



Revista de Matemática: Teoría y Aplicaciones
ISSN: 1409-2433
mta.cimpa@ucr.ac.cr
Universidad de Costa Rica
Costa Rica

de-los-Cobos-Silva, Sergio G.; Goddard-Close, John; Gutiérrez-Andrade, Miguel A.
Regresión borrosa vs. regresión por mínimos cuadrados ordinarios: caso de estudio
Revista de Matemática: Teoría y Aplicaciones, vol. 18, núm. 1, 2011, pp. 33-48
Universidad de Costa Rica
San José, Costa Rica

Disponible en: <http://www.redalyc.org/articulo.oa?id=45326927004>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

REGRESIÓN BORROSA VS. REGRESIÓN POR MÍNIMOS CUADRADOS ORDINARIOS: CASO DE ESTUDIO

FUZZY REGRESSION VS. ORDINARY LEAST SQUARES REGRESSION: CASE STUDY

SERGIO G. DE-LOS-COBOS-SILVA* JOHN GODDARD-CLOSE†
MIGUEL A. GUTIÉRREZ-ANDRADE‡

Received: 18 Feb 2010; Revised: 3 Sep 2010; Accepted: 24 Sep 2010

Resumen

El objetivo del trabajo es presentar la técnica de regresión borrosa y mostrar su aplicación con un ejemplo práctico, para tal propósito, se comparará la ejecución de las técnicas tanto de regresión usual así como la de algunos modelos de regresión borrosa para el estudio del ícono de confianza del consumidor usada como variable respuesta respecto de la cotización del dólar considerada como variable independiente. Se proporciona una pequeña introducción a las diferentes metodologías utilizadas. Se reportan los resultados obtenidos de los algoritmos de

*Departamento de Ingeniería Eléctrica, Universidad Autonómica Metropolitana Iztapalapa, Av. Michoacán y La Purísima s/n, Col. Vicentina, Del. Iztapalapa, México D.F., C.P. 09340 México. E-Mail: cobos@xanum.uam.mx

†Misma dirección que/*same address as* de los Cobos. E-Mail: jgc@xanum.uam.mx

‡Misma dirección que/*same address as* de los Cobos. E-Mail: gamma@xanum.uam.mx

regresión: el usual por mínimos cuadrados ordinarios y 2 de regresión borrosa. Para todos los casos, se reportan las instancias generadas con los datos históricos oficiales y se realiza la comparación entre estos. Finalmente se reporta los resultados numéricos obtenidos por los diferentes métodos.

Palabras clave: regresión lineal, regresión borrosa, programación lineal borrosa.

Mathematics Subject Classification: 65J05, 90C05, 62J86.

Abstract

The objective of this paper is to disseminate the technique of fuzzy regression and to give a practical example of its use. To this end, classical regression is compared to several fuzzy regression models on a problem concerning the consumer confidence index with respect to the dollar rate, the latter taken as the independent variable. A brief introduction is given to each of the different methodologies employed. The results obtained using the regression algorithms, one with ordinary least squares and another two with fuzzy regression, are presented. The instances generated using the official historical data for the problem are given and the numerical results obtained with the regression methods are reported.

Keywords: linear regression, fuzzy regression, fuzzy linear programming.

Mathematics Subject Classification: 65J05, 90C05.

1 Introducción

La crisis económica mundial que se manifestó a partir del último trimestre de 2008 ha provocado gran inestabilidad e incertidumbre, la que se manifiesta en muchos indicadores entre los que se encuentra el índice de confianza del consumidor (ICC) de México. En este trabajo, se presenta un estudio mediante técnicas de regresión usual por mínimos cuadrados y de regresión borrosa considerando el ICC como variable de respuesta respecto de la cotización del dólar, considerada como variable independiente, durante el periodo de enero de 2006 a marzo de 2009.

El ICC, proviene de la Encuesta Nacional sobre la Confianza del Consumidor (ENCO) elaborada conjuntamente por el Instituto Nacional de Estadística y Geografía (INEGI) y Banco de México (BANXICO), y refleja la percepción que los hogares urbanos de México tienen acerca del estado actual y la situación esperada de la economía de los propios hogares, así como del país en general.

La información del ICC se expresa en forma de índice con base en enero de 2003 al que se le asigna el valor de 100, el ICC resulta de promediar cinco índices parciales, de los cuales dos hacen referencia a la situación económica actual y esperada del hogar entrevistado, otros dos atienden a la situación económica actual y esperada del país y el quinto índice refleja qué tan propicio es el momento actual para la compra de bienes de consumo durable, las preguntas para calcular los 5 índices mencionados anteriormente son:

1. Comparada con la situación económica que los miembros de este hogar tenían hace 12 meses, ¿cómo cree que es su situación en este momento?
2. ¿Cómo considera usted que será la situación económica de los miembros de este hogar dentro de 12 meses, respecto a la actual?
3. ¿Cómo considera usted la situación económica del país hoy en día comparada con la de hace 12 meses?
4. ¿Cómo considera usted que será la condición económica del país dentro de 12 meses respecto a la actual situación?
5. Comparando la situación económica actual con la de hace un año, ¿cómo considera en el momento actual las posibilidades de que usted o alguno de los integrantes de este hogar realice compras tales como muebles, televisor, lavadora, otros aparatos electrodomésticos, etc.?

Ahora bien, la política cambiaria en México, es responsabilidad de la Comisión de Cambios, integrada por funcionarios de la Secretaría de Hacienda y Crédito Público y del BANXICO. A finales de 1994, dicha Comisión acordó que el tipo de cambio fuese determinado libremente por las fuerzas del mercado.

La cotización del dólar (CD) para realizar el presente estudio, se obtuvo mediante el promedio mensual de los datos por día proporcionados por BANXICO del tipo de cambio pesos por dólar E.U.A. para solventar obligaciones con fecha de determinación (FIX). El tipo de cambio (FIX) es determinado por el BANXICO con base en un promedio de las cotizaciones del mercado de cambios al mayoreo para operaciones liquidables el segundo día hábil bancario siguiente. Se publica en el Diario Oficial de la Federación un día hábil bancario después de la fecha de determinación y es utilizado para solventar obligaciones denominadas en dólares liquidables en la República Mexicana al día siguiente. Para mayor información sobre este tipo de cambio se puede consultar las disposiciones publicadas en el

Diario Oficial de la Federación del 22 de marzo de 1996 y en el numeral M86 de la circular 2019/95 del BANXICO.

En el presente trabajo se expone la utilización de la regresión lineal por mínimos cuadrados así como un instrumento de regresión alternativo, la regresión borrosa. La regresión borrosa, como se verá más adelante, es más versátil que los convencionales modelos de regresión estadísticos, ya que permite hallar relaciones funcionales cuando la variable dependiente, las variables independientes o ambas no se manifiestan como valores únicos, sino como *intervalos de confianza en posibilidad y no en probabilidad*. Estos instrumentos de tipo borroso se diferencian de los instrumentos econométricos usuales, donde se deben reducir las observaciones a un valor representativo, con la posible pérdida de información que ello implica.

2 Regresión de tipo borroso

Como se indica en [3]: “El instrumento más comúnmente utilizado en economía es la regresión lineal econométrica, que supone la extrapolación al futuro de datos pasados, es decir, presupone una estacionariedad en la evolución de los cursos de acción. Un aspecto importante en la metodología econométrica usual, es que por lo general se desean conocer intervalos de confianza con cierta probabilidad de ocurrencia, más que un valor exacto, de esta manera se hablan de intervalos de confianza en probabilidad”.

Los modelos de regresión borrosa, tienen como objetivo, al igual que toda técnica de regresión el determinar una relación de tipo funcional entre una variable dependiente (llamada también de respuesta) Y respecto de una o varias variables independientes o explicativas, digamos $X = (X_0, X_1, \dots, X_n)$, $X_0 = 1$. En nuestro caso Y será el ICC y X será la CD.

En términos generales, la técnica usual de regresión por mínimos cuadrados ordinarios (MCO), consiste en tratar de ajustar un hiperplano que sea compatible con todos los elementos de la muestra, se supone que la relación entre Y_j y el vector X_j es lineal $\forall j$ (es el modelo más sencillo)¹, pero existe un elemento de perturbación estocástico. Denotemos como ε_j el término de perturbación estocástico y usualmente se acostumbra considerar que sigue una distribución normal. Habitualmente se supone que $E[\varepsilon_j] = 0, \forall j$, la varianza es igual para los errores, σ , y que éstos están no correlacionados, es decir, $E[\varepsilon_k \varepsilon_j] = 0, \forall k \neq j$, el método de estimación de los parámetros es el conocido MCO. En este caso, para hallar la estimación de los parámetros de a_i , y denotados como \hat{a}_i , $i = 0, 1, 2, \dots, m$, se resuelve un programa

¹Rigurosamente hablando, cuando se considera un modelo de regresión lineal, éste es lineal, respecto de los parámetros.

cuadrático cuya función objetivo es: $z = \sum_{j=1}^n (Y_j - \hat{Y}_j)^2$, donde \hat{Y}_j es el valor estimado de $E[Y_j]$ como función de las variables $X_{0,j}, X_{2,j}, \dots, X_{m,j}$, tras haber hallado los valores estimados de a_0, a_1, \dots, a_m , es decir:

$$\hat{Y}_j = \hat{a}_{0,j} + \hat{a}_{1,j}X_{1,j} + \hat{a}_{2,j}X_{2,j} + \dots + \hat{a}_{m,j}X_{m,j}.$$

Finalmente, los parámetros \hat{a}_i expresados como el vector columna, \hat{a} , se pueden encontrar mediante:

$$\hat{a} = (X^T X)^{-1} X^T Y \quad (1)$$

donde X es la matriz que recoge las observaciones de las variables independientes como vectores fila e Y es el vector columna que recoge las observaciones sobre las variables dependientes.

Así, para la j -ésima observación se cumple: $Y_j = a_0X_{0,j} + a_1X_{1,j} + \dots + a_mX_{m,j} + \varepsilon_j, j = 1, 2, \dots, n$, bajo un punto de vista estadístico para el modelo lineal y bajo los supuestos de normalidad, independencia y homocedasticidad, los estimadores $\hat{a}_i, i = 0, 1, 2, \dots, m$, bajo MCO son MELI, i.e., son estimadores lineales insesgados con varianza mínima, pero se debe primeramente verificar los supuestos, lo que en muchas ocasiones (sino es que en la mayoría) es difícil de verificar o se pasa por alto, ya sea porque se cuenta con una pequeña cantidad de observaciones $(Y_j, X_j), j = 1, 2, \dots, n$, o éstas ni siquiera son soportables por el tipo de problema y/o escalas de medición utilizada, esto sin considerar que el modelamiento de tipo lineal (o linealizable) sea el adecuado.

Por lo anterior, si tan sólo contamos con una muestra pequeña, la cual en sí ya puede contener cierto ruido con función de probabilidad desconocida (además es casi imposible el considerar todas las variables que son importantes o indicativas para cualquier modelo), es interesante suponer un modelamiento de tipo borroso, es decir, $Y_j = \hat{a}_{0,j}X_{0,j} + \hat{a}_{1,j}X_{1,j} + \dots + \hat{a}_{m,j}X_{m,j}, j = 1, 2, \dots, n$, donde $\hat{a}_i, i = 0, 1, 2, \dots, m$ son números borrosos. Por supuesto, de manera análoga a la de MCO, una vez que se disponga de la muestra, el objetivo será el de ajustar los coeficientes $\hat{a}_i, i = 0, 1, 2, \dots, m$ bajo cierto criterio de mejor ajuste. Observe que en este caso el término de error no queda introducido como sumando en el hiperplano, sino que se incorpora en los coeficientes al ser estos números borrosos.

Como se indica en [1]: “Respecto a esta forma de modelización, creemos que de alguna forma, ofrece ciertas ventajas sobre la tradicional técnica de regresión, En primer lugar, porque las estimaciones que obtengamos después de ajustar los coeficientes borrosos, no serán variables aleatorias, y por tanto, en muchas ocasiones de difícil tratamiento numérico, sino números borrosos, cuyo tratamiento es más sencillo. Por otra parte, si el fenómeno

de estudio es de carácter económico o social, las observaciones que del mismo se obtienen son consecuencia de la interacción entre las creencias, expectativas, etc., de los agentes que participan en dicho fenómeno, y por tanto, ya hemos señalado que en nuestra opinión, no es del todo adecuado modelizar dicho fenómeno utilizando la teoría de la probabilidad . . . , pero creemos que es más realista modelizar el sesgo que puede darse entre las realizaciones de la variable dependiente y el valor que teóricamente éstas pueden tomar asumiendo que la relación entre variable dependiente y variables explicativas es borrosa, que si damos una naturaleza aleatoria a dicho sesgo”.

Con el fin de facilitar la lectura del presente trabajo, sólo presentamos algunos conceptos, basados principalmente en [1] y [3], otros trabajos que pueden consultarse son entre otros [7] y [10].

2.1 Determinación de los coeficientes de la regresión

Para el fenómeno de estudio suponemos que, el observador dispone de una muestra que representamos como: $(Y_1, X_1), \dots, (Y_n, X_n)$, donde:

1. Y_j es un intervalo real (Y_j^1, Y_j^2) correspondiente a la observación j -ésima de la variable dependiente, $j = 1, 2, \dots, n$, y suponemos que viene dada por un intervalo borroso que está determinado a través de su centro y de su radio como:

$$Y_j = \langle Y_{jC}, Y_{jR} \rangle$$

donde:

$$Y_{jC} = \frac{Y_j^1 + Y_j^2}{2} \quad Y_{jR} = \frac{Y_j^2 - Y_j^1}{2}.$$

Si Y_j es un número cierto o exacto, $Y_{jR} = 0$.

2. Por otra parte, X_j es el vector observado en la j -ésima observación sobre las variables independientes, con $j = 1, 2, \dots, n$. Así, X_j es una variable $m+1$ -dimensional $X_j = (X_{0,j}, X_{1,j}, X_{2,j}, \dots, X_{m,j})$, donde $X_{0,j} = 1 \forall j$, y $X_{i,j}$ es el valor en la j -ésima observación de la muestra para la variable i -ésima. En cualquier caso asumimos que se tratan de observaciones ciertas, es decir de tipo puntual.

Asumimos que la relación existente entre la variable dependiente, Y , que puede venir dada por un intervalo de confianza, y la variable independiente $X = (X_0, X_1, X_2, \dots, X_m)$, que es un vector con componentes ciertos es lineal, de forma que:

$$Y = A_0 + A_1 x_1 + A_2 x_2 + \dots + A_m x_m. \quad (2)$$

donde $A_i, i = 0, 1, \dots, m$ son intervalos de confianza posibilísticos de la forma:

$$A_i = \langle a_{iC}, a_{iR} \rangle, \quad i = 0, 1, \dots, m.$$

De esta forma, si denominamos como $\hat{Y}_j = \langle \hat{Y}_{jC}, \hat{Y}_{jR} \rangle$ al intervalo de confianza correspondiente al valor que estimamos para la j -ésima variable independiente, Y_j , después de hallar los parámetros $A_0, A_1, A_2, \dots, A_m$, como dicha estimación se realizaría mediante la suma: $\hat{Y}_j = A_0 + A_1 X_{1j} + A_2 X_{2j} + \dots + A_m X_{mj}$, $j = 1, 2, \dots, n$, podemos expresarla a través de sus centros y sus radios, que serán funciones de los centros y de los radios de los parámetros A_i , $i = 0, 1, 2, \dots, m$:

$$\begin{aligned} \langle \hat{Y}_{jC}, \hat{Y}_{jR} \rangle &= \sum_{i=0}^m \langle a_{iC}, a_{iR} \rangle X_{ij} = \\ &= \sum_{i=0}^m a_{iC} X_{ij}, \quad \sum_{i=0}^m a_{iR} |X_{ij}|, \quad j = 1, 2, \dots, n. \end{aligned} \quad (3)$$

Para acabar determinando los parámetros $A_0 = \langle a_{0C}, a_{0R} \rangle$, $A_1 = \langle a_{1C}, a_{1R} \rangle, \dots, A_m = \langle a_{mC}, a_{mR} \rangle$, debemos establecer cómo abordar ciertas cuestiones que se tratarán en las siguientes subsecciones.

2.2 Determinación de la bondad del ajuste

La bondad del ajuste es inversa a la incertidumbre (amplitud) de las estimaciones de las observaciones Y_j, \hat{Y}_j . Así, la amplitud de \hat{Y}_j es el radio de dicho intervalo de confianza \hat{Y}_{jR} , el cual se obtiene como:

$$\hat{Y}_{jR} = \sum_{i=0}^m a_{iR} |X_{ij}| = a_{0R} + a_{1R} |X_{1j}| + \dots + a_{mR} |X_{mj}|. \quad (4)$$

Por lo que, la “incertidumbre total” (amplitud total) de todas las estimaciones de la muestra será la suma de los radios de las estimaciones.

$$z = \sum_{j=1}^n \hat{Y}_{jR} = \sum_{j=1}^n \sum_{i=0}^m a_{iR} |X_{ij}| \quad (5)$$

El objetivo final será que, la incertidumbre total de las estimaciones que se obtienen tras ajustar los centros y los radios de los coeficientes A_0, A_1, \dots, A_m sea lo menor posible.

2.3 Medición de la congruencia de las estimaciones

Los parámetros A_0, A_1, \dots, A_m deben procurar, no sólo que la incertidumbre de \hat{Y}_j sea lo menor posible, sino que \hat{Y}_j sea lo más congruente posible con la observación de la variable explicada que pretenden aproximar, Y_j . En este contexto, cabe definir dos aproximaciones a “ \hat{Y}_j congruente con Y_j ”. La primera aproximación [14] consistiría en exigir que la observación esté incluida dentro de su estimación. Otra aproximación [11] consistiría en exigir que la observación sea “igual” a su estimación, es decir, $Y_j = \hat{Y}_j$.

La última cuestión que debe tenerse en cuenta, es *cómo determinar el centro de los parámetros* A_0, A_1, \dots, A_m , es decir $a_{0C}, a_{1C}, \dots, a_{mC}$. En este caso, se puede optar al menos con los siguientes enfoques:

Tanaka e Ishibuchi, determinan los centros $a_{0C}, a_{1C}, \dots, a_{mC}$ a la vez que determinamos los radios $a_{0R}, a_{1R}, \dots, a_{mR}$. En cambio Savic y Pedrycz, determinan el vector de centros $a_C^T = (a_{0C}, a_{1C}, \dots, a_{mC})$ mediante mínimos cuadrados ordinarios (MCO) y tomando como observaciones de las variable dependiente Y_j , su centro, Y_{jC} , si buscamos minimizar la función z dada en la ecuación 5 de forma que las estimaciones \hat{Y}_j contengan las variables independientes observadas, Y_j , es decir: $Y_j \subseteq \hat{Y}_j, \forall j$, para llegar a determinar los parámetros A_i deberemos resolver el siguiente programa lineal:

$$\text{Minimizar } z = \sum_{j=1}^n \hat{Y}_{jR} = \sum_{j=1}^n \sum_{i=0}^m a_{iR} |X_{ij}|, \quad (6)$$

sujeto a las restricciones 7 a 9 dadas por:

$$\hat{Y}_{jC} - \hat{Y}_{jR} = \sum_{i=0}^m a_{iC} X_{ij} - \sum_{i=0}^m a_{iR} |X_{ij}| \leq Y_{jC} - Y_{jR}, \quad (7)$$

$$j = 1, 2, \dots, n,$$

$$\hat{Y}_{jC} + \hat{Y}_{jR} = \sum_{i=0}^m a_{iC} X_{ij} + \sum_{i=0}^m a_{iR} |X_{ij}| \geq Y_{jC} + Y_{jR} \quad (8)$$

$$j = 1, 2, \dots, n,$$

$$a_{iR} \geq 0 \quad i = 0, 1, \dots, m. \quad (9)$$

En este programa, si decidimos determinar los centros y los radios de los parámetros A_0, A_1, \dots, A_m , de forma simultánea, las variables decisión serán tanto $a_{0C}, a_{1C}, \dots, a_{mC}$ como $a_{0R}, a_{1R}, \dots, a_{mR}$, mientras que si los centros se han prefijado previamente mediante MCO, las variables decisión serían únicamente los radios, es decir, $a_{0R}, a_{1R}, \dots, a_{mR}$.

Si únicamente exigimos que las estimaciones que se obtienen de las variables independientes \hat{Y}_j , sean iguales a las realmente observadas, $Y_j, j = 1, 2, \dots, n$, es decir, $\hat{Y}_j = Y_j$, deberemos plantear y resolver:

$$\text{Minimizar } z = \sum_{j=1}^n \hat{Y}_{jR} = \sum_{j=1}^n \sum_{i=0}^m a_{iR} |X_{ij}| \quad (10)$$

sujeto a las restricciones 11 a 13 dadas por:

$$\hat{Y}_{jC} - \hat{Y}_{jR} = \sum_{i=0}^m a_{iC} X_{ij} - \sum_{i=0}^m a_{iR} |X_{ij}| \leq Y_{jC} + Y_{jR} \quad (11)$$

$$j = 1, 2, \dots, n,$$

$$\hat{Y}_{jC} + \hat{Y}_{jR} = \sum_{i=0}^m a_{iC} X_{ij} + \sum_{i=0}^m a_{iR} |X_{ij}| \geq Y_{jC} - Y_{jR} \quad (12)$$

$$j = 1, 2, \dots, n,$$

$$a_{iR} \geq 0 \quad i = 0, 1, \dots, m. \quad (13)$$

Nuevamente, si decidimos determinar los centros y los radios de los parámetros A_0, A_1, \dots, A_m , simultáneamente las variables decisión serán tanto $a_{0C}, a_{1C}, \dots, a_{mC}$ como $a_{0R}, a_{1R}, \dots, a_{mR}$, mientras que si los centros se han prefijado previamente mediante MCO, las variables decisión quedarían reducidas al conjunto $a_{0R}, a_{1R}, \dots, a_{mR}$.

2.4 Predicción con regresión borrosa

Tras haber ajustado los parámetros de la regresión, obtendremos la siguiente relación entre la variable explicada y las explicativas:

$$\hat{Y} = \hat{A}_0 + \hat{A}_1 X_1 + \dots + \hat{A}_m X_m. \quad (14)$$

Por tanto, la predicción de Y , que notaremos como \hat{Y}_e , será:

$$\begin{aligned} \hat{Y} &= \langle \hat{Y}_{eC}, \hat{Y}_{eR} \rangle = \sum_{i=0}^m \langle \hat{a}_{iC}, \hat{a}_{iR} \rangle X_{ie} = \\ &\left\langle \sum_{i=0}^m \hat{a}_{iC} X_{ie}, \sum_{i=0}^m \hat{a}_{iR} |X_{ie}| \right\rangle, \quad \text{con} \quad X_{0,e} = 1. \end{aligned} \quad (15)$$

3 Resultados

Los resultados reportados de los modelos de regresión usual se obtuvieron utilizando el paquete estadístico NCSS y para resolver los modelos programación lineal se usó el paquete de optimización LINDO. Se cuenta con información de tipo histórico de 39 meses, de enero de 2006 a marzo de 2009, se tomaron estos datos considerando que hubo una relativa estabilidad en la CD hasta el último trimestre de 2008. El ICC considerado en este estudio como la variable respuesta Y , se pudo “fuzzificar”, utilizando la regla empírica, que no depende de ningún tipo de distribución de probabilidad para valores de tipo cuantitativo, en el intervalo de una desviación estándar alrededor del valor promedio se concentran alrededor del 65-68% de los datos, por lo que se pudieron obtener los intervalos borrosos $Y_j = \langle Y_{jC}, Y_{jR} \rangle$, donde el subíndice j se refiere al mes, Y_{jC} será igual al índice de confianza del consumidor que es el promedio de los índices explicados en la Introducción del presente trabajo e Y_{jR} es la desviación estándar de los valores de los índices considerados en el mes, los datos utilizados en el presente trabajo se obtuvieron de [17] y [18], y se reportan en Cuadro 1.

Los resultados que se obtuvieron para los diferentes modelos utilizados fueron:

1. Modelo por mínimos cuadrados ordinarios, obteniéndose la recta de regresión:

$$\hat{Y}_j = \hat{a}_0 + \hat{a}_1 X_{1,j} = \\ 167.5055 - 5.9616 X_{1,j}.$$

Se obtuvo un coeficiente de correlación del 66.59%, e intervalos del 95% de confianza(IC) para los parámetros dados por:

$$IC(\hat{a}_0) = (142.3408, 192.6701) \quad IC(\hat{a}_1) = (-8.1862, -3.7370).$$

Hay que recordar que tanto el coeficiente de correlación como el de regresión para el modelo MCO, sólo nos indica qué tan buena puede ser la aproximación de tipo lineal. Que los coeficientes sean cero o cercanos a cero, no significa que no pueda existir algún otro modelo funcional de tipo no lineal.

Cabe mencionar que en un sentido estricto estadístico, con el modelo de regresión por MCO, sólo es permitido pronosticar dentro del

| | Y_c | Y_r | X | $Y_c - Y_r$ | $Y_c + Y_r$ |
|-------------|--------|------------|------------|-------------|-------------|
| Enero(2006) | 107.64 | 2.70795126 | 10.5470045 | 104.932049 | 110.347951 |
| Febrero | 108.02 | 4.07271408 | 10.48128 | 103.947286 | 112.092714 |
| Marzo | 112.56 | 6.15410432 | 10.7467636 | 106.405896 | 118.714104 |
| Abril | 109.46 | 1.65015151 | 11.0421222 | 107.809848 | 111.110152 |
| Mayo | 109.88 | 3.33121599 | 11.0922818 | 106.548784 | 113.211216 |
| Junio | 107.84 | 2.82895741 | 11.3912636 | 105.011043 | 110.668957 |
| Julio | 109.74 | 2.20068171 | 10.9857857 | 107.539318 | 111.940682 |
| Agosto | 110.7 | 3.62353419 | 10.8719652 | 107.076466 | 114.323534 |
| Septiembre | 110.6 | 2.70092577 | 10.9853095 | 107.899074 | 113.300926 |
| Octubre | 108.62 | 2.85954542 | 10.8971455 | 105.760455 | 111.479545 |
| Noviembre | 105.94 | 3.5366651 | 10.91773 | 102.403335 | 109.476665 |
| Diciembre | 109.66 | 4.42696284 | 10.8478667 | 105.233037 | 114.086963 |
| Enero(2007) | 104.26 | 4.56541345 | 10.9528636 | 99.6945865 | 108.825413 |
| Febrero | 103.64 | 3.57603132 | 10.9998 | 100.063969 | 107.216031 |
| Marzo | 106.52 | 3.71846743 | 11.1138619 | 102.801533 | 110.238467 |
| Abril | 106.28 | 1.23166554 | 10.9805842 | 105.048334 | 107.511666 |
| Mayo | 107.28 | 3.42520072 | 10.8166909 | 103.854799 | 110.705201 |
| Junio | 105.26 | 2.966985 | 10.8349762 | 102.293015 | 108.226985 |
| Julio | 104.78 | 1.84580606 | 10.8109318 | 102.934194 | 106.625806 |
| Agosto | 108.02 | 3.14753872 | 11.0456348 | 104.872461 | 111.167539 |
| Septiembre | 105.32 | 1.20706255 | 11.031495 | 104.112937 | 106.527063 |
| Octubre | 101.42 | 5.83155211 | 10.8231435 | 95.5884479 | 107.251552 |
| Noviembre | 99.98 | 2.56651515 | 10.88664 | 97.4134849 | 102.546515 |
| Diciembre | 105.62 | 5.64597202 | 10.8483947 | 99.974028 | 111.265972 |
| Enero(2008) | 103.5 | 3.89551024 | 10.9100136 | 99.6044898 | 107.39551 |
| Febrero | 101.04 | 3.42972302 | 10.76653 | 97.610277 | 104.469723 |
| Marzo | 102.64 | 2.30607892 | 10.7313111 | 100.333921 | 104.946079 |
| Abril | 97.8 | 3.36377764 | 10.51535 | 94.4362224 | 101.163778 |
| Mayo | 94.2 | 4.93558507 | 10.4351524 | 89.2644149 | 99.1355851 |
| Junio | 90.68 | 7.87826123 | 10.3292095 | 82.8017388 | 98.5582612 |
| Julio | 88.38 | 7.33634787 | 10.2155174 | 81.0436521 | 95.7163479 |
| Agosto | 89.62 | 8.36133961 | 10.1095095 | 81.2586604 | 97.9813396 |
| Septiembre | 88.62 | 6.75255507 | 10.6436857 | 81.8674449 | 95.3725551 |
| Octubre | 81.98 | 9.64245819 | 12.631387 | 72.3375418 | 91.6224582 |
| Noviembre | 83.84 | 8.65609612 | 13.1139526 | 75.1839039 | 92.4960961 |
| Diciembre | 84.06 | 8.31101678 | 13.4226333 | 75.7489832 | 92.3710168 |
| Enero(2009) | 81.94 | 10.989677 | 13.8921143 | 70.950323 | 92.929677 |
| Febrero | 78.9 | 13.7308048 | 14.5966421 | 65.1691952 | 92.6308048 |
| Marzo | 79.42 | 12.0134092 | 14.7208 | 67.4065908 | 91.4334092 |

Tabla 1: Valores mensuales de ICC y CD.

| | Y_c | Y_r | X | $Y_c - Y_r$ | $Y_c + Y_r$ |
|----------|-------|-------------|---------|-------------|-------------|
| Abr 2009 | 82.1 | 12.37659081 | 13.4367 | 69.72340919 | 94.47659081 |
| May 2009 | 78.26 | 14.42109566 | 13.1621 | 63.83890434 | 92.68109566 |
| Jun 2009 | 81.04 | 13.35788157 | 13.3418 | 67.68211843 | 94.39788157 |
| Jul 2009 | 85.42 | 13.65877008 | 13.3654 | 71.76122992 | 99.07877008 |
| Ago 2009 | 81.54 | 14.84008086 | 13.0080 | 66.69991914 | 96.38008086 |
| Sep 2009 | 81.88 | 12.37747955 | 13.4212 | 69.50252045 | 94.25747955 |
| Oct 2009 | 76.96 | 13.80771523 | 13.2257 | 63.15228477 | 90.76771523 |
| Nov 2009 | 78.16 | 14.65462384 | 13.1094 | 63.50537616 | 92.81462384 |
| Dic 2009 | 80.14 | 15.13367107 | 12.8631 | 65.00632893 | 95.27367107 |

Tabla 2: Valores mensuales reales de ICC y CD.

rango de valores observados, aunque en la práctica es usual el utilizar este modelo ya sea para interpolar o extrapolar, pero como todo método de pronóstico hay que tener cuidado, puesto que se considera estacionariedad de las observaciones. Un ejemplo interesante, es cuando se tienen valores atípicos (layouts), una práctica muy común es el de eliminarlos o substituirlos ya sea por promedios o por alguna otra cantidad, lo que puede afectar en el modelo final. En cambio, con los modelos borrosos, este tipo de valores se pueden considerar dentro del intervalo de confianza posibilístico, es decir, un intervalo borroso.

2. Modelo de Tanaka-Ishibuchi ($\hat{Y}_j \subseteq Y_j, \forall j$), optimizando \hat{a}_{iC} y optimizando $\hat{a}_{iR}, i = 0, 1$. Para este modelo se encontraron los estimadores borrosos de la forma $(\hat{a}_{iC}, \hat{a}_{iR}), i = 0, 1$, mediante el modelo dado por 6 y las restricciones dadas por 7 a 9, obteniéndose:

$$\begin{aligned}\hat{Y} &= < 96.0085, 0 > + < 0, 2.1127 > X_1 = \\ &< 96.0085, 2.1127X_1 >,\end{aligned}$$

con un valor de la función objetivo de $z = 927.4568$.

3. Modelo de Sakawa-Yano ($\hat{Y}_j = Y_j, \forall j$), optimizando tanto \hat{a}_{iC} como $\hat{a}_{iR}, i = 0, 1$. Para este modelo se encontraron los estimadores borrosos de la forma $(\hat{a}_{iC}, \hat{a}_{iR}), i = 0, 1$, mediante el modelo dado por 10 y las restricciones dadas por 11 a 13, obteniéndose:

$$\begin{aligned}\hat{Y} &= < 100.3277, 0 > + < 0, 0.6893 > X_1 = \\ &< 100.3277, 0.6893X_1 >,\end{aligned}$$

con un valor de la función objetivo de $z = 302.5645$.

4 Conclusiones

Recordemos que un primer paso fue el de fuzzificar la variable respuesta, para poder utilizar la modelación borrosa y se consideró con una relación aproximada de 70-30 al utilizar la regla empírica del tomar el valor promedio más menos una desviación estándar.

Como se puede observar de los resultados, el modelo de regresión por MCO tiene un coeficiente de correlación de tan solo 66.59%, además realizando un estudio sobre este modelo se puede observar que tiene una ordenada al origen de aproximadamente 168 y con una pendiente de aproximadamente -6, lo que no parece ser un buen modelo del comportamiento real del ICC respecto de la CD, puesto que se podría considerar que a partir de un valor del ICC de 168, disminuiría en 6 unidades por cada peso en que se incrementa la CD.

En cambio, los modelos borrosos, muestran ser más apropiados para este estudio, puesto que al considerar que a partir de un valor del ICC de 100(92) existirá una variación de aproximadamente 0.07(2.11) unidades por cada peso en que se incrementa la CD, permitiendo de esta manera ver que existen otros factores diferentes de la cotización del dolar, que influyen en los valores del índice de confianza del consumidor, lo que parece más lógico para este problema en particular.

Ahora bien, para decidir qué modelo borroso es el que mejor aproxima el comportamiento, se podría tomar en cuenta el valor de la función objetivo, por lo que para este estudio se tomara la aproximación Sakawa-Yano. Es un modelo que tiene una buena e interesante interpretación y que permite darle un valor “justo” a la variable CD respecto de su posible influencia para la determinación del ICC.

Se consideraron valores reales de abril a diciembre de 2009 dados en el Cuadro 2 para realizar una comparación de pronóstico. Cabe destacar que para tal fin, se midió la diferencia entre los intervalos, esto se realizó sumando los valores absolutos de las diferencias entre los valores extremos inferiores y de los valores extremos superiores respectivamente de los intervalos de valores observados y los valores predichos obteniéndose, para los meses antes citados un valor de 732.87 considerando el intervalo probabilístico del 95% de confianza y de 354.9 y de 293.04 considerando los intervalos producidos por los modelos de Sakawa-Yano y Tanaka-Ishibuchi respectivamente.

En la Figura 1, se graficaron las siguientes series:

- Serie 1: $\hat{Y}_j = 142.3408 - 8.1862X_j$,
- Serie 2: $\hat{Y}_j = 100.3277 - 0.6893X_j$,

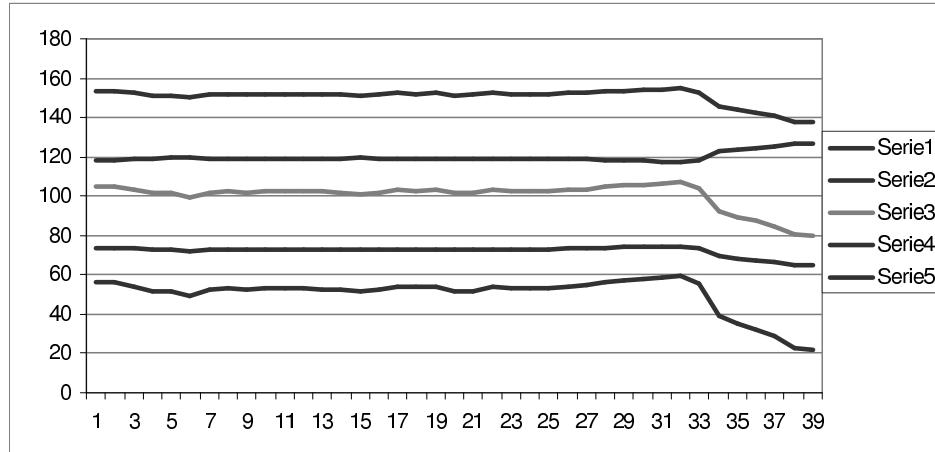


Figura 1: Gráfica de los intervalos borrosos y probabilísticos.

- Serie 3: $\hat{Y}_j = 167.5055 - 5.9616X_j$,
- Serie 4: $\hat{Y}_j = 100.3277 + 0.6893X_j$,
- Serie 5: $\hat{Y}_j = 192.6701 - 3.737X_j$.

Los intervalos borrosos proporcionan valores más conservadores que los intervalos probabilísticos del 95% de confianza, como se puede observar en la Gráfica 1, los valores proporcionados por el modelo de Sakawa-Yano varían entre 90.18 y 110.47 los cuales son más realistas para este problema, es decir, se ajustan más a los valores que podría esperarse que tomará el ICC. Además, hay que recordar que para el caso de pronóstico, los modelos borrosos no tienen ninguna restricción, al contrario de los modelos probabilísticos, lo cual es una ventaja adicional de los primeros.

Cabe mencionar que el estudio se puede ampliar considerando otros criterios de fuzzificación, así como otros tipos de números borrosos y diferentes modelos.

Agradecimientos Los autores desean agradecer los comentarios de los árbitros anónimos, lo que contribuyó a una mejora sustancial del trabajo.

Referencias

- [1] De Andrés Sánchez, J.; Terceño Gómez, A. (2002) “Programación matemática y regresión lineal con instrumentos de la teoría de los subconjuntos borrosos”. Departamento de Gestión de Empresas, Facultad de Ciencias Económicas y Empresariales. Universidad Rovira i Virgili, Tarragona.

- [2] De Andrés Sánchez, J.; Terceño Gómez, A. (2003) “Estimating a term structure of interest rates for fuzzy financial pricing by using 3 fuzzy regression methods”, *Fuzzy Sets and Systems* **139**(2): 313–331.
- [3] De Andrés Sánchez, J.; Terceño Gómez, A. (2003a) “La utilización de la regression borrosa para la predicción económica: Un análisis comparativo con mínimos cuadrados ordinarios”, Departamento de Gestión de Empresas, Facultad de Ciencias Económicas y Empresariales. Universidad Rovira i Virgili, Tarragona.
- [4] de-los-Cobos-Silva, S.G.; Goddard Close, J.; Gutiérrez-Andrade, M.A.; Terceño Gómez, A. (2005) “Comparación de métodos de regresión usual, borrosa y de pronóstico, para tramos carreteros: caso de estudio”, *XII Congreso de la Sociedad Internacional de Gestión Empresarial Fuzzy (SIGEF)*: 517–527.
- [5] de-los-Cobos-Silva, S.G.; Gutiérrez-Andrade, M.A.; Lara Velázquez, P. (2010) “Análisis borroso del impacto del índice de inflación y de la cotización dólar sobre el índice de confianza en México”, *Revista Estocástica: Finanzas y Riesgo* (por publicarse).
- [6] Buckley , J.J; Qu, Y. (1990) “On using α -cuts to evaluate fuzzy equations”, *Fuzzy Sets and Systems* **38**(3): 309–312.
- [7] Dubois, D.; Prade, H. (1980) “Fuzzy Sets and Systems: Theory and Applications”. Academic Press, New York.
- [8] Fedrizzi, M.; Fedrizzi, M.; Ostasiewics, W. (1993) “Towards fuzzy modelling in economics”, *Fuzzy Sets and Systems* **54**(3): 259–268.
- [9] Johnston, J. (1987) *Métodos de Econometría*. Vicens Vives, Barcelona.
- [10] Kaufmann, A.; Gil Aluja, J.; Terceño, A. (1994) *Matemática para la Economía y la Gestión de Empresas*. Foro Científico, Barcelona.
- [11] Sakawa, M.; Yano, H. (1992) “Fuzzy regression and its applications”, in: J. Kacprzyk & M. Fedrizzi (Eds.) *Fuzzy Regression Analysis*, Physica-Verlag, Heidelberg: 91–101.
- [12] Savic, D.; Pedrycz, W. (1992) “Fuzzy linear models: construction and evaluation”, in: J. Kacprzyk & M. Fedrizzi (Eds.) *Fuzzy Regression Analysis*, Physica-Verlag, Heidelberg: 47–60.
- [13] Tanaka, H. (1987) “Fuzzy data analysis by possibilistic linear models”, *Fuzzy Sets and Systems* **24**: 363–375.

- [14] Tanaka, H.; Ishibuchi, H. (1992) “A possibilistic regression analysis based on linear programming”, in: J. Kacprzyk & M. Fedrizzi (Eds.) *Fuzzy Regression Analysis*, Physica-Verlag, Heidelberg: 47–60.
- [15] Wang, H.F.; Tsaur, R.C. (2000) “Insight of a fuzzy regression model”, *Fuzzy Sets and Systems* **112**(3): 355–369.
- [16] Yen, K.K.; Ghoshray, S.; Roig, G. (1999) “A linear regression model using triangular fuzzy number coefficients”, *Fuzzy Sets and Systems* **106**(2): 167–177.
- [17] <http://www.banxico.org.mx/PortalesEspecializados/tiposCambio/> consultado el 10-Ene-2010, 2:25, p.m.
- [18] <http://dgcnesyp.inegi.org.mx/cgi-win/bdieintsi.exe/NIVA050001#ARBOL> consultado el 10-Ene-2010, 3:25, p.m.