

## TECNICAS NO PARAMETRICAS DE ESTIMACION FUNCIONAL, CON OBSERVACIONES DEPENDIENTES\*

Juan M. VILAR FERNANDEZ  
y  
Alejandro QUINTELA DEL RIO

*Universidad de La Coruña*

*En este trabajo se exponen técnicas de estimación no paramétrica de las siguientes curvas de probabilidad: función de densidad, regresión con diseño fijo y diseño aleatorio, y función de autorregresión. Centrando la exposición en los estimadores tipo núcleo o kernel, se investigan las modificaciones que deben realizarse para obtener las estimaciones de estas curvas en el supuesto de que los datos muestrales sean dependientes.*

*Se presentan distintos ejemplos de aplicación de las técnicas descritas a conjuntos de datos económicos.*

### 1. Introducción

Las técnicas de estimación no paramétrica de curvas notables en problemas de probabilidad (densidad, regresión, distribución, razón de fallo...) son métodos de suavización que permiten estimar la curva en estudio, a partir de una muestra o conjunto de observaciones. Referencias clásicas sobre esta metodología son los trabajos de Rosenblatt (1956), Parzen (1962), Nadaraya (1964) y Watson y Leadbetter (1964), aunque ha sido en los últimos años cuando se ha estudiado más en profundidad, tanto en su aspecto teórico como de aplicación en diversos campos (economía, bioestadística, meteorología...). Es por ello que, hoy en día, estas técnicas constituyen una herramienta potente y sencilla para el análisis de datos, siendo en algunos casos complemento y en otros una alternativa válida a las técnicas clásicas de estimación paramétrica, donde se presupone que la curva tiene una determinada forma funcional dependiendo de unos parámetros que hay que estimar. Monografías recientes sobre este tema —que incluyen gran cantidad de ejemplos de aplicación, muchos de los cuales no pueden, de forma eficiente, ser tratados por otros métodos— son las siguientes: Silverman (1986), Bierens (1987) y Hardle (1990). También merece destacarse, por su orientación hacia la Econometría, la recopilación de Ullah (1989).

La mayoría de los estudios sobre estimación no paramétrica de curvas suponen que las observaciones muestrales son independientes, hipótesis no siem-

\* Deseamos agradecer la labor de corrección efectuada por dos evaluadores anónimos, cuya cuidadosa lectura del trabajo original motivó una versión mejorada del mismo.

pre muy realista, principalmente en muchos problemas econométricos, donde los datos son recogidos secuencialmente, lo que lleva a situaciones de dependencia. En cualquier caso, estudios recientes de Roussas (1988), Györfi y otros (1990) y Vilar Fdez. (1989a), entre otros, muestran el buen comportamiento de los estimadores no paramétricos bajo hipótesis de dependencia, presentando las mismas propiedades asintóticas que en el caso de independencia, aunque para obtener buenos resultados con tamaños muestrales finitos su tratamiento debe ser diferente, como se verá más adelante.

En este trabajo se exponen las técnicas de estimación no paramétrica, tipo núcleo o kernel, de la función de densidad en el apartado II, de la función de regresión con diseño fijo y aleatorio en el apartado III, y de la función de predicción para Series de Tiempo en el apartado IV, bajo condiciones de dependencia, indicando los principales problemas en su aplicación y la forma de solventarlos. Para ilustrar el estudio y comprobar el buen comportamiento de los estimadores no paramétricos, se han aplicado los diferentes métodos descritos a dos series de tiempo económicas:  $X$  = «cotización del oro» en pesetas por gramo, e  $Y$  = «cotización de la plata» en pesetas por kilo, ambas consideradas en el mercado de Madrid, durante el período comprendido desde el 1 de abril de 1991 hasta el 24 de enero de 1992 (198 datos)<sup>1</sup>. En el Gráfico 1 pueden verse estas dos series de datos.

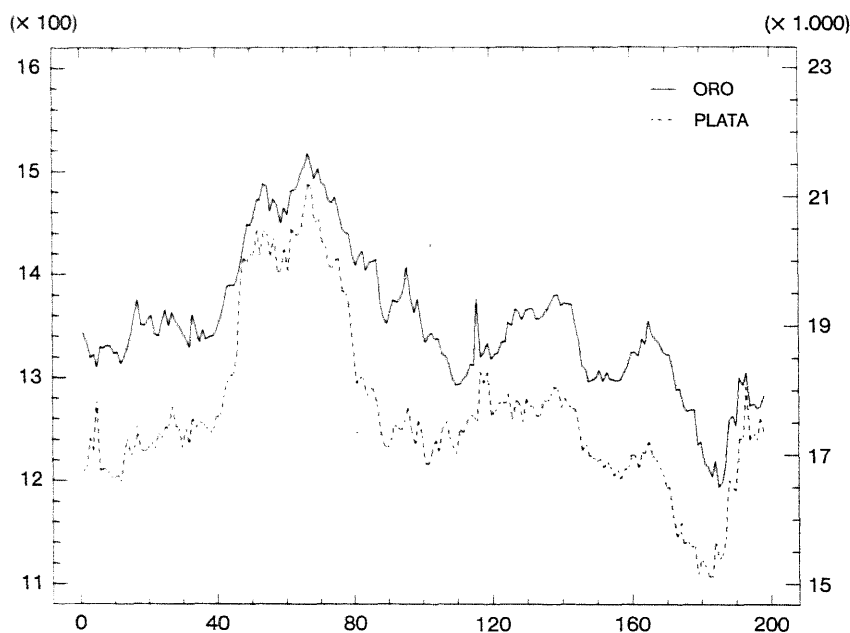


Gráfico 1  
Cotización diaria del Oro y de la Plata.

<sup>1</sup> Para el tratamiento numérico de los datos y su posterior representación gráfica, se han utilizado los paquetes estadísticos STATGRAPHICS y XPLORE, así como software propio, desarrollado por los autores del trabajo.

## 2. Estimación no paramétrica de la función de densidad

Sea  $X$  una variable aleatoria continua real, con función de densidad  $f(x)$ , la cual se desea estimar en un intervalo  $[a, b]$ , a partir de una muestra  $X_1, X_2, \dots, X_n$  de datos no necesariamente independientes. La mayoría de los estimadores no paramétricos de  $f$  pueden escribirse en la forma:

$$\hat{f}_n(x) = n^{-1} \sum_{i=1}^n \delta_h(x, X_i) \quad [1]$$

siendo  $\{\delta_h(x, u)\}$  una sucesión de funciones de ponderación definidas en  $R^2$ , con valores en  $R$ , que indican el peso del dato  $X_i$  en la estimación de  $f(x)$ , tomando valores altos si  $X_i$  está próximo a  $x$ , y cero, o valores próximos a cero, en caso contrario.

Para la elección particular  $\delta_h(x, u) = h(n)^{-1} K[(x-u)/h(n)] = K_h(x-u)$ , se obtiene el *estimador tipo núcleo o kernel*

$$\hat{f}_n(x) = n^{-1} \sum_{i=1}^n K_h(x, X_i) \quad [2]$$

siendo  $K(v)$  una función real de variable real, con  $\int K(v)dv = 1$  (en la mayoría de los casos una función de densidad). Este estimador también se conoce como estimador de Rosenblatt-Parzen y, por ser muy intuitivo y sencillo es, en general, más utilizado y estudiado, frente a otras elecciones del funcional  $\delta_h(x, u)$  que dan lugar a otros estimadores como el histograma, desarrollos ortogonales, spline...

La utilización del estimador [2] exige elegir una función núcleo,  $K$ , y el parámetro de escala  $h = h(n)$ . La elección del primero no tiene una influencia muy grande en los resultados de la estimación, utilizándose normalmente el núcleo gaussiano (densidad de una distribución normal estándar), el núcleo uniforme y, sobre todo, el núcleo de Epanechnikov, definido como sigue:

$$K(u) = 0.75(1-u^2) \text{ si } |u| < 1 \text{ y } 0 \text{ en otro caso} \quad [3]$$

que es óptimo, en el sentido de minimizar la expresión asintótica de la media del error cuadrático integrado entre las funciones núcleo positivas [ver Hodges y Lehman (1956)]. Por ello, es fundamental elegir adecuadamente el parámetro de suavización,  $h(n)$ ; que se considerará óptimo si minimiza algún criterio de error elegido previamente. Normalmente se utilizan criterios globales basados en la norma  $L^2$ , como la media del error cuadrático integrado, *MECI*, definido por:

$$MECI(h) = E \left[ \int (\hat{f}_h(x) - f(x))^2 w(x) dx \right] \quad [4]$$

siendo  $w(x) \geq 0$  una función de peso. También se utilizan otras medidas de error, más sencillas de calcular, como el error cuadrático ponderado

$$ECP(h) = n^{-1} \sum_{i=1}^n [\hat{f}_k(X_i) - f(X_i)]^2 w(X_i) \quad [5]$$

o el error cuadrático integrado

$$ECI(h) = \int (\hat{f}_k(x) - f(x))^2 w(x) dx \quad [6]$$

Bajo condiciones generales, Marron y Härdle (1986) han demostrado que las tres medidas son asintóticamente equivalentes, esto es, que tienden a ser iguales cuando el tamaño muestral tiende a infinito.

Desarrollando la expresión (4), y suponiendo que la función  $w(x) = 1$ , se obtiene

$$MECI(h) = \int (\text{Sesgo}[\hat{f}_k(x)])^2 dx + \int \text{Var}[f_k(x)] dx \quad [7]$$

y haciendo un desarrollo de Taylor en la expresión del sesgo, se obtiene que

$$MECI(h) = AMECI(h) + o(h^4 + n^{-1}h^{-1}) \quad [8]$$

donde se denomina  $AMECI(h)$  a una aproximación del  $MECI(h)$ , asintóticamente equivalente, obtenida como la suma de los dos términos principales del desarrollo anterior, y que para el estimador núcleo viene dada por

$$AMECI(h) = (nh)^{-1} C_k + (1/4)h^4 D_k^2 \left( \int [f'''(u)]^2 du \right) \quad [9]$$

siendo  $C_k = \int K^2(u) du$  y  $D_k = \int u^2 K(u) du$ . El primer sumando es debido a la varianza y el segundo al cuadrado del sesgo (Silverman, 1986). En esta expresión se observa que el parámetro  $h$  sirve de balanza entre la varianza y el sesgo de la estimación, deduciéndose que la elección de un  $h$  pequeño supondrá construir un estimador con poco sesgo y mucha varianza, y elegir un  $h$  grande producirá el efecto contrario: un estimador con poca varianza y mucho sesgo.

Minimizando el  $AMECI(h)$  (derivando respecto a  $h$  e igualando a cero) se obtiene el óptimo, cuya expresión es:

$$h_{AMECI} = n^{-1/5} (4C_k)^{1/5} \left( D_k^2 \int [f'''(u)]^2 du \right)^{-1/5} \quad [10]$$

valor que no se puede calcular, ya que depende de la derivada segunda de la función teórica,  $f$ , que se desea estimar.

Este problema ha motivado que se desarrollen distintas técnicas para obtener la banda directamente de las observaciones, de forma que se minimice algún estimador de los criterios de error más usuales. Existen diferentes metodologías a seguir para buscar la ventana a partir de los datos. Por ejemplo, directa-

mente de la expresión [10], puede obtenerse una estimación del parámetro  $h_{AMECI}$  realizando una estimación del término desconocido  $\int f^{(2)}(u) du$ . Esto se conoce como metodología *plug-in*. Trabajos sobre el tema pueden verse, por ejemplo, en Scott y Factor (1981), y Sheather y Jones (1991).

Otros procedimientos aplican el remuestreo *bootstrap* para estimar el parámetro ventana a partir de los datos. La idea intuitiva de este método es bastante simple: estimar el error cuadrático *MECI* por medio de una versión *bootstrap* de la forma

$$MECI_*(h) = E_* \int [\hat{f}_h^*(t) - \hat{f}_g(t)]^2 dt$$

donde por  $E_*$  se denota la esperanza con respecto a una muestra *bootstrap*  $(X_1^*, \dots, X_n^*)$ , (obtenida a partir de la original mediante algún tipo de muestreo),  $g$  es alguna ventana auxiliar,  $\hat{f}_g$  es un estimador de la densidad que depende de la muestra original  $X_1, X_2, \dots, X_n$ , y  $\hat{f}_h^*$  es un estimador basado en  $(X_1^*, \dots, X_n^*)$ . Se elige entonces el valor de  $h$  que minimice el *MECI*\_. Las distintas versiones de esta metodología se diferencian en el procedimiento de elección del parámetro  $g$  y del procedimiento de generación de la muestra *bootstrap*. Trabajos importantes sobre este tema son los de Cao (1990) y Hall (1990).

Otros métodos de selección de la ventana utilizan puntos de vista diferentes. Son destacables el procedimiento de los *puntos de inflexión* (Cuevas y González Manteiga, 1990), el *método del doble núcleo* (Devroye, 1989), y los distintos métodos de *validación cruzada*, que desarrollaremos inmediatamente. Trabajos muy interesantes, que realizan una descripción de algunos de los procedimientos citados, y además una comparación, mediante estudios de simulación, de los mismos, son los de Jones, Marron y Sheather (1992), y Cao, Cuevas y González Manteiga (1992).

Estos métodos expuestos han tenido su origen en la búsqueda de la banda para estimadores no paramétricos de la densidad, cuando los datos de la muestra en estudio son independientes (muestras aleatorias simples). Las distintas propiedades de optimalidad obtenidas teóricamente para los mismos hacen uso de esta restricción. Únicamente los últimos (métodos de validación cruzada) han sido estudiados bajo la hipótesis de dependencia entre los datos muestrales.

El más utilizado de estos métodos es el de validación cruzada de mínimos cuadrados («least squares cross-validation» o, simplemente, *cross-validation*), introducido por Rudemo (1982) y Bowman (1984), que se basa en minimizar una estimación del *ECI*. En efecto, desarrollando (6), (suponiendo  $w = 1$ ) se obtiene:

$$ECI(h) = \int \hat{f}_h^2 - 2 \int f_h \hat{f} + \int f^2 \quad [11]$$

Dado que el tercer sumando de esta expresión no depende de  $h$ , puede verse [consultar fórmula (3.36) de Silverman (1986)] que minimizar [11] equivale a minimizar la siguiente función de validación cruzada:

$$CV_1(h) = \int \hat{f}_h^2(x) dx - 2n^{-1} \sum_{i=1}^n \hat{f}_h(X_i) \quad [12]$$

siendo

$$\hat{f}_h(X_j) = \frac{1}{n-1} \sum_{i \neq j} h_h(X_j - X_i) \quad [13]$$

ésto es, la estimación de  $f(X_j)$  utilizando toda la muestra excepto el dato  $X_j$ , (técnica denominada «leave-one-out»).

En el supuesto de que los datos muestrales sean dependientes, la técnica descrita no proporciona buenos resultados, ya que si los datos están correlacionados positivamente, al estimar  $f(X_j)$  mediante  $\hat{f}_h^j(X_j)$  según [13], los datos próximos temporalmente a  $X_j$ :  $X_{j-l}, \dots, X_{j-1}, X_{j+1}, X_{j+2}, \dots, X_{j+l}$  proporcionan información sobre  $X_j$  que viene dada por la dependencia de la muestra y no por su estructura probabilística, lo que lleva a elegir en [12] bandas pequeñas y, por tanto, poco suavizadas. Si los datos están correlacionados negativamente, el razonamiento es el inverso y se obtienen estimaciones sobresuavizadas.

Para evitar este problema, Hart-Vieu (1990) proponen calcular la estimación de  $f(X_j)$  eliminando de la muestra los datos próximos (en el tiempo) a  $X_j$ ; en consecuencia, los que están altamente correlacionados con él. Un criterio menos estricto se basa en ponderar los datos con peso inferior cuanto mayor sea su dependencia con  $X_j$ , lo que lleva a elegir la banda que minimice la siguiente función

$$CV_2(h) = \int \hat{f}_h^2(x) w(x) dx - 2n^{-1} \sum_{j=1}^n \hat{f}_h^j(X_j) w(X_j) \quad [14]$$

siendo ahora

$$\hat{f}_h^j(x) = (\bar{\gamma})^{-1} \sum_{i=1}^n K_h(x - X_i) \gamma(|i-j|) \quad [15]$$

con  $\gamma(u)$  una función que ha de verificar:

$$\gamma(u) = \begin{cases} 0 & \text{si } u = 0 \\ 1 & \text{si } u > l_n \end{cases} \quad [16]$$

$0 \leq \gamma(u) \leq 1$  si  $u \leq l_n$  y siendo

$$\bar{\gamma} = \sum_{i=1}^n \gamma(|i-j|)$$

El valor  $l_n$  arriba especificado nos proporcionará la distancia (temporal) en que dos datos pueden ser tratados como si fuesen independientes. Por tanto, es importante la elección de la función  $\gamma(u)$ , ya que clasifica los datos en función de su dependencia (cercanía en el tiempo). Este criterio de validación cruzada, y otro basado en la validación cruzada de Kullback-Leibler, han sido estudiados por Quintela y Vilar (1991), donde se propone la utilización de funciones  $\gamma(u)$  en las que el valor de  $l_n$  se obtiene directamente de los datos, por ejemplo:  $\gamma(u) = 1 - |\hat{r}_n(u)|$  si  $|\hat{r}_n(u)| > 2/\sqrt{n}$ , siendo  $\hat{r}_n(u)$  una estimación de la autocorrelación de orden  $u$ .

La elección de la función  $\gamma(u)$  y del parámetro  $l_n$  tiene una influencia mucho menor en los resultados de la estimación que la elección de la banda  $h$ , no existiendo un método que permita, a partir de la muestra, obtener el parámetro óptimo  $l_n$ , en el sentido de que minimice la función  $CV_2(h)$  [14]. En cualquier caso, si la función  $\gamma(u)$  elegida cumple las propiedades arriba especificadas, se demuestra [ver Quintela y Vilar (1991)], bajo condiciones generales, que el parámetro de suavización elegido minimizando  $CV_2(h)$  es óptimo, en el sentido siguiente:

$$\frac{ERR(h_{CV})}{\inf ERR(h)} \rightarrow 1 \text{ casi seguro}$$

con  $ERR(h) = MECI(h)$ ,  $ECI(h)$  ó  $ECP(h)$ ; es decir, que el error cometido utilizando la ventana de validación cruzada tiende a aproximarse al menor error

posible. Análogamente, se prueba que la razón  $\frac{h_{CV}}{h_{AMECI}} \rightarrow 1$  casi seguro.

De todo lo expuesto se sigue que, aunque existen distintos métodos de selección automática del parámetro de suavizado, no puede afirmarse que esté resuelto satisfactoriamente el problema, sobre todo en un contexto de dependencia. El método de validación cruzada ha sido el más utilizado, pudiéndose modificar en la forma descrita para trabajar con muestras de datos dependientes, aunque presenta serios inconvenientes, tales como su alta variabilidad respecto a la muestra. Además, en los estudios comparativos mediante simulación realizados recientemente [Jones y otros (1992), Cao y otros (1992)], es claramente superado por otros métodos, fundamentalmente los basados en la metodología *plug-in* y *bootstrap*.

Como ejemplo, realizamos una estimación de la función de densidad de la variable  $Y$  «cotización de la plata» a partir de la muestra representada en el Gráfico 1. Para tener una primera idea de la forma de esta densidad se ha calculado el histograma y ajustado una curva gaussiana. Ambas aparecen en el Gráfico 2.

En este gráfico se observa que el ajuste de una distribución normal no es adecuado, ya que hay falta de simetría y dos probables modas en torno a una marcada moda central.

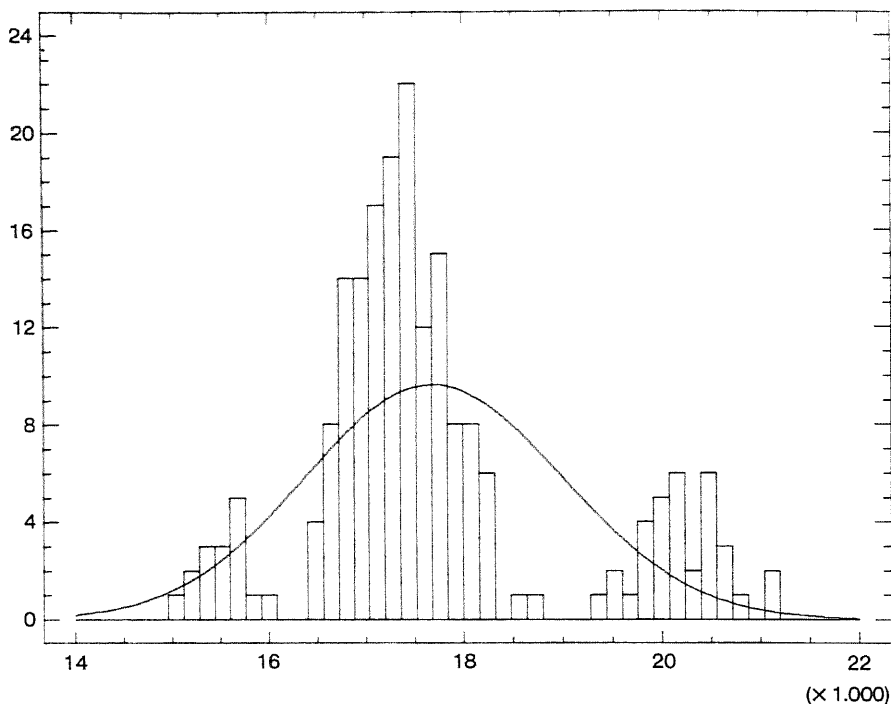


Gráfico 2  
Histograma de la cotización de la plata.

Para utilizar la estimación núcleo definida en [2] se ha obtenido la banda que minimiza la función de validación cruzada  $CV_1(h)$  [12], proporcionando  $h_{CV_1} = 466.196$ , que permite obtener la densidad estimada representada en el Gráfico 3.

De la definición de la variable  $Y$  y del estudio de la función de autocorrelación muestral se deduce la existencia de una fuerte dependencia entre los datos muestrales. Del análisis de los mismos se obtiene que la serie puede ajustarse por un modelo autorregresivo de orden uno, con un polinomio asociado que tendría una raíz próxima a la unidad, lo que permite utilizar la estimación no paramétrica. Un estudio alternativo, también válido, consistiría en diferenciar primero los datos  $y$ , posteriormente, realizar el estudio no paramétrico. Hemos elegido la primera opción, por constituir un buen ejemplo de utilización de la estimación no paramétrica con observaciones fuertemente dependientes, y permitir la comparación entre las técnicas existentes para datos independientes y las mejoras de las mismas en el supuesto de dependencia.

Para eliminar los efectos de la dependencia muestral, se ha calculado la función de validación cruzada  $CV_2(h)$ , utilizando como función de peso  $\gamma(u) = 0$  si  $|u| \leq l_n$  y  $\gamma(u) = 1$  en otro caso, con  $l_n = 5$ . Se ha elegido este valor porque las autocorrelaciones muestrales mayores que 5 son próximas a cero y no signifi-

( $\times 1E-4$ )

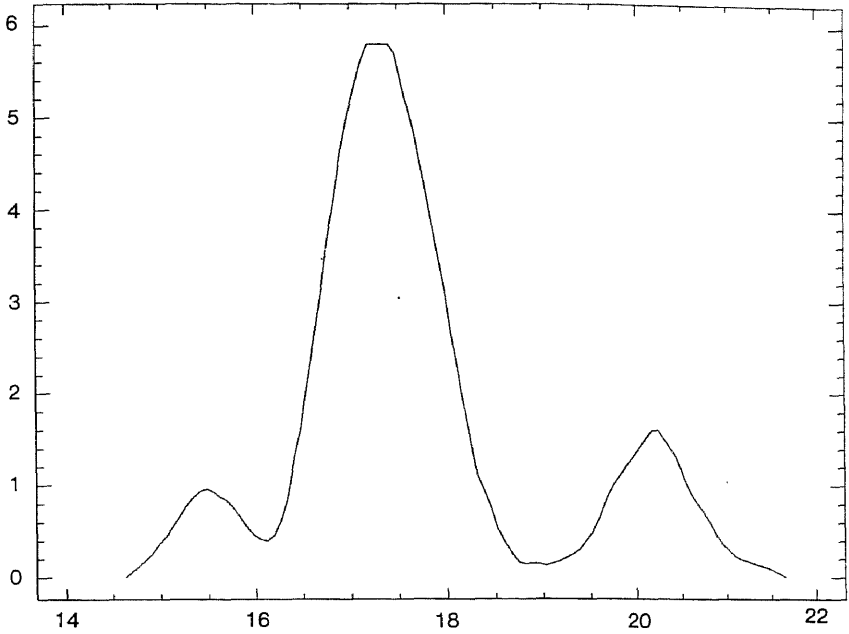


Gráfico 3

( $\times 1.000$ )

Estimación núcleo de la densidad de la plata  $h = 466.196$ .

( $\times 1E-4$ )

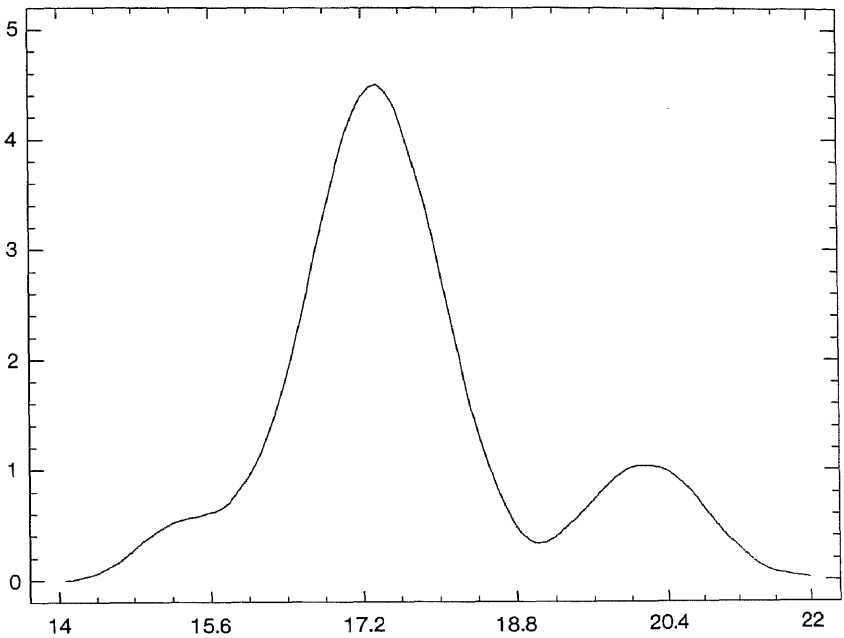


Gráfico 4

( $\times 1.000$ )

Estimación núcleo de la densidad de la plata,  $h = 983.328$ .

cativas. Por su sencillez de cálculo, esta función será la que se utilice en todo el trabajo.

El valor de  $h$  que minimiza la función  $CV_2(h)$  es  $h_{CV_2} = 983.328$ , a partir del cual se obtiene la estimación núcleo de la densidad de la variable, representada en el Gráfico 4, que es más suave que la obtenida anteriormente, en el supuesto de independencia.

### 2.1. Estimación local de la función de densidad

Si se desea estimar la función de densidad en un punto determinado  $x_0 \in [a, b]$ , fácilmente pueden deducirse versiones locales de los criterios para la obtención del parámetro de suavización descritos en [12], para el caso de datos independientes, y en [14], para el caso más general de datos dependientes. Basta con tomar como función peso  $w(y) = W_{n,x_0}(y)$ , una función que dependa del tamaño muestral ( $n$ ) y esté centrada en el punto en estudio  $x_0$ . Generalmente se elige como  $W_{n,x_0}(y) = I_S(y)$ , la función indicadora en el conjunto  $S$ , siendo  $S$  un entorno del punto  $x_0$  de radio tendiendo a cero, o bien funciones gaussianas, esto es  $W_{n,x_0}(y)$  la densidad asociada a una distribución Normal de media  $x_0$  y desviación típica tendiendo a cero. Bajo hipótesis de independencia de la muestra, este algoritmo local ha sido estudiado por Hall-Schucany (1989) y Mielniczuk y otros (1989) y, en un contexto de dependencia, por Quintela-Vilar (1992), quienes demuestran que la estimación local de la densidad obtenida en cada punto del intervalo  $[a, b]$  es asintóticamente mejor (tiene un menor error cuadrático integrado medio) que la obtenida por cualquier método global, esto es utilizando un único parámetro de suavización. Evidentemente, la utilización de ventanas diferentes en cada punto de estimación produce estimadores que no son densidades; son, por tanto, menos suaves que los estimadores con una única banda  $h$  para todos los puntos de estimación, pero tienden a acercarse más a la verdadera densidad en dichos puntos.

A continuación, aplicamos el criterio local descrito a los datos de la variable  $Y =$  «cotización de la plata» utilizados anteriormente. Para ello se calcula una banda local en cada uno de cincuenta puntos equidistantes del intervalo muestral (rango de los datos) considerado, siendo la función de peso local  $W_{n,x_0}(y) = I_S(y)$ , con  $S$  el intervalo de centro  $x_0$  y radio la desviación típica muestral dividida entre dos. Como función  $\gamma(u)$  se tomó la utilizada en el caso global, con  $l_n$  igualmente 5. Las ventanas locales obtenidas se representan en el Gráfico 5, donde se observan cambios notables, tomando valores altos en los extremos del intervalo en estudio y bajos en el centro, lo que es razonable ya que del histograma de la muestra se deduce que la mayor densidad de los datos se encuentra en el centro. Por tanto, valores pequeños de  $h$  permiten tener un número suficiente de observaciones para estimar la densidad; por el contrario, en los extremos del intervalo, al existir menor densidad de datos muestrales, se necesitará una ventana grande para tener un número equivalente de observaciones.

( $\times 1.000$ )

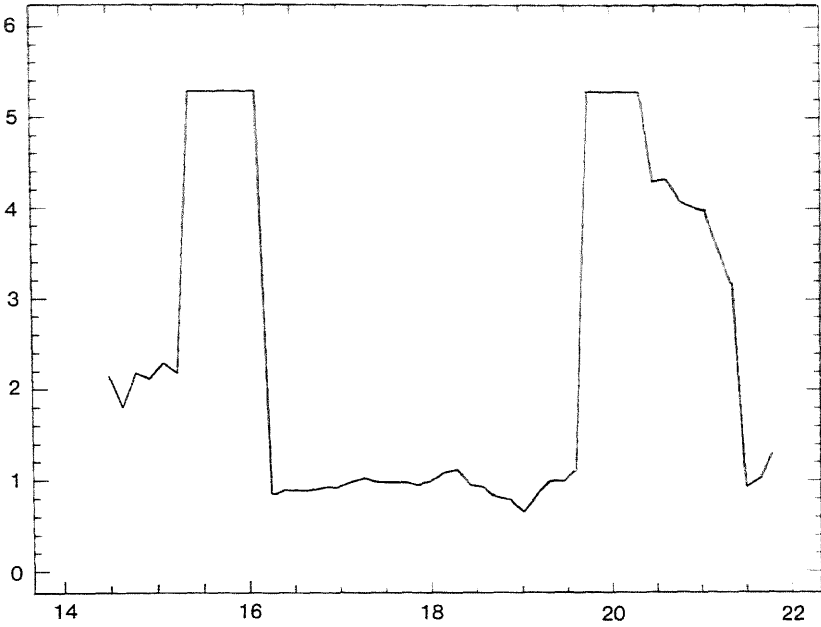


Gráfico 5

( $\times 1.000$ )

Ventanas locales para la estimación de la densidad de la cotización de la plata.

( $\times 1E-4$ )

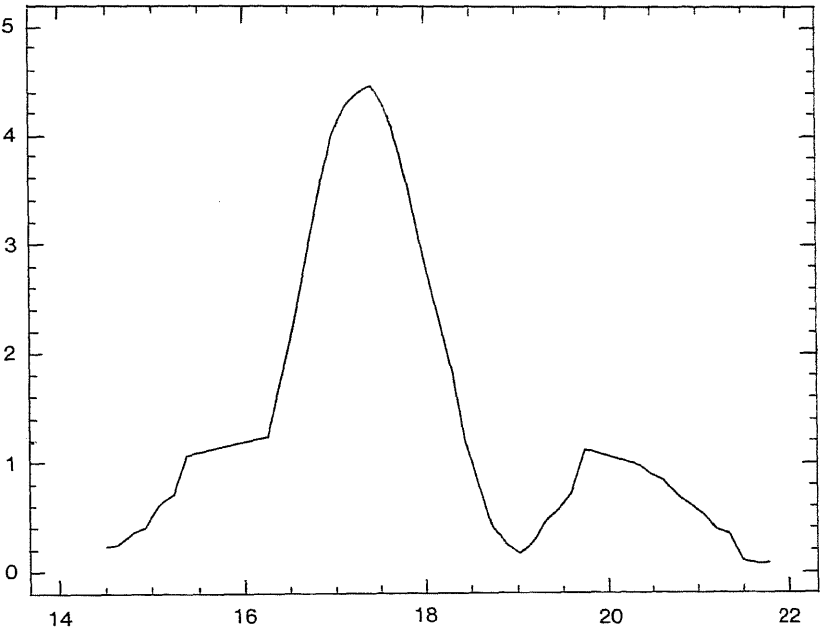


Gráfico 6

( $\times 1.000$ )

Estimación núcleo de la densidad de la plata utilizando ventanas locales.

La densidad de la plata estimada con bandas locales se representa en el Gráfico 6. Globalmente produce una estimación menos suave que la obtenida utilizando un único valor para  $h$  (Gráfico 4) pero, probablemente, más aproximada a la densidad teórica desconocida.

### 3. Estimación no paramétrica de la función de regresión

En este apartado se aborda el problema clásico de estimar una curva de regresión, la cual describe una relación general entre una variable explicativa  $X$  y una variable respuesta  $Y$ , que se supondrán reales. A partir de un conjunto de pares de datos muestrales  $\{(X_i, Y_i)\}_{i=1}^n$ , verificando la relación:  $Y_i = m(X_i) + \varepsilon_i$   $i = 1 \dots n$  siendo  $m(x)$  la función de regresión desconocida y los  $\varepsilon_i$  los errores aleatorios, las técnicas no paramétricas de estimación de  $m(x)$  se basan en suavizar la nube de puntos por procedimientos de medias locales, siendo el formato general de estos estimadores el siguiente:

$$\hat{m}_h(x) = \sum_{i=1}^n \delta_{n,i}(x) Y_i \quad [17]$$

siendo  $\{\delta_{n,i}(x)\}$  una sucesión de pesos que depende del conjunto  $\{X_i\}_{i=1}^n$ , tomando valores altos si  $X_i$  está próximo a  $x$  y próximos a cero, o cero, en caso contrario. Este estimador también puede entenderse como el que se obtiene utilizando «mínimos cuadrados localmente ponderados», esto es,  $\hat{m}_h(x)$  es la solución del siguiente problema de optimización:

$$\text{mín. } \left\{ \theta: \sum_{i=1}^n \delta_{n,i}(x) (Y_i - \theta)^2 \right\} \quad [18]$$

En regresión pueden considerarse dos situaciones diferentes, según como se hayan generado los datos. Si  $X$  e  $Y$  son variables aleatorias, se denomina *modelo de diseño aleatorio* y la función de regresión se define por  $m(x) = E(Y/X = x)$ .

Si se controla la variable regresora  $X$ , se tiene un *modelo de diseño fijo*, de la forma:  $Y_i = m(X_i) + \varepsilon_i$   $i = 1 \dots n$ , siendo las  $\varepsilon_i$  variables aleatorias de media cero y varianza  $\sigma^2$ . Normalmente los  $\{X_i\}_{i=1}^n$  se eligen equidistribuidos en un intervalo  $[a, b]$  y, sin pérdida de generalidad, se puede asumir que  $X_i = i/n$ . Aunque el mecanismo estocástico de ambos modelos es diferente, las técnicas de suavización son análogas en ambos. En cualquier caso se estudiarán los dos modelos de forma diferenciada.

#### 3.1. Modelo de diseño fijo

Un estimador no paramétrico, tipo núcleo, muy utilizado para este modelo, es el propuesto por Priestley-Chao (1972), quienes, bajo el supuesto de que

los datos  $\{X_i\}_{i=1}^n$  están equiespaciados, consideran los pesos  $\delta_{n,i}(x) = (X_{i+1} - X_i) K_h(x - X_i)$ , dando lugar al estimador:

$$\hat{m}_h(x) = h(n)^{-1} \sum_{i=1}^n (X_{i+1} - X_i) K\left(\frac{x - X_i}{h(n)}\right) Y_i \quad [19]$$

siendo  $h(n)$  el parámetro de suavización.

Una generalización de este estimador es la propuesta por Gasser-Müller (1979)

$$\hat{m}_h(x) = \sum_{i=1}^{n-1} Y_i \int_{S_i}^{S_{i+1}} K_h(x - u) du \quad [20]$$

con  $X_i \leq S_i \leq X_{i+1}$ . Este estimador recibe también el nombre de *estimador convolución*, ya que trata de estimar la función convolución

$$(m \star K)(x) = \int m(x - hy) K(y) dy = h^{-1} \int K\left(\frac{x - u}{h}\right) m(u) du$$

El buen comportamiento de este estimador se sigue de su error cuadrático medio dado por (ver Hardle, 1990):

$$\begin{aligned} ECM(x) &= E(\hat{m}_h(x) - m(x))^2 = \\ &= [\text{Sesgo}(\hat{m}_h(x))]^2 + \text{Varianza}(\hat{m}_h(x)) = \\ &= \left(h^2 \frac{m''(x)}{2} D_k\right)^2 + \frac{\sigma^2(x)}{nh} C_k + o\left(h^4 + \frac{1}{nh}\right) \end{aligned} \quad [21]$$

En esta expresión puede observarse, de nuevo, la importancia del parámetro de suavización,  $h(n)$ , que, al igual que en todos los estimadores no paramétricos funcionales, hace de balanza entre el sesgo y la varianza de la estimación, como ya se había discutido en la estimación de la densidad.

Por consiguiente, es fundamental elegir una banda adecuada para obtener buenas estimaciones. Para el estimador de Gasser-Müller [20], minimizando la expresión asintótica de su error cuadrático global:  $MECI = \int ECM(x) dx = \int [m_h(x) - m(x)]^2 dx$  se obtiene la expresión del parámetro  $h$  óptimo, dada por:

$$h_{opt} = n^{-1/5} \left(C_k \int \sigma^2(x) dx\right)^{-1/5} \left(D_k \int m''(x) dx\right)^{-2/5} \quad [22]$$

que, desafortunadamente, no puede calcularse porque depende de la función que deseamos estimar. Esto ha motivado el estudio de técnicas, similares al caso de la densidad (aunque, por el momento, el número de técnicas existentes es sensiblemente menor), para obtener la banda a partir de los datos

muestrales, siendo la más utilizada la de validación cruzada, que se basa en minimizar la función:

$$CV_s(h) = n^{-1} \sum_{i=1}^n (Y_i - \hat{m}_h^i(X_i))^2 w(X_i) \quad [23]$$

con  $w(u)$  una función peso positiva y  $\hat{m}_h^i(X_i)$  la estimación de  $m(X_i)$  a partir de la muestra sin el dato  $(X_i, Y_i)$ .

En el supuesto de que los errores sean dependientes el método anterior no proporciona buenos resultados si las correlaciones son altas, por ello y teniendo en cuenta la similitud entre la estimación de la densidad y de la regresión, Quintela (1992) propone definir  $\hat{m}_h^i(x)$  como sigue:

$$\hat{m}_h^i(x) = \frac{n}{\bar{\gamma}} \sum_{j=1}^{n-1} \left\{ \int_{S_j}^{S_{j+1}} K_h(x-u) du \right\} \gamma(|i-j|) Y_j \quad [24]$$

siendo  $\gamma(u)$  la función definida en [16]. De esta manera se ponderan con menor intensidad las observaciones  $Y_j$  que son más dependientes de  $Y_i$ .

Como ejemplo de utilización de los métodos no paramétricos descritos para estimar la función de regresión en un modelo de diseño fijo, se ha ajustado a la variable  $Z = \text{«Tasa de crecimiento de la cotización de la plata»}$ , definida a partir de la variable  $Y = \text{«cotización de la plata»}$  como sigue

$$Z_t = \frac{Y_{t+1} - Y_t}{Y_t} = \frac{Y_{t+1}}{Y_t} - 1$$

un modelo del tipo  $Z_t = m(t/n) + \varepsilon_t$ ,  $t = 1, 2, \dots, n$ . Como las autocorrelaciones muestrales son, para esta variable, muy pequeñas, se ha calculado la banda minimizando la función de validación cruzada definida en [23], con  $m_h^i(x)$  dado por [24] y  $\gamma(u)$  la misma que se usó en la sección anterior. Se ha obtenido, para  $l_n = 0$  (caso de independencia) un  $h = 26'432$  y para  $l_n = 1$  (caso de dependencia débil) un  $h = 9'563$ . En el Gráfico 7 se ha representado la nube de puntos  $(t, Z_t)$  y las estimaciones no paramétricas de la regresión utilizando el estimador de Gasser-Müller con las bandas calculadas, siendo la estimación con mayor variabilidad la obtenida al hacer  $l_n = 1$  y la más suave (trazo grueso) la construida bajo hipótesis de independencia. Por consiguiente, el utilizar las modificaciones para datos dependientes, cuando estos no lo son, puede empeorar los resultados de la estimación. En la citada figura no se observa una clara tendencia de la función de regresión de la tasa de crecimiento de la cotización de la plata, siendo aceptable la hipótesis de que la serie es ruido blanco.

### 3.2. Modelo de diseño aleatorio

En el supuesto de que  $X$  e  $Y$  sean variables aleatorias, se define la función de regresión  $m(x) = E(Y/X) = \int yf(y/x) dy = f(x)^{-1} \int yf(x, y) dy$  y, aunque tanto el

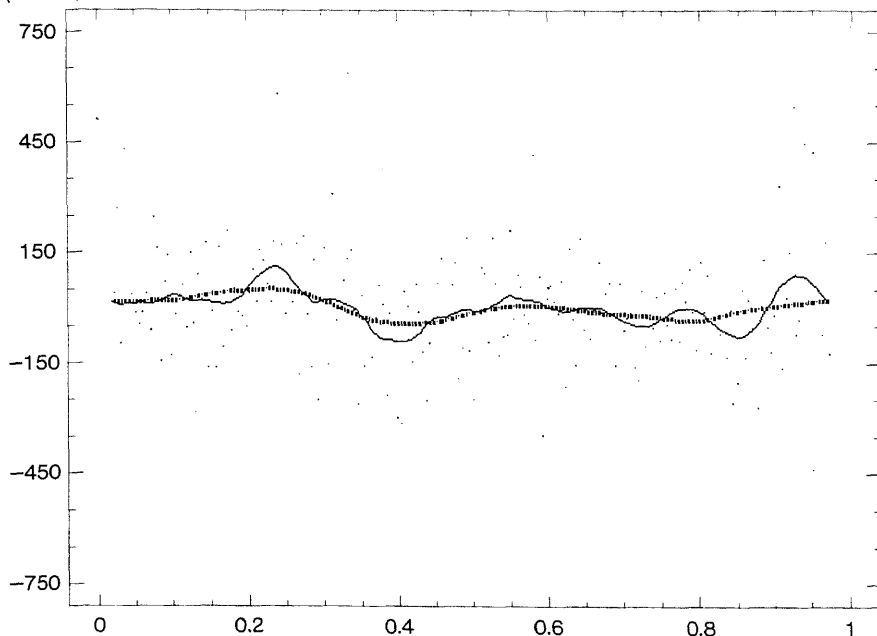
( $\times 1E-4$ )

Gráfico 7

Tasa de crecimiento de la cotización de la plata: datos muestrales y estimaciones de la regresión.

estimador de Priestley-Chao [19], como el de Gasser-Müller [20] admiten generalizaciones muy simples para adaptarse a este contexto, el estimador núcleo más utilizado es el de *Nadaraya-Watson*,<sub>+</sub> definido a partir de los pesos:

$$\delta_{n,i}(x) = \frac{1}{\hat{f}_h(x)} \frac{1}{n} K_h(x - X_i)$$

siendo  $\hat{f}_h(x)$  el estimador núcleo de la densidad de Rosenblatt-Parzen estudiado en la sección 2. El estimador toma así la forma

$$\hat{m}_h(x) = \frac{\sum_{i=1}^n K_h(x - X_i) Y_i}{\sum_{i=1}^n K_h(x - X_i)} = \frac{n^{-1} \sum_{i=1}^n K_h(x - X_i) Y_i}{\hat{f}_h(x)} \quad [25]$$

El error cuadrático medio de este estimador, bajo condiciones generales, viene dado por (Vieu, 1991b):

$$ECM(x) = \left( h^2 \frac{m''(x)f(x) + 2m'(x)f'(x)}{2f(x)} D_k \right)^2 + \frac{\sigma^2(x)}{nhf(x)} C_k + o\left( h^4 + \frac{1}{nh} \right) \quad [26]$$

y de nuevo, se observa en esta expresión la importante influencia de la banda, obteniéndose la óptima (en el sentido de minimizar el error cuadrático) al igualar a cero la primera derivada de la expresión asintótica del error, que va a depender de las funciones desconocidas  $f(x)$  y  $m(x)$ .

Härdle y Vieu (1990) proponen utilizar, para obtener la banda a partir de los datos, la función de validación cruzada definida en [23], adaptándola para el caso de dependencia. Así, definen el estimador  $\hat{m}_h^i(x)$  como sigue:

$$\hat{m}_h^i(x) = \frac{\sum_{j=1}^n K_h(x - X_j) \gamma(|i-j|) Y_j}{\sum_{j=1}^n K_h(x - X_j) \gamma(|i-j|)} \quad [27]$$

Como ejemplo de aplicación, se ha obtenido la estimación de la función de regresión de la «cotización de la plata» en función de la «cotización del oro». Calculando la banda por validación cruzada y suponiendo independencia ( $l_n = 0$ ) se obtiene  $h = 31'68$ . Con el mismo algoritmo pero suponiendo dependencia, lo que es más razonable dada la fuerte correlación de los datos, se obtiene para  $l_n = 5$  una banda mayor:  $h = 57.4$ , con la que se construye una estimación de la regresión más suave, como se observa en el Gráfico 8, en el

( $\times 1.000$ )

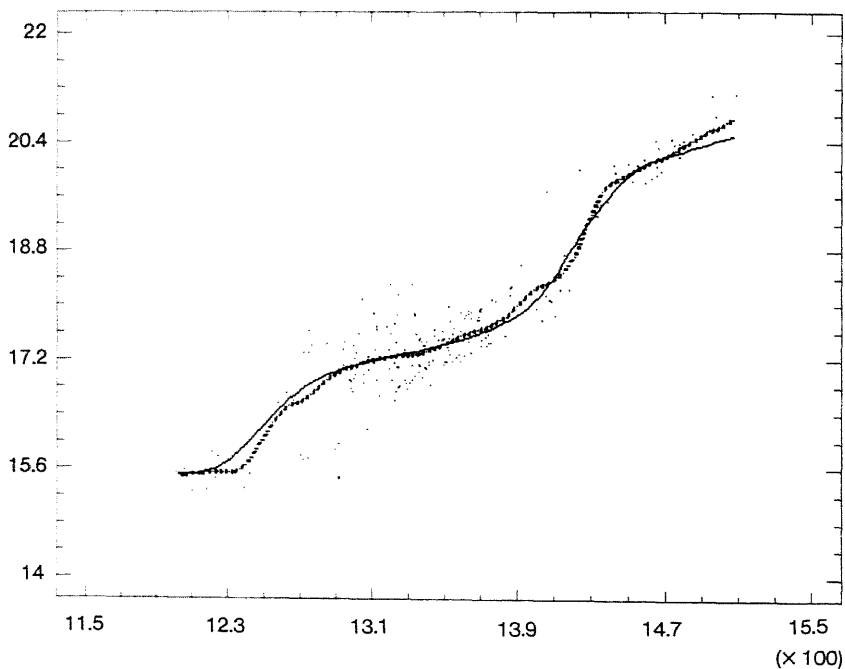


Gráfico 8

Estimación de la regresión de la plata respecto del oro.

que puede verse la nube de puntos y la regresión estimada. En trazo grueso se representa la estimación con  $h = 31'68$  y con trazo suave la otra estimación.

Una versión del estimador convolución para regresión con diseño aleatorio ha sido estudiada por Gasser y Engel (1990), obteniendo un estimador con mayor varianza que el de Nadaraya-Watson (1.5 veces superior), pero con un sesgo más fácil de interpretar, ya que sólo depende de  $m''(x)$  y no de  $f(x)$ . Además, se puede adaptar fácilmente para estimar las derivadas de  $m(x)$ , propiedad que no tiene el de Nadaraya-Watson.

Al igual que en la estimación de la densidad, se pueden obtener estimaciones de la función de regresión en un punto  $x$  utilizando versiones locales de los algoritmos de validación cruzada expuestos. Para ello se toman funciones peso centradas en el punto  $x$ , esto es,  $w(x) = W_{n,x_n}(x)$  [ver Vieu (1991a)].

La generalización de los algoritmos expuestos en los apartados I y II al caso  $p$ -dimensional, esto es, suponiendo que la variable  $X$  toma valores en  $R^p$ , es inmediata desde un punto de vista matemático, pero su aspecto computacional presenta grandes problemas. Esto es debido a que, bajo la hipótesis de que la función a estimar es  $k$  veces diferenciable, la banda óptima es de la forma  $h_{opt} = Cn^{-1/(2k+p)}$ , y, por tanto, la convergencia en media cuadrática es del orden  $n^{-4/(2k+p)}$ , que para valores grandes de  $p$  es muy lenta, lo que obliga a utilizar tamaños muestrales muy grandes para obtener resultados aceptables. Para solventar este problema, en la actualidad se están estudiando técnicas como la «Projection Pursuit Regression» (Hall, 1988) o los «Generalized Additive Models» (Hastie y Tibshirani, 1990).

#### 4. Estimación no paramétrica de la función de predicción

Uno de los principales problemas en el Análisis de Series Temporales es el de la predicción, esto es, estimar a partir de un conjunto de observaciones el valor de futuras observaciones de la serie. Este problema ha sido abordado recientemente utilizando técnicas no paramétricas, por Györfi y otros (1989) y Vilar Fdez. (1989b), entre otros. Todos los estudios se basan en utilizar como predictor la esperanza condicional; por ello, los métodos de estimación no paramétrica de la predicción de una serie temporal estacionaria están estrechamente relacionados con los de estimación no paramétrica de la regresión.

Sea  $\{X_t; t \in Z\}$  una serie de tiempo real y estacionaria. Para estimar el valor de  $X_{N+1}$  a partir de una muestra  $X_1, X_2, \dots, X_N$  se utiliza como función de predicción la función de autorregresión:

$$r(X_N^{(k)}) = E\{X_{N+1}/X_N^{(k)} = (X_{N-k+1} \dots X_N)\} \quad [28]$$

que, teóricamente, es la mejor función de predicción si el proceso  $X_t$  en estudio es markoviano de orden  $k$ . Si se consideran las variables  $X = X_N^{(k)} = (X_{N-k+1} \dots X_N)$  y  $Y = X_{N+1}$ , el problema de predicción radica en estimar la función de autorregresión, que no es más que un caso particular del pro-

blema de estimación de la regresión con diseño aleatorio, bajo la hipótesis de dependencia de los datos, tratado en la sección anterior.

El estimador de Nadaraya-Watson para este caso se deduce de la expresión [25], y es el siguiente:

$$\hat{r}_h(X_N^k) = \frac{\sum_{i=k}^{N-1} K_h(X_i^k - X_N^k) X_{i+1}}{\sum_{i=k}^{N-1} K_h(X_i^k - X_N^k)} \quad [29]$$

siendo  $K_h(u) = h(n)^{-k} K[u/h(n)]$ ,  $u \in R^k$  y  $K(u)$  una función núcleo  $k$ -dimensional, generalmente la densidad asociada a una variable  $k$ -dimensional.

De nuevo, en el cálculo de este estimador aparece como problema básico la elección de la banda,  $h(n)$ , y, dado que se trabaja con la hipótesis de dependencia, deben utilizarse los procedimientos descritos en el apartado anterior, generalizados para el caso  $k$ -dimensional. Por otra parte, si se utiliza un orden de autocorrelación alto, al calcular los estimadores no paramétricos surgen los problemas relacionados con la alta dimensionalidad de la variable regresora, que ya se han comentado.

Además, debe tenerse en cuenta que, en la solución del problema de predicción en una serie de tiempo a partir de una muestra, se pueden cometer dos tipos de errores: uno, el *probabilístico*, en el sentido de que el predictor que se haya elegido  $\{r(X_t)\}$  no sea bueno, y otro el *estadístico*, que se comete al aproximar el predictor por una estimación  $\{r_h(X_t)\}$ , ya que el error cuadrático de predicción puede descomponerse en dos sumandos como sigue:

$$E[X_{t+1} - \hat{r}_h(X_t)]^2 = E[X_{t+1} - r(X_t)]^2 + E[r(X_t) - \hat{r}_h(X_t)]^2 = P + E$$

Si se ha elegido un buen modelo de predicción todo el problema radica en estimar bien  $r(X_t)$ , esto es, en minimizar el error de estimación  $E$ . Empero, si no se ha elegido bien  $r(X_t)$ , aunque su estimación sea buena, los resultados de la predicción no lo serán, ya que se estará cometiendo un error probabilístico,  $P$ , grande. En consecuencia, si la función de autorregresión no es una buena función de predicción o el orden  $k$  no es el adecuado, las técnicas no paramétricas no darán buenos resultados.

Estas técnicas son fácilmente generalizables para obtener las predicciones de  $X_{N+s}$ , con  $s > 1$ , o valores intermedios de la muestra. Para ilustrar el comportamiento de las técnicas no paramétricas en este contexto, se ha estimado la función de autorregresión de la variable «cotización de la plata», de orden uno, que permite predecir el valor, en un instante, de la variable, en función de lo ocurrido en el instante anterior. En el Gráfico 9 se ha representado la nube de puntos muestrales  $(X_{t-1}, X_t)$  y las estimaciones no paramétricas obtenidas mediante el estimador de Nadaraya-Watson de la función de autorregresión. Las bandas se han seleccionado mediante la técnica de validación

cruzada, primero suponiendo independencia ( $l_n = 0$ ), de donde resulta  $h = 506$  (curva con trazo grueso) y después suponiendo dependencia (eligiendo  $l_n = 5$ ) y obteniendo  $h = 708$  (curva con trazo suave). De los resultados puede aceptarse la hipótesis de que la función de autorregresión es lineal.

( $\times 1.000$ )

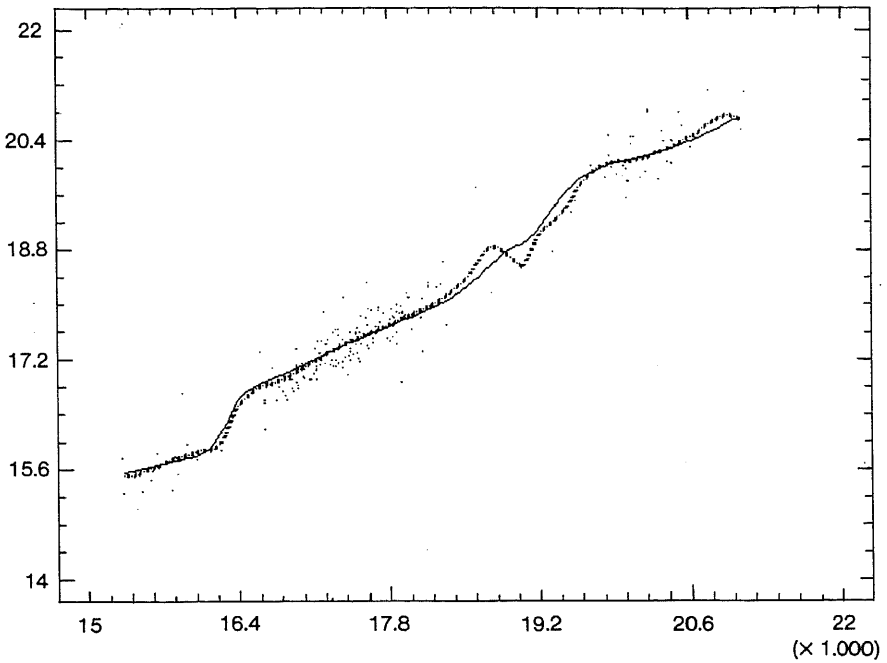


Gráfico 9  
Estimaciones de la autorregresión de la plata.

## Referencias

- Bierens, H. J. (1987): «Kernel estimators of regression functions», en *Advances in Econometrics*, Volume I. Truman F. Bewley, Cambridge University Press.
- Bowman, A. W. (1984): «An alternative method of cross-validation for the smoothing of density estimates», *Biometrika* 71, pp. 353-360.
- Cao Abad, R. (1990): «Aplicaciones y nuevos resultados del método *bootstrap* en la estimación no paramétrica de curvas», Tesis Doctoral. Universidad de Santiago de Compostela.
- Cao Abad, R.; Cuevas, A. y González Manteiga, W. (1992): «A comparative study of several smoothing methods in density estimation». A aparecer en *Computational Statistics and Data Analysis*.
- Cuevas, A. y González Manteiga, W. (1990): «Data-driven smoothing based on convexity properties», en *Nonparametric Functional Estimation and Related Topics*, G. G. Roussas et al. eds., pp. 225-240. Kluwer Academic Publishers.
- Devroye, L. (1989): «The double kernel method in density estimation», *Ann. Inst. Henri Poincaré* 25, pp. 533-580.
- Gasser, T. y Müller, H. G. (1979): «Kernel estimation of regression functions», en *Smoothing techniques for Curve estimation*, eds.: Gasser y Rosenblatt. Heidelberg: Springer-Verlag.
- Gasser, T. y Engel, J. (1990): «The choice of weights in kernel regression estimation», *Biometrika* 77, 2, pp. 377-381.
- Györfi, L.; Härdle, W.; Sarda, P. y Vieu, P. (1990): «Nonparametric Curve Estimation from Time Series», en *Lecture Notes in Statistics*, vol. 60, Springer-Verlag. New York/Berlin.
- Hall, P. (1988): «On Projection Pursuit Regression», Manuscrito no publicado.
- Hall, P. (1990): «Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems», *Journal of Multivariate Analysis* 32, pp. 177-203.
- Hall, P. y Schucany (1989): «A local cross-validation algorithm», *Statistics and Probability letters* 8, pp. 109-117.
- Härdle, W. (1990): *Applied Nonparametric Regression*. Oxford University Press.
- Härdle, W. y Vieu, P. (1990): «Kernel Regression smoothing of time series». Core discussion Paper, número 9031.
- Hart, J. y Vieu, P. (1990): «Data-driven bandwidth choice for density estimation based on dependent data», *The Annals of Statistics* 18, pp. 873-890.
- Hastie, T. J. y Tibshirani, R. J. (1990): *Generalized Additive Models*. Chapman y Hall, Londres.
- Hodges, J. L. y Lehmann, E. L. (1956): «The efficiency of some nonparametric competitors of the *t*-test», *Annals of Mathe. Statistics* 27, pp. 324-335.
- Jones, M. C.; Marron, J. S. y Sheather, S. J. (1992): «Progress in data-based bandwidth selection for kernel density estimation». *Preprint*.
- Marron, J. y Härdle, W. (1986): «Random aproximations of some measures of accuracy in nonparametric curve estimation», *Journal of Multivariate Analysis* 20, pp. 91-113.
- Mielniczuk, J.; Sarda, P. y Vieu, P. (1989): «Local data-driven bandwidth choice for density estimation», *Journal of Statistical Planning and Inference* 23, pp. 53-69.
- Nadaraya, E. A. (1964): «Some new estimates for distribution functions», *Theory Prob. Appl.*, 497-500.
- Parzen, E. (1962): «On estimation of a probability density function and mode», *Annals of Mathematical Statistics* 33, pp. 1065-76.
- Priestley, M. B. y Chao, M. T. (1972): «Nonparametric function fitting», *Journal of the Royal Statistical Society, Series B* 34, pp. 385-92.
- Quintela, A. (1992): «Cálculo del parámetro de suavización en la estimación no paramétrica de curvas con datos dependientes». Tesis Doctoral. Universidad de Santiago de Compostela.

- Quintela, A. y Vilar, J. M. (1991): «Técnicas de validación cruzada en la estimación de la densidad bajo condiciones de dependencia», *Qüestiuo* 15, 1, pp. 21-45.
- Quintela, A. y Vilar, J. (1992): «A local cross-validation algorithm for dependent data», *Test* 1, pp. 123-153.
- Rosenblatt, M. (1956): «A central limit theorem and a strong mixing condition», *Proc. Nat. Acad. Sci.* 42, pp. 43-47.
- Roussas, G. (1988): «Nonparametric estimation in mixing sequences of random variables», *Journal of Statistical Planning and Inference* 18, pp. 135-149.
- Rudemo, M. (1982): «Empirical choices of histograms and kernel density estimates», *Scandinavian Journal of Statistics* 9, pp. 65-78.
- Scott, D. W. y Factor, L. E. (1981): «Monte Carlo study of three data-based nonparametric probability density estimators», *Journal of the American Statistical Society* 76, pp. 9-15.
- Sheather, S. J. y Jones, M. C. (1991): «A reliable data-based bandwidth selection method for kernel density estimation», *Journal of the Royal Statistical Society, Ser. B* 53, pp. 683-690.
- Silverman, B. (1986): *Density estimation for statistics and data analysis*. Chapman and Hall.
- Ullah, A. (1989): *Semiparametric and Nonparametric Econometrics*. Physica Verlag, Heidelberg.
- Vieu, P. (1991a): «Nonparametric Regression: local optimal bandwidth choice», *Journal of the Royal Statistical Society, Serv. B* 53, pp. 453-464.
- Vieu, P. (1991b): «Quadratic errors for nonparametric estimates under dependence», *Journal of Multivariate Analysis* 2, pp. 324-347.
- Vilar Fernández, J. M. (1989a): «Estimación no paramétrica de curvas notables para datos dependientes», *Trabajos de Estadística* 4, pp. 69-88.
- Vilar Fernández, J. M. (1989b): «Estimación recursiva, tipo núcleo, de la función de autorregresión para datos dependientes», *Estadística Española* 31, 121, pp. 207-226.
- Watson, G. S. y Leadbetter, M. R. (1964): «Hazard Analysis II», *Shankhya, Ser. A* 26, pp. 101-116.

## Abstract

In this paper, some nonparametric techniques of estimation are introduced for the following probability curves: density function, fixed and random design regression, and autoregression function. In particular, we study modifications of the usual kernel estimates of these curves for the case that the sample data are dependent. Several examples of applications to economic data are also presented.

*Recepción del original, abril de 1992  
Versión final, diciembre de 1992*