

Un pronóstico no paramétrico de la inflación colombiana

NORBERTO RODRÍGUEZ N.*
PATRICIA SIADO C.**

Resumen

En este trabajo se presentan los resultados de un ejercicio de pronóstico no paramétrico, múltiples pasos adelante, para la inflación colombiana mensual. En particular, se usa estimación kernel para la media condicional de los cambios de la inflación, dada su propia historia. Los resultados de pronóstico se comparan con un modelo ARIMA estacional y un modelo tipo STAR. Se encuentra que, excepto para el pronóstico un mes adelante, el pronóstico no paramétrico mejora a las otras dos metodologías que le compiten; además, de entre las tres alternativas consideradas, el no paramétrico es el único pronóstico que estadísticamente mejora al pronóstico que se hace con un modelo de caminata aleatoria.

Palabras clave: Pronóstico no paramétrico, evaluación y comparación de pronósticos, ancho de banda (bandwidth), estimación kernel.

Abstract

This paper contains the results of a non parametric multi-step ahead forecast for the monthly Colombian inflation, using Mean conditional kernel estimation over inflation changes, with no inclusion of exogenous variables. The results are compared with those from an ARIMA and a non-linear STAR. The nonparametric forecast over perform the others two, as well as being the only, from the three, that statistically improved the naïve forecast given by a random-walk model.

Key words: Nonparametric forecast, Kernel estimation, Forecast evaluation, Bandwidth selection, Rolling forecast.

*Econometrista asociado de la Unidad de Econometría del Banco de la República. Correspondencia nrodrini@banrep.gov.co.

**Estadística de la Universidad Nacional de Colombia.

1. Introducción

El control de la inflación es casi siempre uno de los objetivos prioritarios de la política económica gubernamental, en particular del banco central. Su pronóstico acertado ayudará a atenuar todos los posibles inconvenientes, al permitir tomar medidas remediales anticipadas si es necesario.

Para el estudio de la inflación se han usado varias herramientas estadísticas entre las cuales se encuentran los modelos de series de tiempo a través de métodos paramétricos, con básicamente dos propósitos: uno es construir un modelo que ajuste adecuadamente los datos con la estimación de los parámetros del modelo y así hacer análisis de política; el segundo propósito es usar el modelo identificado y estimado para realizar pronósticos. El presente trabajo se centra exclusivamente en este último objetivo.

La metodología de Box-Jenkins ofrece la manera de lograr estos objetivos a través de la construcción, identificación y predicción de un *proceso autorregresivo de media móvil estacional, SARIMA*; pero la linealidad, que es el soporte fundamental de la teoría Box-Jenkins, es bastante fuerte e inadecuada en muchas situaciones prácticas.

Uno de los problemas que se presenta en predicción, principalmente si el horizonte de pronóstico es largo, es el aumento en el error cuadrático medio de pronóstico conforme aumenta el horizonte de predicción. Consideraciones de modelos paramétricos no lineales tipo STAR, como bi-lineales o procesos ARCH¹ no siempre logran producir un notable mejoramiento en la calidad de la predicción.

Hasta el momento no se habían utilizado métodos no paramétricos para realizar pronósticos en la serie de la inflación colombiana, pero gracias a los desarrollos en la teoría de estadística no paramétrica para series de tiempo, se facilita dicha labor. En el presente trabajo se realizaron pronósticos de la inflación colombiana por medio de los métodos de predicción no paramétricos basados en estimación kernel. No obstante, la aplicación de técnicas no paramétricas a datos de la economía colombiana no es tan poco común como se podría pensar; sólo por mencionar dos trabajos de aplicación a datos de ingresos, están Núñez & Jiménez (1998) y, más novedoso, Zárate (2003).

Los métodos no paramétricos tienen ventajas sobre los paramétricos. Los métodos paramétricos en muchas ocasiones no cumplen con los supuestos acerca de la forma funcional del conjunto de variables aleatorias de las cuales provie-

¹ARCH: *Autoregressive Conditional Heteroskedasticity*. La familia de modelos GARCH, o ARCH generalizado, provee recursos para realizar estimaciones cuando la varianza de un proceso no es constante a través del tiempo.

nen los datos, produciendo así modelos no muy confiables que generan sesgos y deterioran la calidad de los pronósticos. En el campo no paramétrico se evita este problema al permitir una forma funcional flexible, y no un conjunto pequeño de modelos rígidos como lo hacen los paramétricos. Sin embargo, se les hacen dos críticas: la primera se refiere a la demora en el trabajo computacional, y la segunda al amplio error cuadrático medio de predicción. Gracias a desarrollos tecnológicos en el ámbito computacional y su disponibilidad a bajo costo queda sin peso la primera crítica. La segunda ha sido estudiada en la literatura; se ha demostrado que los pronósticos del método no paramétrico de la mediana condicional, al ser comparados con los de modelos ARIMA dan resultados favorables en términos del error cuadrático medio Gannoun (1991)²; resultados similares son mostrados para la media condicional Carbon & Delecroix (1993).

Puesto que se plantea el uso de una nueva técnica de pronóstico, es deseable comparar sus resultados con los de otros modelos existentes. Con eso en mente, y por ubicar al lector, a continuación de esta introducción se presenta una breve sintaxis de dos de los modelos usados por el Banco de la República para generar pronósticos de la inflación. En la sección 3 se introducen los primeros conceptos de estimación no paramétrica de densidades condicionales o funciones de regresión, junto a un ejemplo de datos simulados, para terminar dicha sección con una descripción de algunas de las funciones kernel más usadas en aplicaciones estadísticas de métodos no paramétricos de suavizamiento. Se pasa a la sección 4, donde se resume la metodología de pronósticos no paramétricos múltiples pasos adelante, basados en suavizamiento kernel, y las especificaciones que se requieren del mismo. La sección 5 presenta los resultados del pronóstico no paramétrico de la inflación colombiana, así como los resultados de la evaluación de los mismos. La sección 6 concluye y enuncia algunas recomendaciones para futuro trabajo de aplicación en el área.

2. Modelos paramétricos usados para pronosticar

Actualmente el Banco de la República utiliza, entre otros, dos tipos de modelos para pronosticar mensualmente la inflación; son ellos los tradicionales modelos ARIMA y los modelos no lineales tipo STAR. Una breve descripción de ellos se presenta en seguida.

²Véanse las definiciones más adelante.

2.1. Modelo ARIMA

El modelo que se usa comúnmente para hacer pronósticos a la serie de tiempo de la inflación es un modelo ARIMA estacional de la forma:

$$\phi(L)\Phi(L^{12})(1-L)(1-L^{12})\ln(IPC_t) = \theta(L)\Theta(L^{12})e_t, \quad (1)$$

donde \ln es el logaritmo natural, L es el operador de rezagos³, $\phi(L)$ y $\theta(L)$ son polinomios en L , con coeficientes fijos a través de la muestra, de grados p y q , respectivamente; $\Phi(L^{12})$ y $\Theta(L^{12})$ son polinomios en L^{12} de grados P y Q , respectivamente, los cuales modelan el componente estacional; $\{e_t\}$ es un proceso ruido blanco: variables aleatorias independientes e idénticamente distribuidas con media 0 y varianza σ^2 , usualmente asumidas con distribución normal (gaussiana).

Entre las ventajas de estos modelos se encuentran: son de fácil manejo computacional, están incorporados en la mayoría de paquetes estadísticos, son bastante conocidos y utilizados. Entre sus desventajas, sus pronósticos resultan desmejorados para el mediano o largo plazo.

2.2. Modelos autorregresivos de transición suave: STAR

Los modelos tipo STAR corresponden a una gama de modelos no lineales, presentados como una extensión de los modelos TAR (*Threshold Autoregressive*), donde se supone que el proceso generador de la serie Y_t oscila de forma suave entre dos regímenes:

$$Y_t = \alpha_0 + \sum \alpha_i Y_{t-i} + \left(\beta_0 + \sum \beta_i Y_{t-i} \right) F(Y_{t-d}) + \varepsilon_t, \quad (2)$$

$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, $F(Y_{t-d})$ es una función no lineal de Y_{t-d} que toma valores entre 0 y 1, es no decreciente y continua, se denomina función de transición; los dos modelos más comunes suponen las siguientes funciones:

$$\text{Modelo LSTAR: } F(Y_{t-d}) = [1 + \exp[-\gamma(Y_{t-d} - c)]]^{-1} \quad \gamma > 0 \quad (3)$$

$$\text{Modelo ESTAR: } F(Y_{t-d}) = [1 - \exp[-\gamma(Y_{t-d} - c)^2]] \quad \gamma > 0. \quad (4)$$

Los cambios de la no linealidad que introduce el régimen por medio de la función $F(Y_{t-d})$ dependen de los parámetros γ y c . En particular, para un

³El operador L de rezagos está definido por: $L^k z_t = L \cdot \dots \cdot L z_t = z_{t-k}$.

modelo LSTAR los regímenes de transición ocurren alrededor de $Y_{t-d} = c$ donde el parámetro γ indica el grado de no linealidad, es decir, qué tan rápido ocurre la transición entre los dos regímenes extremos: el paso de cero a uno en $F(Y_{t-d})$. Este modelo ha sido también utilizado con una variable exógena Jalil & Melo (2000).

Entre las desventajas se encuentra que sus intervalos de pronóstico son difíciles de obtener con métodos convencionales y los pronósticos puntuales a mediano y largo plazo deben ser generados vía simulación estocástica. La ventaja que presentan frente a los ARIMA y otras técnicas paramétricas es que resultan pronosticar consistentemente mejor que aquellos a mediano y largo plazo.

2.3. Otros modelos

Existen varios trabajos en los cuales se involucran especificaciones de modelos no lineales para la inflación de Colombia. Uno de los primeros trabajos explica el proceso inflacionario como un modelo *switching* con dos o tres estados (Melo & Misas 1998), pero resulta de difícil implementación y no útil para pronóstico, debido a que bajo cierta especificación necesaria se requieren 20 días continuos de ejecución en computador.

Recientemente se han usado modelos de redes neuronales, los cuales son intensivos en uso de recurso computacional; esto dificulta el proceso de evaluación de sus pronósticos, pero aun así no han dado muestra de mejorar otros modelos en el trabajo de pronosticar (Misas, López & Querubín 2002).

3. Regresión no paramétrica

Con el objeto de introducir conceptos, en este capítulo se presenta la estimación no paramétrica de densidades condicionales, o dicho de otro modo, funciones de regresión, lo cual es básico para entender las estimaciones no paramétricas en modelos de series de tiempo. Con el ánimo de ilustrar las ideas se presenta un ejemplo. La estimación no paramétrica de densidades no condicionales y resultados básicos de estimación kernel son dejados como anexo A.

Como lo escribe Härdle (1990), la aproximación no paramétrica a la estimación de curvas de regresión tiene cuatro propósitos principales. Primero, proveer un método versátil de explorar una relación general entre dos variables. Segundo, generar una predicción de observaciones aún no hechas, sin referencia

a un modelo paramétrico fijo. Tercero, proporcionar una herramienta para encontrar observaciones espurias, mediante el estudio de la influencia de puntos aislados. Cuarto, constituye un método flexible de sustitución de observaciones faltantes o atípicas y permite interpolar entre valores adyacentes de las variables exógenas.

Dentro del marco de regresión no paramétrica se pueden citar además de los métodos basados en kernel o suavizamiento, los de estimación basada en los k vecinos más cercanos (*k-nearest neighbor*), las estimaciones con series ortogonales, regresión cuantílica y los suavizamientos de Splines. El presente trabajo se centra en los basados en metodología kernel, la cual se describe a continuación.

3.1. Estimación de regresión por el método kernel

Se consideran las variables aleatorias bidimensionales, independientes e idénticamente distribuidas $(X_1, Y_1), \dots, (X_n, Y_n)$. Suponiendo que existe una función $r(\cdot)$ la cual modela la relación entre la respuesta Y_i y la co-variable X_i , así: $Y_i = r(X_i) + \varepsilon_i$, y además que $E(\varepsilon_i|X_i) = 0$, se puede mostrar que la aproximación óptima a Y_i , en el sentido de menor error cuadrático medio, está dada por la esperanza condicional,

$$r(x) = E(Y_i|X_i = x), \quad x \in \mathbb{R},$$

nótese que no se hace ningún supuesto sobre la linealidad de dicha relación.

Si se definen las funciones:

$$f(x) = \int f(x, y) dy,$$

la densidad marginal de X , donde $f(x, y)$ es la densidad conjunta, sea:

$$\varphi(x) = \int y f(x, y) dy,$$

entonces, la esperanza condicional de Y_i dado $X_i = x$ es,

$$r(x) = \int \frac{f(x, y)y}{f(x)} dy = \frac{\varphi(x)}{f(x)}, \quad x \in \mathbb{R}.$$

Un estimador kernel de $f(x)$, basado en las n observaciones, es:

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right),$$

donde $K(\cdot)$ es una función simétrica alrededor de cero, continua, acotada, no necesariamente positiva en todo su soporte e integrable (detalles en el anexo A, ecuación 34, haciendo $d = 1$), análogamente, un estimador de $\varphi(x)$ es (Pagan & Ullah 1999, págs. 83-84):

$$\varphi_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) Y_i, \quad x \in \mathbb{R}.$$

El estimador kernel de $r(x)$, basado en n observaciones, está definido como:

$$r_n(x) = \frac{\varphi_n(x)}{f_n(x)}, \quad x \in \mathbb{R};$$

rescribiendo esta función se obtiene:

$$r_n(x) = \sum_{i=1}^n W_{ni}(x) Y_i,$$

donde⁴:

$$W_{ni}(x) = \frac{K\left(\frac{x - X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x - X_s}{h_n}\right)}.$$

Así, el estimador de regresión kernel de $E(Y_i|X_i = x)$ es:

$$r_n(x) = \frac{\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) Y_i}{\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)}, \quad x \in \mathbb{R}, \quad (5)$$

el cual puede verse como un promedio ponderado de los Y , donde el peso depende de la distancia entre X_i y x , la cual es cuantificada por la función $K(\cdot)$; en general el procedimiento de ponderación asigna el mayor peso a puntos cercanos a x , y menor o ningún peso a puntos apartados de x . Este es precisamente el estimador de Nadaraya-Watson; véase Bosq (1998) para detalles adicionales.

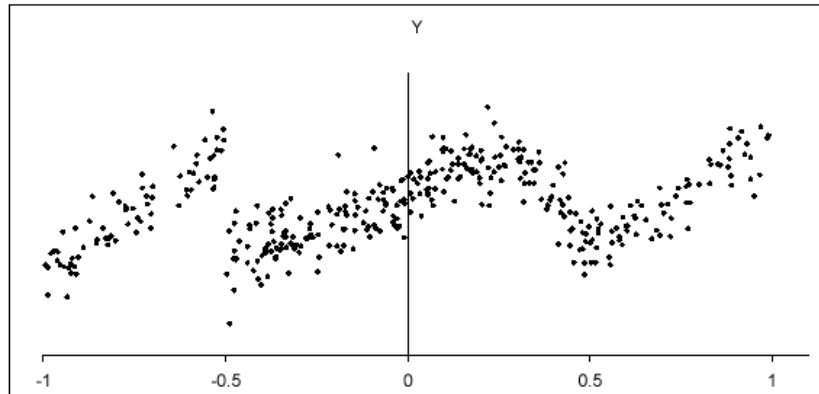
⁴Se debe usar la convención $W_{ni}(x) = 0$ si el denominador es cero.

Las ponderaciones kernel definen una vecindad de puntos alrededor de x . El suavizador kernel puede ser entendido como un polinomio de ajuste local constante. Extensiones naturales a esta idea son las regresiones polinómicas locales, esto es, ajustar una regresión lineal local, un ajuste cuadrático local, etc. Propiedades teóricas deseables son obtenidas con esta estrategia. En la práctica, muchas veces resulta suficiente con la regresión lineal local. Véase Fan & Gijbels (1996).

3.2. Ejemplo de regresión no paramétrica

Este ejemplo, más que detallar el uso de la regresión no paramétrica, pretende ilustrar lo crítico de la selección del ancho de banda. El ejemplo está basado en Györfi, Kohler, Krzizak & Walk (2002); en ese caso $Y_i = r(x_i) + \varepsilon_i$, donde $\varepsilon_i \sim N(0, \text{var}(\varepsilon_i))$, $\text{var}(\varepsilon_i) = 0,2 - 0,1 \cos(2\pi x_i)$ y

$$r(x) = \begin{cases} (x+2)^2/2 & \text{si } -1 \leq x < -0,5, \\ x/2 + 0,875 & \text{si } -0,5 \leq x < 0, \\ -5(x-0,2)^2 + 1,075 & \text{si } 0 \leq x < 0,5, \\ x + 0,125 & \text{si } 0,5 \leq x < 1, \end{cases}$$

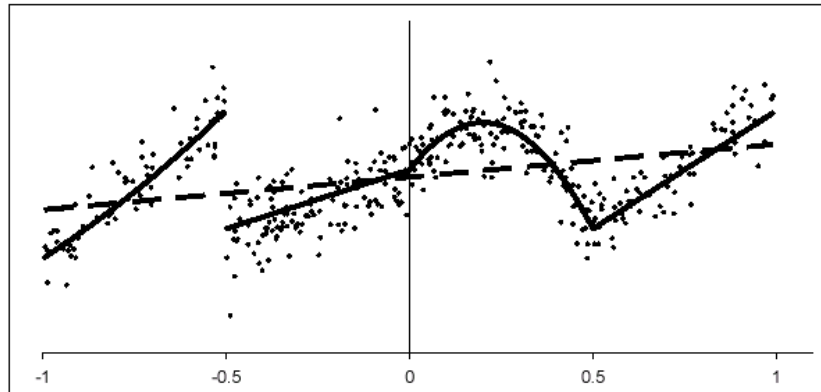


Gráfica 1: Datos simulados.

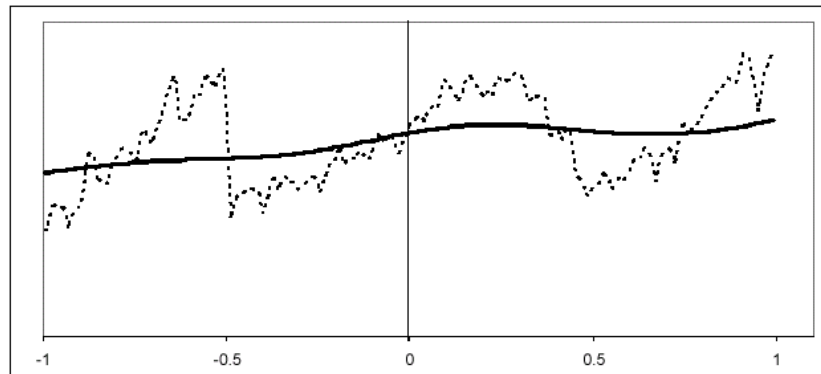
donde x es generado aleatoriamente como una variable normal estándar truncada al intervalo $[-1, 1]$; así, la función $r(x)$ es polinomial a trazos, con discontinuidades; nótese además la heterocedasticidad del término de error. La

gráfica 1 muestra la nube de puntos de los $n = 400$ datos simulados; de allí, se nota que el ojo humano no es capaz de ver con claridad cuál puede ser la función de regresión.

La gráfica 2 muestra el modelo teórico junto con la línea de ajuste que se obtendría con un modelo paramétrico lineal. Obviamente, la aproximación lineal no es una muy buena alternativa.

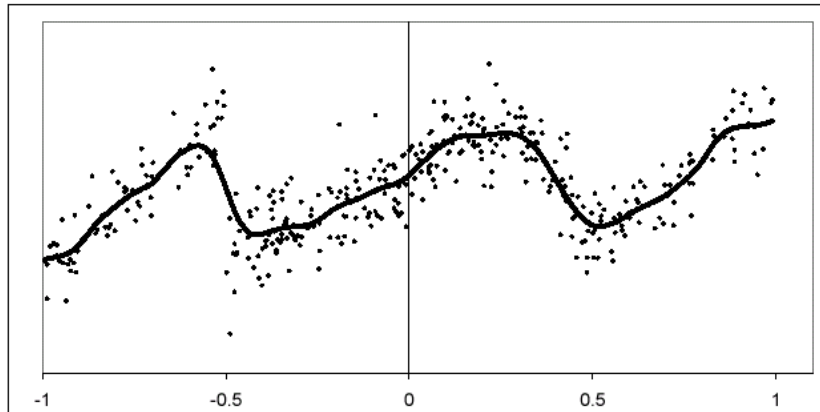


Gráfica 2: Datos simulados, línea continua relación simulada, línea a trazos ajuste lineal.



Gráfica 3: Dos estimaciones kernel, a trazos $h = 0,01$, línea continua $h = 0,4$.

La gráfica 3 muestra dos ajustes obtenidos por el método kernel. El primero, la línea a trazos, se obtiene con el kernel gaussiano y usa $h = 0,01$, lo cual



Gráfica 4: Estimación kernel, con h óptimo, $h = 0,1527$.

produce sub-ajuste (poco suavizamiento), en el sentido de que el ajuste sigue muy de cerca los datos originales. El segundo, la línea continua, se obtiene con $h = 0,4$, generando sobre-ajuste (suavizamiento excesivo).

La gráfica 4 muestra, junto a la relación teórica, el ajuste kernel con selección de ancho de banda que señala el método *plug-in*, en el que se sugiere $h = \hat{\sigma}_x n^{-1/5}$, en este caso $h = 0,1527$. Los expertos recomiendan, en cuanto sea posible, recurrir a la inspección visual del ajuste obtenido para varios anchos de banda y tener en cuenta que es menos dañino excederse un poco en suavizamiento que en sobre-ajuste. Así queda en evidencia la importancia de una adecuada selección del ancho de banda. Los problemas son mucho más graves cuando la regresión es múltiple, esto es, con varios regresores. De la construcción del estimador se encuentra que la metodología es menos influenciada a valores extremos y situaciones atípicas que los modelos paramétricos. Otras aplicaciones, extensiones y detalles se pueden encontrar ilustradas en Härdle (1990).

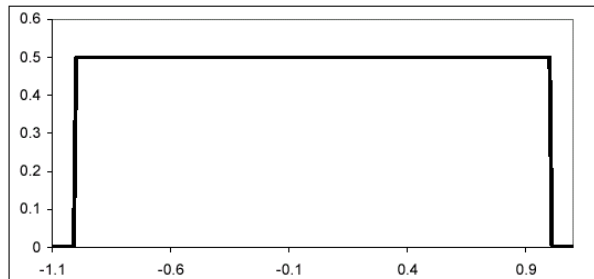
3.3. Funciones kernel

A continuación se mencionan las funciones kernel más usadas en aplicaciones prácticas. Con fines ilustrativos se presentan los kernel estándar, $K(u)$, pero se debe tener en cuenta que en la práctica se utiliza $K_h(u) = h^{-1}K\left(\frac{u}{h}\right)$.

Kernel uniforme: Asigna peso de $\frac{1}{h}$ por igual a todas las observaciones

que están a distancia no mayor de h y cero a las demás.

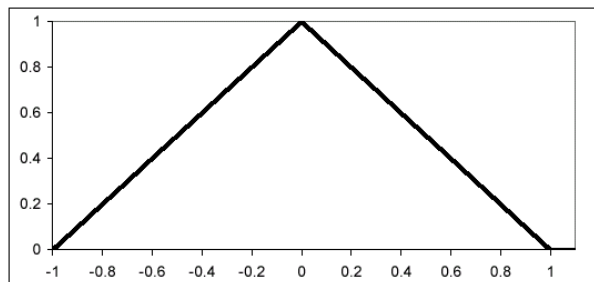
$$K(u) = \frac{1}{2}I_{[-1,1]}(u).$$



Gráfica 5: Kernel uniforme.

Kernel triangular: Asigna pesos de $\frac{1}{h}$ a observaciones coincidentes y el peso de las otras decrece linealmente hasta un peso de cero a las que están a h o más lejos.

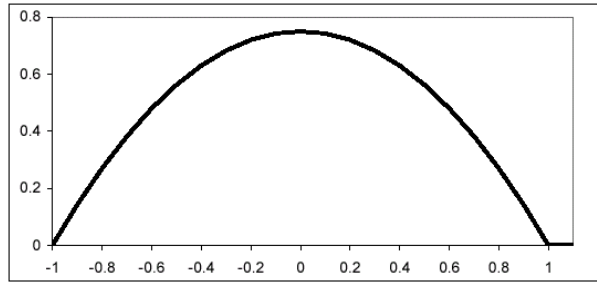
$$K(u) = (1 - |u|)I_{[-1,1]}(u).$$



Gráfica 6: Kernel triangular.

Kernel Epanechnikov: Para las observaciones que están a distancia de 0 a h asigna pesos entre 0,75 y cero, con decrecimiento cuadrático. Las que están a una distancia de h o mayor tienen peso cero.

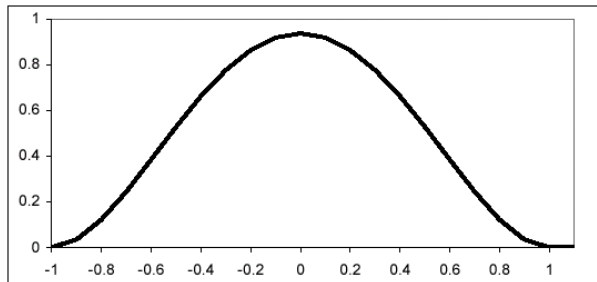
$$K(u) = \frac{3}{4}(1 - u^2)I_{[-1,1]}(u).$$



Gráfica 7: Kernel Epanechnikov.

Kernel bicuadrado: Los pesos para observaciones cercanas son cuando más de 0,93 con decrecimiento polinomial cuártico hasta llegar a cero, cuando la distancia es de h o mayor.

$$K(u) = \frac{15}{16}(1 - 2u^2 + u^4)I_{[-1,1]}(u).$$

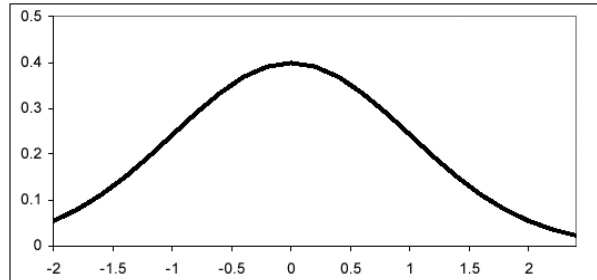


Gráfica 8: Kernel bicuadrado

Kernel gaussiano: Asigna pesos de acuerdo con una densidad normal estándar. Observaciones cuya distancia oscila entre cero y 1 reciben peso entre 0,4 y 0,2; las que están a distancia 3 reciben peso de 0,0039 y prácticamente cero el resto de observaciones. Es un kernel muy popular, con soporte no compacto y diferenciable en todo su soporte o recorrido. En este caso el ancho de banda

en $K_h(u)$ desempeña el papel de desviación estándar.

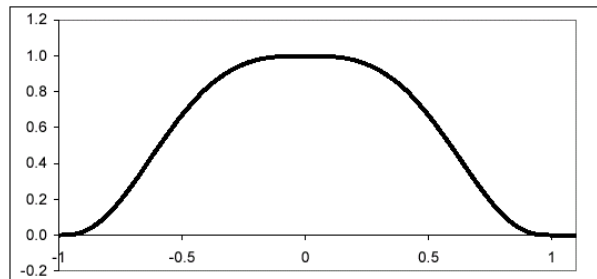
$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right).$$



Gráfica 9: Kernel gaussiano.

Kernel tri-cúbico: Es similar al Epanechnikov pero más plano en la cima, con la ventaja teórica de que es diferenciable en los límites de su soporte ($|u| = 1$)

$$K(u) = (1 - |u|^3)^3 I_{[-1,1]}(u).$$

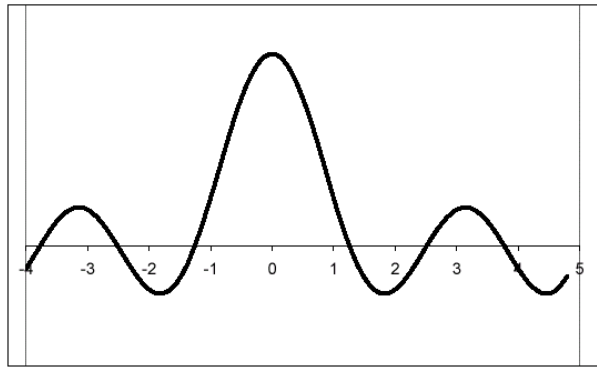


Gráfica 10: Kernel tri-cúbico.

Kernel Dirichlet: Los pesos son asignados según ondas senosoidales de magnitud decrecientes, con un “lóbulo principal” o mayor alrededor de cero y lóbulos laterales o menores a los dos lados. Este kernel resulta de soporte

no acotado. Tiene la particularidad de no ser siempre positivo y es usado en análisis espectral; véase Prietsley (1984).

$$K(u) = \frac{\text{sen}((M + 0,5)u)}{2\pi \text{sen}(0,5u)}.$$



Gráfica 11: Kernel Dirichlet, $M = 2$.

De las anteriores funciones kernel y de otras que existen en la literatura, las más usadas son la función tri-cúbico, Epanechnikov y la gaussiana soportada en sus propiedades estadísticas y asintóticas. Dado que la elección de la función kernel no afecta marcadamente los resultados, como es aceptado ampliamente en la literatura, en este trabajo se usa el kernel gaussiano, como se explica más adelante. Véase Hastie, Tibshirani, Friedman & Friedman (2002).

4. Implementación del método no paramétrico en series de tiempo

Cuando un estimador kernel es aplicado a datos dependientes como en el caso de series de tiempo, el efecto por la dependencia entre las observaciones afecta solamente un pequeño intervalo de tiempo y no a todos los datos (Heiler 1999). Este hecho reduce la dependencia entre las estimaciones, por lo que muchas de las técnicas desarrolladas para datos independientes pueden ser usadas en el caso de series de tiempo. Sin embargo, dentro de las hipótesis se encuentra el que las observaciones sean asintóticamente independientes, lo

que ha sido estudiado ampliamente por varios autores para los casos que se presentan en este capítulo.

4.1. Estacionarización

El primer paso de la implementación de los métodos no paramétricos es la estacionarización de los datos, que se lleva a cabo estabilizando la varianza y removiendo la tendencia, si es del caso; sin embargo, algunos autores afirman que si los cambios en la media o la varianza para las series de tiempo no estacionarias son leves, las técnicas no paramétricas siguen siendo igualmente efectivas⁵.

4.2. Notación y suposiciones generales

Sea $\{Z_t\}$ un proceso observado de serie de tiempo univariado estrictamente estacionario⁶ con $1 \leq t \leq n$. En la práctica, el supuesto de variables independiente e idénticamente distribuidas difícilmente se cumple, mientras que el supuesto de un proceso asintóticamente independiente o *mixing*⁷ y no necesariamente distribuido idénticamente es más simple de mantener. Estos supuestos han sido verificados para procesos gaussianos (Ibragimov & Rozanov 1978) y para procesos ARIMA no gaussianos (Pham & Tran 1985). Sin embargo, esta suposición es muy difícil de verificar en la práctica en situaciones generales.

Como tercer supuesto, $\{Z_t\}$ sigue un proceso d -markoviano⁸ (suposición H), donde d es llamado el coeficiente de Markov. Dada la realización z_1, z_2, \dots, z_n se desea predecir la variable aleatoria no observada Z_{n+m} , $m \geq 1$ donde m es el horizonte de predicción.

Para este propósito se construye el proceso asociado $\{X_t, Y_t\}$, donde

$$X_t = (Z_t, \dots, Z_{t-d+1}), \quad (6)$$

⁵ Véase Bosq (1998, pág. 88), donde se menciona el método *cynical* que consiste en ignorar el componente determinístico de la serie observada, siendo este componente el que podría inducir la no estacionaridad.

⁶ Z_t es estrictamente estacionario si $(Z_t, \dots, Z_{t+n}) \xrightarrow{\text{distribución}} (Z_{t+g}, \dots, Z_{t+n+g})$ para todo entero $t, g, n \geq 1$. Esto indica que los dos vectores aleatorios tienen la misma función de distribución conjunta.

⁷ Véase anexo B.

⁸ Para un proceso Z_t de Markov se cumple: $F(Z_t | Z_{t-s} \geq 1) = F(Z_t | Z_{t-1}, \dots, Z_{t-d})$, con F la función de distribución acumulativa de probabilidad.

y

$$Y_t = Z_{t+m}, \quad t \in \{d, \dots, n\}. \quad (7)$$

Considerando el estimador de regresión kernel r_n , basado en los datos z_1, z_2, \dots, z_n para $E(Y_n|X_n) = E(Z_{n+m}|Z_n, \dots, Z_{n-d+1})$ es

$$r_n(x) = \widehat{E}(Y_n|X_n) = \sum_{t=d}^{n-m} W_t^m(x) Y_t,$$

donde

$$W_t^m(x) = \frac{K_d\left(\frac{x - X_t}{h_n}\right)}{\sum_{t=d}^{n-m} K_d\left(\frac{x - X_t}{h_n}\right)}, \quad x \in \mathbb{R}^d. \quad (8)$$

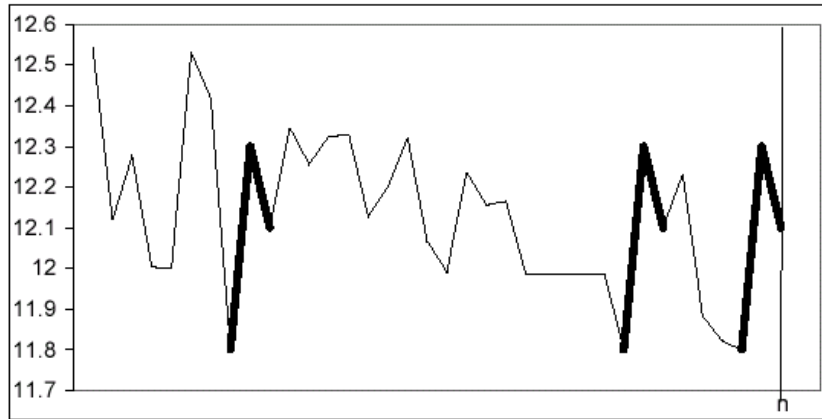
Nótese que la fijación del ancho de banda h_n controla el tamaño de la vecindad local y debe ser un valor real positivo que debe tender a cero cuando n tiende a infinito. La función de $K_d(\cdot)$ es una función kernel d -variada, con integral múltiple igual a uno y ella controla la forma de los pesos. Finalmente, $x = X_n = (Z_n, \dots, Z_{n-d+1})$ se llamará el bloque de referencia con el cual se comparan los otros bloques. Así, una secuencia de observaciones o bloque, tendrá mayor peso en el pronóstico si es más parecido en términos de distancia al bloque de referencia, que un bloque que no lo sea.

La gráfica 12 ilustra esas ideas, cuando se tiene un proceso markoviano de orden $d = 3$. Dependiendo de la amplitud de banda, h_n , y si el soporte del kernel es acotado, las secuencias en negrilla serán consideradas en la generación del pronóstico; las demás no.

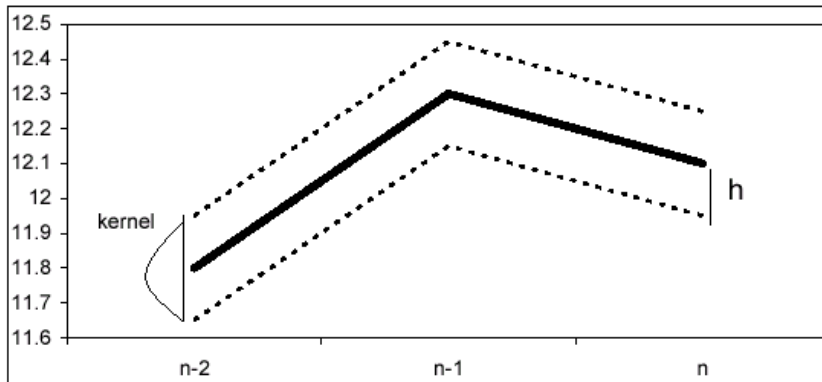
Ahora, fijado un h , las secuencias pasadas que se han de considerar en el pronóstico son aquellas que caen completamente dentro de la banda a trazos que se muestra en la gráfica 13, y los pesos son asignados acorde con la función kernel usada.

4.3. Predictores no paramétricos basados en kernel

La predicción de las variables Z_{n+m} ó Y_n consiste en encontrar la variable aleatoria más cercana (con respecto a cierta norma), conociendo todo el pasado de la serie. Este problema puede ser visto así: suponga que existe una función $r(\cdot)$ que modela la relación entre la respuesta Y y la co-variable X , y que



Gráfica 12: Secuencias por considerar en el pronóstico kernel.



Gráfica 13: Vecindades de sendas pasadas.

$r(\cdot)$ está definida a través de la distribución condicional⁹. Dada una función de pérdida convexa $l(\cdot)$ con mínimo único en cero, se define $\hat{r}(x)$ como la que minimiza la media $E(l(Y - a)|X = x)$, con respecto a a , es decir:

$$\hat{r}(x) = \arg \min_{a \in \mathbb{R}} E(l(Y - a|X = x)).$$

entonces estimando no paramétricamente $r(\cdot)$ por $\hat{r}_n(\cdot)$ y calculando $\hat{r}_n(X_n)$ se genera \hat{Y}_n . De esta forma se obtiene la predicción \hat{Z}_{n+m} . Tres alternativas han sido propuestas en la literatura: media, mediana y moda condicional, dependiendo de la función de pérdida que se considere.

4.3.1. Media condicional

Se puede ver que con $l(u) = u^2$ se llega a la función de media condicional $\hat{r}(x) = E(Y|X = x)$. Entonces usando la suposición H y estimando $\hat{r}(\cdot)$ se obtiene:

$$\hat{r}_n(x) = \sum_{t=d}^{n-m} W_t^m(x) Y_t. \quad (9)$$

Aquí el predictor no paramétrico m -pasos adelante de la media es:

$$\hat{Z}_{n+m}^{\text{med}} = \sum_{t=d}^{n-m} W_t^m(X_n) Y_t. \quad (10)$$

En Collomb (1984) se encuentra que, con alguna condición de regularidad, es posible demostrar que:

$$\left| \hat{Z}_{n+m}^{\text{med}} - Z_{n+m} \right| \xrightarrow{\text{c.s.}} 0.$$

4.3.2. Mediana condicional

Ahora se asume que la distribución condicional de Y dado X es más pesada en un extremo o asimétrica. Entonces es mejor usar la mediana en vez de la media para predecir valores futuros, teniendo en cuenta que la mediana es ampliamente resistente en estos casos. En este ejemplo la función de pérdida

⁹Cuando se ha observado una realización $x_t = (z_1, \dots, z_n)$ de longitud n de un proceso estocástico y se desea obtener la predicción del valor z_{n+m} con el criterio de minimizar el error cuadrático medio (ECM), el predictor \hat{z}_{n+m} que minimiza el ECM es la esperanza de la distribución condicionada $\hat{z}_{n+m} = E[z_{n+m} | x_t]$ como predictor óptimo.

es dada por $l(u) = |u|$, y la solución de $\hat{r}(x)$ lleva a la función mediana condicional $\hat{r}(x) = \inf\{y : F(y|X = x) \geq \frac{1}{2}\}$. Entonces, usando la suposición H y estimando $\hat{r}(\cdot)$ se obtiene:

$$\hat{r}_n(x) = \inf\left\{y : \sum_{t=d}^{n-m} W_t^m(x) I_{\{Y_t \leq y\}} \geq \frac{1}{2}\right\}. \quad (11)$$

Aquí el estimador no paramétrico de la mediana para m pasos adelante está dado por:

$$\hat{Z}_{n+m}^{\text{mediana}} = \inf\left\{y : \sum_{t=d}^{n-m} W_t^m(X_n) I_{\{Y_t \leq y\}} \geq \frac{1}{2}\right\}, \quad (12)$$

bajo algunas condiciones de regularidad se prueba que (Gannoun 1990):

$$\left| \hat{Z}_{n+m}^{\text{mediana}} - Z_{n+m} \right| \xrightarrow{\text{c.s.}} 0.$$

4.3.3. Moda condicional

Algunos autores propusieron un método para producir predictores no paramétricos basados en la función moda $\theta(x) = \arg \max_y f(Y|X = x)$, esta función puede ser estimada como sigue (Collomb, Hardle & Hassani 1987). Suponga que se tiene una función de pérdida no convexa con mínimo único $l(u) = 0$ cuando $u = 0$ y $l(u) = 1$ en otro caso. Entonces la solución de $\hat{r}(x)$ conduce a la función moda condicional $\hat{r}(x) = \arg \max_{y \in \mathbb{R}} f(Y|X = x)$. Después, usando la suposición H y estimando $\hat{r}(\cdot)$, se obtiene:

$$\hat{r}(x) = \arg \min_{y \in \mathbb{R}} h^{-1} \sum_{t=d}^{n-m} W_t^m(x) K\left(\frac{y - Y_t}{h_n}\right). \quad (13)$$

Consecuentemente el predictor no paramétrico m pasos adelante es:

$$\hat{Z}_{n+m}^{\text{moda}} = \arg \max_{y \in \mathbb{R}} \sum_{t=d}^{n-m} W_t^m(X_n) K\left(\frac{y - Y_t}{h_n}\right); \quad (14)$$

bajo algunas condiciones de regularidad se ha demostrado que:

$$\left| \hat{Z}_{n+m}^{\text{moda}} - Z_{n+m} \right| \xrightarrow{\text{c.s.}} 0.$$

Todos los predictores dados pueden ser interpretados con respecto a los pesos $W_t^m(X_n)$. Una observación con peso grande desempeña un papel más

importante en la elaboración de los predictores. Los pesos estarán cerca de cero si el correspondiente bloque X_t contiene un *outlier*; sea éste $Z_{t'}$, ($t' \in \{t, \dots, t-d+1\}$). Así, la metodología de estimación kernel es robusta a la presencia de valores atípicos.

4.4. Elección de coeficientes

4.4.1. Elección de la función kernel K

La función kernel más usada en el caso multivariado es el producto kernel:

$$K_d(x_1, \dots, x_d) = \prod_{j=1}^d K(x_j),$$

donde la función kernel $K(\cdot)$ está definida en la recta real. Será usado el kernel gaussiano, definido como:

$$K_d(x_1, \dots, x_d) = (2\pi)^{-\frac{d}{2}} \exp\left(-\sum_{i=1}^d \frac{x_i^2}{2}\right), \quad x_i \in \mathbb{R}. \quad (15)$$

Esta función corresponde a la idea de dar pesos grandes a X_t cuando está cercano a X_n usando la norma $\|X_n - X_t\|_\infty = \sup_{0 < t' < k+1} |Z_{n-t'} - Z_{t-t'}|$ (Matzner-Løber, Gannoun & Gooijer 1998). En este caso, de no presencia de variables exógenas o explicativas, no se requiere estandarizar las variables x_i (Hastie et al. 2002, pág. 174).

4.4.2. Elección del coeficiente de Markov d

Intuitivamente se podría escoger el d lo más grande posible para no dejar de considerar demasiada información del pasado; no obstante, la elección de d está limitada por la cantidad de datos disponible. Debe ser evidente que cuando el valor del coeficiente de Markov crece, la cantidad de datos disponibles para predicción decrece.

Matzner-Løber et al. (1998) proponen un método empírico para encontrar el d óptimo:

Sea $\widehat{Z}_t(d, h_n)$ la predicción en el tiempo t , dependiendo del coeficiente de Markov d y del ancho de banda h_n obtenido para uno de los tres métodos (media, mediana, moda condicional). Sea $d_{\text{máx}}$ un valor fijo, el cual es bastante

grande para capturar cualquier efecto estacional en los datos. Entonces, para $d \in \{1, \dots, d_{\text{máx}}\}$ se define:

$$f_1(d) = p^{-1} \sum_t |Z_t - \widehat{Z}_t(d, h_{nd})| \quad (16)$$

$$f_2(d) = p^{-1} \sum_t \{Z_t - \widehat{Z}_t(d, h_{nd})\}^2 \quad (17)$$

$$f_3(d) = \sup |Z_t - \widehat{Z}_t(d, h_{nd})|, \quad (18)$$

donde $t = n - p, \dots, n$. El parámetro de suavizamiento h_{nd} será definido después. El valor de p se toma como $p = \lfloor \frac{n}{4} \rfloor$ para series de tiempo con $n < 100$, y para series con tamaño igual o superior a cien se usa $p = \lfloor \frac{n}{5} \rfloor$, donde $\lfloor \cdot \rfloor$ representa la parte entera del argumento. Si las funciones $f_j(d)$, $j \in \{1, 2, 3\}$ no se incrementan después de cierto valor d , se escoge el valor de $f_j(d)$ con el cual ocurre la estabilización.

En otro caso se escoge $\arg \min_d f_j(d)$; de esta forma se encuentran tres valores, d_1, d_2, d_3 , y se toma $d = \max_j d_j$. Esta aproximación es rápida pero puede ocasionar problemas para series de tiempo que tienen pocas observaciones (Auestad & Tjøstheim 1990).

4.4.3. Elección del ancho de banda h_n

El ancho de banda determina el suavizamiento del pronóstico. Un ancho de banda pequeño prácticamente reproduce los datos, mientras que uno extremadamente grande produce como estimación una constante.

Si h crece, la varianza del estimador decrece mientras que el sesgo se incrementa y viceversa. Usualmente el ancho de banda óptimo es seleccionado al balancear el trance entre varianza y el cuadrado del sesgo, mediante la minimización de alguna medida global de error.

Generalmente para la estimación de la densidad se usa la siguiente elección de h (Deheuvels 1977):

$$h_{nd} : \widehat{\sigma}_n n^{-1/(d+4)}, \quad (19)$$

donde $\widehat{\sigma}_n$ es el estimador de la desviación estándar de la serie de tiempo $\{Z_t; t = 1, \dots, n\}$. Esta elección garantiza una tasa óptima de convergencia con respecto al criterio del error cuadrado medio. Sin embargo, no es óptima en todos los casos, puesto que no se toman en cuenta las condiciones *mixing*. Otra elección usada es:

$$\widehat{h}_n = c \times h_{nd}, \quad c \in (0, 5]. \quad (20)$$

Dos procedimientos son utilizados para encontrar c . El primero es el de validación cruzada y el segundo es el método empírico.

Validación cruzada: Este procedimiento ha sido estudiado por muchos autores (Györfi, Härdle, Sarda & View 1989); su objetivo es encontrar c que minimice alguna medida de error cuadrático. Dentro de esta categoría existen dos posibilidades.

Un procedimiento se llama *validación cruzada local* (VCL), el cual encuentra un valor de h para cada horizonte de pronóstico, y el otro es la *validación cruzada global* (VCG) donde se encuentra un valor común de h para todos los horizontes de predicción. Las formas de cómputo son las siguientes:

$$VCL(c, m) = (n - k - c + 1)^{-1} \sum_{t=d}^{n-m} \left(\widehat{Z}_{t+m|t}(c) - Z_{t+m} \right)^2, \quad (21)$$

$$\widehat{c} = \arg \min_{c \in [0,5]} VCL(c, m),$$

y

$$VCG(c) = (n - M - k + 1)^{-1} \sum_{t=k}^{n-M} M^{-1} \sum_{m=1}^M \left(\widehat{Z}_{t+m|t}(c) - Z_{t+m} \right)^2 \quad (22)$$

$$\widehat{c} = \arg \min_{c \in [0,5]} VCG(c),$$

donde $\widehat{Z}_{t+m|t}(c)$ denota la versión de dejar la t -ésima observación por fuera y M es el horizonte máximo de predicción.

Método empírico: Este procedimiento usa parte de los datos pasados para encontrar el ancho de banda. El *método empírico local* (MEL) y el *método empírico global* (MEG) son los siguientes:

$$MEL(c, m) = p^{-1} \sum_{t=n-p-m+1}^{n-m} \frac{|\widehat{Z}_{t+m}(c) - Z_{t+m}|}{|Z_{t+m}|} \times 100 \quad (23)$$

$$\widehat{c} = \arg \min_{c \in [0,5]} MEL(c, m),$$

$$MEG(c) = p^{-1} \sum_{t=n-p-M+1}^{n-M} \frac{1}{M} \sum_{m=1}^M \frac{|\widehat{Z}_{t+m}(c) - Z_{t+m}|}{|Z_{t+m}|} \times 100 \quad (24)$$

$$\widehat{c} = \arg \min_{c \in [0,5]} MEG(c).$$

El primer método mencionado encuentra un valor de c para cada horizonte de predicción, y el segundo encuentra un valor de c para todo horizonte de predicción.

4.5. Comparación de los métodos

Seis medidas se usan para medir la capacidad de pronóstico dentro de muestra: el error medio (EM), el error absoluto medio (EAM), el error absoluto porcentual medio (EAPM), la raíz del error cuadrático medio (RECM), la raíz del error cuadrático medio porcentual (RECOMP) y la estadística U de Theil. Sus fórmulas aparecen a continuación:

$$EM(m) = (p - m + 1)^{-1} \sum_{r=0}^{p-m} (Z_{n-p+r+m} - \hat{Z}_{n-p+r+m}), \quad (25)$$

$$EAM(m) = (p - m + 1)^{-1} \sum_{r=0}^{p-m} |\hat{Z}_{n-p+r+m} - Z_{n-p+r+m}|, \quad (26)$$

$$EAPM(m) = (p - m + 1)^{-1} \sum_{r=0}^{p-m} \frac{|\hat{Z}_{n-p+r+m} - Z_{n-p+r+m}|}{|Z_{n-p+r+m}|} \times 100, \quad (27)$$

$$RECM(m) = \sqrt{(p - m + 1)^{-1} \sum_{r=0}^{p-m} \left(\hat{Z}_{n-p+r+m} - Z_{n-p+r+m} \right)^2}, \quad (28)$$

$$RECOMP(m) = \sqrt{(p - m + 1)^{-1} \sum_{r=0}^{p-m} \left(\frac{\hat{Z}_{n-p+r+m} - Z_{n-p+r+m}}{Z_{n-p+r+m}} \times 100 \right)^2}, \quad (29)$$

$$U - Theil(m) = \frac{\sum_{r=0}^{p-m} (Z_{n-p+r+m} - \hat{Z}_{n-p+r+m})^2}{\sum_{r=0}^{p-m} (Z_{n-p+r+m} - Z_{n-p+r})^2}, \quad (30)$$

donde n es el tamaño de la serie, $p = \lceil \frac{n}{5} \rceil$ indica el número de observaciones que son quitadas al final de la serie para hacer comparaciones dentro de muestra, $r = 1, \dots, p$ indica el aumento de tamaño de muestra.

4.6. Comparación estadística de los métodos

Bajo el enfoque de Diebold & Mariano (1995) –DM en adelante–, se parte de la hipótesis nula de que el nuevo modelo no mejora al modelo existente o modelo referencia, contra la alternativa de que el modelo nuevo mejora al modelo referencia. Interesa por tanto rechazar la hipótesis nula.

$$H_0 : DM_i \geq 0,$$

$$H_1 : DM_i < 0.$$

Para $i = 1, 2, 3$, usando

$$DM_1 = (O - PN)^2 - (O - PE)^2,$$

$$DM_2 = |O - PN| - |O - PE|,$$

$$DM_3 = \left| \frac{PN}{obs_{PN} - 1} \right| - \left| \frac{PE}{obs_{PE} - 1} \right|,$$

donde: O = Dato observado, PE = Pronóstico modelo referencia, PN = Pronóstico modelo nuevo, Obs = Número de observaciones usadas para pronosticar.

Bajo la hipótesis nula, tanto DM_1 , DM_2 como DM_3 se distribuyen cada una como una normal estándar. Así, lo deseable es obtener valores para estas estadísticas menores que cero (0) con p -valores pequeños.

5. Aplicación práctica

5.1. Datos

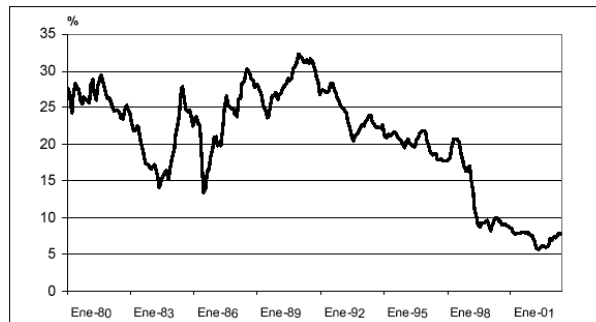
Para la selección de coeficientes, inicialmente se tomó la serie de inflación colombiana medida mensualmente a través del IPC calculado por el Departamento Administrativo Nacional de Estadística, DANE, a partir de enero de 1980 hasta septiembre de 2002; contiene un total de 273 datos. Fuente: Archivos del Banco de la República Bogotá Colombia¹⁰. Computacionalmente se trabajan cifras en escala (0,1).

Con el objeto de encontrar coeficientes óptimos y comparar los pronósticos de este modelo con los de los modelos paramétricos, se tomaron datos desde febrero de 1980 hasta marzo de 1998 ($n - p + r$ datos, con $p = \lceil \frac{n}{5} \rceil$ y $r = 1, \dots, n - p$), es decir, iniciando con 219 observaciones; luego se aumenta el tamaño de muestra hasta septiembre de 2002.

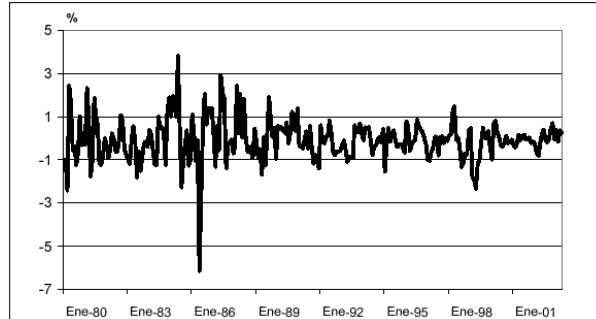
¹⁰A pesar de ser deseable en la práctica estadística —especialmente la no paramétrica— utilizar la mayor cantidad de información posible, en este trabajo se descartan cifras previas por presentar cambios muy bruscos en niveles, además de hacer comparables los resultados con otros modelos que usan la misma información.

5.2. Estacionarización

En la gráfica 14 se presenta la evolución de la inflación colombiana. Se puede apreciar que la serie no tiene media constante, y aunque parece presentar cambio de varianza, se ha considerado que de existir y modelar los mismos¹¹, esto no afecta significativamente los pronósticos bajo los métodos no paramétricos, al igual que no lo hace en los métodos paramétricos tipo GARCH. Para estabilizar la media se realizó una diferenciación de orden uno; en la gráfica 15 se ve la inflación diferenciada, que muestra un comportamiento constante en la media. Se deja para futuro trabajo la detección y el modelaje de los posibles cambios en varianza.



Gráfica 14: Inflación colombiana.



Gráfica 15: Inflación diferenciada.

¹¹Esto se puede llevar a cabo, entre otras alternativas, en el espíritu de Hardle & Yang (1996).

5.3. Predictor usado

El predictor usado es el de la media condicional mencionado en la sección 4.3.1, ecuación 10. Este predictor fue utilizado porque la serie no presenta valores atípicos; además, el uso de los diferentes predictores no paramétricos basados en kernel mencionados en la sección 4.3 genera resultados similares, según conclusiones de estudios realizados para series de tiempo con más de 100 observaciones (Matzner-Løber et al. 1998, Gooijer & Zerom 2000).

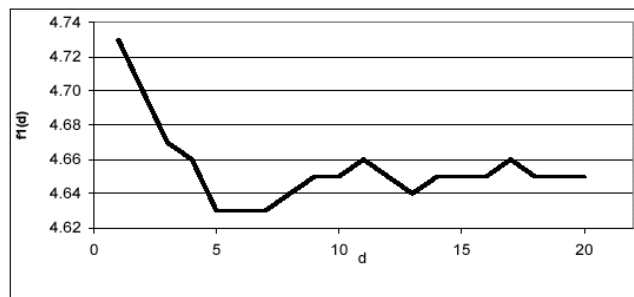
Para obtener los pesos que se le da a cada uno de los vectores de orden d comparados con el último vector de los $n - p + r$ elementos de la serie, se ha hecho uso de la función kernel gaussiana multivariada no correlacionada de la ecuación 15, siguiendo conclusiones que aseguran que cualquier elección de la función continua usada presenta resultados similares.

5.4. Elección de coeficientes

5.4.1. Elección inicial de coeficientes

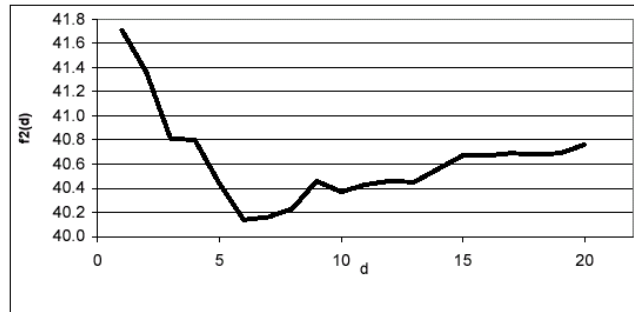
Elección del coeficiente de Markov óptimo d : Se hizo por medio de las tres funciones $f_j(d)$ $j = 1, 2, 3$ vistas en la sección 4.4.2, con anchos de banda de acuerdo con cada d , dados por el h de la fórmula 19. Las funciones tomaron valores $d = 1, \dots, 20$. Tomándose 20 como el valor máximo para alcanzar a capturar la estacionalidad en los datos, no se consideran valores mayores por la excesiva pérdida en información.

Las gráficas para estas tres funciones se encuentran a continuación:

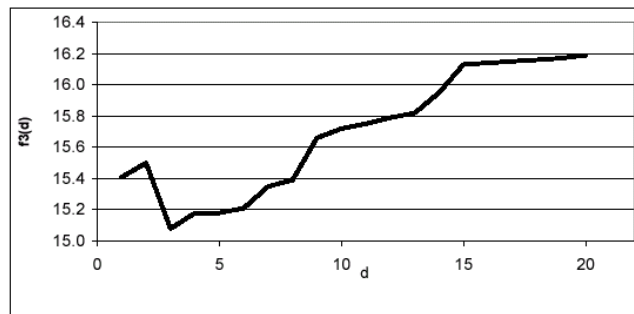


Gráfica 16: Función $f_1(d)$.

Se puede ver que las funciones $f_1(d)$ y $f_2(d)$ decrecen antes e incrementan



Gráfica 17: Función $f_2(d)$.



Gráfica 18: Función $f_3(d)$.

a partir de $d = 6$ y $f_3(d)$ hace lo mismo en $d = 4$. Como se tienen dos valores para d , se tomó el mayor de ellos como el coeficiente óptimo de Markov, esto es, $d = 6$.

Elección del ancho de banda óptimo h_n : Se obtuvo por medio de la fórmula mostrada en la sección 4.4.3, ecuación 19; el valor encontrado para h con $d = 6$ es $h_n = 0,56$. Los anteriores valores de d y h_n fueron usados para todos los horizontes de predicción. El resultado de pronósticos con dichos coeficientes no mejoraban los de un modelo ARIMA; se omite su presentación aquí.

5.4.2. Elección posterior de coeficientes

Pensando en mejorar estos resultados se hallaron coeficientes d_m y h_{nm} para cada horizonte de predicción, utilizando una medida modificada del MEL que se presentó en la ecuación 23.

Este procedimiento se llevó a cabo obteniendo errores de pronóstico para cada $m = 1, \dots, 12$, tamaños de muestra $n - d, \dots, n$ y para valores de $h_n = 0,001, \dots, 0,1, \dots, 1$ y de $d = 1, \dots, 20$. Para comparar la calidad de pronóstico se usó la siguiente medida:

$$MELM(h_n, m, d) = (p - m + 1)^{-1} \sum_{r=0}^{p-m} \left| \hat{z}_{n-p+r+m}(h_n, m, d) - z_{n-p+r+m} \right|,$$

$$\hat{h}_{nm} = \arg \min_{h \in (0,1]} MELM(h_n, m, d).$$

Esta medida permitió hacer un resumen para todos los tamaños de muestra y encontrar los d y h_n en cada horizonte de predicción para los cuales esta medida era mínima; los resultados se encuentran en la tabla 1.

Llama la atención el hecho de que para horizontes largos –léase 11 y 12 meses–, el orden del proceso markoviano resulta extremadamente pequeño; esto puede estar siendo causado porque para dicha cantidad de rezagos comienza a ser evidente el problema del curso de la dimensionalidad. Dicho de otra forma, se disponen de pocas observaciones para estimar esas relaciones, haciendo que la porción de puntos que caen en los límites sea muy grande cuando d crece. Véase, entre otros, Hastie et al. (2002).

5.5. Comparación de los métodos

Para la elección de d_m y h_{nm} mostrada en la tabla 1 se obtienen las medidas de bondad de pronóstico como se presenta en la tabla 2; ordenadas éstas por

Tabla 1: Resultados d_m y h_{nm} óptimos.

m	d_m	h_{nm}
1	12	0.007
2	11	0.007
3	10	0.007
4	9	0.007
5	8	0.006
6	14	0.009
7	6	0.006
8	15	0.009
9	4	0.004
10	13	0.008
11	2	0.005
12	2	0.005

RECM, los resultados son alentadores. El período de evaluación es de enero de 2000 a mayo de 2003. Detallando el error medio se encuentra que el método no paramétrico, excepto para cinco y seis meses, es el que consistentemente genera pronósticos con menores sesgos (absolutos) que las otras dos metodologías. Según el ordenamiento presentado, el pronóstico no paramétrico supera a los otros dos en 9 de las 12 ocasiones, pero si los resultados se ordenaran por EAM, EAPM o U-Theil, el NP supera a los ARIMA y no lineal en todos los horizontes, y el RECM favorece al no paramétrico en 11 de los 12 horizontes. Todo lo anterior muestra además la importancia de la correcta selección de los coeficientes d y h .

Los valores de las estadísticas DM_i y sus p -valores, $P-DM_i$, se presentan en la tabla 3; el modelo referencia es el de caminata aleatoria. En este caso, los resultados no son tan favorables. No obstante que en general ninguno de los pronósticos muestra mejorar estadísticamente a los de un pronóstico ingenuo (caminata aleatoria), cuando de pronosticar un mes adelante se trata, el NP es el único que al 10% de significación mejora el pronóstico de ingenuidad. Además, el NP es el que muestra menores valores de probabilidad para todas las estadísticas. No se descarta el posible problema de insuficiente tamaño de muestra para validar el resultado teórico de DM.

Queda abierta la discusión de si los pronósticos no paramétricos resultan o no menos persistentes, esto es, menos influenciados por la información más reciente utilizada, que los modelos ARIMA y STAR. Dados los resultados de la anterior comparación, se podría conjeturar que sí, aunque es recomendable hacerles un seguimiento detallado en tiempo real, antes de aseverar conclusiones al respecto.

Tabla 2: Resultados d_m y h_{nm} óptimos.

Modelo	m	Obs.	EM	EAM	EAPM	RECM	RECOMP	U-THEIL
NP	1	41	0.01	0.22	3.01	0.30	4.20	0.89
ARIMA	1	41	-0.16	0.27	3.48	0.35	4.51	1.05
STR	1	41	0.02	0.34	4.50	0.42	5.52	1.27
NP	2	40	0.01	0.38	5.23	0.50	7.44	0.87
ARIMA	2	40	-0.38	0.54	7.00	0.69	9.08	1.20
STR	2	40	0.03	0.57	7.59	0.70	9.41	1.22
NP	3	39	0.00	0.51	7.24	0.68	10.35	0.92
STR	3	39	0.05	0.77	10.11	0.93	12.43	1.25
ARIMA	3	39	-0.64	0.83	10.96	1.01	13.73	1.37
NP	4	38	-0.04	0.63	9.00	0.83	12.65	0.97
STR	4	38	0.04	0.84	11.11	1.02	13.82	1.19
ARIMA	4	38	-0.94	1.12	14.62	1.34	18.09	1.56
NP	5	37	-0.09	0.77	11.03	0.99	14.98	1.03
STR	5	37	0.06	0.92	12.30	1.11	15.08	1.15
ARIMA	5	37	-1.20	1.39	18.30	1.63	22.07	1.69
STR	6	36	0.09	1.07	14.35	1.22	16.80	1.13
NP	6	36	-0.12	0.94	13.24	1.17	17.21	1.08
ARIMA	6	36	-1.47	1.67	22.23	1.93	26.13	1.79
STR	7	35	0.17	1.13	15.28	1.33	18.37	1.10
NP	7	35	-0.11	1.05	14.68	1.28	18.63	1.06
ARIMA	7	35	-1.71	1.93	25.95	2.21	30.05	1.82
STR	8	34	0.25	1.25	16.90	1.45	20.00	1.11
NP	8	34	-0.15	1.21	16.67	1.46	20.31	1.11
ARIMA	8	34	-1.98	2.20	29.85	2.48	34.03	1.89
NP	9	33	-0.19	1.26	17.30	1.51	20.68	1.10
STR	9	33	0.33	1.35	18.41	1.57	21.82	1.15
ARIMA	9	33	-2.29	2.49	34.03	2.77	38.03	2.01
NP	10	32	-0.23	1.33	18.09	1.62	21.61	1.13
STR	10	32	0.40	1.41	19.53	1.71	23.72	1.20
ARIMA	10	32	-2.61	2.75	37.78	3.07	42.29	2.15
NP	11	31	-0.29	1.33	18.21	1.65	22.00	1.11
STR	11	31	0.48	1.54	21.41	1.83	25.63	1.23
ARIMA	11	31	-2.95	3.03	42.05	3.36	46.77	2.27
NP	12	30	-0.33	1.33	18.21	1.65	22.08	1.08
STR	12	30	0.57	1.62	22.69	1.96	27.54	1.29
ARIMA	12	30	-3.28	3.34	46.62	3.68	51.48	2.41

Tabla 3: Comparación estadística de los métodos.

Modelo	m	DM_1	P_DM_1	DM_1	P_DM_2	DM_1	P_DM_3
NP	1	-1.49	0.07	-1.92	0.03	-1.88	0.03
ARIMA	1	-0.42	0.34	-0.35	0.36	-0.35	0.36
STR	1	1.68	0.95	1.85	0.97	1.83	0.97
NP	2	-0.22	0.41	-0.38	0.35	-0.40	0.34
ARIMA	2	1.13	0.87	1.28	0.90	1.26	0.90
STR	2	1.70	0.96	1.83	0.97	1.88	0.97
NP	3	0.05	0.52	-0.49	0.31	-0.37	0.36
STR	3	1.82	0.97	1.89	0.97	1.91	0.97
ARIMA	3	3.05	1.00	3.10	1.00	3.02	1.00
NP	4	0.28	0.61	-0.23	0.41	-0.04	0.48
STR	4	1.64	0.95	1.46	0.93	1.47	0.93
ARIMA	4	3.67	1.00	3.66	1.00	3.61	1.00
NP	5	0.79	0.78	0.06	0.53	0.25	0.60
STR	5	1.35	0.91	1.07	0.86	1.16	0.88
ARIMA	5	3.94	1.00	3.77	1.00	3.71	1.00
STR	6	1.08	0.86	1.10	0.86	1.03	0.85
NP	6	1.34	0.91	0.43	0.66	0.46	0.68
ARIMA	6	3.88	1.00	3.65	1.00	3.67	1.00
STR	7	0.11	0.54	0.26	0.60	0.08	0.53
NP	7	0.26	0.60	-0.33	0.37	-0.38	0.35
ARIMA	7	3.91	1.00	3.51	1.00	3.51	1.00
STR	8	-0.31	0.38	-0.06	0.48	-0.13	0.45
NP	8	0.35	0.64	-0.01	0.50	-0.24	0.40
ARIMA	8	4.14	1.00	3.44	1.00	3.55	1.00
NP	9	0.49	0.69	-0.04	0.49	-0.34	0.37
STR	9	-0.26	0.40	0.00	0.50	-0.22	0.41
ARIMA	9	4.34	1.00	3.93	1.00	4.10	1.00
NP	10	0.24	0.59	-0.06	0.48	-0.39	0.35
STR	10	0.18	0.57	-0.06	0.48	-0.18	0.43
ARIMA	10	4.35	1.00	3.88	1.00	4.12	1.00
NP	11	-0.11	0.45	-0.52	0.30	-0.96	0.17
STR	11	0.85	0.80	0.38	0.65	0.30	0.62
ARIMA	11	4.35	1.00	3.98	1.00	4.21	1.00
NP	12	-0.85	0.20	-0.92	0.18	-1.29	0.10
STR	12	1.52	0.94	0.79	0.79	0.66	0.75
ARIMA	12	4.38	1.00	4.12	1.00	4.29	1.00

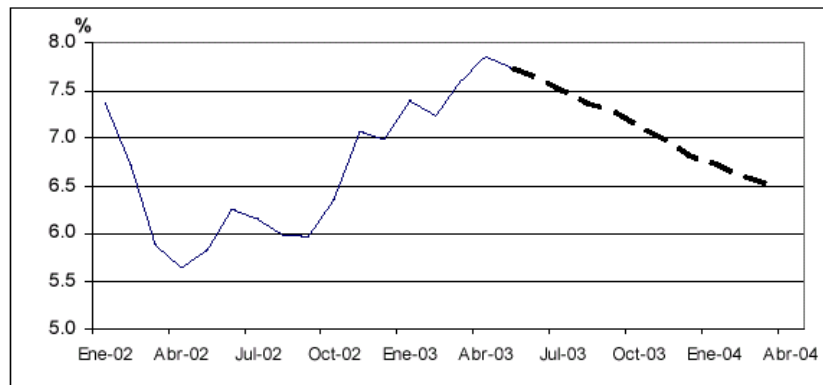
Tabla 4: Pronósticos usando la serie completa.

Mes	m	d_m	h_m	Pronóstico
Jun-03	1	12	0.007	7.6
Jul-03	2	11	0.007	7.5
Ago-03	3	10	0.007	7.4
Sep-03	4	9	0.007	7.3
Oct-03	5	8	0.006	7.1
Nov-03	6	14	0.009	7.0
Dic-03	7	6	0.006	6.8
Ene-04	8	15	0.009	6.7
Feb-04	9	4	0.004	6.6
Mar-04	10	13	0.008	6.5
Abr-04	11	2	0.005	6.4
May-04	12	2	0.005	6.3

5.6. Pronósticos usando información de la serie completa

Se realizaron pronósticos para los 12 meses siguientes al final de la muestra disponible a la fecha. Para ello se usaron los mismos coeficientes y rezagos reportados en la sección 5.4.2. Los resultados se pueden ver en la tabla 4 y en la gráfica 19.

Se nota que los pronósticos presentan tendencia a decaer, y al parecer no son influenciados por el comportamiento de la información más reciente utilizada, en este caso, febrero a mayo de 2003.



Gráfica 19: Pronósticos de la inflación enero de 2002 a abril de 2004.

6. Conclusiones y recomendaciones

En este trabajo se aplicaron técnicas no paramétricas tipo kernel para generar y evaluar pronósticos de la inflación colombiana. Los resultados de la evaluación de pronósticos son alentadores, pues en general mejoran, en sentido estadístico y computacional, a los de las metodologías actualmente aplicadas.

La elección del ancho de banda h y del coeficiente de Markov d desempeña un papel muy importante en los pronósticos de series de tiempo, en particular para la inflación colombiana, usando métodos no paramétricos basados en kernel; se encontró que la elección final de los coeficientes para cada horizonte de predicción permite mejorar los pronósticos dados bajo los modelos paramétricos ARIMA y no lineal STAR.

En la literatura especializada se encuentran diversas ideas para encontrar intervalos de pronósticos bajo estos métodos no paramétricos, los cuales pueden aplicarse para tener, además del pronóstico puntual, un rango de su confiabilidad. Johnston (1982), usando Bickel & Rosenblatt (1973); los *bootstrapping* son otra alternativa aunque intensiva en cómputo.

Existen también en la literatura no paramétrica reportes de los efectos adversos de las condiciones de borde (*“boundary conditions”*) o agrupamiento en alguno(s) de los límites internos de las bandas de suavizamiento, lo cual puede generar sesgos en las estimaciones. El kernel especial de Gasser & Muller (1979) puede ser usado para sobrellevar este problema; otra alternativa de solución es usar regresión lineal local o más aún polinómica local; se puede consultar Fan & Gijbels (1996). Otra opción sería usar funciones kernel, aunque normales, no independientes.

No se desea terminar sin antes resaltar el hecho de que el tema de análisis no paramétrico de series en el tiempo es un área de extensa aplicación no explotada aún en nuestro país; por ello se invita a otros investigadores y académicos a examinar en esta prometedora área.

Nota: Este documento se basa en el trabajo de grado presentado por Patricia Siado para obtener el título de Estadística, del Departamento de Estadística de la Universidad Nacional de Colombia. Se agradecen los comentarios de los jurados calificadores y la ayuda del profesor Fabio H. Nieto. De mucha ayuda fue el curso “Estimación no paramétrica y robusta en series de tiempo” dictado por Dr. Siegfried Heiler, profesor de la Universidad de Konstanz, Alemania, para el doctorado de Estadística de la Universidad Nacional de Colombia, Bogotá, agosto a septiembre de 2001. No obstante, cualquier error que persista es de nuestra exclusiva responsabilidad. Esta versión del trabajo se vio nutrida de las discusiones con Héctor Zárate, a quien también se le agradece. Los errores y omisiones son únicamente nuestros. Cualquier posible opinión expresada aquí no compromete la posición oficial del Banco de la República ni tampoco la de ninguno de los miembros de su Junta Directiva.

A. ANEXO A

La estimación no paramétrica de densidades puede ser útil en el análisis exploratorio de datos, pero puede ser también usada para problemas más estándar en pruebas de hipótesis e inferencia estadística.

A.1. Estimador de densidades usando histogramas

Cuando el objetivo de estudio es construir un modelo de distribución de probabilidad para un conjunto de datos, no se necesita hacer suposiciones sobre la posible distribución de la cual provienen los datos, sino que se puede estimar directamente la función de densidad a partir de los datos, siempre y cuando se tenga una muestra grande.

A continuación se muestra la construcción del estimador no paramétrico para funciones de densidad de variables aleatorias, que ha sido base de muchos predictores no paramétricos.

El estimador natural es el *histograma* con n datos e intervalos de amplitud h . La estimación del histograma de la función de densidad en el punto x es:

$$\hat{f}(x) = \frac{1}{h} \frac{n_{(0)}}{n}, \quad (31)$$

donde $n_{(0)}$ es el número de datos en el intervalo $\left[\dot{x} - \frac{h}{2}, \dot{x} + \frac{h}{2}\right)$, donde \dot{x} es una marca de clase, h es la amplitud de intervalo y n el tamaño total de la muestra.

Esta estimación es fácil de calcular pero tiene la desventaja de ser constante dentro del intervalo, y las estimaciones son muy dependientes del origen y de la amplitud del intervalo, pues considera únicamente los datos dentro de cada uno ignorando los datos adyacentes por próximos que estén; para resolver este último problema se da cierto peso a los datos de intervalos contiguos al que se está estimando.

Tomando como $n_{(0)}$ el número de datos en el intervalo $\left(\dot{x} \pm \frac{h}{2}\right)$, $n_{(h)}$, $n_{(-h)}$ al número de datos en los intervalos adyacentes $\left(\dot{x} + h \pm \frac{h}{2}\right)$ y $\left(\dot{x} - h \pm \frac{h}{2}\right)$, respectivamente, se construye un estimador que asigna cierto peso a los datos que están en estos intervalos, y para los intervalos adyacentes el peso es el mismo, por simetría, obteniéndose

$$\hat{f}(x) = \frac{1}{hn} \left[\alpha_0 n_{(0)} + \alpha_1 \{n_{(h)} + n_{(-h)}\} \right],$$

para $x \in \left[\hat{x} - \frac{h}{2}, \hat{x} + \frac{h}{2} \right)$, donde $\alpha_0, \alpha_1 > 0$ y $\alpha_0 + 2\alpha_1 = 1$.

Esta idea puede generalizarse incluyendo el resto de los intervalos con peso decreciente para obtener:

$$\hat{f}(x) = \frac{1}{hn} \left[a_0 n_{(0)} + \sum_{i=1}^m \alpha_i [n_{(hi)} + n_{(-hi)}] \right], \quad (32)$$

donde $\alpha_0 + 2 \sum \alpha_i = 1$, para valores de i tales que los intervalos formados contengan elementos de la muestra.

El anterior estimador puede aplicarse si se divide el rango o soporte de valores de la variable x en k puntos x_1, \dots, x_k , para k tan grande como se quiera; se elige un valor de h y se aplica la ecuación 32 a cada punto. La estimación de $\hat{f}(x_i)$ equivale a construir un histograma con centros de clase:

$$x_i - mh; \quad x_i - (m-1)h, \dots, x_i - h, \quad x_i, \quad x_i + h, \dots, x_i + (m-1)h, \quad x_i + mh,$$

y estimar la densidad en el punto x_i aplicando la ponderación simétrica 32. Para calcular $\hat{f}(x_{i+1})$ se toma x_{i+1} como nuevo punto central y se aplica de nuevo 32.

Este proceso equivale a calcular la frecuencia absoluta en cada punto dando ciertos coeficientes de ponderación a cada uno de los datos, la cual depende de la distancia a dicho punto.

A.2. Estimación de densidad por el método kernel

El estimador encontrado anteriormente es llamado el estimador de densidad kernel. Formalmente se supone que los datos x_1, \dots, x_n vienen de una secuencia de variables reales aleatorias independientes con una densidad común f perteneciente a alguna familia \mathfrak{S} .

Si \mathfrak{S} es grande (por ejemplo si \mathfrak{S} contiene las densidades continuas), se sabe que el estimador insesgado de f puede no existir y que el supremo de la función máximo verosímil es infinito (Bosq 1998, pág. 3).

Entonces un estimador de densidad primario es el del histograma mencionado en la sección anterior, definido formalmente como:

$$\hat{f}_n(x) = \frac{v_{nj}}{n(a_{n,j} - a_{n,j-1})}, \quad x \in I_{nj}, \quad j \in Z,$$

donde $I_{nj} = [a_{n,j-1}, a_{n,j})$ y $(a_{nj}, j \in Z)$ es una secuencia estrictamente creciente tal que $|a_{jn}| \rightarrow \infty$ cuando $|j| \rightarrow \infty$ y donde $v_{nj} = \sum_{i=1}^n 1_{[a_{n,j-1}, a_{n,j})}(X_i)$.

Si f es continua en el intervalo I_{nj} y si $a_{n,j} - a_{n,j-1}$ es pequeño, entonces $\hat{f}_n(x)$ está cercano a $f(x)$ para cada x que pertenece a I_{nj} .

Dado que este estimador no utiliza la información de toda la muestra, solamente los datos contenidos en el intervalo I_{nj} , se construye el estimador histograma adaptable definido como:

$$f_n^* = \frac{v_n(x)}{nh_n}, \quad x \in R,$$

donde

$$v_n(x) = \sum_{i=1}^n I_{\left[x - \frac{h_n}{2}, x + \frac{h_n}{2}\right]}(X_i).$$

Así $f_n^*(x)$ se puede escribir en términos de una función continua, simétrica y acotada $K(\cdot)$ que da pesos a los datos en el intervalo dependiendo de la distancia al punto x :

$$f_n^*(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right), \quad x \in R. \quad (33)$$

Considerando el caso en que las observaciones sean la realización de un proceso estocástico $\{X_t\}$, por la extensión del teorema de Kolmogorov la distribución v de un proceso estocástico está completamente especificada por las distribuciones finito dimensionales (Bhat 1933). El problema de estimación de v se reduce a las estimaciones de las densidades finito dimensionales asociadas.

Así si $(X_t, t \in Z)$ es un proceso estocástico de dimensión R^d con función de densidad común f , el estimador de densidad kernel d -dimensional se puede escribir como:

$$\begin{aligned} f_n(x) &= \frac{1}{nh_n^d} \sum_{t=1}^n K_d\left(\frac{x - X_t}{h_n}\right) \\ &= \frac{1}{nh_n^d} \sum_{t=1}^n K_d\left(\frac{x_1 - X_{1t}}{h_n}, \dots, \frac{x_d - X_{dt}}{h_n}\right), \quad x \in R^d, \end{aligned} \quad (34)$$

donde $K_d(\cdot)$ es una función kernel d -variada.

B. ANEXO B

B.1. Condiciones *mixing*

Las condiciones *mixing* son herramientas matemáticas planteadas para dar propiedades asintóticas a los estimadores basadas en kernel para datos dependientes. Básicamente éstas prueban el control de la dependencia entre X_t y X_s cuando la distancia en el tiempo $t - s$ se incrementa. Generalmente estas condiciones son difíciles de chequear, sin embargo si el proceso es una cadena de Markov estacionaria, entonces la ergodicidad geométrica implica regularidad absoluta, lo cual implica las condiciones *strong mixing* (Härdle, Lütkepohl & Chen 1997, págs. 53-54).

B.1.1. Condición *strong mixing* o α -*mixing*

Para A un elemento de la σ -álgebra φ_k generada por $[\{X_s\}, s \leq t]$ y B un elemento de la σ -álgebra φ_{k+s} generada por $[\{X_s\}, s \geq t+k]$, se dice que una secuencia es α -*mixing* si:

$$\sup_{\varphi_k, \varphi_{k+s}} |P(A \cap B) - P(A)P(B)| \leq \alpha_k,$$

donde $\alpha_k \rightarrow 0$ cuando $k \rightarrow \infty$.

B.1.2. Condición *uniformly mixing* ó ϕ -*mixing*

Se dice que una secuencia es ϕ -*mixing* si:

$$|P(A \cap B) - P(A)P(B)| \leq \phi_k P(A),$$

donde $\phi_k \rightarrow 0$ cuando $k \rightarrow \infty$.

Bibliografía

- Auestad, B. & Tjøstheim, D. (1990), 'Identification of nonlinear time series: First order characterization and order determination', *Biometrika* **77**, 669–687.
- Bhat, U.Ñ. (1933), *Elements of Applied Stochastic Processes*, John Wiley & Sons, New York.
- Bickel, P. J. & Rosenblat, M. (1973), 'On some global measures of the deviations of density function estimates', *Annals of Statistics* **1**, 1071–1091.
- Bosq, D. (1998), *Nonparametric Statistics for Stochastic Processes*, Springer-Verlag, New York.
- Carbon, M. & Delecroix, M. (1993), 'Nonparametric vs parametric forecasting in time series: a computational point of view', *Applied Stochastic Models and Data Analysis* **9**, 215–229.
- Collomb, G. (1984), 'Propriétés de convergence presque complète du prédicteur à noyau', *Zeitschrift für Wahrscheinlichkeitstheorie* **66**, 441–460.
- Collomb, G., Hardle, W. & Hassani, S. (1987), 'A note on prediction via estimation of the conditional mode function', *Journal of Statistical Planning and Inference* **15**, 227–236.
- Deheuvels, P. (1977), 'Estimation non paramétrique de la densité par histogramme généralisé', *Revue de Statistique Appliquée* **35**, 5–42.
- Diebold, F. X. & Mariano, R. S. (1995), 'Comparing predictive accuracy', *Journal of Business and Economic Statistics* **13**, 253–263.
- Fan, J. & Gijbels, I. (1996), *Local Polynomial Modeling and Its Applications*, Chapman and Hall, London.
- Gannoun, A. (1990), 'Estimation non paramétrique de la médiane conditionnelle: médiano gramme et méthode du noyau', *Publication de l'Institut de Statistique de l'Université de Paris* **45**, 11–22.
- Gannoun, A. (1991), 'Prédiction non paramétrique: médianogramme et méthode du noyau en estimation de la médiane conditionnelle', *Statistique et Analyse des Données* **16**, 23–42.
- Gasser, T. & Muller, H. G. (1979), Kernel estimation of regression functions, in 'Smoothing Techniques for Curve Estimation', Springer-Verlag, Heidelberg, pp. 23–68.

- Gooijer, J. D. & Zerom, D. (2000), 'Kernel-based multistep-ahead predictions of the us short-term interest rate', *Journal of Forecasting* **19**, 335–353.
- Györfi, L., Härdle, W., Sarda, P. & View, P. (1989), *Nonparametric Curve Estimation from Time Series*, Springer-Verlang, New York.
- Györfi, L., Kohler, M., Krzizak, A. & Walk, H. (2002), *A Distribution Free Theory on Nonparametric Regression*, Springer-Verlang, New York.
- Härdle, W. (1990), *Applied Non-parametric Regression*, Cambridge University Press, New York.
- Härdle, W., Lütkepohl, H. & Chen, R. (1997), 'A review of nonparametric time series analysis', *International Statistical Review* **65**, **1**, 49–72.
- Hardle, W. & Yang (1996), *Nonparametric Time Series Model Selection*, Humboldt-Universität zu Berlin, Discussion paper.
- Hart, J. D. & Wherly, T. E. (1986), 'Kernel regression estimation using repeated measurement data', *Journal of the American Statistical Association* **81**, 1080–1088.
- Hastie, T., Tibshirani, R., Friedman, J. H. & Friedman, J. (2002), *Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, New York.
- Heiler, S. (1999), A survey on nonparametric time series analysis, Working paper, Universität Konstanz Fakultät für Wirtschaften.
- Ibragimov, I. A. & Rozanov, Y. (1978), *Gaussian Random Processes*, Spring Verlag, New York.
- Jalil, M. & Melo, V. (2000), 'Una relación no lineal entre inflación y los medios de pago', *Borradores de economía. Banco de la República* **145**, Bogotá.
- Johnston, G. J. (1982), 'Probabilities of maximal deviations for nonparametric regression functions estimates', *Journal of Multivariate Analysis* **12**, 402–414.
- Kreiss, J. P. & Franke (1992), 'Bootstrapping stationary autoregressive moving-average models', *Journal of Time Series Analysis* **13**, 297–317.
- Liu, R. Y. & Singh (1992), 'Moving blocks jackknife and bootstrap capture weak dependence', *Exploring the Limits of Bootstrap* pp. 225–248.

- Matzner-Løber, E., Gannoun, A. & Gooijer, J. G. D. (1998), 'Nonparametric forecasting: a comparison of three kernel-based methods', *Communications in Statistics: Theory and Methods* **27**, 1593–1617.
- Melo, L. F. & Misas, M. A. (1998), 'Análisis del comportamiento de la inflación trimestral en Colombia bajo cambios de régimen: una evidencia a través del modelo "switching de hamilton"', *Borradores de Economía. Banco de la República* **086**.
- Misas, M. A., López, E. & Querubín, P. (2002), 'La inflación en Colombia: Una aproximación desde las redes neuronales', *Borradores de Economía. Banco de la República* **199**.
- Núñez, J. & Jiménez, J. (1998), Correcciones a los ingresos de las encuestas de hogares y distribución del ingreso urbano, *in* c. F. Sánchez, ed., 'La distribución del ingreso en Colombia, DNP', Tercer Mundo Editores.
- Pagan, A. & Ullah, A. (1999), *Nonparametric Econometrics*, Cambridge University Press, UK.
- Pham, T. D. & Tran (1985), 'Some strong mixing properties of time series models', *Stochastic Processes and their Applications* **19**, 297–303.
- Priestsley, M. B. (1984), *Spectral Analysis and Time Series*, Academic Press, London.
- Zárate, H. M. (2003), 'Cambios en la estructura salarial: Una historia desde la regresión cuantílica', *Borradores de Economía, Banco de la República* **245**.