

## TEST DE BONDAD DE AJUSTE DEL MODELO LINEAL GENERAL BAJO CORRELACIÓN SERIAL DE LOS ERRORES\*

J.M. VILAR FERNÁNDEZ\* y W. GONZÁLEZ MANTEIGA†

*Dado el siguiente modelo de regresión de diseño fijo, con correlación serial en los errores:  $Y_i = m(x_i) + \epsilon_i$ , donde  $x_i \in \mathcal{C}$ ,  $i = 1, \dots, n$ , siendo  $\mathcal{C}$  un conjunto compacto de  $\mathbb{R}$ , con error aleatorio  $\epsilon_i$  siguiendo una estructura lineal de tipo  $MA(\infty)$ , se propone un nuevo método para contrastar la hipótesis de que la función de regresión siga un modelo lineal, de la forma:  $m_\theta(-) = \mathcal{A}^t(-)\theta$ , con  $\theta \in \Theta \subset \mathbb{R}^q$ , y  $\mathcal{A}$  es un funcional de  $\mathbb{R}$  en  $\mathbb{R}^q$ .*

*El estadístico propuesto para contrastar la hipótesis de linealidad, que denominamos  $d^2$ , se obtiene como la distancia del tipo Cramer-von Mises entre el estimador no paramétrico de Gasser-Müller,  $\hat{m}$  de la función de regresión y el estimador paramétrico de mínima distancia bajo la hipótesis de linealidad:  $m_{\hat{\theta}}$ .*

*Los resultados presentados de normalidad asintótica para ambos estimadores:  $\hat{\theta}_n$  y  $d^2$ , y los estudios de simulación llevados a cabo sirven para ilustrar el efecto de la dependencia e indicar algunas formas de elegir el parámetro de suavización. Finalmente se incluyen ejemplos con datos reales.*

**Testing the hypothesis of a General Linear Model when errors are non-independent.**

**Key words:** Test de hipótesis paramétrica, modelos de regresión, series de tiempo, estimadores no paramétricos.

**Clasificación A.M.S.:** 62G07, 62G10, 62J05, 62M10.

\* J.M. Vilar Fernández. Departamento de Matemáticas. Universidad de La Coruña. España.

† W. González Manteiga. Departamento de Estadística e Investigación Operativa. Universidad de Santiago de Compostela. España.

✱ Este trabajo ha sido parcialmente subvencionado por la DGYCYT (PB91-0794) y Xunta de Galicia (X4GA20701B92).

-Article rebut el setembre de 1993.

-Acceptat el febrer de 1994.

## 1. INTRODUCCIÓN

Tradicionalmente, a lo largo de este siglo, ha sido de interés en el campo de la Estadística, el estudio de la relación entre las variables observadas en los diversos problemas de la vida real. Siendo el objetivo principal la creación de modelos de tipo estocástico, que sirvan para interpretar dicha relación.

De los modelos diseñados para dicho análisis, destacan por su importancia, los llamados modelos de regresión. Así, si  $Y$  es la variable de interés (variable respuesta, variable predecible, etc.) y  $x^t = (x_1, \dots, x_p)$  es el conjunto de variables regresoras (variables predictoras, ..., etc.), de las que se desea estudiar su influencia sobre la variable anterior, se define un modelo de regresión, como el dado por:

$$(1.1) \quad Y_i = m_\theta(x_{1i}, \dots, x_{pi}) + \epsilon_i, \quad i = 1, \dots, n; \quad \theta \in \Theta \subset \mathbb{R}^q$$

en el que  $\{\epsilon_i\}_{i=1}^n$  es una sucesión de variables aleatorias de media cero y  $\{(x_{1i}, \dots, x_{pi}), Y_i\}_{i=1}^n$  representa la muestra inicial observada.

Uno de los tópicos más estudiados en la teoría de los modelos de regresión, como el (1.1) antes descrito, es el relativo a los distintos contrastes de hipótesis acerca de la forma  $m_\theta$  que mejor explica la dependencia entre las variables predictoras y la variable respuesta. Existe una extensa literatura al respecto, de la que destacamos: Seber (1977), para modelos de regresión lineal, esto es,  $m_\theta(x) = \mathcal{A}^t(x)\theta$ , con  $\mathcal{A}(x) \in \mathbb{R}^q$  y  $x \in \mathbb{R}^p$ ; y, Seber y Wild (1989), para situaciones en las que, en general,  $m_\theta$  es una función no lineal de  $\theta$ . En cualquiera de estos casos, el planteamiento general del contraste de hipótesis viene dado por la hipótesis nula:

$$H_0: Y_i = m_\theta(x_{1i}, \dots, x_{pi}) + \epsilon_i, \quad i = 1, \dots, n; \quad \theta \in \Theta_1 \subset \Theta, \text{ con } \dim(\Theta_1) = q_1 < q$$

frente a la alternativa general  $H_1$ , que representa la validez del modelo (1.1).

Una perspectiva más general que la descrita en los párrafos anteriores consistiría en suponer que el espacio asociado a  $H_1$  no estuviese restringido a una familia paramétrica de funciones, tratándose de contrastar:

$$(1.2) \quad H_0: m \in \{m_\theta(-): \theta \in \Theta \subset \mathbb{R}^q\}$$

frente a la alternativa:

$$H_1 = \text{“}m \text{ es una función con cierto grado de suavidad.”}$$

De este modo estaríamos realmente contrastando la bondad de ajuste de una familia paramétrica de funciones de regresión; haciéndose, por tanto, necesaria

la estimación no paramétrica de la función de regresión  $m$  en la hipótesis general  $H_1$ . El estudio de estos contrastes es bastante reciente, apareciendo los primeros desarrollos formales fundamentalmente en los tres últimos años. Raz (1990), Firth-Glosup-Hinkley (1991), Staniswalis-Severini (1991), Eubank-Spiegelman (1990), Kozek (1991), Hart-Wehrly (1992) y Eubank-Hart (1992 y 1993) entre otros, son algunos ejemplos del trabajo desarrollado en esa línea cuando se toman, principalmente, como estimadores no paramétricos los de tipo núcleo o spline.

Una extensión natural del modelo (1.1) viene dada por la suposición de que los errores  $\{\epsilon_i\}_{i=1}^n$  siguen una estructura de dependencia. Lo que sucede frecuentemente, en el estudio de muestras de datos económicos, ya que en muchos casos estos son recogidos secuencialmente a lo largo del tiempo (índices de desempleo, precios de ciertos productos, . . . , etc.) mostrando una clara correlación serial.

Como consecuencia de la presencia de correlación entre los errores en el modelo, los clásicos estimadores  $\hat{\theta}$  de máxima verosimilitud, cuando no se considera la dependencia de las observaciones, o los de mínimos cuadrados, pueden llegar a ser muy ineficientes. (Ver el capítulo 6 de Seber y Wild (1989) para un análisis detallado del efecto de la correlación en la estimación de los parámetros). Además la dependencia en los errores del modelo da lugar a estimaciones infra o sobresuavizadas de  $m$ , en la hipótesis general  $H_1$ , cuando la ventana (o parámetro de suavización) es elegida adecuadamente para situaciones de errores incorrelados (ver Altman (1990) o Chu-Marron (1991) para más detalles).

En lo que sigue se tratará el contraste de la hipótesis general (1.2) bajo la hipótesis de que  $\epsilon_i$ , el error aleatorio en el modelo (1.1), sigue una estructura de tipo lineal: "medias móviles de orden infinito". Es decir,  $\epsilon_i = \sum_{j=0}^{\infty} b_j e_{i-j}$ , con  $\{e_i\}$  una sucesión de ruido blanco, o sea, variables aleatorias independientes de media cero y varianza  $\sigma_e^2$ .

El contraste de este tipo de hipótesis con alternativas generales no paramétricas y con suposición de dependencia en los errores es, desde nuestra perspectiva, un planteamiento más general que no había sido abordado hasta la actualidad. La suposición anterior, del tipo de medias móviles, incluye, entre otros, a los modelos denominados ARMA.

En este trabajo se estudia un contraste de la hipótesis nula, cuando ésta es de tipo lineal, esto es,  $H_0: m \in \{m_\theta(-) = \mathcal{A}^t(-)\theta: \theta \in \Theta \subset \mathbb{R}^q\}$ , siendo  $\mathcal{A}$  un funcional de  $\mathbb{R}$  en  $\mathbb{R}^q$ . Para ello se mide la distancia entre estimador paramétrico bajo  $H_0$  y un estimador no paramétrico, tipo núcleo.

En la sección 2 se define el estimador de mínima distancia,  $\hat{\theta}_n$ , del parámetro del modelo, y el estadístico del contraste de bondad de ajuste,  $d^2$ , obteniéndose la distribución asintótica de ambos estimadores bajo la hipótesis nula  $H_0$ . En el apartado 3 se analizan estudios de simulación que ilustran el efecto de la dependencia e indican una forma razonable y eficiente de elegir el parámetro de suavización. En el apartado 4, se estudian algunos ejemplos con datos reales. Finalmente, en el Apéndice se presenta la demostración de los resultados obtenidos.

## 2. DEFINICIONES Y RESULTADOS ASINTÓTICOS

### 2.1. Definición de los estimadores

En lo que sigue consideraremos el siguiente modelo de diseño fijo:

$$(2.1) \quad Y_i = m(x_i) + \epsilon_i$$

donde  $m$  es una función suave, que sin pérdida de generalidad, suponemos definida en  $[0,1]$ , los  $\{x_i\}_{i=1}^n$  están equiespaciados, es decir,  $x_i = i/n$ , los errores  $\{\epsilon_i\}_{i=1}^n$  constituyen un proceso causal como se comentó en el apartado anterior:

$\epsilon_i = \sum_{j=0}^{\infty} b_j \epsilon_{i-j}$ , con  $\{\epsilon_t\}$  ruido blanco de varianza  $\sigma_\epsilon^2$ . (Ver Brockwell y Davis (1991), Def. 3.1.3 para más detalles) y la muestra  $\{Y_i\}_{i=1}^n$  representa a las observaciones obtenidas en los puntos de diseño.

Para contrastar la hipótesis nula de linealidad

$$H_0: m \in \{m_\theta(-) = \mathcal{A}^t \theta: \theta \in \Theta \subset \mathbb{R}^q\}, \mathcal{A} \text{ un funcional de } \mathbb{R} \text{ en } \mathbb{R}^q$$

frente a la alternativa:

$$H_1: "m \text{ es una función con cierto grado de suavidad.}"$$

Se calculará la distancia entre una estimación no paramétrica,  $\hat{m}_n(x)$ , de la función de regresión,  $m(x)$ , y un estimador paramétrico de la misma función, bajo el supuesto de que se verifique  $H_0$ . En particular, se tomará una distancia del tipo Cramer-von Mises:

$$(2.2) \quad d^2(\hat{m}_n, H_0) = \min_{\theta \in \Theta} \int (\hat{m}_n(x) - \mathcal{A}^t(x)\theta)^2 \omega(x) d\Omega_n(x) = d^2(\hat{m}_n, \mathcal{A}^t(x)\hat{\theta}_n)$$

donde  $\omega$  es una función de ponderación que se utiliza para evitar el efecto frontera y  $\Omega_n$  es la distribución empírica de los puntos del diseño  $\{x_i\}_{i=1}^n$ .

Para estimar la función de regresión utilizaremos estimadores no paramétricos núcleo del tipo Gasser-Müller (1979):

$$(2.3) \quad \hat{m}_n(x) = \sum_{j=1}^n h^{-1} \left( \int_{s_{j-1}}^{s_j} K \left( \frac{x-s}{h} \right) ds \right) Y_j = \sum_{j=1}^n W_{n,j}(x) Y_j$$

con  $s_0 = 0$ ,  $s_{j-1} \leq x_j \leq s_j$ ,  $s_n = 1$ , donde  $K$  es una función núcleo: una función de densidad simétrica y de soporte compacto.

Bajo la hipótesis nula  $H_0$  de linealidad, es posible estimar  $\theta$  con diversos criterios “root- $n$ ”, esto es, que la convergencia en probabilidad del estimador al parámetro sea de orden  $n^{-\frac{1}{2}}$  ( $\sqrt{n}(\hat{\theta} - \theta) = O_{\mathbb{P}}(1)$ ), por ejemplo, el mínimo cuadrático bajo ciertas condiciones de regularidad (Seber-Wild (1989), cap. 6). Pero hemos considerado más conveniente utilizar un criterio de mínima distancia, que resulta de la optimización de (2.2). Una estimación de este tipo parece más natural para un posterior contraste de la hipótesis  $H_0$ .

Por tanto, el estimador  $\hat{\theta}_n$  en (2.2) representa el estimador de mínima distancia asociado, y viene dado por la expresión:

$$(2.4) \quad \hat{\theta}_n = \mathbf{B}^{-1} \left( \sum_{\tau=1}^n \mathcal{A}(x_\tau) \omega(x_\tau) \hat{m}_n(x_\tau) \right) \quad \text{con } \mathbf{B} = \sum_{j=1}^n \mathcal{A}(x_j) \mathcal{A}^t(x_j) \omega(x_j)$$

obtenida de minimizar la función  $\Psi(\theta) = \int (\hat{m}_n(x) - \mathcal{A}^t(x)\theta)^2 \omega(x) d\Omega_n(x)$  en el conjunto  $\Theta \subset \mathbb{R}^q$ , siendo el valor de esta expresión en el mínimo, el estadístico del contraste de bondad de ajuste,  $d^2$ . Además, como se prueba en el teorema 1, abajo indicado, la estimación de mínima distancia también es root- $n$  bajo condiciones no muy restrictivas.

## 2.2. Resultados asintóticos

Para el estudio asintótico de los estimadores de mínima distancia y del estadístico  $d^2$ , consideraremos las siguientes hipótesis además de las ya indicadas:

- A1.** La función  $m$  es dos veces diferenciable, con derivadas acotadas, y  $\omega$  tiene soporte compacto contenido en  $(0,1)$ .
- A2.** La función núcleo  $K$  es derivable continuamente.

**A3.** La serie de errores  $\{\epsilon_i\}$  tiene función de autocovarianza:

$$\gamma(|i-j|) = \sigma_\epsilon^2 \sum_{\ell=0}^{\infty} b_\ell b_{\ell+|i-j|} \text{ absolutamente sumable, esto es, } \sum_{k=-\infty}^{+\infty} |\gamma(k)| < \infty$$

**A4.** Para algún  $\delta > 0$  se verifica que:  $E[|e|^{4+2\delta}] < \infty$ .

**A5.** El coeficiente de curtosis de las variables del ruido blanco es cero.

**A6.**  $nh \rightarrow \infty$  y  $h = o\left(n^{-\frac{\delta+2}{2\delta+2}}\right)$  cuando  $n \rightarrow \infty$ .

Utilizando estas hipótesis y el Teorema Central del Límite para sucesiones con parte principal  $m(n)$ -dependiente debido a Nieuwenhuis (1992) se ha obtenido la distribución asintótica del estimador de mínima distancia,  $\hat{\theta}_n$  y del estadístico  $d^2$ .

### Teorema 1

Bajo la hipótesis nula,  $H_0$ , y si se cumplen las hipótesis anteriores, se verifica:

a)  $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma_\epsilon^2 \mathbf{Q}^{-1} \mathbf{Q}^* \mathbf{Q}^{-1})$  donde  $\mathbf{Q} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathcal{A}(x_i) \mathcal{A}^t(x_i) \omega(x_i)$   
y  $\mathbf{Q}^* = \sum_{u=-\infty}^{\infty} \gamma(u) R(u)$ , con  $R(u) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^{n-u} \mathcal{A}(x_j) \mathcal{A}^t(x_{j+u}) \omega(x_j) \omega(x_{j+u})$   
(límites que se supone que existen).

b)  $\sqrt{n^2 h} \left( d^2(\hat{m}, \mathcal{A}^t(-)\hat{\theta}_n) - \frac{\bar{\omega} \sigma_\epsilon^2 \int K^2}{nh} \right) \xrightarrow{d} N(0, \sigma_d^2)$  con  $\sigma_\epsilon^2 = E(\epsilon^2)$ ;

$\sigma_d^2 = 2 \left( \sum_{k=-\infty}^{\infty} \gamma(k) \right)^2 \int (K * K)^2 \int \omega^2$ ; y  $\bar{\omega} = \frac{1}{n} \sum_{i=1}^n \omega(x_i)$  siendo  $K * K$  la convolución de la función núcleo consigo misma.

### 2.3. Comentarios

**Comentario 1.** Las hipótesis A1. y A2. son clásicas en la literatura relativa a la estimación no paramétrica. El carácter absolutamente sumable de la función de autocovarianzas es verificado, por ejemplo, por todos los modelos ARMA (ver

(Brockwell-Davis (1991), pág. 219). La hipótesis A5. se incluye únicamente para una simplificación de los cálculos de la parte b) del teorema que se expone más adelante. Finalmente la hipótesis A6. indica que la ventana  $h$  ha de ser elegida con un orden que fluctúa entre  $n^{-1}$  y  $n^{-\frac{\epsilon+2}{2\epsilon+2}}$ . En particular, esta hipótesis implica que  $nh^4 \rightarrow 0$ , lo que es necesario para la cancelación de sesgos.

**Comentario 2.** Obsérvese que bajo la ausencia de correlación serial,  $\gamma(k) = 0$  para todo  $k \neq 0$  y, por tanto,  $\sigma_d^2 = 2\sigma_\epsilon^4 \int (K * K)^2 \int \omega^2$ , extendiéndose el resultado de González Manteiga y Cao Abad (1991) obtenido en el contexto de datos independientes. En este trabajo se utilizan hipótesis análogas a las utilizadas aquí, siendo las diferencias en el orden de elección del parámetro de suavización y el orden del momento del error muestral que tiene que ser acotado, y sobre todo, que allí se trabaja bajo hipótesis de independencia de las observaciones, lo que hace que la demostración de los resultados obtenidos en este trabajo sea más compleja.

**Comentario 3.** El estimador mínimo cuadrático,  $\hat{\theta}_{mc}$ , es un caso particular del estimador de mínima distancia,  $\hat{\theta}$ , que corresponde a una elección degenerada del parámetro de suavización:  $h \cong 1/n$  en (2.3), para el estimador piloto usado en (2.2), y produce una mayor frecuencia de rechazo de las hipótesis cuando  $d^2(\hat{m}_n, H_0)$  es usado como estadístico para el contraste de  $H_0$ . Además, existen elecciones del parámetro de suavización no degeneradas para las que se verifica:

$$d^2(\hat{m}_n, A^t(-)\hat{\theta}_n) < d^2(\hat{m}_n, A^t(-)\hat{\theta}_{mc})$$

lo que justifica la utilización del estimador  $\hat{\theta}$  en lugar de  $\hat{\theta}_{mc}$  para el cálculo de  $d^2$ . Por otra parte, una elección del parámetro de suavización del orden  $1/n$  convierte a  $d^2$  en el clásico test de bondad de ajuste basado en los residuos (F-test).

### 3. ESTUDIOS DE SIMULACIÓN

#### 3.1. Influencia de la dependencia. Un ejemplo de simulación

La consideración de la dependencia en los errores asociados a los modelos (1.1) o (2.1) cuando ésta existe es de notable importancia. Por ejemplo, si se utiliza el resultado asintótico dado en b) para contrastar la hipótesis nula de linealidad  $H_0$  se rechaza al nivel  $\alpha$  si:

$$(3.1) \quad \sqrt{n^2 h} \left( d^2 \left( \hat{m}_n, \mathcal{A}^t(-)\hat{\theta}_n \right) - \frac{\bar{w}\sigma_\epsilon^2 \int K^2}{nh} \right) > z_\alpha \sigma_d$$

donde  $z_\alpha$  es tal que  $\Phi(z_\alpha) = 1 - \alpha$  ( $\Phi$  es la función de distribución de una  $N(0,1)$ ). Es evidente, por tanto, que la omisión de la posible estructura de dependencia de los errores puede hacer que se rechace la hipótesis más veces que las correspondientes al nivel que si, por ejemplo, hay correlación serial positiva.

**Tabla 1**

$\sigma_\epsilon^2 = 0.1$								
$\alpha = 0.10$					$\alpha = 0.05$			
$h$	.018	.020	.024	.025	.0150	.020	.025	.028
$\rho = 0$	.9426	.8662	.8116	.7872	1.0000	.9490	.8742	.8014
$\rho = .2$	.8588	.7362	.6580	.6254	1.0000	.8768	.7462	.6500
$\rho = .4$	.7278	.5394	.4672	.4386	1.0000	.7294	.5660	.4654
$\rho = .6$	.5938	.3702	.2980	.2716	1.0000	.5438	.3790	.2938
$\rho = .8$	.5646	.2484	.1754	.1540	1.0000	.3646	.2220	.1602

$\sigma_\epsilon^2 = 1$								
$\alpha = 0.10$					$\alpha = 0.05$			
$h$	.0100	.0200	.0500	.0600	.0300	.0350	.0500	.0600
$\rho = 0$	1.0000	.9918	.9726	.9714	.9962	.9932	.9874	.9882
$\rho = .2$	1.0000	.9494	.9076	.9190	.9742	.9598	.9480	.9530
$\rho = .4$	.9998	.8040	.7572	.7882	.8826	.8512	.8454	.8658
$\rho = .6$	.9994	.5456	.5248	.5834	.6616	.6176	.6356	.6786
$\rho = .8$	.9992	.3290	.3066	.3652	.4048	.3724	.3964	.4540

Este efecto se observa en el siguiente estudio de simulación: se han generado 5.000 muestras de tamaño  $n = 25$  del modelo  $Y_1 = m(x_i) + \epsilon_i = 1 + 2x_i + x_i^2 + \epsilon_i$ ,  $i = 1, \dots, n$ ; en el que  $x_i = i/n$  y  $\epsilon_i = \rho\epsilon_{i-1} + e_i$  con  $e_i \in N(0, \sigma_e^2)$ . Es decir,  $\epsilon_i$  sigue una estructura del tipo  $AR(1)$  con  $|\rho| < 1$ . Para cada una de las muestras se contrasta a nivel  $\alpha$ , suponiendo que los errores son independientes, la hipótesis

nula  $H_0$ : “ $m$  es un polinomio de grado menor o igual que dos”, presentándose en la Tabla 1 adjunta la frecuencia de aceptación de la hipótesis para distintos valores de  $\alpha, \rho, h$  y  $\sigma^2$ . Aquí y en lo que sigue  $\hat{m}_n$  es un estimador tipo Priestley-Chao (1972) con núcleo gaussiano y función de ponderación,  $\omega(x) = 1$ .

En esta tabla se observa claramente que al aumentar la dependencia de los errores del modelo disminuye progresivamente la frecuencia de aceptación, ocurriendo de forma drástica para valores altos de la correlación,  $\rho = 0.6$  ó  $0.8$ , lo que confirma lo expuesto anteriormente. También destacar la gran influencia en los resultados del parámetro de suavización,  $h$ , elegido, problema que tratamos en el apartado siguiente. Para la obtención de la Tabla 1, se ha elegido  $h$  de forma empírica, después de distintas pruebas se han tomado los  $h$  que mejores resultados proporcionaban y se exponen resultados con otros valores de  $h$  para comparar.

En todo caso debe de tenerse en cuenta que se están calculando proporciones de rechazo utilizando la distribución asintótica y estimando parámetros (las covarianzas) lo que hace que trabajando con muestras finitas los resultados no coincidan con los teóricos.

### 3.2. Consideraciones en base a estudios de simulación sobre la selección del parámetro de suavización

En este apartado se exponen los resultados de un amplio estudio de simulación realizado con un objetivo doble. En primer lugar analizar el buen comportamiento del estadístico introducido en el apartado 2 y en segundo lugar buscar orientaciones sobre la forma de elegir el parámetro de suavización,  $h$ .

En la primera parte del estudio de simulación se han generado 2.000 muestras de tamaño 50 de los modelos siguientes:

$$Y_i = m_j(x_i) + \epsilon_i, \quad x_i = i/50, \quad i = 1, \dots, 50 \quad \text{para } j = 1, 2, 3$$

siendo

$$\begin{aligned} m_1(x) &= 1 + 2x + x^2 && \text{(verifica la hipótesis } H_0) \\ m_2(x) &= 1 + 2x + x^2 + 5x^3 && \text{(no se verifica la hipótesis } H_0) \\ m_3(x) &= 2 + 2\text{sen}(3\pi x/2) && \text{(no se verifica la hipótesis } H_0) \end{aligned}$$

y  $\{\epsilon_i\}$  es el ruido con una correlación serial siguiendo una estructura del tipo MA con varianza  $\sigma^2 = 0.10$  y  $0.50$ . Como modelos de dependencia MA se han utilizado los siguientes:

$$\begin{aligned} \text{DEP.1.:} \quad \epsilon_t &= e_t + 0.9e_{t-1}, && \text{por tanto } \rho_1 = 0.497 \\ \text{DEP.2.:} \quad \epsilon_t &= e_t - 1.5e_{t-1} + 0.9e_{t-2}, && \text{por tanto } \rho_1 = -0.7; \quad \rho_2 = 0.22 \end{aligned}$$

siendo  $\{\epsilon_t\}$  ruido blanco gaussiano y  $\{\rho_1, \rho_2\}$  los coeficientes de correlación a uno y dos retardos.

Se contrasta la hipótesis nula  $H_0$ : “ $m$  es un polinomio de grado menor o igual que dos”. El contraste utilizado es el asintótico dado por (3.1) donde  $\sigma_d$  es el valor teórico dado por el Teorema 1 y  $\hat{\theta}_n$  es el estimador de mínima distancia. En las Tablas 2 y 3 se presentan las proporciones de aceptación de  $H_0$  para distintos  $h$ , los dos tipos de dependencias, y niveles de significación  $\alpha = 0.10$  y  $0.01$ . También se ha obtenido la media del nivel crítico del test.

**Tabla 2**  
**Dependencia 1**

$\sigma^2 = 0.10$		Niv. Sign. $\alpha = 0.10$	Niv. Sign. $\alpha = 0.01$	Media Niv. Crít.
$h = .005$	$m_1$	0.103	0.316	0.0328
	$m_2$	0.053	0.208	0.0188
	$m_3$	0.000	0.000	0.0000
$h = .009$	$m_1$	0.759	0.967	0.2874
	$m_2$	0.591	0.891	0.2005
	$m_3$	0.001	0.012	0.0006
$h = .013$	$m_1$	0.627	0.934	0.2071
	$m_2$	0.053	0.358	0.0208
	$m_3$	0.000	0.005	0.0002
$h = .017$	$m_1$	0.392	0.786	0.1102
	$m_2$	0.000	0.003	0.0001
	$m_3$	0.000	0.000	0.0000

$\sigma^2 = 0.50$		Niv. Sign. $\alpha = 0.10$	Niv. Sign. $\alpha = 0.01$	Media Niv. Crít.
$h = .006$	$m_1$	0.303	0.665	0.0988
	$m_2$	0.273	0.625	0.0902
	$m_3$	0.068	0.240	0.0243
$h = .010$	$m_1$	0.809	0.980	0.3180
	$m_2$	0.759	0.968	0.2858
	$m_3$	0.368	0.728	0.1154
$h = .014$	$m_1$	0.774	0.972	0.2900
	$m_2$	0.586	0.911	0.1867
	$m_3$	0.295	0.640	0.0925
$h = .018$	$m_1$	0.732	0.960	0.2620
	$m_2$	0.392	0.781	0.1073
	$m_3$	0.231	0.567	0.0746

**Tabla 3**  
**Dependencia 2**

$\sigma^2 = 0.10$		Niv. Sign. $\alpha = 0.10$	Niv. Sign. $\alpha = 0.01$	Media Niv. Crít.
$h = .008$	$m_1$	0.273	0.284	0.2588
	$m_2$	0.149	0.150	0.1345
	$m_3$	0.000	0.000	0.0000
$h = .012$	$m_1$	0.994	0.995	0.9926
	$m_2$	0.117	0.176	0.1556
	$m_3$	0.000	0.000	0.0000
$h = .016$	$m_1$	0.995	0.997	0.9935
	$m_2$	0.000	0.000	0.0000
	$m_3$	0.000	0.000	0.0000
$h = .017$	$m_1$	0.726	0.756	0.6871
	$m_2$	0.000	0.000	0.0000
	$m_3$	0.000	0.000	0.0000

$\sigma^2 = 0.50$		Niv. Sign. $\alpha = 0.10$	Niv. Sign. $\alpha = 0.01$	Media Niv. Crít.
$h = .008$	$m_1$	0.281	0.293	0.2658
	$m_2$	0.248	0.272	0.2360
	$m_3$	0.009	0.010	0.0080
$h = .023$	$m_1$	1.000	1.000	1.0000
	$m_2$	0.236	0.255	0.1952
	$m_3$	0.006	0.008	0.0037
$h = .038$	$m_1$	1.000	1.000	1.0000
	$m_2$	0.000	0.000	0.0000
	$m_3$	0.000	0.000	0.0000
$h = .053$	$m_1$	0.772	0.828	0.6919
	$m_2$	0.000	0.000	0.0000
	$m_3$	0.000	0.000	0.0000

En los resultados obtenidos en el estudio de simulación y detallados en las tablas anteriores se aprecia gran sensibilidad del test descrito respecto al parámetro de

suavización,  $h$ , ya que pequeñas variaciones de este parámetro llevan a resultados diferentes. Nuevamente, en las tablas obtenidas, se ha elegido el  $h$  de forma empírica, utilizando el que hemos considerado más adecuado, después de distintas pruebas, y otros valores próximos de forma que se pueda observar cómo influyen en los resultados.

Para estudiar con mayor detalle la influencia de este parámetro se ha hecho un segundo estudio de simulación en el que utilizando los mismos modelos de funciones y de dependencia de los errores que en la simulación anterior, se ha calculado la media del estadístico,  $d^2$ , en función de  $h$ , para 100 muestras simuladas de tamaño 50, que se denota por  $\bar{d}^2(h)$  y los niveles críticos medios del test en dos situaciones: primero, utilizando las autocovarianzas teóricas de los errores, obteniendo la función  $CLT(h)$ . Y, en segundo lugar, estimando las autocovarianzas de los errores a partir de los datos de cada muestra siguiendo la técnica de Müller-Stadmüller (1988), obteniendo la función  $CLE(h)$ .

En las figuras 1-2-3-4-5 se han representado las funciones  $\bar{d}^2(h)$ ,  $CLT(h)$  y  $CLE(h)$  bajo distintas situaciones y se observa que una buena elección de  $h$  para la hipótesis nula puede consistir en tomar el mínimo local más pequeño resultante de optimizar:

$$\min_h \left\{ d^2 \left( \hat{m}_{n,h}, \mathcal{A}^t(-)\hat{\theta}_{n,h} \right) \right\}$$

ya que a este valor de  $h$  le corresponde el mayor nivel crítico. Obsérvese que el mínimo es local ya que cuando  $h \rightarrow \infty$ , se verifica que  $d^2 \rightarrow 0$ , pues en ese caso  $\hat{m}(x)$  es una curva sobresuavizada que tiende a ser constante. También de las gráficas representadas se deduce que cuando se utiliza el test con autocovarianzas estimadas los resultados son bastantes similares a los obtenidos con autocovarianzas teóricas.

En hipótesis alternativas, como la dada por los modelos  $m_2(x)$  y  $m_3(x)$  la  $h$  elegida por el criterio anterior puede no ser la adecuada como se puede apreciar en las Figuras 4-5, donde se observa una tendencia de aceptación de la hipótesis para dicho valor de  $h$ .

En todo caso, las conclusiones anteriores se han obtenido en base a un estudio de simulación, que aunque amplio, no es exhaustivo, y permanecen abiertas muchas cuestiones sobre el tema. Por ejemplo, si se puede asegurar la existencia de un mínimo local en la curva  $d^2(h)$ . Por todo ello y teniendo en cuenta que una correcta elección de  $h$  es crucial, será necesaria una amplia investigación sobre este particular, principalmente, bajo hipótesis alternativas, ya que bajo  $H_0$  la sugerencia hecha parece funcionar adecuadamente, aunque es costosa desde un punto de vista computacional.

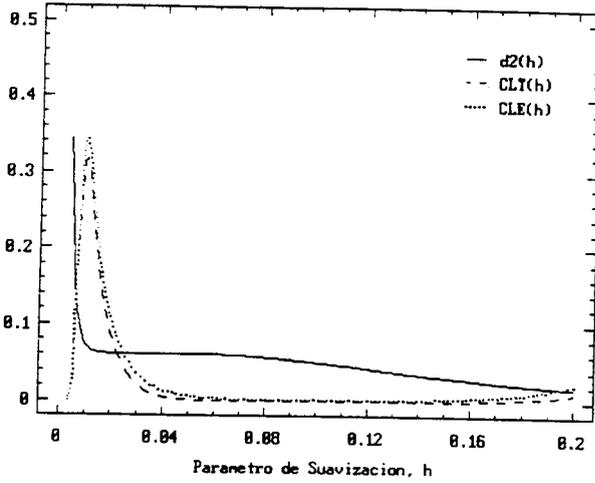


Figura 1.

Curvas  $\bar{d}^2(h)$  (media de las  $d^2$  para 100 muestras),  $CLT(h)$  (nivel crítico medio con autocovarianzas de los errores conocidas) y  $CLE(h)$  (autocovar. estimadas) para  $m_1(x) = 1 + 2x + x^2$ , con **DEP.1.** y  $\sigma^2 = 0.10$ .

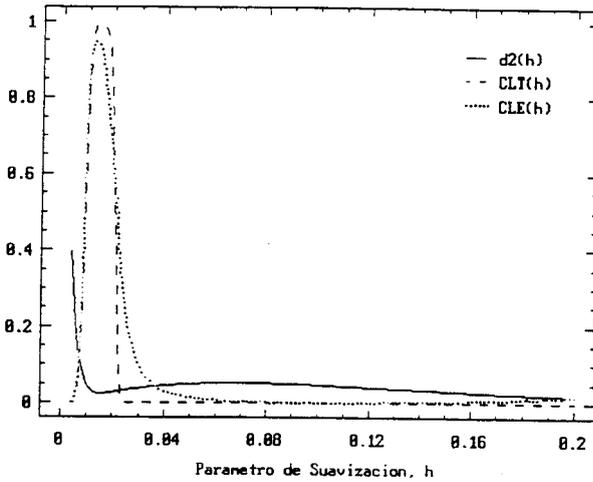
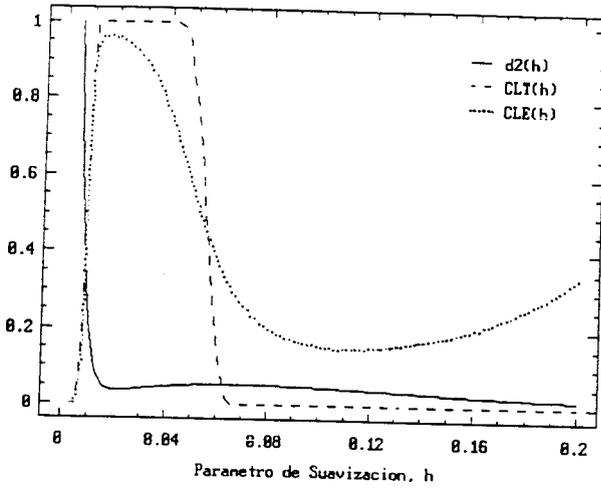


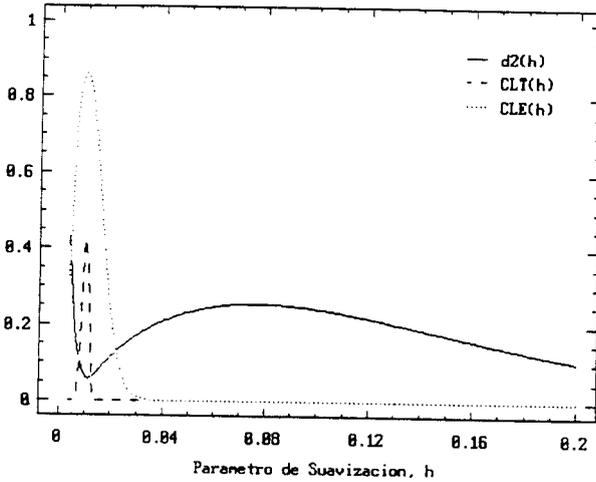
Figura 2.

Curvas  $\bar{d}^2(h)$  (media de las  $d^2$  para 100 muestras),  $CLT(h)$  (nivel crítico medio con autocovarianzas de los errores conocidas) y  $CLE(h)$  (autocovar. estimadas) para  $m_1(x) = 1 + 2x + x^2$ , con **DEP.2.** y  $\sigma^2 = 0.10$ .



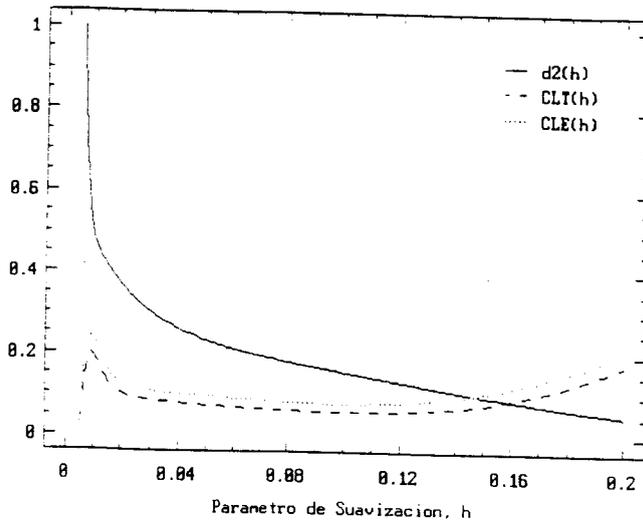
**Figura 3.**

Curvas  $\bar{d}^2(h)$  (media de las  $d^2$  para 100 muestras),  $CLT(h)$  (nivel crítico medio con autocovarianzas de los errores conocidas) y  $CLE(h)$  (autocovar. estimadas) para  $m_1(x) = 1 + 2x + x^2$ , con **DEP.2.** y  $\sigma^2 = 0.50$ .



**Figura 4.**

Curvas  $\bar{d}^2(h)$  (media de las  $d^2$  para 100 muestras),  $CLT(h)$  (nivel crítico medio con autocovarianzas de los errores conocidas) y  $CLE(h)$  (autocovar. estimadas) para  $m_2(x) = 1 + 2x + x^2 + 5x^3$ , con **DEP.2.** y  $\sigma^2 = 0.10$ .



**Figura 5.**

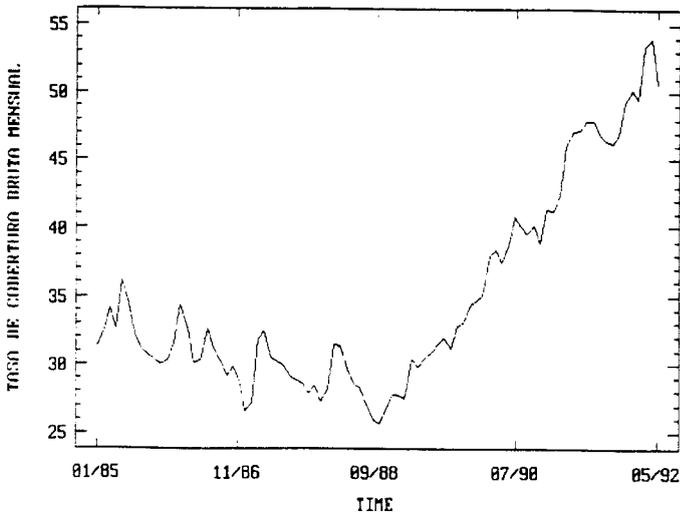
Curvas  $\bar{d}^2(h)$  (media de las  $d^2$  para 100 muestras),  $CLT(h)$  (nivel crítico medio con autocovarianzas de los errores conocidas) y  $CLE(h)$  (autocovar. estimadas) para  $m_3(x) = 2 + 2\text{sen}(3\pi x/2)$ , con **DEP.1.** y  $\sigma^2 = 0.50$ .

#### 4. ESTUDIO DE UN CASO CON DATOS REALES

Se estudiará en este apartado un ejemplo en el que a partir de una muestra de datos reales se aplica el contraste diseñado para modelos con posible correlación serial entre los errores. Como variable respuesta  $Y_i$  se considera “la tasa de cobertura bruta de las prestaciones por desempleo total en Galicia”, definida como sigue:

$$T.C.B. = \frac{\text{número de beneficiarios de prestaciones por desempleo total y asistencias económicas}}{\text{Paro registrado}} \times 100$$

La muestra utilizada abarca el período de Enero de 1985 a Abril de 1992 (88 datos, facilitados por el Instituto Galego de Estadística, Xunta de Galicia y publicados en los Boletines de Series Estadísticas de Galicia). La variable regresora,  $x_i = i/88$ ,  $i = 1, \dots, 88$ , es el tiempo normalizado en el intervalo  $[0,1]$ . En la figura 6 se ha representado la serie  $\{Y_i\}_{i=1}^{88}$  de datos observados a lo largo del tiempo. Y del análisis de las funciones de autocorrelación y autocorrelación parcial de la muestra se deduce utilizando la metodología Box-Jenkins que se puede realizar una buena aproximación por un modelo ARIMA con componente estacional y, por tanto, por un modelo  $MA(\infty)$ .



**Figura 6.**

Datos de la Tasa de Cobertura Bruta mensual en Galicia de Enero de 1985 a Abril de 1992.

Si se realiza un ajuste mediante regresión polinómica de grado cinco entre la variable  $Y_i$  y el tiempo  $x_i = i/88$ ,  $i = 1, \dots, 88$ :

$$Y_i = m_5(x_i) + \epsilon_i = \theta_1 + \theta_2 x_i^2 + \theta_3 x_i^3 + \theta_4 x_i^4 + \theta_5 x_i^5 + \epsilon_i, \quad i = 1, \dots, 88$$

suponiendo una estructura de independencia y homocedasticidad para los errores con estructura gaussiana, resulta la siguiente tabla adjunta:

**Tabla 4**

Variables Independientes	Estimación	t-valor	Nivel de Significación
Término Constante	32.804	28.12	0.0000
$x_i$	-1.438	-0.06	0.9500
$x_i^2$	-33.380	-0.24	0.8095
$x_i^3$	-93.376	-0.27	0.7866
$x_i^4$	379.177	1.01	0.3135
$x_i^5$	-232.475	-1.57	0.1180
Coeficiente de correlación Ajustado = 0.9508		Durbin-Watson: 0.863	

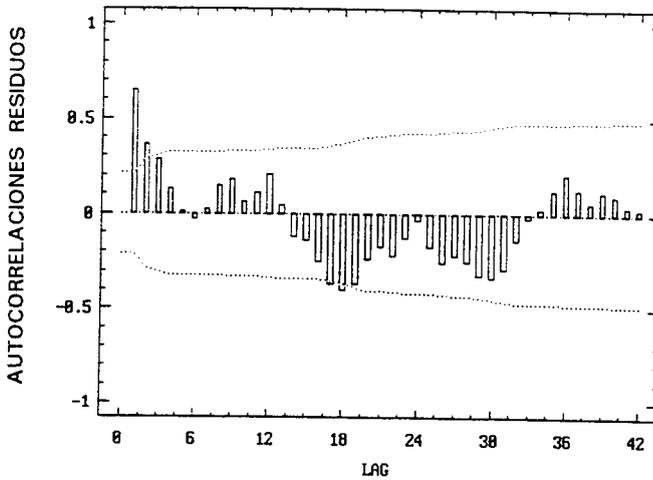
De los resultados expuestos en esta tabla se deduce que el modelo explica de forma significativa la variable respuesta aunque los contrastes individuales (test  $t$ ) no sean significativos, lo que se debe a que tienen varianzas muy altas. Esto ocurre con frecuencia en los modelos polinómicos ya que las variables  $x, x^2, \dots$  son fuertemente dependientes y provoca problemas de multicolinealidad. (Ver Peña, 1987).

Además tampoco es recomendable en la teoría de regresión polinómica utilizar modelos de grado elevado y en este ejemplo también se puede obtener un ajuste adecuado utilizando un polinomio de grado dos, esto es, haciendo un ajuste del tipo:  $Y_i = m_\theta(x_i) + \epsilon_i = \theta_1 + \theta_2 x_i + \theta_3 x_i^2$  se obtienen los siguientes resultados (Tabla 5):

**Tabla 5**

Variables Independientes	Estimación	t-valor	Nivel de Significación
Término Constante	35.828	56.25	0.0000
$x_i$	-43.204	-14.86	0.0000
$x_i^2$	61.436	22.06	0.0000
Coeficiente de correlación Ajustado = 0.9312		Durbin-Watson: 0.615	

Como indica el Coeficiente de Correlación Ajustado el modelo de grado dos proporciona un ajuste casi tan bueno como el de grado cinco pero con la ventaja de que los parámetros del modelo son significativos ya que el problema de multicolinealidad no se presenta como al ajustar un polinomio de grado mayor. Por otra parte, del resultado obtenido para el contraste de Durbin-Watson se obtiene 0.615 valor muy inferior al 1.60 donde se acepta la hipótesis de independencia de los residuos al nivel de significación del 0.05. Además el contraste de Box-Ljung, cuyo valor es 118.58 nos confirma la dependencia de los residuos, lo que gráficamente puede observarse en la figura 7 en la que se ha representado la gráfica de la función de autocorrelación de los residuos.



**Figura 7.**

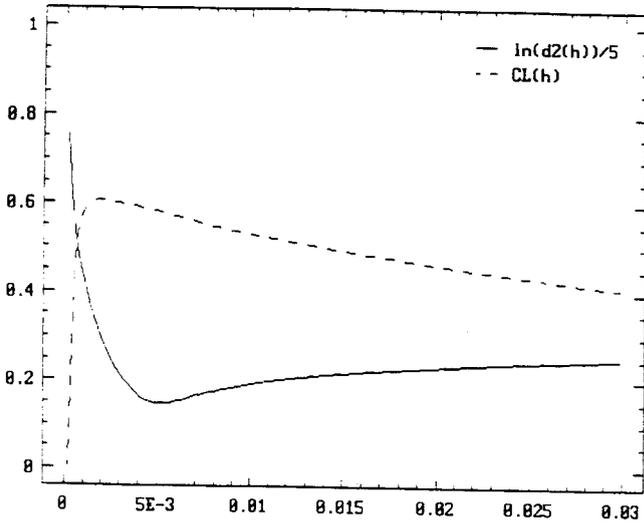
Función de autocorrelación de los residuos al ajustar un polinomio de grado dos a la Tasa de Cobertura Bruta.

Para realizar un análisis mediante el test asintótico (3.1), con el que se contrasta la hipótesis:

$$H_0: "m_\theta \text{ es un polinomio de grado menor o igual que dos}"$$

se ha calculado el valor del estadístico  $d^2(h)$  para valores del parámetro de suavizado,  $h$ , desde 0.0002 hasta 0.0300 con incrementos de 0.0002 suponiendo que los residuos siguen un modelo MA(20) y se ha calculado el nivel crítico  $(CL(h))$  asociado a los valores del estadístico.

En la figura 8 se han representado las dos funciones obtenidas,  $d^2(h)$  reescalada de manera adecuada y  $CL(h)$ .



**Figura 8.**

Funciones  $\ln(d^2(h))/5$  ( $d^2(h)$ : valor del estadístico) y  $CL(h)$  (Nivel crítico asociado al valor del estadístico).

De los resultados obtenidos se deduce que para un amplio intervalo de valores de  $h$  se aceptaría la hipótesis  $H_0$  con un alto nivel de significación y si deseamos apoyar la hipótesis  $H_0$  elegiríamos el  $h$  que minimiza la función  $d^2(h)$  que coincide, aproximadamente, con el máximo de la función  $CL(h)$ . En nuestro estudio esto ocurre para el valor de  $h = 0.0052$ . Para este valor de  $h$  y otros próximos se ha calculado el valor del estadístico, el del estadístico estandarizado en el supuesto de dependencia de los residuos, estimando las autocovarianzas según Müller-Stad Müller y el valor del estadístico estandarizado suponiendo estructura de independencia en los errores y en el que  $\sigma_\epsilon^2$  es estimada de forma estándar mediante la suma de residuos al cuadrado ajustada por el número de parámetros. Finalmente, se han calculado los niveles de significación asociados a los dos estadísticos tipificados. Los resultados obtenidos se recogen en la tabla 6.

**Tabla 6**

$H_0$ : " $m_\theta$  es un polinomio de grado  $\leq 2$ "

Banda: $h$	Estadístico: $d^2$	Dependencia		Independencia	
		Estad. tip.	Niv. Sign.	Est. tip.	Niv. Sig.
0.0010	75.215	-0.199	0.579	-14.102	1
0.0030	8.387	-0.254	0.601	-7.963	1
<b>0.0052</b>	<b>4.315</b>	<b>-0.199</b>	<b>0.579</b>	<b>2.905</b>	<b>0.002</b>
0.0070	5.165	-0.146	0.559	10.458	0
0.0090	6.332	-0.095	0.538	16.028	0
0.0110	7.262	-0.053	0.522	20.449	0

De los resultados obtenidos se deduce que bajo la hipótesis de dependencia de los residuos se aceptaría la hipótesis de que  $m_\theta$  es un polinomio de grado menor o igual que dos para distintos valores de  $h$  y claramente para el  $h$  que minimiza el estadístico  $d^2$ , sin embargo si utilizásemos la hipótesis de independencia de los residuos se rechazaría la hipótesis  $H_0$  para la mayoría de los valores de  $h$  aceptándose solamente para valores muy pequeños de  $h$ . Esto pone de manifiesto la importancia de considerar la posible estructura de dependencia de los residuos así como la sensibilidad del estadístico respecto al parámetro  $h$ , donde mucha teoría sobre su elección queda todavía por desarrollar.

Finalmente, comentar que el polinomio estimado, de grado dos, por mínima distancia, para  $h = 0.0052$  viene dado por:

$$m(x) = 34.89 - 36.06x + 52.67x^2$$

los residuos de este ajuste pueden modelarse adecuadamente, según la metodología de Box-Jenkins por un AR(1) de ecuación:

$$\epsilon_t = 0.29\epsilon_{t-1} + e_t$$

## 5. ANEXO. DEMOSTRACIONES

En este apartado se realiza una descripción, no detallada, de las demostraciones de los apartados a) y b) del teorema 1, que son bastante laboriosas y

técnicas. Se basan en la utilización del siguiente teorema de Nieuwenhuis (1992) sobre la normalidad asintótica de disposiciones triangulares con parte principal  $m(n)$ -dependiente.

## Teorema 2

Considérese la disposición triangular  $\{X_{\tau n}\}_{\tau=1}^n$  que admite la siguiente descomposición:  $X_{\tau n} = X_{\tau n, m(n)} + \bar{X}_{\tau n, m(n)}$ .

Siendo la parte principal  $(X_{\tau n, m(n)})$   $m(n)$ -dependiente Y  $(\bar{X}_{\tau n, m(n)})$  es la parte residual.

Denominemos  $B_n^2 = \text{Var} \left( \sum_{\tau=1}^n X_{\tau n} \right)$ . Y supongamos que las disposiciones triangulares  $\left( \frac{X_{\tau n}}{B_n} \right)$  y  $\left( \frac{\bar{X}_{\tau n, m(n)}}{B_n} \right)$  verifican, respectivamente, las condiciones  $C_1$  y  $C_1^*$ :

$$\text{Condición } C_1: \max_{i < j \leq n} \frac{1}{j-i} \text{Var} \left( \frac{1}{B_n} \sum_{k=i+1}^j X_{kn} \right) = O \left( \frac{1}{n} \right) \quad \text{cuando } n \rightarrow \infty$$

$$\text{Condición } C_1^*: \max_{i < j \leq n} \frac{1}{j-i} \text{Var} \left( \frac{1}{B_n} \sum_{k=i+1}^j \bar{X}_{kn, m(n)} \right) = o \left( \frac{1}{n} \right) \quad \text{cuando } n \rightarrow \infty$$

Y ambas satisfacen la condición  $C_2$  para algún  $\delta > 0$ .

$$\text{Condición } C_2: \max_{i < \tau \leq n} \frac{1}{B_n^{2+\delta}} \text{E} |Y_{\tau n}|^{2+\delta} = O \left( \frac{1}{n^{1+(\delta/2)}} \right) \quad \text{cuando } n \rightarrow \infty$$

(siendo  $Y_{\tau n} = x_{\tau n}$  o  $Y_{\tau n} = x_{\tau n, m(n)}$ ). Entonces:

$$\frac{1}{B_n} \sum_{\tau=1}^n (X_{\tau n} - \text{E} X_{\tau, n}) \xrightarrow{d} N(0, 1) \quad \text{cuando } n \rightarrow \infty$$

### *Demostración del Teorema 1.a)*

Dado que el estimador  $\hat{\theta}_n$  asociado a (2.2) viene dado por:

$$\hat{\theta}_n = B^{-1} \left( \sum_{\tau=1}^n \mathcal{A}(x_\tau) \omega(x_\tau) \hat{m}_n(x_\tau) \right) \quad \text{con } B = \sum_{j=1}^n \mathcal{A}(x_j) \mathcal{A}^t(x_j) \omega(x_j)$$

bajo la hipótesis  $H_0$  se verifica que:

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta) &= \sqrt{n} \left\{ B^{-1} \left( \sum_{i=1}^n \sum_{\tau=1}^n \mathcal{A}(x_i) \mathcal{A}^t(x_\tau) \omega(x_i) W_{n,\tau}(x_i) \theta \right) - \theta \right\} + \\ &+ \sqrt{n} \left\{ B^{-1} \left( \sum_{i=1}^n \sum_{\tau=1}^n \mathcal{A}(x_i) \omega(x_i) W_{n,\tau}(x_i) \epsilon_\tau \right) \right\} = \Delta_1 + \Delta_2 \end{aligned}$$

donde  $\Delta_1$  es el término debido al sesgo.

Bajo las hipótesis A1-A2 y A6. y usando las expresiones para el sesgo del estimador Gasser-Müller:

$\sum_{\tau=1}^n \mathcal{A}^t(x_\tau) W_{n,\tau}(x_i) - \mathcal{A}^t(x_i) = O(h^2)$ , de forma uniforme en  $i$ , en el soporte de  $\omega$ , y, por tanto:

$$\Delta_1 = \sqrt{n} \left\{ n B^{-1} \left( \frac{1}{n} \sum_{i=1}^n \mathcal{A}(x_i) \omega(x_i) \right) \right\} O(h^2) = O(\sqrt{nh^2}) \rightarrow 0$$

ya que  $nh^4 \rightarrow 0$ .

Por otro lado para  $\Delta_2$  se obtiene utilizando A6.,

$$\begin{aligned} \Delta_2 &= \sqrt{nb} B^{-1} \sum_{\tau=1}^n \left( \sum_{i=1}^n \mathcal{A}(x_i) \omega(x_i) W_{n,\tau}(x_i) \right) \epsilon_\tau = \\ &= \sqrt{nb} B^{-1} \sum_{\tau=1}^n \left( \mathcal{A}(x_\tau) \omega(x_\tau) + O(h^2) + O\left(\frac{1}{nh}\right) \right) \theta_\tau = \\ &= \sqrt{nb} B^{-1} \sum_{\tau=1}^n \mathcal{A}(x_\tau) \omega(x_\tau) \epsilon_\tau + o_{\mathbb{P}}(1) \end{aligned}$$

Por tanto basta con probar la normalidad asintótica de

$$\sqrt{nb} B^{-1} \sum_{\tau=1}^n \mathcal{A}(x_\tau) \omega(x_\tau) \theta_\tau$$

Para ello utilizando el Teorema 2 se demuestra la normalidad asintótica de

$$\begin{aligned} \sum_{\tau=1}^n c^t \mathcal{A}(x_\tau) \omega(x_\tau) \epsilon_\tau &= \sum_{\tau=1}^n d_{\tau n} \epsilon_\tau = \sum_{\tau=1}^n d_{\tau n} \left( \sum_{j=0}^{\infty} b_j e_{\tau-j} \right) = \\ (5.1) \quad &= \sum_{\tau=1}^n X_{\tau n}, \quad \forall c \in \mathbb{R}^q \end{aligned}$$

Realicemos la siguiente descomposición  $X_{\tau n} = X_{\tau n, m(n)} + \bar{X}_{\tau n, m(n)}$ , siendo

$$X_{\tau n, m(n)} = \sum_{j=0}^{m(n)} d_{\tau n} b_j e_{\tau-j} \quad (\text{parte principal})$$

$$\bar{X}_{\tau n, m(n)} = \sum_{j=m(n)+1}^{\infty} d_{\tau n} b_j e_{\tau-j} \quad (\text{parte residual})$$

Calculemos, en primer lugar,  $B_n^2$

$$B_n^2 = \text{Var} \left( \sum_{\tau=1}^n X_{\tau n} \right) = \sum_{\tau=1}^n \sum_{\tau'=1}^n d_{\tau} d_{\tau'} \gamma(\tau - \tau') = c^t \sum_{j=1}^n \sum_{\ell=j-n}^{n-1} a_{j\ell} c$$

Y utilizando A3. y el teorema de la convergencia dominada se obtiene,

$$(5.2) \quad \frac{B_n^2}{n} \longrightarrow c^t \sum_{n=-\infty}^{\infty} \gamma(n) R(n) c$$

Se prueba a continuación que las sucesiones  $\left\{ \frac{X_{\tau n}}{B_n} \right\}$  y  $\left\{ \frac{\bar{X}_{\tau n, m(n)}}{B_n} \right\}$  verifican las condiciones del teorema 2.

$$\boxed{(C_1)} \quad \frac{1}{j-i} \text{Var} \left( \sum_{k=i+1}^j \frac{X_{\tau n}}{B_n} \right) \leq \frac{1}{j-i} \sum_{k=i+1}^j \sum_{k'=i+1}^j |\gamma(k-k')| O \left( \frac{1}{B_n^2} \right) \leq$$

$$\leq \sum_{h=-(j-i-1)}^{j-(i+1)} \left[ 1 - \frac{|h|}{j-i} \right] |\gamma(h)| O \left( \frac{1}{B_n^2} \right)$$

y dado que cuando  $(j-i) \rightarrow \infty$  se sigue que

$$\sum_{h=-(j-i-1)}^{j-(i+1)} \left[ 1 - \frac{|h|}{j-i} \right] |\gamma(h)| \longrightarrow \sum_{h=-\infty}^{\infty} |\gamma(h)|$$

se obtiene que cuando  $n \rightarrow \infty$

$$\max_{1 < j \leq n} \frac{1}{j-i} \text{Var} \left( \sum_{k=i+1}^j \frac{X_{kn}}{b_n} \right) = O \left( \frac{1}{B_n^2} \right) = O \left( \frac{1}{n} \right)$$

$$\begin{aligned}
\boxed{(C_1^*)} \left| \frac{1}{j-i} \text{Var} \left( \sum_{k=i+1}^j \frac{\bar{X}_{kn,m(n)}}{B_n} \right) \right| &\leq \\
&\leq \sum_{h \in \mathbb{Z}} \left| \text{Cov} \left( \sum_{j=m(n)+1}^{\infty} b_j e_{1-j}, \sum_{j=m(n)+1}^{\infty} b_j e_{1+h-j} \right) \right| O(1) \leq \\
&\leq \sigma_\varepsilon^2 O(1) \sum_{h \in \mathbb{Z}} \sum_{j > m(n)} \sum_{\substack{\ell > m(n) \\ \{\ell - j = h\}}} |b_j| |b_\ell| = \sigma_\varepsilon^2 O(1) \left( \sum_{k > m(n)} |b_k| \right)^2 \rightarrow 0
\end{aligned}$$

cuando  $n \rightarrow \infty$ , , uniformemente en  $i < j \leq n$ .

$\boxed{(C_2)}$  Aplicando el Lema de Fatou, la desigualdad de Jensen y A4., se obtiene

$$E|X_{\tau_n}|^{2+\delta} \leq O(1) \underline{\lim} E \left( \left| \sum_{j=0}^{m(n)} |b_j| |e_{\tau-j}| \right|^{2+\delta} \right) < \infty$$

Y, dado que,  $\frac{1}{B_n^{2+\delta}} = O\left(\frac{1}{n^{1+\delta/2}}\right)$  se obtiene que la sucesión  $\left\{ \frac{X_{\tau_n}}{B_n} \right\}$  verifica la condición  $C_2$ . Y utilizando la desigualdad de Minkowsky se obtiene que dicha condición también la verifica la sucesión  $\left\{ \frac{\bar{X}_{\tau_n, m(n)}}{B_n} \right\}$ .

De todo lo anterior se sigue que  $\frac{1}{B} \sum_{\tau=1}^n X_{\tau_n}$  converge en distribución a una  $N(0,1)$ . Ahora, aplicando el teorema de Cramer-Wold se sigue la conclusión del teorema 1.a). ■

*Demostración del Teorema 1.b)*

Por el resultado anterior, Teorema 1.a) se verifica:

$$\begin{aligned}
d^2 \left( \hat{m}_n, \mathcal{A}^t(-)\hat{\theta}_n \right) &= \frac{1}{n} \sum_{j=1}^n \left( \sum_{i=1}^n W_{ni}(x_j) Y_i - \mathcal{A}^t(x_j)\theta + O_{\mathbb{P}}(n^{-1/2}) \right)^2 \omega(x_j) = \\
&= \frac{1}{n} \sum_{j=1}^n \left( \sum_{i=1}^n W_{ni}(x_j) Y_i - \mathcal{A}^t(x_j)\theta \right)^2 \omega(x_j) +
\end{aligned}$$

$$\begin{aligned}
& + \frac{2}{n} \sum_{j=1}^n \left( \sum_{i=1}^n W_{ni}(x_j) Y_i - \mathcal{A}^t(x_j) \theta \right)^2 \omega(x_j) O_{\mathbb{P}}(n^{-1/2}) + O_{\mathbb{P}}(n^{-1/2}) = \\
& = \Delta_1 + \Delta
\end{aligned}$$

donde  $\Delta$  engloba a los dos últimos sumandos y será un término despreciable para el estudio asintótico.

Por lo que respecta a  $\Delta_1$ :

$$\begin{aligned}
\Delta_1 & = \frac{1}{n} \sum_{j=1}^n b^2(\theta, x_j) \omega(x_j) + \sum_{i=1}^n b_{nii} \epsilon_i^2 + 2 \sum_{\tau < k} b_{n\tau k} \epsilon_{\tau} \epsilon_k + \\
& + \frac{2}{n} \sum_{j=1}^n b(\theta, x_j) \left( \sum_{i=1}^n W_{ni}(x_j) \epsilon_i \right) \omega(x_j) = \Delta_{11} + \Delta_{12} + \Delta_{13} + \Delta_{14}
\end{aligned}$$

donde

$$b(\theta, x_j) = \sum_{i=1}^n W_{ni}(x_j) \mathcal{A}^t(x_i) \theta - \mathcal{A}^t \theta$$

y

$$b_{n\tau k} = \frac{1}{n} \sum_{j=1}^n W_{n\tau}(x_j) W_{nk}(x_j) \omega(x_j)$$

Bajo las hipótesis A1., A2. y A6.,

$$(5.3) \quad \sqrt{n^2 h} \Delta_{11} = O(nh^{9/2}) \longrightarrow 0,$$

$$E(\Delta_{12}) = \frac{\sigma_{\epsilon}^2}{nh} \int K^2 \bar{\omega} + O\left(\frac{1}{n}\right)$$

Además, por A5.,

$$\text{Var}(\Delta_{12}) = 2 \sum_i \sum_j b_{nii} b_{njj} \gamma^2(i-j)$$

Por otro lado y teniendo en cuenta que las hipótesis sobre  $K$  permiten acotar las diferencias entre los pesos del tipo Gasser-Müller y Priestley-Chao y utilizando sumas de Riemann se obtiene:

$$\text{Var}(\Delta_{12}) \simeq 2 \cdot \frac{1}{n^4 h^2} \sum_i \sum_j \omega(x_i) \omega(x_j) \gamma^2(i-j) \left( \int K^2 \right)^2 \simeq O\left(\frac{1}{n^3 h^2}\right)$$

De esto y como consecuencia de (5.3), se sigue:

$$(5.4) \quad \Delta_{12} = \frac{\sigma_\epsilon^2}{nh} \int K^2 \bar{\omega} + O_{\mathbb{P}} \left( \frac{1}{n^{(3/2)h}} \right)$$

Por lo que respecta a  $\Delta_{13}$ , con razonamientos similares a los utilizados anteriormente se obtiene:

$$(5.5) \quad \text{Var}(\Delta_{13}) \simeq \frac{2}{n^2 h} \left( \sum_{k=-\infty}^{\infty} \gamma(k) \right)^2 (K * K)^2 \int \omega^2 = \sigma_d^2$$

Finalmente para  $\Delta_{14}$ , aplicando la desigualdad de Hölder se obtiene:

$$\text{Var}(\Delta_{14}) = \mathbb{E} \left( \frac{2}{n} \sum_{j=1}^n b(\theta, x_j) \left( \sum_{i=1}^n W_{ni}(x_j) \epsilon_i \right) \omega(x_j) \right)^2 \leq O(h^4) \sum_i \sum_\tau b_{ni\tau} \gamma(i - \tau)$$

Y utilizando aproximaciones de Riemann se obtiene:

$$\begin{aligned} \text{Var}(\Delta_{14}) &\simeq O(h^4) \sum_i \sum_\tau \frac{1}{n^2 h} \left( (K * K) \left( \frac{x_\tau - x_i}{h} \right) \right) \gamma(i - \tau) \omega(x_i) \leq \\ &\leq O(h^4) \sum_i \sum_\tau \frac{1}{n^2 h} \cdot \gamma(i - \tau) = O \left( \frac{h^4}{nh} \right) \end{aligned}$$

por la sumabilidad de  $\gamma$  y acotación de  $K$  y  $\omega$ . De este modo:

$$\sqrt{n^2 h} \Delta_{14} = O_{\mathbb{P}}(\sqrt{nh^2}) \rightarrow 0 \quad \text{dado que } nh^4 \rightarrow 0$$

En lo que sigue se obtiene la distribución asintótica de

$$\sqrt{n^2 h} \left[ d^2 \left( \hat{m}_n, \mathcal{A}^t(-) \hat{\theta}_n \right) - \frac{\bar{\omega} \sigma_\epsilon^2 \int K^2}{nh} \right]$$

que por los cálculos realizados es la misma que la de  $\sqrt{n^2 h} \Delta_{13}$ .

Obsérvese además que  $\sqrt{n^2 h} \Delta = O_{\mathbb{P}}(\sqrt{h}) = o_{\mathbb{P}}(1)$ .

Teniendo en cuenta que

$$\sum_{j=1}^n W_{n\tau}(x_j) W_{nk}(x_j) \omega(x_j) \simeq \frac{1}{n^2 h} \left( (K * K) \left( \frac{x_\tau - x_k}{h} \right) \right) \omega(x_\tau)$$

sin más que aplicar aproximaciones de Riemann, se deduce que la distribución asintótica de  $\Delta_{13}$  es la misma que la de:

$$\begin{aligned} \frac{1}{n^2 h} \sum_{\tau \neq k} \left( (K * K) \left( \frac{x_\tau - x_k}{h} \right) \right) \omega(x_\tau) \epsilon_\tau \epsilon_k &= \\ &= \frac{1}{n^2 h} \sum_{\tau=1}^n \left\{ \sum_{\substack{k \\ \tau \neq k}} \left( (K * K) \left( \frac{x_\tau - x_k}{h} \right) \right) \epsilon_k \right\} \omega(x_\tau) \epsilon_\tau = \sum_{\tau=1}^n X_{\tau n} \end{aligned}$$

Nuevamente, se obtiene la distribución asintótica de expresión utilizando el resultado de Nieuwenhuis (1992) (Teorema 2).

Para ello, denotemos  $\tilde{\epsilon}_i = \sum_{j=0}^{\tilde{m}(n)} b_j e_{i-j}$ , entonces  $X_{\tau n}$  se puede descomponer en  $X_{\tau n, m(n)} + \bar{X}_{\tau n, m(n)}$  siendo:

$$X_{\tau n, m(n)} = \frac{1}{n^2 h} \sum_{\tau \neq k} \left( (K * K) \left( \frac{x_\tau - x_k}{h} \right) \right) \omega(x_\tau) \tilde{\epsilon}_\tau \tilde{\epsilon}_k \quad (\text{parte principal})$$

$$\bar{X}_{\tau n, m(n)} = \frac{1}{n} \sum_{\tau \neq k} \left( (K * K) \left( \frac{x_\tau - x_k}{h} \right) \right) \omega(x_\tau) (\epsilon_\tau \epsilon_k - \tilde{\epsilon}_\tau \tilde{\epsilon}_k) \quad (\text{parte residual})$$

Eligiendo  $\tilde{m}(n) = o(nh)$ , se tiene que

$$\frac{[nh + \tilde{m}(n)]}{n} \longrightarrow 0 \quad \Longleftrightarrow \quad n^{1+(2/j)} h^{2+(2/j)} \longrightarrow 0$$

condición que se cumple por la hipótesis A6.

La demostración de que las disposiciones triangulares así definidas verifican las condiciones  $C_1$ ,  $C_1^*$  y  $C_2$  del teorema 2, concluyen la demostración del teorema 1.b). ■

## 5. BIBLIOGRAFÍA

- [1] **Altman, N.S.** (1990). "Kernel smoothing of data with correlated errors". *J.A.S.A.*, **85**, **411**, 739–749.
- [2] **Brockwell, P.J. & Davis, R.A.** (1991). *Time series theory and methods*. 2<sup>a</sup> ed. Springer.
- [3] **Chu, C.K. & Marron, S.** (1991). "Comparison of two bandwidth selectors with dependent errors". *The Annals of Stat.*, **19**, **4**, 1906–1918.
- [4] **Eubank, R.L. & Spiegelman, C.H.** (1990). "Testing the goodness of fit of a linear model via nonparametric regression techniques". *J.A.S.A.*, **85**, **410**, 387–392.
- [5] **Eubank, R.L. & Hart, J.D.** (1992). "Commonality of cusum, Von Neumann and smoothing based goodness of fit tests". *Biometrika*, **80**, **1**, 89–98.
- [6] **Eubank, R.L. & Hart, J.D.** (1993). "Testing goodness-of-fit in regression via order selection criterio". *Annals of Stat.*, **20**, **3**, 1412–1425.
- [7] **Firth, D., Glosup, J. & Hinkley, D.** (1991). "Model checking with nonparametric curves". *Biometrika*, **78**, **2**, 245–252.
- [8] **Gasser, T. & Muller, H.G.** (1979). "Kernel estimation of regression functions in Smoothing Techniques for Curve Estimation". *Lect. Notes in Math Springer*, 23–68.
- [9] **González Manteiga, W. & Cai Abad, R.** (1991). "Testing hypothesis of general linear model using nonparametric regression estimation". (Aparecerá en *Test*.)
- [10] **Hart, J.D. & Wehrly, T.E.** (1992). "Kernel regression when the boundary region is large, with an application to testing the adquacy of polynomial models". *J.A.S.A.*, **87**, **420**, 1018–1024.
- [11] **Kozek, A.S.** (1991). "A nonparametric test of fit of a parametric model". *Journal of Multiv. Analysis*, **37**, 66–75.
- [12] **Muller, H.G. & Statmuller, U.** (1988). "Detecting dependencies in smooth regression models". *Biometrika*, **75**, **4**, 639–650.
- [13] **Nieuwenhuis, G.** (1992). "Central limit theorems for sequences with  $m(n)$ -dependent main part". *Journal of Stat. Planning and Inference*, **32**, 229–241.
- [14] **Peña Sánchez de Rivera, D.** (1987). *Estadística, Métodos y Modelos*. Vol. 2, Alianza Universidad Textos.
- [15] **Priestley, M.B. & Chao, M.T.** (1972). "Nonparametric function fitting". *J. Roy. Stat. Soc. B*, 385–392.
- [16] **Raz, J.** (1990). "Testing for no effect when estimating a smooth function by nonparametric regression: A randomization approach". *J.A.S.A.*, **85**, **409**, 132–138.

- [17] **Seber, G.** (1977). *Linear Regression Analysis*. John Wiley.
- [18] **Seber, G. & Wild** (1989). *Nonlinear Regression*. John Wiley.
- [19] **Staniswalis, J. & Severini, T.** (1991). "Diagnostics for assessing regression models". *J.A.S.A.*, **86**, 415, 684–692.

## ENGLISH SUMMARY:

### TESTING THE HYPOTHESIS OF A GENERAL LINEAR MODEL WHEN ERRORS ARE NON-INDEPENDENT

Vilar Fernández, J.M. and González Manteiga, W.

#### 1. INTRODUCTION

One of the most extensively studied questions in the theory of parametric regression models:

$$(1.1) \quad Y_i = m_\theta(x_{1i}, \dots, x_{pi}) + \epsilon_i, \quad i = 1, \dots, n; \quad \theta \in \Theta \subset \mathbb{R}^q$$

with  $x_i \in \mathcal{C}$ ,  $i = 1, \dots, n$ , a compact set in  $\mathbb{R}$ , and  $\{\epsilon_i\}_{i=1}^n$  a sequence of random variables with zero mean, is how to determine whether a model  $m_\theta$  is better supported by the data than another.

In this paper we have designed a new method for testing the hypothesis that the regression function follows a general linear model,

$$H_0: m \in \{m_\theta(\bullet) = \mathcal{A}^t(\bullet)\theta: \theta \in \Theta \subset \mathbb{R}^q\},$$

with  $\mathcal{A}$  a function from  $\mathbb{R}$  to  $\mathbb{R}^q$ , against the alternative

$$(1.2) \quad H_1 = \text{"}m \text{ belongs to one class of smooth functions."}$$

The statistic used for testing the given hypothesis ( $d^2$ ) is defined to be the difference between the average squared errors (ASE) associated with the non-parametric estimator  $\hat{m}$  of  $m$  the minimum distance estimators  $m_{\hat{\theta}}$  of  $m$ .

The asymptotic normality of both  $d^2$  and  $\hat{\theta}_n$  are obtained under general conditions and the random errors  $\{\epsilon_i\}$  have an MA( $\infty$ ) structure. Simulations and examples with real data show how the non independence of errors can affect decision on  $H_0$ .

## 2. DEFINITIONS AND ASYMPTOTIC RESULTS

We consider the fixed design model  $Y_i = m(x_i) + \epsilon_i$ , ( $i = 1, \dots, n$ ) where  $m$  is a smooth function defined in  $[0,1]$ , the  $\{x_i\}_{i=1}^n$  are equally spaced ( $x_i = i/n$ ), the  $Y_i$  are observations obtained at the  $x_i$ , and the errors  $\epsilon_i$  are generated by an infinite order moving average:  $\epsilon_i = \sum_{j=0}^{\infty} b_j e_{i-j}$ , where  $\{\epsilon_i\}$  is zero mean white noise of variance  $\sigma_e^2$ .

The functional distance we shall use to test  $H_0: m \in \{m_\theta = \mathcal{A}^t(\cdot)\theta | \theta \in \Theta\}$  is a Crámer-von-Mises type distance given by:

$$\begin{aligned}
 d^2(\hat{m}_n, H_0) &= \min_{\theta \in \Theta} \int (\hat{m}_n(x) - \mathcal{A}^t(x)\theta)^2 \omega(x) d\Omega_n(x) = \\
 (2.1) \qquad &= \frac{1}{n} \sum_{i=1}^n \left( \hat{m}_n(x_i) - \mathcal{A}^t(x_i)\hat{\theta}_n \right)^2 \omega(x_i)
 \end{aligned}$$

where  $\omega$  is a weighing function and  $\Omega_n$  empirical distribution of the points of the design. As non-parametric estimator  $\hat{m}_n$  we use a Gasser-Müller estimator:

$$\hat{m}_n(x) = \sum_{j=1}^n h^{-1} \left\{ \int_{s_{j-1}}^{s_j} K\left(\frac{x-s}{h}\right) ds \right\} Y_j = \sum_{j=1}^n W_{n,j}(x) Y_j$$

The estimator  $\hat{\theta}_n$  obtained by the minimization of the function:  $\Psi(\theta) = \int (\hat{m}_n(x) - \mathcal{A}^t(x)\theta)^2 \omega(x) d\Omega_n(x)$  becomes the classical least squares estimator  $\hat{\theta}_{ls}$  if a degenerate smoothing parameter ( $h_n$ ) is used, being, in general, the probability of rejecting  $H_0$  smaller for  $\hat{\theta}_n$  than for  $\hat{\theta}_{ls}$ , since as a rule

$$d^2(\hat{m}_n, \mathcal{A}^t(-)\hat{\theta}_n) < d^2(\hat{m}_n, \mathcal{A}^t(-)\hat{\theta}_{ls})$$

Like  $\hat{\theta}_{ls}$ ,  $\hat{\theta}_n$  is a “**root-n**” estimator for  $\theta$  if  $H_0$  holds, i.e.

$$n^{1/2} (\hat{\theta}_n - \theta) = O_{\mathbb{P}}(1).$$

This is a consequence (part a) of the following theorem, which we prove with the help of the Central Limit Theorems for sequences with  $m_n$ -dependent main part due to Nieuwenhuis (1992).

**Theorem 1**

Under assumption  $H_0$  and subject to general premises

a)  $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma_\epsilon^2 \mathbf{Q}^{-1} \mathbf{Q}^* \mathbf{Q}^{-1})$

where  $\mathbf{Q} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathcal{A}(x_i) \mathcal{A}^t(x_i) \omega(x_i)$

and  $\mathbf{Q}^* = \sum_{u=-\infty}^{\infty} \gamma(u) R(u)$ ,

with  $R(u) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^{n-u} \mathcal{A}(x_j) \mathcal{A}^t(x_{j+u}) \omega^2(x_j) \omega^2(x_{j+u})$

b)  $\sqrt{n^2 h} \left( d^2(\hat{m}, \mathcal{A}^t(\cdot) \hat{\theta}_n) - \frac{\bar{\omega} \sigma_\epsilon^2 \int K^2}{nh} \right) \xrightarrow{d} N(0, \sigma_d^2)$

where  $\sigma_d^2 = 2 \left( \sum_{k=-\infty}^{\infty} \gamma(k) \right)^2 \int (K * K)^2 \int \omega^2$

and  $\gamma$  is the autocovariance function of the errors.

Note that for independent errors (i.e. when  $\gamma(k) = 0$  for all non-zero  $k$ ,  $\sigma_d^2 = 2\sigma_\epsilon^4 \int (K * K)^2 \int \omega^2$ , which is González-Manteiga and Cao Abad's (1994) result.

**3. SIMULATION STUDIES**

The Theorem 1.a) shows that the variance of the least distance estimator  $\hat{\theta}_n$  is asymptotically larger if the sample errors are positively correlated than if they are uncorrelated. It is also clear from Th.1.b) that ignoring positive correlation when it exists can lead to  $H_0$  being unduly rejected. This behaviour is illustrated by the results of a first simulation study.

To study the behaviour of test  $d^2$  as smoothing parameter  $h$  is varied, we performed a second simulation experiments, and we observed that the choice of bandwidth is crucial and the simulation study suggest that a good selection for  $h$ , under  $H_0$  consists in taking the first minimum of the optimization:

$$\min_h \left[ d^2 \left( \hat{m}_{n,h}, \mathcal{A}^t(-)\hat{\theta}_{n,h} \right) \right]$$

suggestion made above works well under  $H_0$ . But this is not a good selection under other alternatives. More research is necessary in order to investigate the selection of  $h$  under other alternatives.

The influence of correlation among the errors is also illustrated by analysis of a real example studied in the section 4. And, finally, the Appendix is devoted to a sketch of the proof of theorem 1.