retardo 1	0.8237	0.8964	0.8254	0.8222	0.8965	0.8450	0.8916
Retardo 2	0.8642	0.8799	0.8582	0.8638	0.8807	0.8616	0.8824
Retardo 3	0.8765	0.8974	0.8679	0.8753	0.8977	0.8694	0.8784
Retardo 4	0.8732	0.9012	0.8720	0.8734	0.9015	0.8731	0.8742
Retardo 5	0.8740	0.8895	0.8781	0.8747	0.8898	0.8791	0.8751
Retardo 6	0.8720	0.9132	0.8781	0.8723	0.9130	0.8784	0.8728

Nº de series sin fallo: 100 100 100

Tabla 7
MODELO MA (1)

Ventana X factor	Cuad. B-J	Cuad. SP			Cuad. NP		
		—	0.1	1	10	0.1	1
Retardo 1	1.0132	1.1092	1.0552	1.0067	1.1458	1.1311	1.2090
Retardo 2	1.2407	1.2868	1.2670	1.2639	1.2750	1.2552	1.2518
Retardo 3	1.2481	1.2765	1.2557	1.2495	1.2762	1.2535	1.2475
Retardo 4	1.2506	1.2690	1.2574	1.2514	1.2691	1.2569	1.2507
Retardo 5	1.2592	1.2890	1.2641	1.2594	1.2890	1.2642	1.2592
Retardo 6	1.2547	1.2858	1.2601	1.2548	1.2858	1.2599	1.2547

Ventana X factor	Abs. B-J	Abs. SP			Abs. NP		
		—	0.1	1	10	0.1	1
Retardo 1	0.8099	0.8923	0.8465	0.8037	0.9258	0.9121	0.9863
Retardo 2	0.9989	1.0295	1.0121	1.0092	1.0184	1.0011	0.9980
Retardo 3	0.9945	1.0196	1.0008	0.9956	1.0193	0.9990	0.9939
Retardo 4	0.9984	1.0144	1.0042	0.9991	1.0143	1.0038	0.9985
Retardo 5	1.0041	1.0291	1.0079	1.0042	1.0291	1.0079	1.0041
Retardo 6	1.0011	1.0281	1.0054	1.0013	1.0280	1.0052	1.0011

Nº de series sin fallo: 100 100 100

Las Figuras 1 a 7, que también se incluyen, representan la variación del error cuadrático en función de los factores multiplicadores de la ventana para las tres metodologías.

Figura 1
ERROR CUADRÁTICO INCONDICIONAL COMPARADO
DE LOS TRES METODOS. MODELO AR (1)

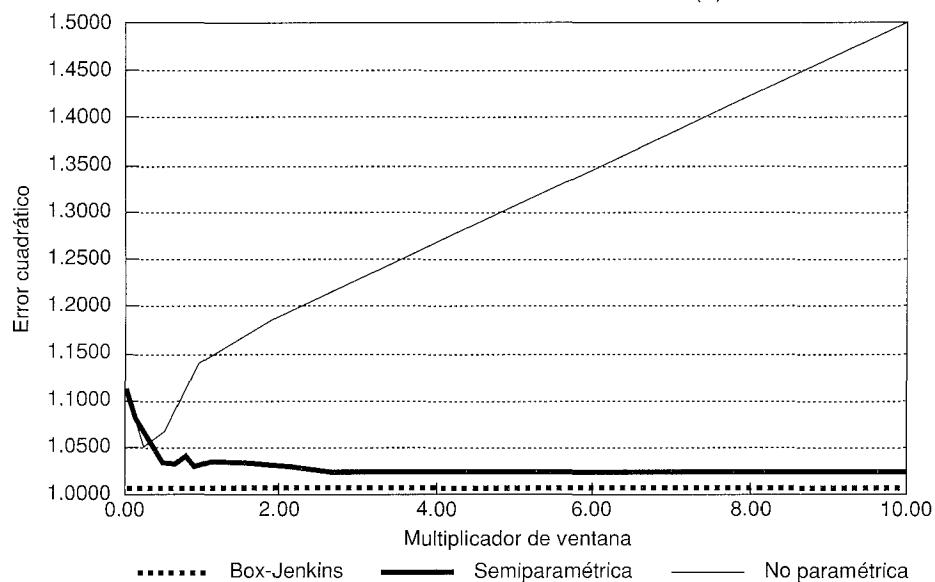


Figura 2
ERROR CUADRATICO INCONDICIONAL COMPARADO
DE LOS TRES METODOS. MODELO ERGODICO

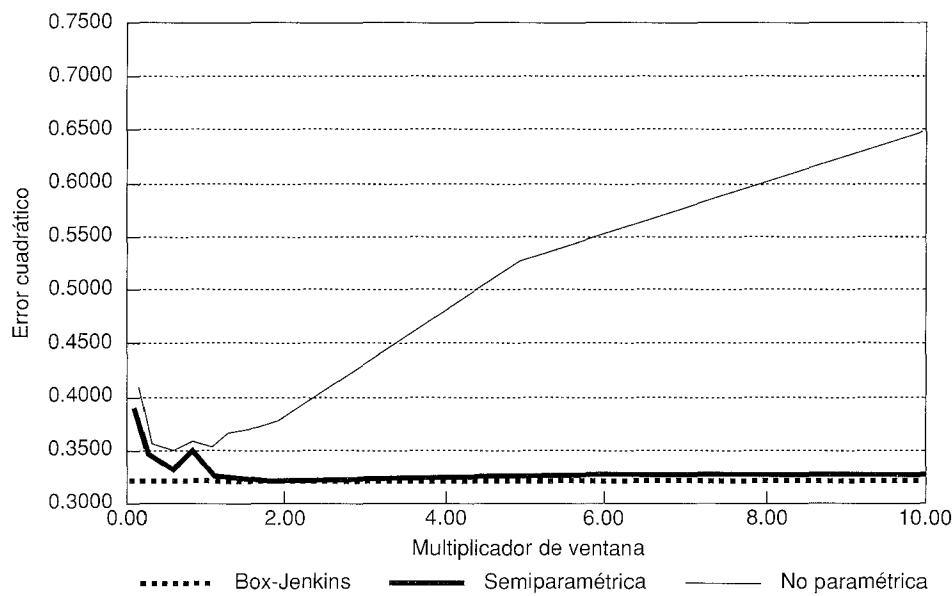


Figura 3
ERROR CUADRATICO INCONDICIONAL COMPARADO
DE LOS TRES METODOS. MODELO PSEUDO-ESTACIONARIO

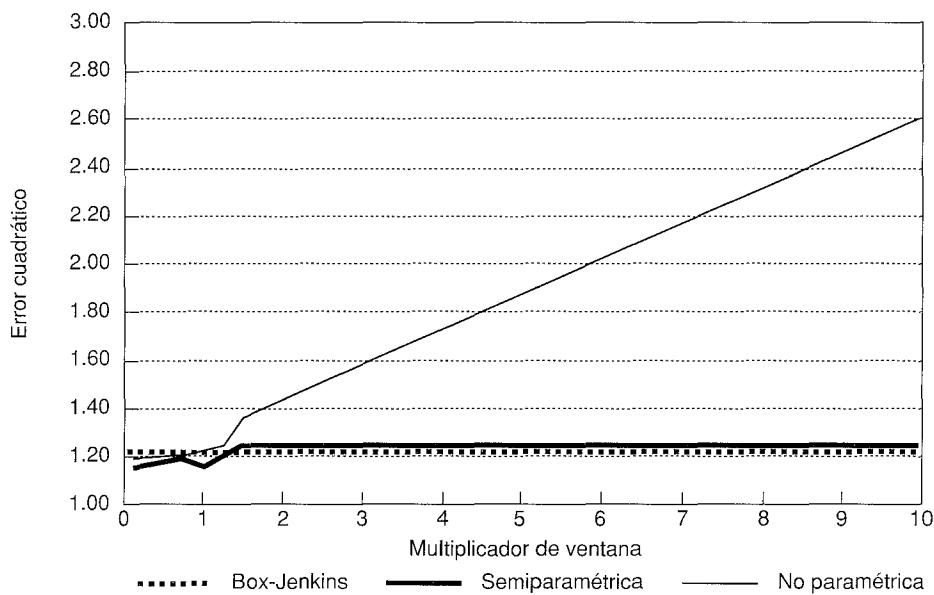


Figura 4
ERROR CUADRATICO INCONDICIONAL COMPARADO
DE LOS TRES METODOS. MODELO AR NO LINEAL

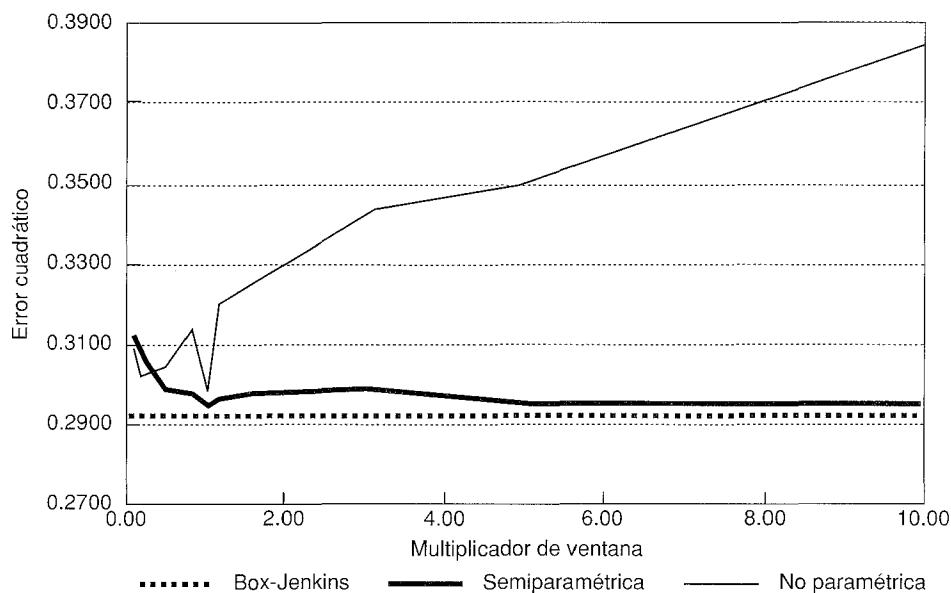


Figura 5
ERROR CUADRATICO INCONDICIONAL COMPARADO
DE LOS TRES METODOS. MODELO AR NO LINEAL II

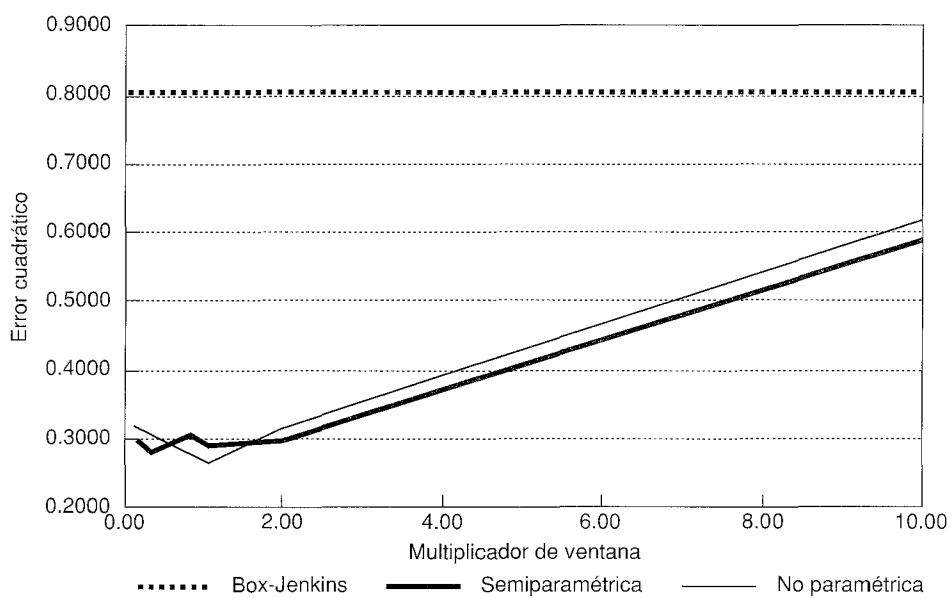


Figura 6
ERROR CUADRÁTICO INCONDICIONAL COMPARADO
DE LOS TRES MÉTODOS. MODELO ARMA (1,1)

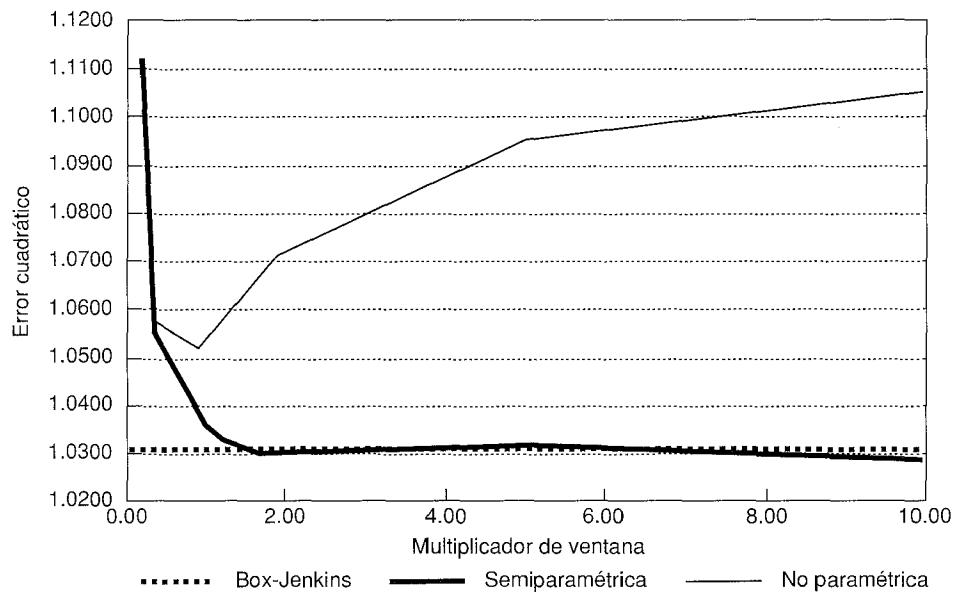
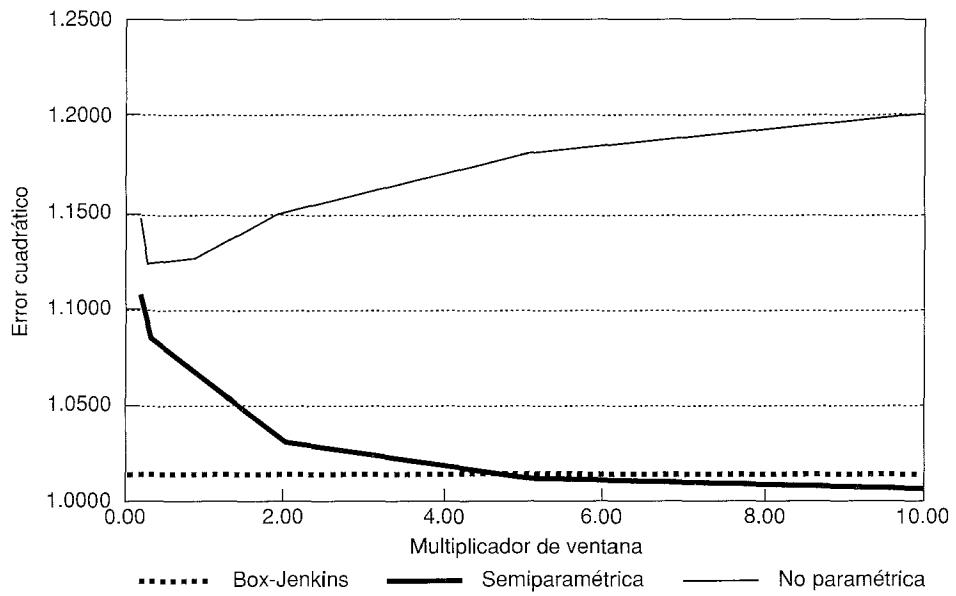


Figura 7
ERROR CUADRÁTICO INCONDICIONAL COMPARADO
DE LOS TRES MÉTODOS. MODELO MA (1)

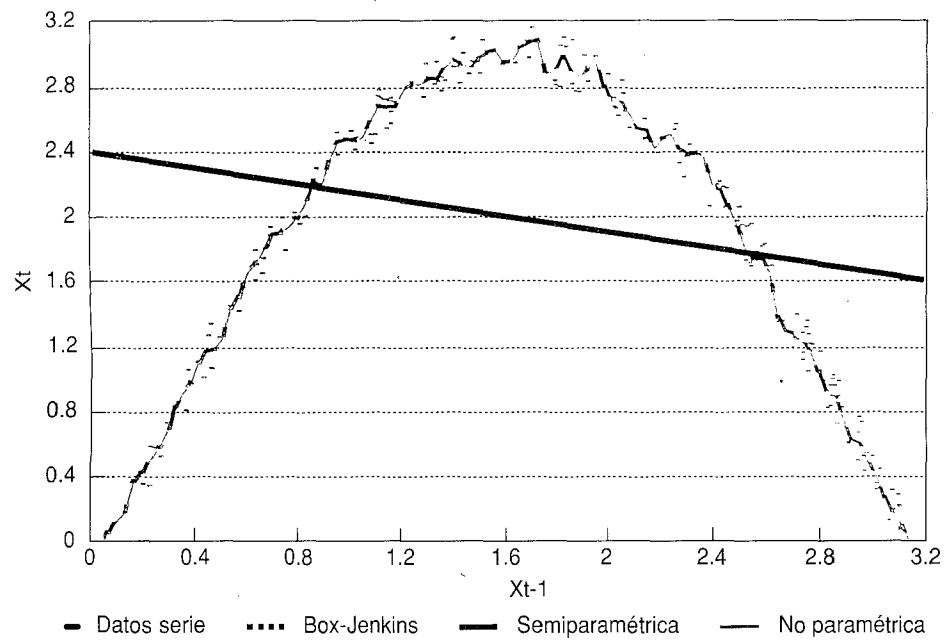


3. CONCLUSIONES

1. Los modelos Box-Jenkins son competitivos. La predicción no paramétrica presenta unos resultados moderadamente buenos pero, en general, peores que los de la predicción semiparamétrica.
2. La elección de la ventana es muy importante tanto para la metodología no paramétrica como para la semiparamétrica. En general, la ventana óptima local de validación cruzada no parece ser la óptima semiparamétrica. Esta necesita, en ocasiones, un grado de sobresuavización (ver, por ejemplo, Figuras 6 y 7) sobre aquélla. El estudio de en qué medida debe realizarse esa sobresuavización puede ser objeto de futura investigación. En este trabajo se seleccionaron de modo subjetivo diversos multiplicadores del parámetro ventana.
3. En algunos modelos no lineales (Serie 5), el funcionamiento Box-Jenkins es deficiente en comparación con las otras modelizaciones. A título ilustrativo, para esta serie se adjunta en la Figura 8 la estimación no paramétrica de la función de autorregresión a un retardo y la estimación lineal AR(1) Box-Jenkins; este último fue el modelo más veces seleccionado por el mecanismo anteriormente descrito en las cien simulaciones de la Serie 5.

Figura 8

FUNCIONES DE AUTORREGRESIÓN ESTIMADAS EN MODELO AR NO LINEAL II



4. En el amplio espectro de estos modelos todavía no se da cabida a todos los posibles, para algunos de los cuales la metodología Box-Jenkins se comporta muy mal. Como ilustración, en las Figuras 9 y 10 se presenta la estimación no paramétrica de las funciones de autorregresión $E(X_{t+1}|X_t, X_{t-1})$ y $E(X_{t+6}|X_t, X_{t-1})$ asociadas a los niveles de inmigración X_t (relativos a la aplicación citada en la introducción de este trabajo) y construidas con un banco de datos de 500 observaciones. Como se puede observar, la aproximación AR(2) Box-Jenkins da resultados especialmente malos cuando, pasando de un nivel alto X_{t-1} a uno bajo X_t , se trata de predecir X_{t+1} .

Figura 9
SUPERFICIE DE REGRESIÓN DE DATOS DE INMIGRACIÓN

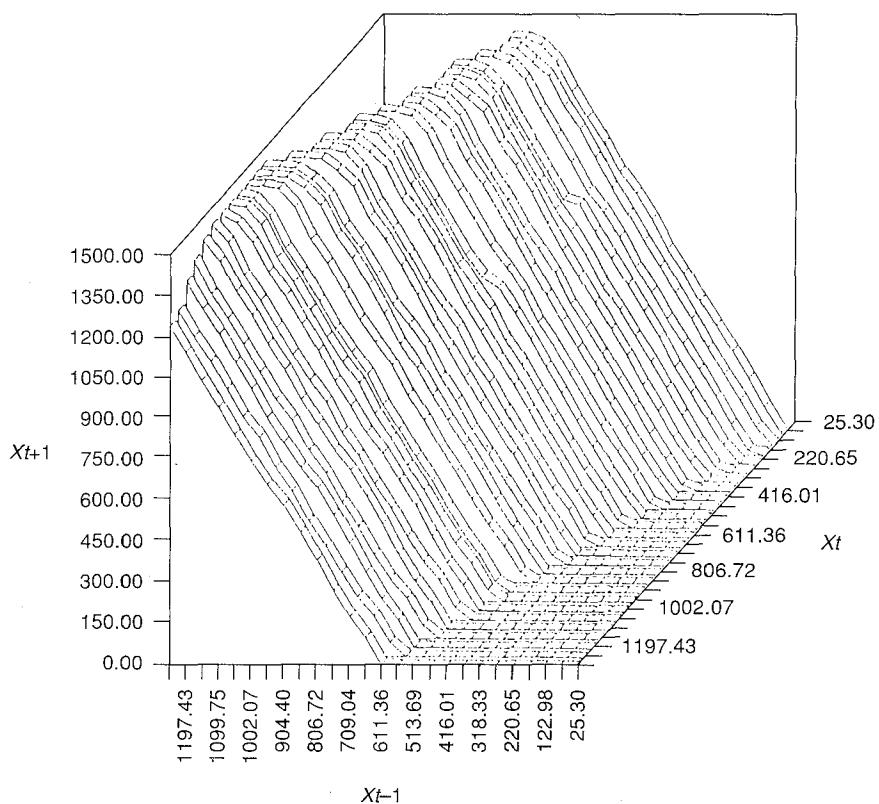
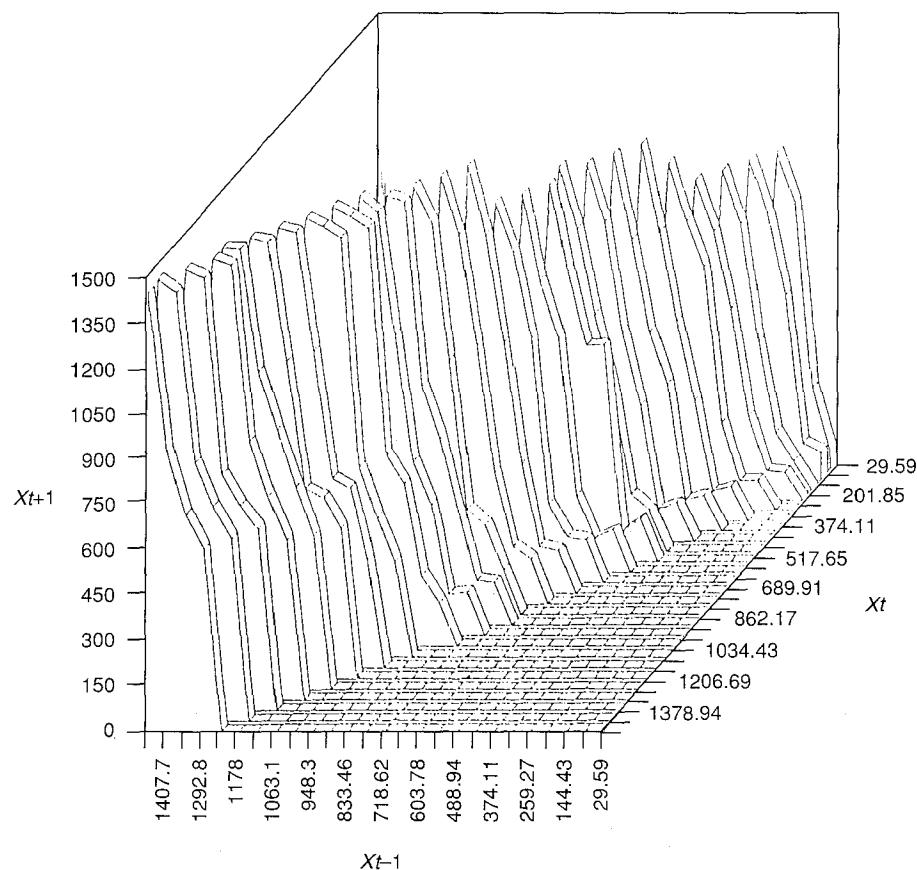


Figura 10
SUPERFICIE DE REGRESION DE DATOS DE INMISION



Finalmente, queremos comentar que, aunque el estudio desarrollado se llevó a cabo con pesos tipo núcleo, de forma análoga se podría haber realizado con pesos k_n -puntos más próximos [3] con resultados previsiblemente parecidos.

4. REFERENCIAS

- BROCKWELL, P. J., y DAVIS, R. A. (1991): *Time Series: Theory and Methods*, 2.^a ed. revisada, Springer-Verlag.
- CAO, R.; GONZÁLEZ-MANTEIGA, W.; PRADA SÁNCHEZ, J. M.; GARCÍA-JURADO, I., y FEBRERO-BANDE, M. (1992): «Forecasting Using a Semiparametric Model», *Computational Statistics*, Vol. 1, Y. Dodge y J. Whittakoe (eds.), Physica-Verlag, 327-330.
- GONZÁLEZ-MANTEIGA, W.; PRADA SÁNCHEZ, J. M.; CAO, R.; GARCÍA-JURADO, I.; FEBRERO-BANDE, M., y LUCAS-DOMÍNGUEZ, T. (1993): «Time-Series Analysis for Ambient Concentrations», *Atmospheric Environment*, 27 A, 153-158.
- HÄRDLE, W., y VIEU, P. (1992): «Kernel Regression Smoothing of Time Series», *Journal of Time Series*, Vol. 13, 3, 208-232.
- VIEU, P. (1991): «Nonparametric Regression: Local Optimal Bandwidth Choice», *Journal of the Royal Statistical Society, Series B*, Vol. 53, 2, 453-464.
- YAKOWITZ, S. J. (1985): «Nonparametric Density Estimation, Prediction and Regression for Markov Sequences», *JASA*, Vol. 80, 339, 215-221.
- YAKOWITZ, S. J. (1987): «Nearest-Neighbour Methods for Time Series Analysis», *Journal of Time Series*, Vol. 8, 2, 235-247.

SUMMARY

Given a sample $\{X_{t-n+1}, \dots, X_{t-1}, X_t\}$ of a stationary time series $\{X_t\}_{t \in \mathbb{Z}}$, the problem of giving a k -lags prediction \hat{X}_{t+k} of such a series is addressed. To that aim, a simulation study comparing Box-Jenkins techniques, nonparametric techniques (based on the nonparametric kernel estimation of the autoregression function) and semiparametric techniques (designed using the sum of the nonparametric prediction mentioned above and the Box-Jenkins prediction constructed with the residual series) is performed.

Key Words: Time Series, Prediction, Nonparametric Estimation, Semiparametric Estimation, Box-Jenkins methodology, Simulation.

AMS Classification: 62M20, 62G99.