
Intervalos de predicción para pronósticos no paramétricos de la inflación colombiana

Prediction intervals for nonparametric forecasting of colombian inflation

Fabio Guacaneme^a
fabio.guacaneme@ui.colpatria.com

Resumen

Este trabajo contiene los resultados de algunas aplicaciones del suavizamiento Kernel. Se presentan resultados del suavizamiento para la serie de la inflación total colombiana; predicciones múltiples pasos adelante en base a los predictores denominados media y mediana condicional, las predicciones generadas son comparadas con un modelo ARIMA, un modelo STAR y con redes neuronales; finalmente, se comparan los intervalos de predicción del modelo ARIMA con la técnica no paramétrica. Se encuentra que los intervalos de la segunda técnica son mejores en el periodo de tiempo evaluado.

Palabras clave: Ancho de banda, estimador Kernel, media y mediana condicional, regresión no paramétrica, regresión polinómica local.

Abstract

This paper contains one of the Kernel smoothing applications. It shows Colombian total inflation's smoothing results; multi-steps-ahead forecast, based on Mean and Median conditional predictors, the forecast generated are compared with an ARIMA and STAR models and with neuronal networks; finally the intervals of prediction of ARIMA model using nonparametric methods are compared. It found that the intervals using the second method are better in the period of evaluated time.

Key words: Bandwidth, Kernel Estimator, Local Polynomial Regression, Mean and Median Conditional, Non-parametric Regression.

^aProfesional estadístico. Seguros Colpatria.

1. Introducción

La inflación, en especial cuando está fuera de control, es catalogada como uno de los mayores problemas de la economía. Consiste en un aumento general y continuo de los niveles de precios y se manifiesta a través de una relativa disminución del valor del dinero, es decir, la reducción en el poder adquisitivo a causa de una elevación de los precios con relación a la cantidad de bienes y servicios que se puedan comprar con un mismo monto.

En términos generales, un proceso inflacionario se inicia por diversas causas, siendo una de la más comunes la que se origina por un desequilibrio marcado entre la oferta y la demanda, es decir, cuando por diversas presiones económicas la demanda de bienes y servicios es superior a la oferta disponible de los mismos a los precios actuales, o cuando la oferta está limitada por la baja productividad o por restricciones del mercado.

La reducción del poder adquisitivo del dinero afecta generalmente a los individuos más pobres; en general, desacelera e incluso frena el crecimiento del país y reduce la eficiencia de la economía, razones por las cuales, los principales objetivos del Banco de la República, establecidos por la Constitución Política de Colombia en el año 1991, son reducir y controlar los niveles de la inflación. Pero mientras se consigue llegar a un nivel bajo y estable de esta, se pueden presentar problemas en materia de desempleo. Debido a lo anterior, la Junta Directiva del Banco ha optado por tomar medidas encaminadas a reducirla gradualmente. Una vez el objetivo de una inflación baja y estable se alcance, se podrá contar con una economía más confiable y atractiva para la inversión tanto nacional como extranjera, y por tanto con mayor capacidad para generar empleo y crecer a niveles aceptables. Con todo lo anterior en mente, es importante tener pronósticos confiables tanto puntual como por intervalo.

2. Justificación

Rodríguez & Siado (2003) comparan los resultados de pronóstico de un modelo ARIMA estacional, un modelo tipo STAR y aquellos que se obtienen con los métodos no paramétricos. En dicho trabajo, que se constituye como el primero de su género para Colombia, se concluye que los pronósticos generados con el modelo no paramétrico son estadísticamente mejores de acuerdo con las medidas de bondad de pronóstico y un periodo de evaluación particular.

Los métodos no paramétricos han evidenciado sus bondades pronosticando la inflación al menos cuando dichos pronósticos mensuales se hacen a corto plazo (de uno hasta cuatro o cinco meses); sin embargo, falta la construcción de los intervalos de pronóstico que proporcionen un rango de credibilidad y pronósticos que sean más confiables en el mediano plazo, considerando un horizonte desde seis hasta doce o más meses.

Incrementar el rango de confiabilidad mencionado es la razón principal de este trabajo, además por que obteniendo buenos pronósticos de la inflación le permite al Banco de la Republica controlar y tomar las medidas necesarias, si se evidencia que la tasa objetivo, en determinado momento, no se está logrando.

3. Modelos paramétricos para series de tiempo

En este trabajo, con fines comparativos, hace uso de dos modelos paramétricos, el primero es un modelo lineal ARIMA y el segundo es un modelo no lineal de transición suave STAR.

3.1. Modelos ARIMA

Este tipo de modelos pueden ser expresados de la siguiente manera,

$$\phi(B) \Phi(B^{12}) (1 - B)Y_t = \theta(B) \Theta(B^{12}) e_t \quad (1)$$

Donde Y_t es la serie de la inflación colombiana, B es el operador de rezagos, $\phi(B)$ y $\theta(B)$ son polinomios en B , de grados p y q , respectivamente; $\Phi(B^{12})$ y $\Theta(B^{12})$ son polinomios en B^{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 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 se puede mencionar que sus pronósticos resultan auto-correlacionados y generalmente desmejorados para el mediano y largo plazo.

3.2. Modelos STAR

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

$$Y_t = \alpha_0 + \sum \alpha_i Y_{t-i} + \{\beta_0 + \sum \beta_j Y_{t-j}\} F(Y_{t-d}) + \varepsilon_t \quad (2)$$

$\varepsilon_t \sim N(0, \sigma^2)$, $F(Y_{t-d})$ es una función no lineal en 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:

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

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

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 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. La aplicación al caso de la inflación colombiana se puede encontrar en Jalil y Melo (2000).

Entre las desventajas de esta metodología 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 pronósticos consistentemente mejores que aquellos a mediano y largo plazo.

4. Método no paramétrico para series de tiempo

En esta sección se presentan algunos detalles del modelo no paramétrico basado en la función Kernel y en modelos de regresión generalizados, se define como se hace suavizamiento de una serie de tiempo, además algunos predictores y finalmente la construcción de los intervalos de confianza (de Gooijer & Zerom 1999, Matzner-L et al. 1998). En Rodríguez & Siado (2003) se encuentra cómo se da origen a este método no paramétrico mediante aproximación basada en un histograma.

4.1. Una aproximación mediante un histograma

Para dar la idea inicial de las técnicas de estimación de las funciones de densidades que se van a utilizar se comienza con la situación en la cual X es una variable aleatoria continua. La estimación de la función se hará promediando los x_i que se encuentren alrededor de un intervalo de longitud h , es decir, los valores que pertenezcan a $\left[x - \frac{h}{2}, x + \frac{h}{2} \right]$. De esta manera, la estimación empírica de la función de densidad se puede escribir como:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n I \left(\frac{-1}{2} \leq \frac{x - x_i}{h} \leq \frac{1}{2} \right) \quad (5)$$

donde la función $I(A)$ es 1 si la condición A se cumple y 0 si no. Lo anterior se puede ver como la construcción de histogramas basados en observaciones locales x , donde cada observación x está en el centro de un intervalo muestral. La amplitud h de los intervalos controlan la cantidad de observaciones por las que serán suavizadas, en este caso suavizadas por su promedio.

Como es una función de densidad se debe garantizar que la integral en todo su recorrido sea igual a uno, por lo que es necesario que la función $I(A)$ tenga esta propiedad, lo cual se cumple, pues se trata de una distribución de densidad uniforme en el intervalo de longitud h .

4.2. Estimador Kernel

Rosenblant (1956) y Parzen (1969), proponen utilizar la función Kernel $K(u)$ en lugar de la función de densidad uniforme $I(u)$, pues notan que esta tiene algunas propiedades que no la favorecen. La nueva función debe cumplir las siguientes propiedades.

$$\begin{aligned} K(u) &\geq 0 & K(u) &= K(-u) \text{ (Simetría)} \\ \lim_{|u| \rightarrow \infty} K(u) &= 0 & K(-\infty) &= K(\infty) = 0 \\ \int_{-\infty}^{\infty} K(u) &= 1 & \int_{-\infty}^{\infty} uK(u) &= 0 \text{ (Esperanza)} \end{aligned}$$

Para el caso continuo, la forma general del estimador Kernel es:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{-1}{2} \leq \frac{x - x_i}{h} \leq \frac{1}{2}\right) \quad (6)$$

Existe una gran gama de funciones kernel univariado que vale la pena resaltar como el kernel uniforme, triangular, Epanechnikov, tri-cubico, bicuadrado y la gaussiana; ésta última es la que se considera en este trabajo (Rodríguez & Siado 2003).

4.3. Estimador de densidad multivariado

Una herramienta útil a la hora de realizar los pronósticos es la función kernel multivariada. Sea $X = (X_1, X_2, \dots, X_d)$ que denota un vector d -dimensional con densidad $f(x)$ definida sobre \mathbb{R}^d y sea (x_1, x_2, \dots, x_n) una muestra aleatoria independiente tomada de $f(x)$. La forma general del estimador Kernel de $f(x)$ es:

$$\hat{f}_H(x) = \frac{1}{n} \sum_{i=1}^n K_H(x - x_i) \quad (7)$$

donde $K_H(x) = |H|^{-1/2} K(H^{-1/2}x)$, $K(\cdot)$ es una función Kernel multivariada y H es una matriz simétrica definida positiva de tamaño $d \times d$, conocida como matriz de anchos de banda. Cuando H es diagonal el estimador de densidad Kernel de $f(x)$ es:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_1 h_2 \dots h_d} K\left(\frac{x_1 - x_{1i}}{h_1}, \frac{x_2 - x_{2i}}{h_2}, \dots, \frac{x_d - x_{di}}{h_d}\right)$$

donde $h = (h_1, h_2, \dots, h_d)'$. Un método común para la construcción de funciones Kernel multivariada es el uso de Kernel producto que expresa de la siguiente forma:

$$K_d(x_1, x_2, \dots, x_d) = \prod_{j=1}^d K_{d_j}(x_j) \quad (8)$$

Es usual emplear la función Kernel gaussiana estándar multivariada, que se define como:

$$K_d(x_1, x_2, \dots, x_d) = (2\pi)^{-d/2} \exp\left(-\sum_{t=1}^d \frac{x_t^2}{2}\right) \quad (9)$$

Haciendo uso de la definición anterior, el estimador de la función de densidad Kernel es:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{t=1}^n \frac{1}{h_1 h_2 \dots h_d} \prod_{j=1}^d K_j\left(\frac{x_j - x_{jt}}{h_j}\right) \quad (10)$$

4.4. Regresión polinómica local para series de tiempo

Supongamos que se tiene un conjunto de datos de la siguiente forma $(Z_t, Z_{t-1}) = (Y_t, X_t)$ con $t = 2, 3, \dots, n$, donde Z_t es una serie de tiempo y Z_{t-1} es el regresor. El interés radica en ajustar un modelo de la siguiente manera:

$$\begin{aligned} Y_t &= E(Y_t | X_t = x) + u_t \\ Y_t &= E(Z_t | Z_{t-1} = z_0) + u_t \\ Y_t &= m(z_0) + u_t \end{aligned} \quad (11)$$

donde u es el termino del error y con media cero y varianza σ^2 . La siguiente estimación de $m(x)$ es el caso más simple de un estimador que se construye haciendo un promedio ponderado usando la función kernel y se conoce con el nombre de estimador *Nadayaara – Watson*.

$$\hat{m}(z_0) = \sum_{t=1}^n w_t(\psi_t) y_t \quad (12)$$

donde,

$$w_t(\psi_t) = \frac{K(\psi_t)}{\sum_{s=1}^n K(\psi_s)} \quad y \quad K(\psi_t) = K\left(\frac{Z_{t-1} - z_0}{h}\right) \quad (13)$$

Esta estimación se puede ver como la combinación lineal ponderada de observaciones locales similares a un punto dado z_0 . Se conoce a h como el ancho de banda que genera un intervalo que determina cuantos y con que ponderaciones de los Z_t alrededor de cada z_0 son usados para calcular el promedio. Cuando h es pequeño

proporciona la mejor estimación local pero esto genera problemas que se discuten en la siguiente sección.

El estimador *Nadayaara - Watson* ajusta una constante a los datos que se encuentran cercanos a un punto z_0 . El ajuste lineal local aproxima una línea recta de pendiente mayor que cero y dictada por el comportamiento de los datos. En general, es posible hacer un ajuste polinómico local de grado p , con el cual se puede obtener otro estimador para $m(z_0)$, esto es posible por medio de mínimos cuadrados ponderados, como se muestra a continuación en forma matricial.

Se considera el modelo $Y = X\beta + U$, donde U es un vector de errores con media 0 y matriz de varianza $\sigma^2 W^{-1}$ los vectores \mathbf{Y} y β son:

$$\mathbf{Y} = \begin{pmatrix} z_2 \\ z_3 \\ \cdot \\ \cdot \\ z_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_p \end{pmatrix}$$

además la matriz X se define como:

$$\mathbf{X} = \begin{pmatrix} 1 & (z_1 - z_0) & (z_1 - z_0)^2 & \cdot & \cdot & (z_1 - z_0)^p \\ 1 & (z_2 - z_0) & (z_2 - z_0)^2 & \cdot & \cdot & (z_2 - z_0)^p \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & (z_{n-1} - z_0) & (z_{n-1} - z_0)^2 & \cdot & \cdot & (z_{n-1} - z_0)^p \end{pmatrix}$$

y finalmente la matriz \mathbf{W} es diagonal y contiene las ponderaciones propuestas anteriormente.

$$\mathbf{W} = \begin{pmatrix} K(\psi_1) & 0 & 0 & \cdot & \cdot & 0 \\ 0 & K(\psi_3) & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & K(\psi_{n-1}) \end{pmatrix}$$

De acuerdo con la metodología de mínimos cuadrados ponderados, la estimación de los parámetros se obtiene mediante:

$$\beta = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y} \quad (14)$$

Por lo tanto, el estimador de $m(z_o)$ se consigue al realizar $\hat{m}(z_o) = \mathbf{e}'_1\beta = \beta_0^{(p)}$, donde $\mathbf{e}'_1 = (1, 0, 0, \dots, 0)$, entonces $\beta_0^{(p)}$ corresponde al nivel alrededor de z_0 que resulta de hacer RPL de grado p . Cabe notar que el estimador propuesto por *Nadayaara - Watson* es un caso particular de RPL cuando $p = 0$.

4.5. Ancho de banda

Escoger el ancho de banda h es importante porque es necesario controlar el balance entre dos factores que afectan el error cuadrático medio, los cuales son el sesgo y la varianza de la estimación. Una pequeña longitud de h implica que el valor esperado de la estimación $\hat{m}(x)$ sea cercano a $m(x)$, lo cual refleja una alta varianza para el estimador, ya que se involucran pocas observaciones en la comparación. Por otra parte, la variabilidad puede ser reducida incrementando h , aunque es inevitable que el sesgo sea más alto.

Hay una variedad de técnicas para escoger un valor apropiado de h , usualmente motivado por la necesidad de minimizar el error cuadrático medio. En este trabajo se utiliza el siguiente.

$$h = \sigma_y \left\{ \frac{4}{(d+2)n} \right\}^{1/d+4}, \quad (15)$$

donde σ_y es la varianza del proceso, d determina la dimensión de la función Kernel multivariada y n es la longitud de la serie.

4.6. Pronósticos múltiples pasos adelante

Con el fin de incorporar dos de los tres tipos de predictores no paramétricos propuestos por Matzner-L et al. (1998), denominados *media y mediana condicional*, se define un proceso Markoviano de orden d para una serie de tiempo estrictamente estacionaria $\{Z_t \mid t \geq 1\}$ como:

$$F_{Z_t \mid Z_{t-1}, \dots, Z_1}(z_t \mid z_{t-1}, \dots, z_1) = F_{Z_t \mid Z_{t-1}, \dots, Z_{t-d}}(z_t \mid z_{t-1}, \dots, z_{t-d}) \quad (16)$$

Lo anterior se puede entender como la probabilidad de que un valor Z_t sea explicado por toda la historia, que es igual a la probabilidad de que el mismo Z_t sea explicado por tan solo d observaciones anteriores. De acuerdo con lo anterior, si $\{Z_i\}_{i=1}^n$ son la observaciones de un proceso, la idea es predecir Z_{n+m} , donde m es conocido como el horizonte de pronóstico ($1 \leq m \leq n - d$) y d es el orden del mencionado proceso markoviano.

Para este propósito, siguiendo a de Gooijer & Zerom (1999), se construye el siguiente proceso $\{X_t, Y_t\}$, donde $X_t = \{Z_t, \dots, Z_{t+d-1}\}$ y $Y_t = Z_{t+m}$ para $t \in \{d, \dots, n - m\}$. Predecir consiste en encontrar la variable aleatoria más cercana, conociendo todo o parte de su pasado. Usando este principio, el predictor denominado *media condicional* para $X_t = x_n$ se define como:

$$\arg \min_{a \in \mathbb{R}} E\{(Y_n - a)^2 \mid x_n\} \quad (17)$$

Si se cumple que $\{Z_i\}_{i=1}^n$ es un proceso markoviano de orden d , el estimador \hat{Z}_{n+m}^{mean} del predictor se define como una combinación lineal de Y_t haciendo uso de

las ponderaciones que sugiere Nadayara - Watson, dependiendo de los valores X_t cercanos al punto x_n :

$$\hat{Z}_{n+m}^{mean} = \sum_{t=d}^{n-m} w_t(x_n) y_{t+m} \quad (18)$$

donde

$$w_t(x_n) = \frac{K_d\left(\frac{x_n - x_t}{h}\right)}{\sum_{t=d}^{n-m} K_d\left(\frac{x_n - x_t}{h}\right)} \quad x_n \in \mathbb{R}^d \quad (19)$$

Es posible generalizar el predictor *media condicional*, haciendo uso de regresión polinómica local de grado p , de manera similar a como se mostró en la sección 3.4, asumiendo el siguiente modelo.

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}$$

donde, \mathbf{Y} es el vector de tamaño $(n - d - m) \times 1$; \mathbf{X} es una matriz que tiene en la primera columna un vector de $\mathbf{1}_{n-d-m}$, n número de observaciones, d coeficiente de Markov y m horizonte de pronóstico,

$$\mathbf{Y} = \begin{pmatrix} z_{d+m} \\ z_{d+m+1} \\ \cdot \\ \cdot \\ z_n \end{pmatrix} \quad \mathbf{X} = \left(\mathbf{1}_{n-d-m} \mid \mathbf{A}_d^{(1)} \mid \mathbf{A}_d^{(2)} \mid \cdots \mid \mathbf{A}_d^{(p)} \right)$$

la matriz $\mathbf{A}_d^{(j)}$ es la matriz de rezagos que actúan como regresores, (j) indica que se está considerando RPL de grado $j = 1, 2, 3, \dots, p$ y d indica que se están considerando d rezagos como regresores, es decir que la observación Z_t depende de d observaciones anteriores.

$$\mathbf{A}_d^{(j)} = \begin{pmatrix} (z_d - z_n)^j & (z_{d-1} - z_{n-1})^j & \cdot & \cdot & (z_1 - z_{n-d+1})^j \\ (z_{d+1} - z_n)^j & (z_d - z_{n-1})^j & \cdot & \cdot & (z_2 - z_{n-d+1})^j \\ (z_{d+2} - z_n)^j & (z_{d+1} - z_{n-1})^j & \cdot & \cdot & (z_3 - z_{n-d+1})^j \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ (z_{n-m} - z_n)^j & (z_{n-m-1} - z_{n-1})^j & \cdot & \cdot & (z_{n-d-m+1} - z_{n-d+1})^j \end{pmatrix}$$

El vector $\boldsymbol{\beta}$ contiene $(pd + 1)$ parámetros que se quieren estimar de la siguiente forma:

$$\boldsymbol{\beta} = \left(\beta_0 \mid \boldsymbol{\beta}^{(1)'} \mid \boldsymbol{\beta}^{(2)'} \mid \cdots \mid \boldsymbol{\beta}^{(p)'} \right)'$$

donde

$$\boldsymbol{\beta}^{(p)} = \left(\beta_1^p, \beta_2^p, \dots, \beta_{d+m}^p \right)'$$

Finalmente, $U \sim N(0, \sigma^2 W^{-1})$, donde $W = \text{diag}(K(\psi_d), K(\psi_{d+1}), \dots, K(\psi_{n-m}))$. A través de mínimos cuadrados ponderados se obtiene la estimación de β .

$$\beta = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y} \quad (20)$$

En concordancia con la sección de regresión polinómica local para series de tiempo, la estimación de los pronósticos m pasos adelante Z_{n+m} , se obtiene haciendo $\hat{Z}_{n+m} = \mathbf{e}'_1 \hat{\beta} = \hat{\beta}_0$, donde $\mathbf{e}'_1 = (1, 0, 0, \dots, 0)$.

La *mediana condicional* debe satisfacer para $X_t = x_n$ que:

$$\arg \min_{a \in \mathbb{R}} E\{|Y_n - a| | x_n\} \quad (21)$$

La estimación de la mediana condicional, se construye con base en la estimación de la función de distribución condicional, es decir,

$$\hat{F}(y|X = x_n) = \sum_{t=d}^{n-m} \mathbf{1}_{\{Y_t \leq y\}} W_t(x_n) \quad (22)$$

donde $\mathbf{1}_A$ denota la función indicadora para el conjunto A . El predictor no paramétrico está dado por

$$\hat{Z}_{n+m}^{median} = \inf\{y | \hat{F}(y|X = x_n) \geq 1/2\} \quad (23)$$

De nuevo, es posible generalizar este predictor teniendo en cuenta información de variables rezagadas.

4.7. Selección del coeficiente de Markov

El funcionamiento de los métodos de predicción es altamente dependiente de d . Intuitivamente se toma d tan grande como sea posible para no perder mucha información del pasado. Para escoger el coeficiente, es basada en dos métodos empíricos

$$f_1(d) = \frac{1}{n} \sum_t |Z_t - Z_{n+m}^{mean}(d)| \quad f_2(d) = \frac{1}{n} \sum_t (Z_t - Z_{n+m}^{mean}(d))^2 \quad (24)$$

donde $t \in \{n-p, \dots, n\}$ y se puede notar que el predictor depende de d y además que el valor de p es la parte entera de $n/5$. Como se mencionó anteriormente, si el predictor proviene de RPL entonces la funciones se pueden reescribir, tomando $\beta_{0,n+m}(d)$ en lugar de $Z_{n+m}^{mean}(d)$

4.8. Intervalos de predicción

Simple puntos de predicción, sin especificar su precisión, son usualmente inadecuados en la práctica. Para evaluar futura incertidumbre es importante dar intervalos

de predicción (PI). En el contexto de modelos lineales para series de tiempo con distribución normal de los errores, los intervalos de predicción consisten en construir intervalos simétricos alrededor de los puntos de predicción usualmente con la media condicional. Sin embargo, estos intervalos pueden ser bastante engañosos cuando la distribución predictiva no es normal o el proceso generador de los datos no se puede asumir lineal. En la última situación, es necesario construir una conveniente distribución predictiva condicional.

La función de distribución predictiva condicional de Z_{n+m} dado $\{Z_t, t \leq n\}$ puede ser usada directamente para construir los intervalos de predicción para Z_{n+m} . Supongamos que $\Omega_\alpha \equiv \Omega_\alpha(x_n)$ es un (PI) con una probabilidad de cobertura de $\alpha \in (0, 1)$, es decir $P\{Y_n \in \Omega_\alpha | X = x_n\} = \alpha$. Este intervalo se puede construir en infinito número de formas. Una obvia manera de calcular un IP para Y_n dado $X = x_n$ es:

$$|\xi_{0.5-\alpha/2}(x_n), \xi_{0.5+\alpha/2}(x_n)| \quad \alpha \in [0, 1] \quad (25)$$

donde $\xi_\alpha(x_n)$ es el percentil α -ésimo de Y_n dado X_n , es decir, $P\{Y_n \leq \xi_\alpha(x) | X = x_n\} = \alpha$. El PI anterior será referido como intervalo percentílico condicional (CPI). La amplitud de los intervalos no es constante, como en el caso de modelos lineales y puede variar con respecto a la posición en el estado espacio.

Infortunadamente, los CPI no son siempre apropiados cuando la distribución predictiva es asimétrica o multimodal y el horizonte de predicción es largo. Para evitar este problema, se propone el uso de los llamados *intervalos modales condicionales más pequeños* (SCMI). Los SCMI, para Y_n dado $X = x_n$ son:

$$[a, b] = \arg \min\{Leb\{[c, d]\} \mid F(d|X = x_n) - F(c|X = x_n) \geq \alpha\} \quad (26)$$

donde $Leb(C)$ denota la medida de Lebesgue para el conjunto C . Esta definición fue originalmente propuesta por Lientz (1970,1972) para funciones de distribución no-condicional. En la práctica, un estimador para los SCMI es obtenido reemplazando $F(\cdot|\cdot)$ en (26) por la estimación de la función de distribución definida anteriormente.

4.9. Criterios de selección para mejores pronósticos y bondad de ajuste

Los indicadores de localización y dispersión que se utilizan para evaluar los pronósticos y bondad de ajuste de la inflación colombiana arrojados por cada modelo son:

| Estadísticos de localización y dispersión | | |
|---|------|--|
| Error Medio | EM | $\frac{1}{m} \sum_{t=n+1}^m (Z_t - \hat{Z}_t)$ |
| Error Porcentual Medio | EPM | $\frac{1}{m} \sum_{t=n+1}^m \frac{Z_t - \hat{Z}_t}{Z_t}$ |
| Error Cuadrático Medio | ECM | $\frac{1}{m} \sum_{t=n+1}^m (Z_t - \hat{Z}_t)^2$ |
| Error Absoluto Medio | EAM | $\frac{1}{m} \sum_{t=n+1}^m Z_t - \hat{Z}_t $ |
| Error Porcentual Absoluto Medio | EPAM | $\frac{1}{m} \sum_{t=n+1}^m \frac{ Z_t - \hat{Z}_t }{ Z_t }$ |

(27)

5. Aplicación

Se toma serie de tiempo de la inflación mensual total colombiana, desde enero del año 1974 hasta octubre del año 2006 (figura 1), la cual se obtiene a partir del IPC¹ mensual.

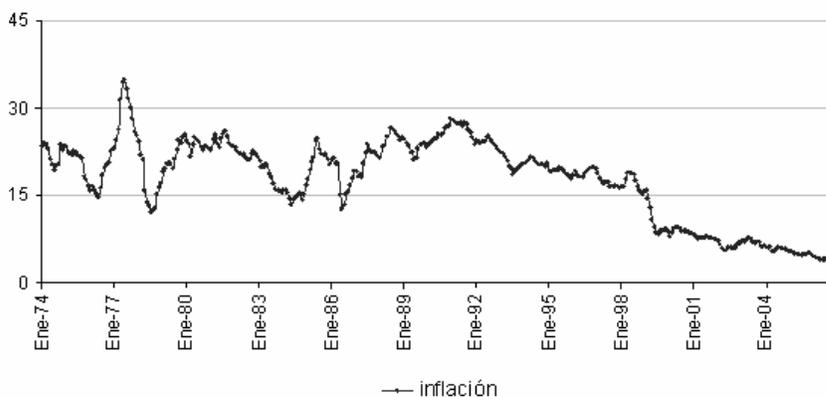


Figura 1: *Inflación total Colombiana.*

Debido a la tendencia que se observa en la serie y a que la media del proceso no es constante, se considera la diferencia de ésta para realizar los siguientes tres ejercicios.

¹Fuente: www.dane.gov.co

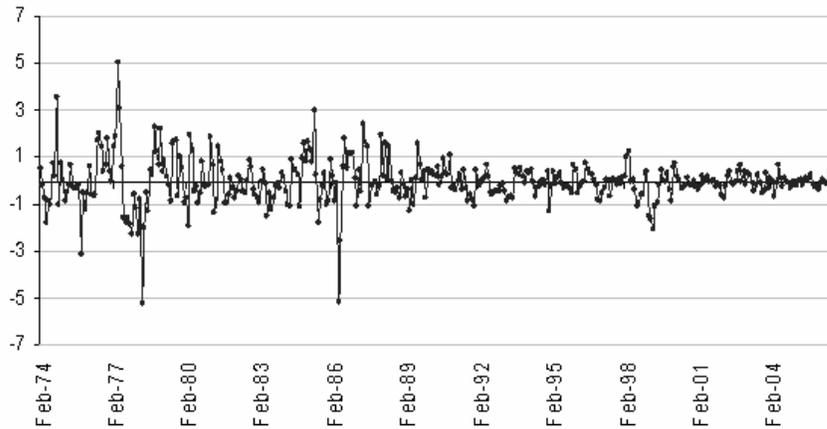


Figura 2: *Diferencia de la serie de la inflación.*

El primero es la comparación del suavizamiento de la serie con un modelo ARIMA y con la técnica no paramétrica; el segundo es realizar, para dos periodos, predicciones con el modelo lineal ARIMA, con el no lineal STAR, con redes neuronales y finalmente con la técnica no paramétrica haciendo uso de los predictores media y mediana condicional, finalmente se hace comparación de la longitud de los intervalos de pronóstico generados con el modelo ARIMA y con la técnica no paramétrica.

6. Resultados

6.1. Suavizamiento de la serie de la inflación total colombiana

Comparando el suavizamiento generado por el modelo ARIMA ² y con la técnica no paramétrica, por medio de los estadísticos de bondad de ajuste, que se observan en Tabla 1, el suavizamiento generado por el segundo método presenta menos sesgo pero más variabilidad respecto al primero que se menciona.

Se puede observar también que a medida que aumenta el grado de la regresión polinómica local se experimenta un mejor ajuste.

²ARIMA(0, 1, 2)(0, 1, 2)₁₂: Una diferencia estacional y otra de orden 1, MA(2) estacional y otra no estacional para la inflación total colombiana.

| Tabla 1 | | | | | |
|---------|----------|----------|---------|---------|---------|
| | EM | EPM | ECM | EAM | EAPM |
| ARIMA | -0,07510 | -0,00906 | 0,46846 | 0,48427 | 0,02973 |
| RPL(0) | 0,00468 | 0,00020 | 0,71784 | 0,54487 | 0,03188 |
| RPL(1) | -0,00567 | -0,00040 | 0,71044 | 0,54254 | 0,03186 |
| RPL(2) | 0,00186 | 0,00001 | 0,70407 | 0,54009 | 0,03178 |

En la Figura 3 se observa el suavizamiento que se logra con el método no paramétrico, un comportamiento similar se encuentra con el mencionado modelo ARIMA.

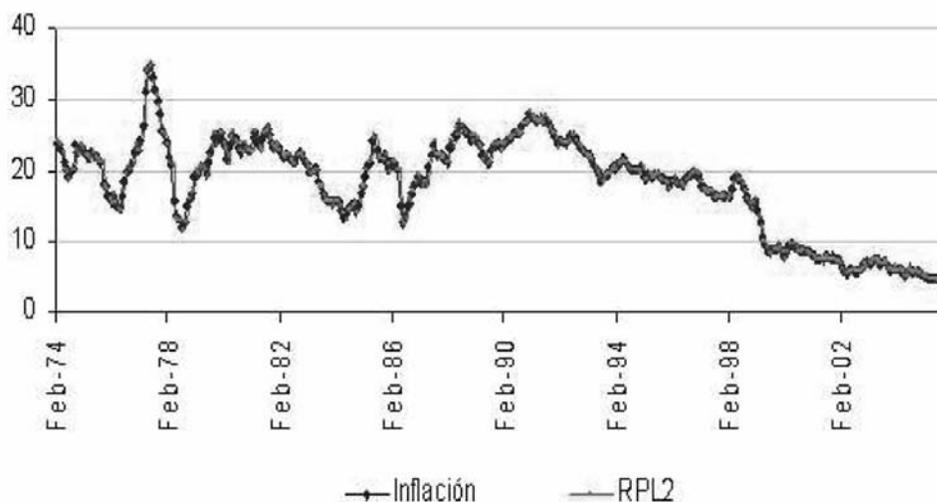


Figura 3: Suavizamiento de la serie de la inflación.

6.2. Pronósticos

Se seleccionan dos periodos de tiempo para hacer la evaluación de los pronósticos generados mediante cuatro técnicas: ARIMA, STAR, redes neuronales y el método no paramétrico. Para el primer periodo se selecciona el coeficiente de Markov $d = 13$ y para el segundo periodo $d = 15$, de acuerdo con la funciones definidas en la sección 3.7.

El primer periodo de evaluación que se considera es: enero de 2004 hasta diciembre de 2004. El segundo periodo que se considera es: noviembre de 2005 hasta octubre de 2006. Se presenta, en esta sección los pronósticos generados mediante la mediana condicional, pues este predictor presenta mejores características que la media condicional.

Los pronósticos para el primer periodo se muestran en la Tabla 2:

| Inflación | No paramétrico | STAR | Red Neuronal | ARIMA |
|-----------|----------------|------|--------------|-------|
| 6,00 | 6,68 | 5,96 | 6,20 | 6,46 |
| 6,09 | 6,72 | 5,96 | 6,34 | 7,36 |
| 6,03 | 6,53 | 5,60 | 6,23 | 7,14 |
| 5,34 | 6,44 | 5,22 | 5,80 | 6,76 |
| 5,23 | 6,32 | 5,07 | 5,78 | 6,47 |
| 5,89 | 6,16 | 4,78 | 6,07 | 6,39 |
| 6,00 | 5,84 | 4,75 | 6,60 | 6,37 |
| 5,73 | 5,35 | 4,44 | 6,73 | 6,15 |
| 5,80 | 4,85 | 4,40 | 6,74 | 5,94 |
| 5,73 | 4,31 | 4,94 | 7,00 | 5,73 |
| 5,66 | 4,12 | 5,19 | 7,08 | 5,31 |
| 5,35 | 3,93 | 4,86 | 6,83 | 4,83 |

En la Figura 4 se observan la inflación y los pronósticos para el periodo en consideración, haciendo uso de los cuatro modelos.

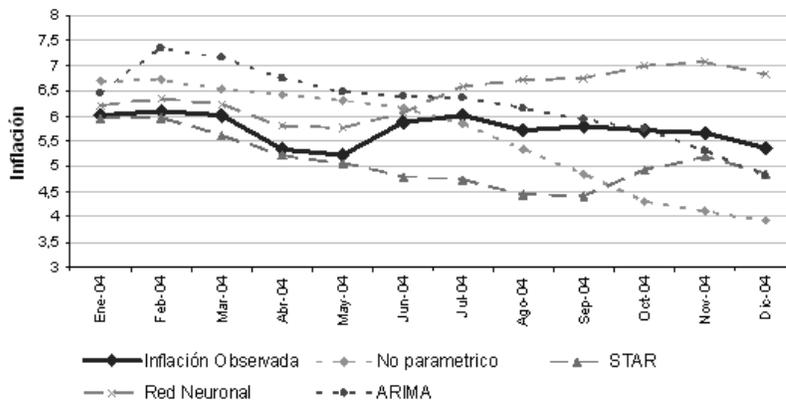


Figura 4: *Primer periodo de evaluación.*

En la Tabla 3 se encuentran los estadísticos de bondad de pronóstico para los cuatro métodos, de manera similar que con el suavizamiento se observan las predicciones de la técnica no paramétrica que resultan menos sesgadas que los otros modelos

| | EM | EPM | EAM | ECM | EPAM |
|-------|--------|--------|-------|-------|-------|
| ARIMA | -0,505 | -0,088 | 0,650 | 0,631 | 0,115 |
| STAR | 0,641 | 0,111 | 0,641 | 0,646 | 0,111 |
| REDES | -0,711 | -0,126 | 0,711 | 0,728 | 0,126 |
| NP | 0,133 | 0,024 | 0,844 | 0,918 | 0,150 |

A continuación se observan los pronósticos generados por los cuatro métodos mencionados para el segundo periodo de evaluación.

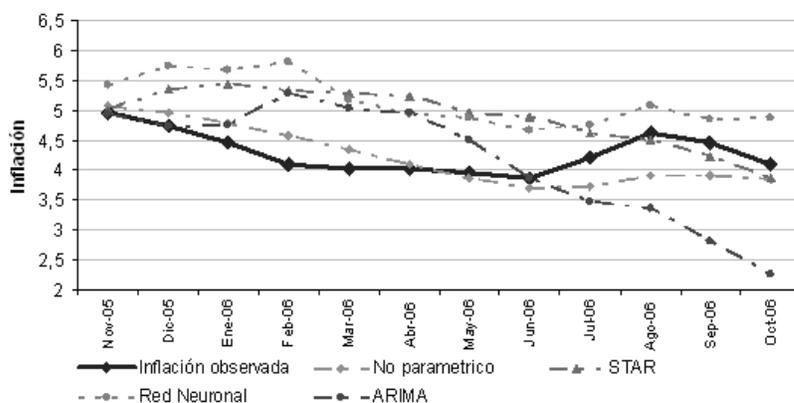


Figura 5: Segundo periodo de evaluación.

| Inflación | No paramétrico | STAR | Red Neuronal | ARIMA |
|-----------|----------------|------|--------------|-------|
| 4,97 | 5,09 | 5,02 | 5,42 | 4,95 |
| 4,74 | 4,98 | 5,36 | 5,76 | 4,75 |
| 4,46 | 4,78 | 5,45 | 5,69 | 4,77 |
| 4,10 | 4,59 | 5,33 | 5,82 | 5,30 |
| 4,03 | 4,35 | 5,28 | 5,17 | 5,05 |
| 4,04 | 4,11 | 5,25 | 4,94 | 4,96 |
| 3,96 | 3,88 | 4,97 | 4,89 | 4,50 |
| 3,86 | 3,72 | 4,90 | 4,68 | 3,88 |
| 4,23 | 3,73 | 4,63 | 4,76 | 3,49 |
| 4,62 | 3,92 | 4,52 | 5,07 | 3,36 |
| 4,48 | 3,92 | 4,23 | 4,86 | 2,82 |
| 4,10 | 3,85 | 3,88 | 4,88 | 2,26 |

En los estadísticos de bondad de pronóstico se observa que el método no paramétrico presenta mayor precisión para los pronósticos de la inflación de este periodo, por otra parte los generados por el modelo ARIMA resultan ser menos sesgados.

| | EM | EPM | EAM | ECM | EPAM |
|-------|--------|--------|-------|-------|-------|
| ARIMA | 0,124 | 0,023 | 0,793 | 0,995 | 0,187 |
| STAR | -0,603 | -0,146 | 0,696 | 0,690 | 0,168 |
| REDES | -0,862 | -0,204 | 0,862 | 0,882 | 0,204 |
| NP | -0,335 | -0,082 | 0,424 | 0,237 | 0,102 |

6.3. Intervalos de predicción

Se generan los intervalos de predicción para el último periodo en consideración, teniendo en cuenta el modelo ARIMA y la técnica no paramétrica. En el siguiente gráfico se observan conjuntamente los intervalos de predicción y los pronósticos generados por los dos métodos.

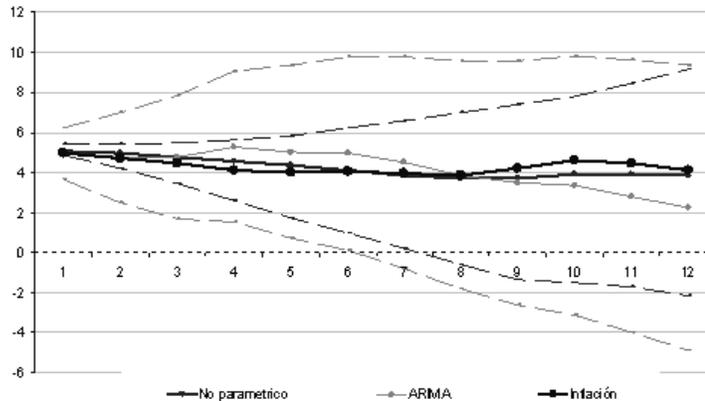


Figura 6: *Intervalos de predicción.*

Se observa que en todos los casos que los intervalos de predicción de la técnica no paramétrica, tienen la característica de ser de menor longitud con respecto a los intervalos del modelo ARIMA.

7. Conclusiones

Los suavizamientos generados por los dos métodos no presentan grandes diferencia de acuerdo con los estadísticos de bondad de ajuste. De manera similar, los pronósticos generados para los dos periodos de evaluación son parecidos, pero se destacan siempre los generados por la técnica no paramétrica.

La elección del coeficiente de Markov d y el ancho de banda h es importante a la hora de realizar pronósticos. El segundo parámetro lo es, porque se encontró que si se tomaba un ancho de banda más pequeño la longitud de los intervalos disminuía, pero en consecuencia los pronósticos no eran buenos, por el contrario, si el ancho de banda era más grande, la longitud de los intervalos era más grande pero con mejores pronósticos.

En cuanto a los dos predictores media y mediana condicional, se encuentran mejores resultados con la mediana condicional, esto en términos de los estadísticos de bondad de pronóstico. Para los horizontes de predicción $m = 1, 2, \dots, 12$, se

encontró que la longitud de los intervalos de predicción generados por el modelo ARIMA es mayor que los intervalos de predicción de la técnica no paramétrica.

Con lo anterior, se puede decir que la técnica no paramétrica es una buena herramienta para pronosticar y generar intervalos de predicción que debe ser explorada.

Recibido: 25 de febrero de 2010

Aceptado: 16 de abril de 2010

Referencias

- de Gooijer, J. G. & Zerom, D. (1999), Kernel-based multistep-ahead predictions of the us short-term interest rate, Tinbergen Institute Discussion Papers 99-015/4, Tinbergen Institute.
- Draper, N. & Smith, H. (1998), *Applied Regression Analysis*, Wiley.
- Jalil, M. & Melo, V. (2000), 'Una relación no lineal entre la inflación y los medios de pago', *Borradores de Economía* (145).
- Matzner-L, E., Gannoun, A. & DeGooijer, J. (1998), 'Nonparametric forecasting: a comparison of three kernel-based methods', *Communications in Statistics - Theory and Methods* **27**(7), 1593 – 1617.
- Pagan, A. & Ullah, A. (1999), *Nonparametric Econometrics*, Oxford Univ. Press.
- Ramsay, J. & Silverman, B. (2005), *Funcional Data Analysis*, Springer.
- Rodríguez, N. & Siado, P. (2003), 'Un pronóstico no paramétrico de la inflación colombiana', *Revista Colombiana de Estadística* **26**(2), 89 – 128.