QÜESTIIÓ, vol. 24, 2, p. 267-291, 2000

SELECCIÓN DE LA VENTANA EN SUAVIZACIÓN TIPO NÚCLEO DE LA PARTE NO PARAMÉTRICA DE UN MODELO PARCIALMENTE LINEAL CON ERRORES AUTORREGRESIVOS⋆

GERMÁN ANEIROS-PÉREZ*
Universidad de La Coruña

Supongamos que $y_i = \zeta_i^T\beta + m(t_i) + \varepsilon_i$, $i = 1, \ldots, n$, *donde el vector* $(p\times 1)$ β *y la función* $m(\cdot)$ *son desconocidos, y los errores* ε_i *provienen de un proceso autorregresivo de orden uno (AR(1)) estacionario. Discutimos aquí el problema de la selección del parámetro ventana de un estimador tipo núcleo de la función* $m(\cdot)$ *basado en un estimador Generalizado de Mínimos Cuadrados de* β. *Obtenemos la expresión asintótica de una ventana óptima y proponemos un método para estimarla, de modo que dé lugar a un estimador óptimo de* $m(\cdot)$. *Los resultados obtenidos generalizan aquellos obtenidos por Quintela (1994b) en regresión no paramétrica.*

Bandwidth selection in kernel smoothing of the nonparametric part in a partial linear model with autoregressive errors

Palabras clave: modelos parcialmente lineales, suavización núcleo, selección de la ventana, series de tiempo

Clasificación AMS (MSC 2000): 62G07, 62G20, 62M10

⋆ Investigación financiada por la Xunta de Galicia (España) bajo el Proyecto de investigación XUGA 10503A98.

* Departamento de Matemáticas. Facultad de Informática. Universidad de La Coruña. Campus de Elviña, s/n. La Coruña. Spain. E-mail: ganeiros@udc.es

– Recibido en octubre de 1999.
– Aceptado en marzo de 2000.

1. INTRODUCCIÓN

Supondremos que tenemos un conjunto de datos observados $\{(y_i, \zeta_i^T, t_i)^T\}_{i=1}^n$ de acuerdo al modelo

$$(1) \qquad y_i = \zeta_i^T \beta + m(t_i) + \varepsilon_i \quad i = 1, \ldots, n,$$

donde $(\zeta_i^T, t_i) = ((x_{i1}, x_{i2}, \ldots, x_{ip}), t_i) \in A \times [0,1] \subset \mathbb{R}^p \times \mathbb{R}$ son puntos del diseño fijos, β es un vector $(p \times 1)$ de parámetros desconocidos, $m(\cdot)$ es una función real desconocida definida en $[0,1]$ (sin pérdida de generalidad) y $\{\varepsilon_i\}_{i=1}^n$ es una muestra de errores idénticamente distribuidos.

El anterior modelo es un modelo semiparamétrico conocido como un modelo «parcialmente lineal». Es a la vez una generalización de un modelo de regresión lineal múltiple y una restricción de un modelo no paramétrico puro (ambos con $p+1$ variables regresoras), con lo que combinará la flexibilidad de los modelos no paramétricos con la sencillez y buena interpretación de los modelos paramétricos. Fue propuesto por Engle y otros (1986) para tratar de explicar la relación entre el tiempo y el consumo de electricidad. A partir de aquí, diversos estudios de distintos estimadores de β y $m(.)$ fueron hechos, entre los que cabe citar los trabajos de Heckman (1986) y Rice (1986) referentes a suavización spline, así como el trabajo de Speckman (1988) referente a suavización núcleo. Si nos centramos en estimadores del parámetro β, destacaríamos el trabajo de Linton (1995) quien, basándose en suavización por medio de polinomios locales, propone un estimador de Mínimos Cuadrados para β. Una vez estandarizado, obtiene aproximaciones de segundo orden para sus dos primeros momentos, las cuales son posteriormente utilizadas para obtener la expresión de una ventana asintóticamente óptima. Además, propone un método para estimar dicha ventana.

En todos los trabajos anteriormente citados se supone que los errores ε_i son i.i.d.. En muchos casos, esta suposición no es realista pues, por ejemplo, es muy posible que las observaciones sean obtenidas secuencialmente en el tiempo. Es por tanto interesante el disponer de estudios en los que los errores verifiquen alguna condición de dependencia. En este campo podemos citar el trabajo de Gao (1995), que presupone que los errores provienen de un proceso lineal, así como el de Schick (1996), que trabaja con un modelo AR(1) en los errores. Ambos estudios se centran en estimadores de β. Aneiros y Quintela (1998, 1999a), bajo condiciones α-mixing (Doukhan (1994)) en la estructura de los errores, proponen y estudian estimadores para β y $m(\cdot)$, obteniendo propiedades de consistencia para ambos y de normalidad asintótica para el estimador de β. Además, proponen un método de validación cruzada modificada para la elección de la ventana común a utilizar por ambos estimadores, demostrando que da lugar a una ventana asintóticamente óptima. Posteriormente, Aneiros y Quintela (1999b), centrándose en el estimador de β y suponiendo para los errores un modelo AR(1), generalizan resultados de Linton (1995), obteniendo la expresión de una ventana asintóticamente óptima para dicho estimador.

El objetivo del presente trabajo es generalizar resultados de normalidad asintótica y expresiones de ventanas óptimas existentes en estimación no paramétrica pura tanto en el caso de errores independientes (Gasser y Müller (1984)) como en el caso de errores dependientes (Hart (1991), Quintela (1994a,b)) al caso de la estimación de $m(\cdot)$ en un modelo parcialmente lineal con errores dependientes. Concretamente, el estudio se hará con errores que provengan de un modelo AR(1), si bien es generalizable a distintos procesos lineales.

Nuestro trabajo está organizado de la siguiente manera: en la sección 2 presentamos el estimador, junto con las hipótesis empleadas y los resultados obtenidos. En la sección 3 probamos dichos resultados.

Unas palabras sobre notación. C denotará una constante genérica mayor que cero, que puede tomar diferentes valores de línea a línea e incluso de fórmula a fórmula. Para $\mathbf{u} = (u_1, \ldots, u_n)^T \in \mathbb{R}^n$, $||\mathbf{u}||^2$ significa $\sum u_i^2$. Acerca de la notación $o(\cdot)$, $O(\cdot)$, $o_p(\cdot)$ y $O_p(\cdot)$, véase Fuller (1976).

2. EL ESTIMADOR Y LOS RESULTADOS

En forma matricial, el modelo (1) puede expresarse como:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{m} + \varepsilon, \tag{2}$$

donde $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$, $\mathbf{X} = (\zeta_1, \ldots, \zeta_n)^T$ (siendo $\zeta_i^T = (x_{i1}, \ldots, x_{ip})$, $i = 1, \ldots, n$), $\mathbf{m} = (m(t_1), \ldots, m(t_n))^T$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$. Además, supondremos que existe una relación entre las variables regresoras del tipo:

$$x_{ij} = g_j(t_i) + \eta_{ij} \quad i = 1, \ldots, n \quad j = 1, \ldots, p, \tag{3}$$

con lo que a su vez tenemos que:

$$\mathbf{X} = \mathbf{G} + \eta, \tag{4}$$

donde $\mathbf{G} = (\mathbf{g}_1, \ldots, \mathbf{g}_p)$ (siendo $\mathbf{g}_j = (g_j(t_1), \ldots, g_j(t_n))^T$, $j = 1, \ldots, p$) y $\eta = (\eta_1, \ldots, \eta_n)^T$ (donde $\eta_i = (\eta_{i1}, \ldots, \eta_{ip})^T$, $i = 1, \ldots, n$). Relaciones como (3) junto con hipótesis acerca de η_{ij} pueden verse en Speckman (1988), Linton (1995) o Gao (1995), entre otros. Como hemos hecho notar en la introducción, consideraremos que la matriz $\mathbf{X}$ es fija, con lo que también lo será la matriz η. No obstante, si consideramos que η_{ij} son observaciones de variables aleatorias i.i.d. de media cero e independientes de ε_i, los resultados que obtendremos seguirán siendo válidos, entendiendo en este caso que las esperanzas se consideran respecto de la distribución de ε y condicionadas por η. En este caso, las hipótesis en las que intervenga η deben verificarse en probabilidad con respecto a η.

A lo largo de este trabajo, nos basaremos en estimadores tipo núcleo. Si consideramos el modelo (1) sin la componente lineal,

$$y_i = m(t_i) + \varepsilon_i, \tag{5}$$

un estimador tipo núcleo puede ser escrito de la forma:

$$m_{n,h}(t) = \sum_{i=1}^{n} w_{n,h}(t, t_i) y_i, \tag{6}$$

donde $w_{n,h}(\cdot, t_i)$ es una función de ponderación (que se escribe a través de una función $K(\cdot)$, el núcleo), y dependiente de un parámetro de escala (el parámetro ventana h). Es de sobras conocida la importancia que tiene la elección de dicho parámetro ventana a la hora de obtener buenas propiedades para el estimador (ver, por ejemplo, Quintela (1996) para una revisión de distintos métodos de selección del parámetro ventana, y una comparación de los mismos bajo condiciones de dependencia).

En este trabajo, nos centraremos en el estimador de Gasser y Müller (1979), y además consideraremos que los puntos t_i están igualmente separados. Por tanto, trabajaremos con pesos de la forma $w_{n,h}(t, t_i) = h^{-1} \int_{(i-1)/n}^{i/n} K(\frac{t-u}{h}) du$, siendo $h > 0$, $t_i = (i - 1/2)/n$ y K una función cuyo soporte es $[-1, 1]$. En el caso de querer obtener estimaciones en puntos de la frontera (entiéndase como frontera el conjunto $[0, h) \cup (1 - h, 1]$), el estimador de Gasser y Müller se verá afectado (su sesgo tiene orden distinto (superior) cuando estima en puntos de la frontera que cuando estima en el interior). Esto puede evitarse si, en estos casos, se trabaja con los llamados núcleos frontera. En concreto, si $t = qh \in [0, h)$ o $t = 1 - qh \in (1 - h, 1]$, donde $h < 1/2$, entonces en lugar de utilizar $K(\cdot)$ utilizaremos un núcleo frontera $K_q(\cdot)$ (ver Gasser y Müller (1979, 1984)). Las suposiciones acerca de estos núcleos serán especificadas después.

Supondremos que

$$E(\varepsilon) = 0, Var(\varepsilon) = E[\varepsilon\varepsilon^T] = \sigma_\varepsilon^2 \Psi,\ \Psi \neq \mathbf{I} \text{ y definida positiva.} \tag{7}$$

Debido a que Ψ es definida positiva, existe una matriz $n \times n$ $\mathbf{P}$ tal que $\mathbf{P}\Psi\mathbf{P}^T = \mathbf{I}$. Por tanto, $\mathbf{P}^T\mathbf{P} = \Psi^{-1}$. $\mathbf{P}$ no es única (Judge y otros (1985)).

Comentario 2.1

Para errores estacionarios autorregresivos de orden uno (AR(1)) ($\varepsilon_i = \rho\varepsilon_{i-1} + e_i$, donde $|\rho| < 1$ y los e_i son variables aleatorias i.i.d. con $E[e_i] = 0$ y $E[e_i^2] = \sigma_e^2$), tenemos que la matriz de correlaciones Ψ tiene la forma:

$$\Psi = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}, \tag{8}$$

y por lo tanto, su inversa:

$$\Psi^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & \ddots & 0 \\ 0 & -\rho & 1+\rho^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -\rho \\ 0 & 0 & \dots & -\rho & 1 \end{pmatrix}. \tag{9}$$

Elegimos (Judge y otros (1985)):

$$\mathbf{P} = \frac{1}{\sqrt{1-\rho^2}} \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & \ddots & 0 \\ 0 & -\rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -\rho & 1 \end{pmatrix}. \tag{10}$$

Si suponemos que Ψ es conocida y que $\mathbf{P}\widetilde{\mathbf{X}}$ tiene rango máximo, usando Mínimos Cuadrados Generalizados, podemos estimar β por (Aneiros y Quintela (1999a)):

$$\hat{\beta}_h = \left(\widetilde{\mathbf{X}}^T \Psi^{-1} \widetilde{\mathbf{X}}\right)^{-1} \widetilde{\mathbf{X}}^T \Psi^{-1} \widetilde{\mathbf{y}}, \tag{11}$$

siendo

$$\begin{aligned} \widetilde{\mathbf{y}} &= (\mathbf{I} - \mathbf{W})\mathbf{y}, \\ \widetilde{\mathbf{X}} &= (\mathbf{I} - \mathbf{W})\mathbf{X}, \end{aligned} \tag{12}$$

y $\mathbf{W}$ una matriz de suavización con elementos $\{w_{n,h}(t_i,t_j)\} = \{w_{ij}\}$ (hemos suprimido la dependencia de W con n y h). Ahora, a partir de (1), (5) y (6), parece intuitivo el proponer como estimador de $m(t)$:

$$\hat{m}_h(t) = \sum_{i=1}^{n} w_{n,h}(t,t_i)(y_i - \zeta_i^T \hat{\beta}_h). \tag{13}$$

Propiedades de consistencia de este estimador pueden verse en Aneiros y Quintela (1999a). Como hemos dicho antes, lo que vamos a hacer ahora es estudiar su normalidad asintótica, así como la expresión y estimación de una ventana global asintóticamente óptima. Para ello, algunas de las hipótesis que utilizaremos son:

2.1. Hipótesis

(K.1) **a)** $K(\cdot)$ es Hölder continua, simétrica en torno al cero, con soporte $[-1,1]$.

b) $\int K(u)du = 1$, $\int u^\nu K(u)du \neq 0$ y $\int u^z K(u)du = 0$, $z = 1, \ldots, \nu - 1$.

(K.2) **a)** Si $t = qh \in [0,h)$ ($t = 1 - qh \in (1-h,1]$), entonces $K_q(\cdot)$ tiene soporte $[-1,q]$ ($[-q,1]$) y satisface **(K.1.b)** con momento de orden ν uniformemente acotado en $q \in [0,1]$.

b) $K_q(\cdot)$ es tal que su varianza está uniformemente acotada en $q \in [0,1]$.

(D.1) Los puntos del diseño $\{t_i\}_{i=1}^n$ son fijos y están igualmente separados ($t_i = (i - 1/2)/n$).

(E.1) Los errores $\{\varepsilon_i\}_{i=1}^n$ provienen de un proceso estacionario $AR(1)$ dado por $\varepsilon_i = \rho\varepsilon_{i-1} + e_i$, $(|\rho| < 1)$, donde $\{e_i\}$ es una sucesión de variables aleatorias i.i.d., absolutamente continuas, con media cero y varianza finita.

(M.1) Las funciones $m(\cdot), g_1(\cdot), \ldots, g_p(\cdot)$ tienen $\nu \geq 2$ derivadas continuas en $[0,1]$.

(G.1) $n^{-1}\eta^T\Psi^{-1}\eta \longrightarrow \mathbf{V}$ donde $\mathbf{V} = \{V_{ij}\}$ es una matriz definida positiva.

(G.2) $||\mathbf{W}\eta_j||^2 = ||\mathbf{W}^T\eta_j||^2 = O(h^{-1})$, $1 \leq j \leq p$.

(G.3) $n^{-1}\eta^T\Psi^{-1}\widetilde{\mathbf{m}} = O(n^{-1/2}h^\nu)$.

(G.4) $||\mathbf{W}^T\Psi^{-1}\eta_j|| = O(ne(n))$, $1 \leq j \leq p$, donde $e(n) = c_1h^{2\nu} + c_2(nh)^{-1}$, siendo c_1 y c_2 constantes positivas.

(W.1) $w(\cdot)$ es una función continua no negativa con soporte compacto $[w_1, w_2] \subset (0,1)$, $w_1 < w_2$.

Otras hipótesis serán introducidas a medida que las vayamos necesitando.

Comentario 2.2

Las hipótesis **(G.1)-(G.4)** pueden parecer un tanto artificiales. Sin embargo, puede verse en Aneiros y Quintela (1999a) que son verificadas bajo las demás hipótesis, si suponemos que las filas de η son vectores aleatorios i.i.d. con media cero y matriz de varianzas-covarianzas finita. Hipótesis de este tipo pueden verse en Speckman (1988), y en Gao (1995) tenemos otras alternativas. Las demás condiciones son usuales en el conjunto de suavización tipo núcleo. **(K.2)** es suficiente para remediar efectos frontera (ver Gasser y Müller (1979), (1984)). La existencia de tales núcleos modificados para v arbitrario es establecida en Gasser y otros (1985). En cuanto a **(E.1)**, el presente trabajo podría generalizarse a otras estructuras paramétricas en los errores, tales como estructuras ARMA(p,q) si bien, debido a que la matriz de correlaciones forma parte del estimador, esto aumentaría considerablemente los cáculos.

Denotemos $\overline{m}_h(t) = \sum_{i=1}^{n} w_{n,h}(t,t_i)(y_i - \zeta_i^T \beta)$; es decir, $\overline{m}_h(t)$ sería el estimador de $m(t)$ suponiendo que conociésemos β.

Bajo las hipótesis **(D.1)**, **(E.1)**, **(K.1)** y **(M.1)** se tiene que, si $nh \to \infty$ y $h \to 0$ cuando $n \to \infty$, entonces (ver Hart (1991) para la expresión de la varianza y Gasser-Müller (1984) para la del sesgo):

(14)
$$E(\overline{m}_h(t) - m(t))^2 = \frac{1}{nh}\sigma_\varepsilon^2 \frac{1+\rho}{1-\rho} \int K^2(u)du + \frac{h^{2v}\alpha_v^2}{(v!)^2}(m^{(v)}(t))^2 + o((nh)^{-1} + h^{2v}),$$

uniformemente en $t \in [h, 1-h]$, donde $\alpha_v = \int u^v K(u)du$.

Por tanto, junto con las hipótesis **(K.2)**, **(G.1)-(G.4)**, si además $nh^{4v} \to 0$ y $nh^2 \to \infty$, se tiene que (Aneiros y Quintela (1999a)):

(15)
$$E(\hat{m}_h(t) - m(t))^2 = \frac{1}{nh}\sigma_\varepsilon^2 \frac{1+\rho}{1-\rho} \int K^2(u)du + \frac{h^{2v}\alpha_v^2}{(v!)^2}(m^{(v)}(t))^2 + o((nh)^{-1} + h^{2v}),$$

uniformemente en $t \in [h, 1-h]$.

Si consideramos como una medida global del error de estimación el Error Cuadrático Medio Integrado, a saber:

$$MISE(\hat{m}_h) = E\left[\int (\hat{m}_h(t) - m(t))^2 w(t)dt\right],$$

es fácil obtener, por medio de (15) y de la hipótesis **(W.1)**, que:

$$\begin{aligned} MISE(\hat{m}_h) &= \frac{1}{nh}\sigma_\varepsilon^2 \frac{1+\rho}{1-\rho} \int K^2(u)du \int w(u)du + \frac{h^{2v}\alpha_v^2}{(v!)^2} \times \\ &\quad \times \int (m^{(v)}(u))^2 w(u)du + o((nh)^{-1} + h^{2v}). \end{aligned} \tag{16}$$

Por tanto, suponiendo que

$$\int (m^{(\nu)}(u))^2 w(u)du \neq 0, \tag{17}$$

una ventana global asintóticamente óptima para el estimador $\hat{m}_h(\cdot)$ es

$$h_0 = C_0 n^{-\frac{1}{2\nu+1}} = \left[\frac{C_K \sigma_\varepsilon^2 \frac{1+\rho}{1-\rho}}{\int (m^{(\nu)}(u))^2 w(u)du} \right]^{\frac{1}{2\nu+1}} n^{-\frac{1}{2\nu+1}}, \tag{18}$$

donde $C_K = \frac{(\nu!)^2 \quad K^2(u)du \quad w(u)du}{2\nu\alpha_\nu^2}$. Como vemos, la expresión (18) coincide con la que se obtendría en el caso de conocer el valor de β, es decir, en el caso de trabajar con un modelo no paramétrico puro, lo cual era de esperar una vez conocido el trabajo de Aneiros y Quintela (1999a).

Supongamos ahora que además se verifica la hipótesis:

(E.2) Existe algún $\delta > 0$ tal que $E\left[|\varepsilon_i|^{2+\delta}\right] < \infty$.

Teniendo presente (14) y la normalidad asintótica de $\overline{m}_{h_0}(t)$ (Quintela (1994a)), se obtiene que, si $t \in (0,1)$, entonces:

$$n^{\nu/(2\nu+1)}(\overline{m}_{h_0}(t) - m(t)) \xrightarrow{D} N(B,V), \tag{19}$$

siendo

$$\begin{aligned} B &= \tfrac{(-1)^\nu}{\nu!} C_0^\nu \alpha_\nu m^{(\nu)}(t), \\ V &= \sigma_\varepsilon^2 \tfrac{1+\rho}{1-\rho} C_0^{-1} \int K^2(u)du. \end{aligned} \tag{20}$$

Quintela (1994b) demuestra que, bajo ciertas hipótesis, si consideramos una sucesión de ventanas $\left\{\widehat{h}_n\right\}$ verificando:

$$\frac{\widehat{h}_n}{h_0} \xrightarrow{P} 1, \tag{21}$$

entonces, si $t \in (0,1)$ y $\nu = 2$, se tiene que:

$$n^{\nu/(2\nu+1)}(\overline{m}_{\widehat{h}_n}(t) - m(t)) \xrightarrow{D} N(B,V). \tag{22}$$

Utilizando (19), sin más que cambiar $\nu = 2$ por un $\nu > 2$ genérico en dicha demostración de Quintela (1994b), se obtiene que (22) se mantiene. Basándonos en (19) y (22), demostraremos el siguiente resultado:

2.2. Resultados

Teorema 2.1

Bajo las hipótesis **(D.1)**, **(E.1)**, **(E.2)**, **(G.1)-(G.4)**, **(K.1)**, **(K.2)**, **(M.1)**, **(W.1)**, *junto con (17) y (21), si además se verifica:*

(D.2) *las componentes de* **X** *están uniformemente acotadas*

o

(D.2*) *se verifica el modelo (3) con* $E\eta_{ij} = 0$ *y* $E\left|\eta_{ij}\right|^{2(2\nu+1)+\delta'} < C < \infty$ *para algún* $\delta' > 0$ $(1 \leq i \leq n, 1 \leq j \leq p)$,

se tiene que:

$$a)\ n^{\nu/(2\nu+1)}(\hat{m}_{h_0}(t) - m(t)) \xrightarrow{D} N(B,V).$$

$$b)\ n^{\nu/(2\nu+1)}(\hat{m}_{\widehat{h}_n}(t) - m(t)) \xrightarrow{D} N(B,V).$$

Por tanto, si somos capaces de obtener una sucesión $\{\widehat{h}_n\}$ verificando (21), tendremos una forma de construir un estimador asintóticamente equivalente a un estimador óptimo (ambos tienen la misma distribución asintótica).

Suponiendo que conocemos ρ, y considerando una sucesión de estimadores del tipo $\widehat{h}_n = \widehat{C}_0 n^{-\frac{1}{2\nu+1}}$, todo lo que necesitamos para verificar (21) es que $\widehat{C}_0$ sea un estimador consistente para estimar C_0. Para ello basta, debido a su expresión, con estimar de forma consistente $\int (m^{(\nu)}(t))^2 w(t)dt$ y σ_ε^2. Para estimar la integral, parece intuitivo construir un estimador de $m^{(\nu)}$, llamémosle $\widehat{m}_{h,b,\nu}$ (siendo b un nuevo parámetro ventana), y después estimar $\int (m^{(\nu)}(t))^2 w(t)dt$ por medio de $\int (\widehat{m}_{h,b,\nu}(t))^2 w(t)dt$. De esta forma, aparece un nuevo problema debido a que nuevamente deberíamos buscar métodos para seleccionar sus parámetros ventana. Sin embargo, esta elección es de menor importancia, de la misma forma que ocurre en la estimación de la densidad (Scott, Tapia y Thompson (1977), Scott y Factor (1981), Sheather (1986)).

Estudiaremos ahora condiciones que nos permitan estimar C_0 de forma consistente.

Sea:

$$\widehat{m}_{h,b,\nu}(t) = \frac{1}{b^{\nu+1}} \sum_{i=1}^{n} \int_{(i-1)/n}^{i/n} \overline{K}^{(\nu)}(\frac{t-u}{b})du(y_i - \zeta_i^T \hat{\beta}_h), \quad (23)$$

donde:

(K.3) $\overline{K}^{(\nu)}$ es la derivada de orden ν de una función ν veces derivable $\overline{K}$ que verifica:

a) soporte de $\overline{K} \subset [-1,1]$

b) $\int \overline{K}(x)dx = 1$

c) $\overline{K}^{(j)}(-1) = \overline{K}^{(j)}(1) = 0,\ j = 0,\dots,\nu-1$ y $\overline{K}^{(\nu)}$ está acotada.

Teorema 2.2

Supongamos que se verifican **(D.1)**, **(D.2)** *(en cuyo caso además exigimos $nb^{\nu+1/2} \to \infty$) o* **(D.2*)** *(en cuyo caso además $nb^{\nu+3\nu/(4\nu-1)} > C > 0$),* **(E.1)**, **(K.1)**- **(K.3)**, **(M.1)**, **(G.1)-(G.3)**, *y $m^{(\nu)}$ verifica una condición de Lipschitz. Entonces, si $h \in \left[a_1 n^{-1/(2\nu+1)}, a_2 n^{-1/(2\nu+1)}\right]$, $0 < a_1 < a_2 < \infty$ constantes, y suponiendo que $b \to 0$ cuando $n \to \infty$, se tiene que:*

$$\sup_{t\in[w_1,w_2]} \left| E\left[\widehat{m}_{h,b,\nu}(t)\right] - m^{(\nu)}(t) \right| \underset{n\to\infty}{\longrightarrow} 0$$

Teorema 2.3

Bajo las suposiciones del Teorema 2.2, si además se verifica **(E.2)** *y $nb^{2\nu+1} \to \infty$, se tiene:*

$$\sup_{t\in[w_1,w_2]} \left| \widehat{m}_{h,b,\nu}(t) - m^{(\nu)}(t) \right| \xrightarrow{P} 0.$$

Comentario 2.3

En el Teorema 2.2 es posible dar un mayor margen a la elección de h, permitiendo que h pertenezca a un intervalo del tipo $\left[a_1 n^{-1/(2\nu+1)-\gamma}, a_2 n^{-1/(2\nu+1)+\gamma}\right]$, siendo γ un número real no negativo verificando ciertas condiciones. Sin embargo esto conllevará a complicar las condiciones que deberá verificar b. Es por ello que hemos preferido fijar $\gamma = 0$ en el anterior intervalo.

Comentario 2.4

Las hipótesis hechas acerca del parámetro b en el Teorema 2.2 son satisfechas si consideramos $b = cn^{-\theta}$, siendo $0 < \theta \leq (4\nu-1)/(2\nu(2\nu+1))$ y c una constante positiva. Por tanto podemos considerar el caso $b = h$. En cuanto a las del Teorema 2.3, se verifican si $0 < \theta < 1/(2\nu+1)$. Por tanto, en este caso, no cabe la posibilidad de considerar parámetros iguales.

Comentario 2.5

Del Teorema 2.3 se deduce que $\int(\widehat{m}_{h,b,v}(t))^2 w(t)dt$ es un estimador consistente de $\int(m^{(v)}(t))^2 w(t)dt$. En cuanto a σ_ε^2, puede verse en Gao (1995) un método para estimarlo consistentemente basado en un estimador de Mínimos Cuadrados de β. Por tanto, junto con el Teorema 2.1, si ρ es conocido, tenemos ya una forma de construir un estimador de $m(t)$ asintóticamente equivalente a un estimador óptimo (ambos tienen la misma distribución asintótica).

Supongamos ahora que ρ es desconocido. En este caso, puesto que nuestro estimador $\hat{m}_h(t)$ depende de ρ a través de $\hat{\beta}_h$, no podemos trabajar con él. Sea $\widehat{\rho}$ un estimador de ρ, y consideremos el estimador:

$$\widehat{\widehat{m}}_h(t) = \sum_{i=1}^{n} w_{n,h}(t,t_i)(y_i - \zeta_i^T \widehat{\widehat{\beta}}_h), \tag{24}$$

donde

$$\widehat{\widehat{\beta}}_h = \left(\widetilde{\mathbf{X}}^T \Psi_*^{-1}(\widehat{\rho})\widetilde{\mathbf{X}}\right)^{-1} \widetilde{\mathbf{X}}^T \Psi_*^{-1}(\widehat{\rho})\widetilde{\mathbf{y}}, \tag{25}$$

siendo

$$\Psi_*^{-1}(\rho) = \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & \ddots & 0 \\ 0 & -\rho & 1+\rho^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -\rho \\ 0 & 0 & \dots & -\rho & 1 \end{pmatrix}.$$

Como puede observarse, hemos hecho una pequeña modificación de la matriz Ψ^{-1} (9), motivada por el carácter aleatorio que tendrá a partir de ahora. De esta forma, nos evitamos problemas con el denominador cuando introducimos $\widehat{\rho}$. En el caso de conocer ρ, el estimador no varía.

Ahora imponemos las siguientes condiciones, que son una extensión de las condiciones de regularidad de Fuller y Battese (1973). Hemos introducido subíndices para hacer notar la dependencia con n.

(G.5) La sucesión de matrices $\left\{\widetilde{\mathbf{X}}_n\right\}$ es tal que

$$\lim_{n\to\infty} \frac{1}{n} \widetilde{\mathbf{X}}_n^T \mathbf{S}_n(\rho) \widetilde{\mathbf{X}}_n = \mathbf{H}(\rho) \tag{26}$$

y

$$\frac{1}{n}\widetilde{\mathbf{X}}_n^T \mathbf{S}_n(\rho)(\widetilde{\mathbf{m}}_n + \widetilde{\varepsilon}_n) = O_p(n^{-\frac{1}{2}}), \tag{27}$$

donde $\mathbf{S}_n(\rho) = \frac{\partial}{\partial\rho}\Psi_{*n}^{-1}(\rho)$ y $\mathbf{H}(\rho)$ es una matriz cuyos elementos son funciones continuas de ρ.

(R.1) Disponemos de un estimador $\widehat{\rho}$ de ρ que satisface la condición:

$$\widehat{\rho} = \rho + O_p(n^{-\omega}), \quad \omega > 0. \tag{28}$$

Teorema 2.4

Bajo las hipótesis del Teorema 2.1, junto con **(G.5)** *y* **(R.1)**, *se verifica:*

a) $n^{\nu/(2\nu+1)}(\widehat{\widehat{m}}_{h_0}(t) - m(t)) \xrightarrow{D} N(B,V)$.

b) $n^{\nu/(2\nu+1)}(\widehat{\widehat{m}}_{\widehat{h}_n}(t) - m(t)) \xrightarrow{D} N(B,V)$.

Comentario 2.6

Es fácil ver que, bajo nuestras hipótesis acerca de los pesos w_{ij}, las funciones g_j y los errores ε_i, si las filas de η son vectores aleatorios i.i.d. con media cero y matriz de varianzas-covarianzas finita, entonces se verifica **(G.5)** (ver Aneiros y Quintela (1999a)).

Comentario 2.7

Acerca de la condición **(R.1)**, puede verse en Schick (1994) un ejemplo de un estimador $\widehat{\rho}_*$ de ρ que, bajo ciertas condiciones, verifica que $n^{1/2}(\widehat{\rho}_* - \rho)$ es asintóticamente normal con media cero y varianza $1-\rho^2$.

Puesto que ahora desconocemos ρ, no podemos utilizar $\widehat{m}_{h,b,\nu}(t)$ como estimador consistente de $m^{(\nu)}(t)$ para ser luego utilizado en la expresión de $\widehat{h}_n$. Proponemos en su lugar el estimador

$$\widehat{m}^*_{h,b,\nu}(t) = \frac{1}{b^{\nu+1}}\sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}\overline{K}^{(\nu)}(\frac{t-u}{b})du(y_i - \zeta_i^T\hat{\beta}_h^*), \tag{29}$$

siendo

$$\hat{\beta}_h^* = \left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{y}}. \tag{30}$$

Es decir, a la hora de estimar el vector de parámetros β para posteriormente obtener un estimador consistente de $m^{(\nu)}(t)$, lo hacemos basándonos en residuos de estimación por Mínimos Cuadrados Ordinarios. Debemos introducir dos nuevas hipótesis:

(G.1*) $n^{-1}\eta^T\eta \longrightarrow \mathbf{V}^*$ donde $\mathbf{V}^* = \{V^*_{ij}\}$ es una matriz definida positiva.

(G.3*) $n^{-1}\eta^T\widetilde{\mathbf{m}} = O(n^{-1/2}h^\nu)$.

Comentario 2.8

Estas hipótesis son utilizadas por Speckman (1988). En su artículo puede verse un caso en el que se verifican.

Teorema 2.5

Supongamos que se verifican **(D.1)**, **(D.2)** *(en cuyo caso además exigimos* $nb^{\nu+1/2}\to\infty$*) o* **(D.2*)** *(en cuyo caso además* $nb^{\nu+3\nu/(4\nu-1)} > C > 0$*),* **(E.1)**, **(K.1)- (K.3)**, **(M.1)**,**(G.1*)**, **(G.2)**, **(G.3*)**, *y* $m^{(\nu)}$ *verifica una condición de Lipschitz. Entonces, si* $h \in \left[a_1 n^{-1/(2\nu+1)}, a_2 n^{-1/(2\nu+1)}\right]$, $0 < a_1 < a_2 < \infty$ *constantes, y suponiendo que* $b \to 0$ *cuando* $n\to\infty$*, se tiene que:*

$$\sup_{t\in[w_1,w_2]} \left| E\left[\widehat{m}^*_{h,b,\nu}(t)\right] - m^{(\nu)}(t)\right| \underset{n\to\infty}{\longrightarrow} 0$$

Teorema 2.6

Bajo las suposiciones del Teorema 2.5, si además se verifica **(E.2)** *y* $nb^{2\nu+1}\to\infty$*, se tiene:*

$$\sup_{t\in[w_1,w_2]} \left|\widehat{m}^*_{h,b,\nu}(t) - m^{(\nu)}(t)\right| \xrightarrow{P} 0.$$

Comentario 2.9

Por tanto, en el caso de no conocer ρ, tenemos también un método para estimar de forma consistente $\int (m^{(\nu)}(t))^2 w(t)dt$ con lo que, como ya sabemos estimar consistentemente σ^2_ε (ver el Comentario 2.5) y ρ (ver el Comentario 2.7), tenemos un método para obtener una sucesión de ventanas $\left\{\widehat{h}_n\right\}$ verificando (21).

3. DEMOSTRACIONES

A lo largo de las demostraciones, será utilizado el siguiente Lema.

Lema 3.1

Sean $g_j : [0,1] \to \mathbb{R}$ *funciones acotadas,* $j = 1, \dots, p$*, y supongamos que se verifica* **(D.2***)*. Entonces:*

$$\max_{ij} |x_{ij}| = o_p(n^{1/(2(2\nu+1))}).$$

Demostración

Sea $M_n = n^{1/(2(2\nu+1))}/\log n$. Utilizando **(D.2*)** y la desigualdad de Markov, se tiene que:

$$P(\max_{i,j} |\eta_{ij}| \geq M_n) \leq CnM_n^{-2(2\nu+1)-\delta'}.$$

Por tanto, se verifica que $\max_{i,j} |x_{ij}| = o_p(n^{1/(2(2\nu+1))})$. Ahora, utilizando la acotación de g_j, $j = 1, \dots, p$, junto con la relación (3), se obtiene directamente la tesis de este Lema.

También se utilizará que $\max_i \#\{j \,/\, w_{ij} \neq 0\} = O(nh)$, como consecuencia del tipo de pesos utilizados y de que el núcleo $K(\cdot)$ tiene soporte compacto.

Demostración del Teorema 2.1

Parte a)

Debido a (19), es suficiente demostrar que $n^{\nu/(2\nu+1)}(\hat{m}_{h_0}(t) - m(t))$ y $n^{\nu/(2\nu+1)}(\overline{m}_{h_0}(t) - m(t))$ tienen la misma distribución asintótica, para lo cual basta con probar que la diferencia de ambos tiende a cero en probabilidad.

Supongamos que se verifica **(D.2)**. Escribamos

$$\left|n^{\nu/(2\nu+1)}(\hat{m}_{h_0}(t) - \overline{m}_{h_0}(t))\right| = \left|n^{\nu/(2\nu+1)}(\sum_{i=1}^{n} w_{n,h_0}(t,t_i)\zeta_i^T(\hat{\beta}_{h_0} - \beta))\right| =$$

$$= O_p\left(n^{\nu/(2\nu+1)} nh_0 (nh_0)^{-1} n^{-1/2}\right) = O_p(n^{-1/(2(2\nu+1))}).$$

Si la que se verifica es **(D.2*)**, razonando como en el caso anterior y utilizando el Lema 3.1, obtendríamos

$$\left| n^{\nu/(2\nu+1)}(\hat{m}_{h_0}(t) - \overline{m}_{h_0}(t)) \right| = o_p(1).$$

Hemos utilizado que $\left\| \hat{\beta}_{h_0} - \beta \right\| = O_p(n^{-1/2})$, lo cual es consecuencia del Teorema 2.1 de Aneiros y Quintela (1999a).

Parte b)

De la misma forma, ahora como consecuencia de (22), es suficiente demostrar que $n^{\nu/(2\nu+1)}(\hat{m}_{\widehat{h}_n}(t) - m(t))$ y $n^{\nu/(2\nu+1)}(\overline{m}_{\widehat{h}_n}(t) - m(t))$ tienen la misma distribución asintótica.

Escribamos:

$$n^{\nu/(2\nu+1)}(\hat{m}_{\widehat{h}_n}(t) - \overline{m}_{\widehat{h}_n}(t)) = -n^{\nu/(2\nu+1)}(\sum_{i=1}^{n} w_{n,\widehat{h}_n}(t,t_i)\zeta_i^T(\hat{\beta}_{\widehat{h}_n} - \beta)).$$

Sean $\varepsilon > 0$ y $\lambda > 0$. Se tiene que

$$P\left\{ \left| n^{\nu/(2\nu+1)}(\hat{m}_{\widehat{h}_n}(t) - \overline{m}_{\widehat{h}_n}(t)) \right| \geq \varepsilon \right\} \leq$$

$$\leq P\left\{ \left(\left| n^{\nu/(2\nu+1)}(\hat{m}_{\widehat{h}_n}(t) - \overline{m}_{\widehat{h}_n}(t)) \right| \geq \varepsilon \right) \cap \left(\left| \frac{\widehat{h}_n}{h_0} - 1 \right| \leq \lambda \right) \right\} + P\left\{ \left| \frac{\widehat{h}_n}{h_0} - 1 \right| > \lambda \right\} \leq$$

$$\leq P\left\{ \sup_{|z|\leq\lambda} \left| n^{\nu/(2\nu+1)}(\sum_{i=1}^{n} w_{n,h_0(1+z)}(t,t_i)\zeta_i^T(\hat{\beta}_{h_0(1+z)} - \beta)) \right| \geq \varepsilon \right\} + P\left\{ \left| \frac{\widehat{h}_n}{h_0} - 1 \right| > \lambda \right\}$$

Por hipótesis, se tiene que el segundo sumando converge a cero. En cuanto al primero, se tiene que

$$\left| n^{\nu/(2\nu+1)}(\sum_{i=1}^{n} w_{n,h_0(1+z)}(t,t_i)\zeta_i^T(\hat{\beta}_{h_0(1+z)} - \beta)) \right| =$$

$$= O_p\left(n^{\nu/(2\nu+1)} n h_0(1+z)(n h_0(1+z))^{-1} n^{-1/2} \right) = O_p(n^{-1/2(2\nu+1)}),$$

uniformemente en $|z| \leq \lambda$. Por tanto,

$$\sup_{|z|\leq\lambda} \left| n^{\nu/(2\nu+1)}(\sum_{i=1}^{n} w_{n,h_0(1+z)}(t,t_i)\zeta_i^T(\hat{\beta}_{h_0(1+z)} - \beta)) \right| = O_p(n^{-1/2(2\nu+1)}).$$

Si la que se verifica es (D.2*), razonando como en el caso anterior y utilizando el Lema 3.1, obtendríamos

$$\sup_{|z|\leq\lambda}\left|n^{\nu/(2\nu+1)}(\sum_{i=1}^{n} w_{n,h_0(1+z)}(t,t_i)\zeta_i^T(\hat{\beta}_{h_0(1+z)}-\beta))\right| = o_p(1).$$

Por tanto, en cualquiera de los dos casos, tenemos que el primer sumando también converge a cero. En los anteriores razonamientos hemos utilizado que $||\hat{\beta}_{h_0(1+z)}-\beta|| = O_p(n^{-1/2})$ uniformemente en $|z|\leq\lambda$ (ver Aneiros y Quintela (1999a)).

Demostración del Teorema 2.2

Utilizando que $m^{(\nu)}$ verifica una condición de Lipschitz y que los errores tienen media cero, junto con **(D.1)**, **(K.3)** y que $b\to 0$ y $nb^{\nu}\to\infty$, se obtiene, por medio de una demostración similar a la del Teorema 2.1 de Gasser y Müller (1984), que:

$$\sup_{t\in[w_1,w_2]}\left|E\left[\overline{m}_{b,\nu}(t)\right]-m^{(\nu)}(t)\right| \underset{n\to\infty}{\rightarrow} 0,$$

siendo $\overline{m}_{b,\nu}(t)$ el estimador (23) cambiando $\hat{\beta}_h$ por β, es decir:

$$\overline{m}_{b,\nu}(t) = \frac{1}{b^{\nu+1}}\sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}\overline{K}^{(\nu)}(\frac{t-u}{b})du(y_i-\zeta_i^T\beta).$$

Resta por tanto demostrar que:

$$\sup_{t\in[w_1,w_2]}\left|E\left[\widehat{m}_{h,b,\nu}(t)-\overline{m}_{b,\nu}(t)\right]\right| \underset{n\to\infty}{\rightarrow} 0. \tag{31}$$

Supongamos en primer lugar que se verifica **(D.2)**. Entonces:

$$\left|E\left[\widehat{m}_{h,b,\nu}(t)-\overline{m}_{b,\nu}(t)\right]\right| = \left|\frac{1}{b^{\nu+1}}\sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}\overline{K}^{(\nu)}(\frac{t-u}{b})du\zeta_i^T\,\mathrm{Sesgo}(\hat{\beta}_h)\right| =$$

$$= O(\frac{1}{b^{\nu+1}}nbn^{-1}(h^{2\nu}+h^{\nu}(nh)^{-1/2})) = O(\frac{1}{b^{\nu}}(h^{2\nu}+h^{\nu}(nh)^{-1/2})),$$

donde se ha utilizado que $\overline{K}^{(\nu)}$ está acotada y tiene soporte acotado, y que se verifica que $Sesgo(\hat{\beta}_h) = O\left(h^{2\nu}\right)+O(h^{\nu}(nh)^{-1/2})$ (como puede verse en Aneiros y Quintela (1999a)).

En caso de verificarse **(D.2*)**, siguiendo el mismo razonamiento y utilizando el Lema 3.1, se obtendría:

$$\left|E\left[\widehat{m}_{h,b,v}(t)-\overline{m}_{b,v}(t)\right]\right| = o_p(n^{1/(2(2v+1))}\frac{1}{b^v}(h^{2v}+h^v(nh)^{-1/2})).$$

En cualquiera de los dos casos, como consecuencia de las hipótesis hechas acerca de b y h, se obtiene (31).

Demostración del Teorema 2.3

Debido al Teorema 2.2, basta demostrar que:

$$\sup_{t\in[w_1,w_2]}\left|\widehat{m}_{h,b,v}(t)-E\left[\widehat{m}_{h,b,v}(t)\right]\right| \xrightarrow{P} 0.$$

Se tiene que:

$$\sup_{t\in[w_1,w_2]}\left|\widehat{m}_{h,b,v}(t)-E\left[\widehat{m}_{h,b,v}(t)\right]\right| \leq \sup_{t\in[w_1,w_2]}\left|\frac{1}{b^{v+1}}\sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}\overline{K}^{(v)}(\frac{t-u}{b})du\varepsilon_i\right| +$$

$$+\sup_{t\in[w_1,w_2]}\left|\frac{1}{b^{v+1}}\sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}\overline{K}^{(v)}(\frac{t-u}{b})du\zeta_i^T(\hat{\beta}_h-E(\hat{\beta}_h))\right|.$$

Estudiemos cada sumando por separado.

Por **(E.1)**, **(E.2)** y **(K.3)**, aplicando la modificación de la desigualdad de Whittle (1960) hecha por Chu (1989) (expresión (3.6.3) de Chu) se obtiene:

$$E\left[\left(\sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}\overline{K}^{(v)}(\frac{t-u}{b})du\varepsilon_i\right)^2\right] \leq C\sum_{i=1}^{n}\left(\int_{(i-1)/n}^{i/n}\overline{K}^{(v)}(\frac{t-u}{b})\right)^2 = \tag{32}$$

$$= O(nbn^{-2}) = O(n^{-1}b) \tag{33}$$

uniformemente en t. Por tanto,

$$\sup_{t\in[w_1,w_2]}\left|\sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}\overline{K}^{(v)}(\frac{t-u}{b})du\varepsilon_i\right| = O_p((n^{-1}b)^{1/2})$$

y

$$\sup_{t\in[w_1,w_2]}\left|\frac{1}{b^{v+1}}\sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}\overline{K}^{(v)}(\frac{t-u}{b})du\varepsilon_i\right| = O_p((nb^{2v+1})^{-1/2}) = o_p(1).$$

Estudiemos el segundo sumando, suponiendo en primer lugar que se verifica **(D.2)**.

$$\sup_{t\in[w_1,w_2]}\left|\frac{1}{b^{\nu+1}}\sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}K_\nu(\frac{t-u}{b})du\zeta_i^T(\hat{\beta}_h-E(\hat{\beta}_h))\right|=$$

$$=O_p(\frac{1}{b^{\nu+1}}nbn^{-1}n^{-1/2})=O_p((nb^{2\nu})^{-1/2}).$$

En caso de verificarse **(D.2*)**, siguiendo el mismo razonamiento y utilizando el Lema 3.1, se obtendría:

$$\sup_{t\in[w_1,w_2]}\left|\frac{1}{b^{\nu+1}}\sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}K_\nu(\frac{t-u}{b})du\zeta_i^T(\hat{\beta}_h-E(\hat{\beta}_h))\right|=o_p(n^{1/(2(2\nu+1))}(nb^{2\nu})^{-1/2}).$$

En cualquiera de los dos casos, como consecuencia de las hipótesis hechas acerca de b y h, se obtiene que la anterior expresión converge a cero en probabilidad.

Demostración del Teorema 2.4

Parte a)

Utilizando el Teorema 2.1.a, basta demostrar que $n^{\nu/(2\nu+1)}(\widehat{\hat{m}}_{h_0}(t)-m(t))$ y $n^{\nu/(2\nu+1)}\times$ $\times(\hat{m}_{h_0}(t)-m(t))$ tienen la misma distribución asintótica.

Supongamos que se verifica **(D.2)**. Escribamos

$$\left|n^{\nu/(2\nu+1)}(\widehat{\hat{m}}_{h_0}(t)-\hat{m}_{h_0}(t))\right|=\left|n^{\nu/(2\nu+1)}(\sum_{i=1}^{n}w_{n,h_0}(t,t_i)\zeta_i^T(\widehat{\hat{\beta}}_{h_0}-\hat{\beta}_{h_0}))\right|=$$

$$=O_p(n^{\nu/(2\nu+1)}nh_0(nh_0)^{-1}n^{-1/2})=O_p(n^{-1/(2(2\nu+1))}).$$

Hemos utilizado que $\left\|\widehat{\hat{\beta}}_{h_0}-\hat{\beta}_{h_0}\right\|=O_p(n^{-1/2})$ (lo cual es obtenido en la demostración del Teorema 2.6 de Aneiros y Quintela (1999a)).

Si la que se verifica es **(D.2*)**, razonando como en el caso anterior y utilizando el Lema 3.1, obtendríamos

$$\left|n^{\nu/(2\nu+1)}(\widehat{\hat{m}}_{h_0}(t)-\hat{m}_{h_0}(t))\right|=o_p(1).$$

Parte b)

Utilizando el Teorema 2.1.b, es suficiente demostrar que $n^{\nu/(2\nu+1)}(\widehat{\widehat{m}}_{\widehat{h}_n}(t)-\hat{m}_{\widehat{h}_n}(t)) \xrightarrow{P} 0$. El siguiente razonamiento es similar al hecho para demostrar el Teorema 2.1.b.

Sean $\varepsilon > 0$ y $\lambda > 0$. Se tiene que:

$$P\left\{\left|n^{\nu/(2\nu+1)}(\widehat{\widehat{m}}_{\widehat{h}_n}(t)-\hat{m}_{\widehat{h}_n}(t))\right| \geq \varepsilon\right\} \leq$$

$$\leq P\left\{\left(\left|n^{\nu/(2\nu+1)}(\widehat{\widehat{m}}_{\widehat{h}_n}(t)-\hat{m}_{\widehat{h}_n}(t))\right| \geq \varepsilon\right) \cap \left(\left|\frac{\widehat{h}_n}{h_0}-1\right| \leq \lambda\right)\right\} + P\left\{\left|\frac{\widehat{h}_n}{h_0}-1\right| > \lambda\right\} \leq$$

$$\leq P\left\{\sup_{|z|\leq\lambda}\left|n^{\nu/(2\nu+1)}\sum_{i=1}^{n} w_{n,h_0(1+z)}(t,t_i)\zeta_i^T(\hat{\beta}_{h_0(1+z)}-\hat{\hat{\beta}}_{h_0(1+z)})\right| \geq \varepsilon\right\} + P\left\{\left|\frac{\widehat{h}_n}{h_0}-1\right| > \lambda\right\}.$$

Por hipótesis, se tiene que el segundo sumando converge a cero.

En cuanto al primero, siguiendo el mismo razonamiento que el hecho en la demostración del Teorema 2.1.b, y utilizando que $\left\|\hat{\beta}_{h_0(1+z)}-\hat{\hat{\beta}}_{h_0(1+z)}\right\| = O_p(n^{-1/2})$ uniformemente en $|z| \leq \lambda$ (lo cual es obtenido en la demostración del Teorema 2.6 de Aneiros y Quintela (1999a)), se obtiene su convergencia a cero en probabilidad.

Demostración del Teorema 2.5

Como consecuencia de nuestras hipótesis, se puede utilizar el Lema 1 de Speckman (1988). Por medio de dicho Lema, junto con **(E.1)** y **(G.3*)** se obtiene, sin más que seguir la demostración del Teorema 2.1.a de Aneiros y Quintela (1999a), que $Sesgo(\hat{\beta}_h^*) = O\left(h^{2\nu}\right) + O(h^{\nu}(nh)^{-1/2})$. Ahora, utilizando lo anterior, el resto de la demostración es similar a la del Teorema 2.2.

Demostración del Teorema 2.6

Siguiendo la demostración del Teorema 2.2.b de Speckman (1988), se tiene que $\hat{\beta}_h^* - E(\hat{\beta}_h^*) = O_p(n^{-1/2})$. Utilizando este resultado, el resto de la demostración es similar a la del Teorema 2.3.

4. REFERENCIAS

Aneiros-Pérez, G. and Quintela del Río, A. (1998). «Selección de la ventana en modelos parcialmente lineales». *Proc. XXIV Congreso Nacional de Estadística e I.O.* (Almería, Spain), 353-354.

Aneiros, G. and Quintela, A. (1999a). «Asymptotic properties in partial linear models under dependence». *En revisión.*

Aneiros, G. and Quintela, A. (1999b). «Bandwidth choice for a GLS estimator of the parametric part in a partial linear model with autoregressive errors». *En revisión.*

Chu, C. K. (1989). *Some results in nonparametric regression*. Ph. D. dissertation, Dept. Statistics, Univ. North Carolina, Chapel Hill.

Doukhan, P. (1994). *Mixing: properties and examples*. Lecture Notes in Statistics, 85. Springer-Verlag.

Engle, R., Granger, C., Rice, J. and Weiss, A. (1986). «Nonparametric estimates of the relation between weather and electricity sales». *J. Amer. Statist. Assoc.*, 81, 310-320.

Fuller, W. A. and Battese, G. E. (1973). «Transformations for estimation of linear models with nested-error structure». *J. Amer. Statist. Assoc.*, 68, 626-632.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*. Wiley.

Gao, J. (1995). «Asymptotic theory for partly linear models». *Commun. Statist.-Theory Meth.*, 24, 1985-2009.

Gasser, T. and Müller, H. G. (1979). *Kernel estimation of regression functions*. Smoothing techniques for curve estimation (eds Th. Gasser and M. Rosenblatt), Springer-Verlag.

Gasser, T. and Müller, H. G. (1984). «Estimating regression functions and their derivatives by the kernel method». *Scandinavian Journal of Statistics*, 11, 171-185.

Gasser, T., Müller, H. G. and Mammitzsch, V. (1985). «Kernels for nonparametric curve estimation». *J. R. Statist. Soc.* B, 238-252.

Härdle, W. and Vieu, P. (1992). «Kernel regression smoothing of time series». *J. Time Ser. Anal.*, 13, 209-232.

Hart, J. (1991). «Kernel Regression Estimation with Time Series Errors». *J. R. Statist. Soc.*, B, 53, 173-187.

Hart, J. and Vieu, P. (1990). «Data-driven bandwidth choice for density estimation based on dependent data». *Ann. Statist.*, 18, 873-890.

Heckman, N. (1986). «Spline smoothing in partly linear model». *J. R. Statist. Soc.*, B, 48, 244-248.

Judge, G., Griffiths, W., Carter, R., Lütkepohl, H. and Lee, T. C. (1985). *The theory and practice of econometrics*. Wiley.

Linton, O. (1995). «Second order approximation in the partially linear regression model». *Econometrica*, 63, 1079-1112.

Quintela del Río, A. (1994a). «Nonparametric estimation of regression functions based on dependent data». *New Progress in Probability and Statistics*, (International Science Publishers, Zeits, Holland), 139-154.

Quintela del Río, A. (1994b). «A plug-in technique in nonparametric regression with dependence». *Comm. Statist. Theory Methods*, 23, 2581-2603.

Quintela del Río, A. (1996). «Comparison of bandwidth selectors in nonparametric regression under dependence». *Comp. Statist. Data Anal.*, 21, 563-580.

Rice, J. (1984b). «Boundary modification for kernel regression». *Communications in Statistics*, Series A, 13, 893-900.

Rice, J. (1986). «Convergence rates for partially splined models». *Statist. Probabil. Lett.*, 4, 203-208.

Schick, A. (1994). «Estimation of the autocorrelation coefficient in the presence of a regression trend». *Statist. Probabil. Lett.*, 21, 371-380.

Speckman, P. (1988). «Kernel smoothing in partial linear models». *J. R. Statist. Soc.*, B, 50, 413-436.

Vieu, P. (1991). «Quadratic errors for nonparametric estimates under dependence». *Journal of Multivariate Analysis*, 2, 324-347.

Whittle, P. (1960). «Bounds for the moments of linear and quadratic forms in independent variables». *Theory of Probability and Applications*, 5, 302-305.

ENGLISH SUMMARY

BANDWIDTH SELECTION IN KERNEL SMOOTHING OF THE NONPARAMETRIC PART IN A PARTIAL LINEAR MODEL WITH AUTOREGRESSIVE ERRORS

GERMÁN ANEIROS-PÉREZ*
Universidad de La Coruña

Suppose that $y_i = \zeta_i^T \beta + m(t_i) + \varepsilon_i$, $i = 1, \ldots, n$, *where the* $(p \times 1)$*-vector* β *and the function* $m(\cdot)$ *are unknown, and the errors* ε_i *pertain to a stationary* $AR(1)$ *process. The problem of bandwidth selection for a kernel estimator for* $m(\cdot)$*, based on a Generalized Least Squares estimator for* β*, is addressed here. We obtain an asymptotic expression of an optimal bandwidth, and we propose to use a plug-in methodology in order to obtain an optimum estimation of* $m(\cdot)$*. The obtained results are a generalization of those obtained by Quintela (1994) in nonparametric regression.*

Keywords: partial linear models, kernel smoothing, bandwidth selection, time series

AMS Classification (MSC 2000): 62G07, 62G20, 62M10

⋆ Research supported by the Xunta de Galicia (Spain) under research project XUGA 10503A98.
* Departamento de Matemáticas. Facultad de Informática. Universidad de La Coruña. Campus de Elviña, s/n. La Coruña. Spain. E-mail: ganeiros@udc.es
– Received October 1999.
– Accepted March 2000.

1. INTRODUCTION

We consider that we have an observed data set $\{(y_i, \zeta_i^T, t_i)^T\}_{i=1}^n$ generated by the semiparametric model

$$y_i = \zeta_i^T \beta + m(t_i) + \varepsilon_i \quad i = 1, \ldots, n, \tag{1}$$

where $(\zeta_i^T, t_i) = ((x_{i1}, x_{i2}, \ldots, x_{ip}), t_i) \in A \times [0,1] \subset \mathbb{R}^p \times \mathbb{R}$ are fixed design points, β is a $(p \times 1)$-vector of unknown parameters, $m(\cdot)$ is an unknown real-valued function defined on $[0,1]$ (without loss of generality) and $\{\varepsilon_i\}_{i=1}^n$ a sample of errors identically distributed. This semiparametric model is called a "partial linear model". In this paper we suppose that the errors are generated by an AR(1) process, i. e., $\varepsilon_i = \rho\varepsilon_{i-1} + e_i$, where the e_i are i.i.d. and $|\rho| < 1$.

If we suppose that ρ is known, we propose to estimate $m(\cdot)$ by $\hat{m}_h(t) = \sum_{i=1}^n w_{n,h}(t, t_i) \times \times (y_i - \zeta_i^T \hat{\beta}_h)$, where $\{w_{n,h}(t, t_i)\}$ are the weights of Gasser and Müller (1979), $\hat{\beta}_h = \left(\widetilde{\mathbf{X}}^T \Psi^{-1} \widetilde{\mathbf{X}}\right)^{-1} \widetilde{\mathbf{X}}^T \Psi^{-1} \widetilde{\mathbf{y}}$, Ψ is the correlation matrix of the errors, $\widetilde{\mathbf{y}} = (\mathbf{I} - \mathbf{W})\mathbf{y}$, $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$, $\widetilde{\mathbf{X}} = (\mathbf{I} - \mathbf{W})\mathbf{X}$, $\mathbf{X} = (\zeta_1^T, .., \zeta_n^T)$ and $\mathbf{W} = \{w_{n,h}(t_i, t_j)\}$. Under suitable conditions, Aneiros and Quintela (1999) show that

$$h_0 = C_0 n^{-\frac{1}{2\nu+1}} = \left[\frac{C_K \sigma_\varepsilon^2 \frac{1+\rho}{1-\rho}}{\int (m^{(\nu)}(u))^2 w(u) du}\right]^{\frac{1}{2\nu+1}} n^{-\frac{1}{2\nu+1}}$$

is an optimal bandwidth for the estimator $\hat{m}_h(\cdot)$, where

$$C_K = \frac{(\nu!)^2 \int K^2(u) du \int w(u) du}{2\nu \alpha_\nu^2}, \ \alpha_\nu = \int u^\nu K(u) du$$

and $w(\cdot)$ is a continuous function defined on $[w_1, w_2] \subset (0,1)$.

2. THE RESULTS

Under suitable conditions, we obtain:

Theorem 2.1

a) $n^{\nu/(2\nu+1)}(\hat{m}_{h_0}(t) - m(t)) \xrightarrow{D} N(B, V)$.

b) $n^{\nu/(2\nu+1)}(\hat{m}_{\widehat{h}_n}(t) - m(t)) \xrightarrow{D} N(B, V)$.

Here we denote $B=\frac{(-1)^{\nu}}{\nu!}C_0^{\nu}\alpha_{\nu}m^{(\nu)}(t)$, $V=\sigma_{\varepsilon}^{2}\frac{1+\rho}{1-\rho}C_0^{-1}\int K^2(u)du$, and $\widehat{h}_n$ is a consistent estimator of h_0. A way of constructing an asymptotically equivalent estimate to an optimum estimate (they have the same asymptotic distribution) is obtained by means of this theorem. In order to achieve this, the following expression must be used: $\widehat{h}_n=\widehat{C}_0 n^{-\frac{1}{2\nu+1}}$, being $\widehat{C}_0$ a consistent estimation of C_0. Gao (1995) proposes a consistent estimation of σ_{ε}^2. Furthermore, we propose to estimate $m^{(\nu)}(\cdot)$ by $\widehat{m}_{h,b,\nu}(t)=\frac{1}{b^{\nu+1}}\sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}\overline{K}^{(\nu)}(\frac{t-u}{b})du(y_i-\zeta_i^T\hat{\beta}_h)$, and we obtain:

Theorem 2.2

$$\sup_{t\in[w_1,w_2]}\left|E\left[\widehat{m}_{h,b,\nu}(t)\right]-m^{(\nu)}(t)\right| \underset{n\to\infty}{\rightarrow} 0$$

Theorem 2.3

$$\sup_{t\in[w_1,w_2]}\left|\widehat{m}_{h,b,\nu}(t)-m^{(\nu)}(t)\right| \xrightarrow{P} 0.$$

If ρ is unknown, we propose to estimate $m(\cdot)$ by $\widehat{\widehat{m}}_h(t)=\sum_{i=1}^{n}w_{n,h}(t,t_i)(y_i-\zeta_i^T\widehat{\widehat{\beta}}_h)$, where $\widehat{\widehat{\beta}}_h=\left(\widetilde{\mathbf{X}}^T\Psi_*^{-1}(\widehat{\rho})\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^T\Psi_*^{-1}(\widehat{\rho})\widetilde{\mathbf{y}}$, $\widehat{\rho}$ is a consistent estimator of ρ and $\Psi_*^{-1}(\rho)=(1-\rho^2)\Psi^{-1}(\rho)$. We obtain:

Theorem 2.4

a) $n^{\nu/(2\nu+1)}(\widehat{\widehat{m}}_{h_0}(t)-m(t))\xrightarrow{D}N(B,V)$.

b) $n^{\nu/(2\nu+1)}(\widehat{\widehat{m}}_{\widehat{h}_n}(t)-m(t))\xrightarrow{D}N(B,V)$.

In this case (ρ unknown), we propose to estimate $m^{(\nu)}(\cdot)$ by $\widehat{m}^*_{h,b,\nu}(t)=\frac{1}{b^{\nu+1}}\sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}\overline{K}^{(\nu)}(\frac{t-u}{b})du(y_i-\zeta_i^T\hat{\beta}_h^*)$, where $\hat{\beta}_h^*=\left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{y}}$, and we obtain:

Theorem 2.5

$$\sup_{t\in[w_1,w_2]}\left|E\left[\widehat{m}^*_{h,b,\nu}(t)\right]-m^{(\nu)}(t)\right| \underset{n\to\infty}{\rightarrow} 0$$

Theorem 2.6

$$\sup_{t \in [w_1, w_2]} \left| \widehat{m}^{*}_{h,b,v}(t) - m^{(v)}(t) \right| \xrightarrow{P} 0.$$