



BANCO DE LA REPÚBLICA
SUBGERENCIA DE ESTUDIOS ECONÓMICOS

LA DEMANDA DE EFECTIVO EN COLOMBIA:
UNA CAJA NEGRA A LA LUZ DE LAS REDES NEURONALES

Martha A. Misas Arango*
Enrique López E.
Carlos A. Arango A.
Juan Nicolás Hernández A.

Noviembre 2003

* Agradecemos la colaboración de Sergio Olarte en la estimación de los modelos lineales univariados.

1 Introducción

En Colombia, como en muchos países que han optado por un régimen de inflación objetivo, se continúa monitoreando el comportamiento de algunas variables monetarias, dentro de ellas el efectivo, y pronosticando su senda futura, utilizando para ello modelos econométricos de diversos tipos. Sin embargo, la historia reciente del efectivo ha mostrado que se trata de una variable muy difícil de modelar mediante especificaciones lineales. De otro lado, no son satisfactorios los resultados obtenidos con modelos univariados de series de tiempo (ARIMA y ARIMA con intervención y transferencia) en la medida en que se han presentado elevados errores de pronóstico.

La dificultad que enfrentan los modelos tradicionales para predecir la futura demanda de efectivo es un fenómeno que se ha observado en muchas economías¹. La limitada bondad de pronóstico, se produce no solo cuando estos modelos se enfrentan a cambios estructurales, sino también, por su naturaleza no lineal. Por esta razón en la literatura contemporánea se explora la posibilidad de una relación no lineal entre las variables que conforman la demanda de efectivo.

Para el caso colombiano, esta posibilidad resulta de especial relevancia ya que durante la última década la economía ha sufrido importantes transformaciones que han traído consigo una mayor volatilidad en las variables que determinan la función de demanda de dinero como el ingreso y la tasa de interés. En particular, recientemente se ha dado una fuerte reducción en la rentabilidad de activos alternativos al efectivo debido a la reducción en la tasa de inflación y a la introducción del impuesto a las transacciones financieras. Asimismo, los ciclos de la actividad económica parecen ser más pronunciados en el pasado reciente.

Evidencia empírica preliminar muestra la posible existencia de una relación no lineal entre el efectivo y sus determinantes. Autores como Arango y Gonzalez (2000) encontraron que si bien en la economía colombiana existe una relación lineal estable en la función de demanda por efectivo en el largo plazo, el modelo lineal de corrección de errores es estadísticamente superado por una versión no-lineal conocida como modelo de transición suave o STM (smooth transition model). Es posible, sin embargo, que la relación no lineal sea de un tipo diferente a la especificación STM. Por esa razón, la estrategia definida en este documento, consiste en explorar el potencial de otro tipo de modelos no lineales.

¹ Véase por ejemplo el trabajo de Jansenn N. (1998) "The Demand for M0 in the United Kingdom Reconsidered: Some Specification Issues" Mimeo, Bank of England

El objetivo principal de este trabajo es el de mejorar los pronósticos del efectivo para el caso colombiano, reconociendo la naturaleza no lineal de la relación entre el efectivo y sus variables explicativas, como también, explorar los modelos de redes neuronales ANN (artificial neural network) como modelos no lineales. Estos últimos permiten modelar cambios de régimen, observaciones atípicas y estacionalidad. Como lo sugiere Frances (2000) la ventaja de ANN es su capacidad para aproximar cualquier función no lineal. Aplicadas a series de tiempo caracterizadas por relaciones no lineales, las redes neuronales proveen de pronósticos superiores a los obtenidos a través de modelos lineales.

El documento se organiza en siete secciones incluida esta introducción. En la segunda sección se presenta evidencia gráfica para ilustrar el comportamiento reciente del efectivo en Colombia. Existen, como se muestra en la sección siguiente, argumentos teóricos sobre la existencia potencial de no linealidades en la demanda de dinero tanto en el largo como en el corto plazo así como posibilidades empíricas para incorporar no linealidades en la estimación de la demanda de efectivo. En la cuarta sección se presentan los modelos lineales y no-lineales que van a ser utilizados con el fin de examinar y comparar su desempeño de pronóstico. La quinta sección recoge los resultados de la estimación de dichos modelos y se presentan los criterios de selección de la mejor red neuronal. En la sexta sección se compara el desempeño de pronóstico. La última sección concluye.

2 Comportamiento reciente del efectivo.

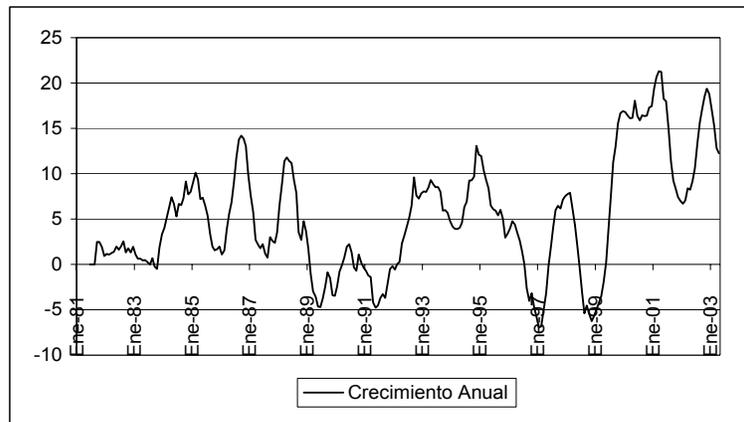
El comportamiento del efectivo real durante las últimas dos décadas ha estado marcado por tres periodos claramente diferenciados. Durante los 80, su comportamiento se caracterizó por bajos niveles de crecimiento anual, para la variable medida con frecuencia mensual, aunque también por una alta volatilidad. A partir de 1991 se dio un ritmo de crecimiento sostenido, el cual se mantuvo hasta 1995 seguido por tres años de desaceleración de éste, hasta que en 1999 reaparece de nuevo una fuerte y persistente tendencia de crecimiento que alcanza el orden del 12% anual y se mantiene durante los últimos 4 años (Grafico 1).

Ese comportamiento reciente del efectivo es inesperado por varias razones. La primera de ellas es que muchos autores coinciden en pronosticar una reducción secular de la demanda de efectivo en el largo plazo como consecuencia del uso cada vez más frecuente de los sistemas de pagos electrónicos (Markose y Loke, 2002). Algunos autores plantean incluso que puede llegarse a

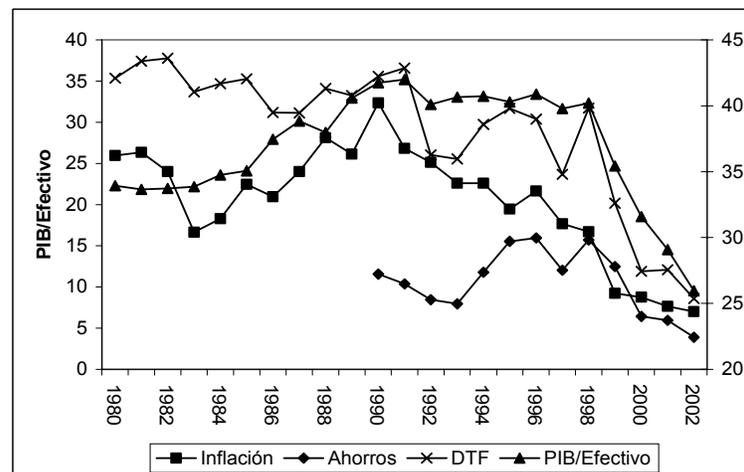
la desaparición de la base monetaria con consecuencias sobre la política monetaria (Dowd, 1990 ; King, 1999 y Friedman, 1999)².

De otro lado, la historia difícilmente puede ser conciliada con el comportamiento de la actividad económica. El indicador de velocidad del efectivo (gráfico 2), muestra que dichos episodios representan innovaciones a la demanda de efectivo que van más allá del comportamiento de la actividad económica. En efecto, la velocidad aumentó continuamente entre 1983 y 1991 para luego permanecer más o menos estable hasta 1999 cuando se da una fuerte caída alcanzando niveles muy por debajo de los observados durante las dos décadas.

Gáfico 1: Crecimiento Anual del Efectivo Real



Gáfico 2: Velocidad del Dinero y Costo de Oportunidad



² Los trabajos de Goodhart (2000) y Goodhart and Krueger (2001) muestran, por el contrario, un gran escepticismo frente al pronóstico de un declive secular del efectivo. Se afirma que el efectivo es insustituible por el anonimato que brinda a quien lo posee. En economías con economías subterráneas de gran tamaño su uso se encuentra muy justificado.

El gráfico 2 alcanza a revelar que un componente importante del comportamiento de la velocidad esta asociado a los movimientos en el costo de oportunidad de las tenencias de efectivo. Por un lado, la escalada inflacionaria de los 80 parece haber inducido una mayor velocidad del dinero probablemente asociada con innovaciones tecnológicas ahorradoras de efectivo, como cajeros electrónicos, y tarjetas crédito-debito. No obstante, la tendencia creciente en la velocidad se trunca en 1991 coincidiendo con la nueva senda decreciente en la inflación observada durante la década de los 90, hasta que en 1999 la fuerte disminución en la velocidad coincide con una profunda caída en los retornos de los activos mas cercanos al efectivo como la tasa DTF y las tasas de depósitos de ahorro³. Esta última, en particular parece coincidir estrechamente con el comportamiento de la velocidad aun cuando solo se tiene una serie relativamente reciente.

El gráfico 3, sin embargo, muestra que la relación entre el efectivo real y las variables que eventualmente entrarían en la estimación de una ecuación de demanda de efectivo (variables que midan el costo de oportunidad o que capturen como variables de escala la actividad económica) no puede ser totalmente caracterizada por una especificación lineal. En este gráfico se presentan las relaciones no-condicionales entre la variación anual en el efectivo real por un lado y la tasa de interés, la inflación, el índice de producción manufacturera y su volatilidad. Claramente, mayores niveles de tasa de interés/inflación inducen menores ritmos de crecimiento de los saldos reales de efectivo, no obstante dicha relación parece acentuarse en el caso de la DTF y atenuarse en el caso de la inflación una vez el costo de oportunidad cruza un umbral que en el caso de la DTF podría ser del 22% y en de la inflación del 16%.

La variable de escala que en este caso es el índice de producción industrial⁴ parece ser la única que entraría eventualmente en forma lineal en la demanda por efectivo. Este índice se considera en algunas especificaciones de demanda de efectivo como una variable proxy de la volatilidad en los flujos de ingresos y gastos⁵. Por esa razón, en el recuadro inferior derecho se presenta la relación entre efectivo y volatilidad en el ingreso. La relación positiva entre los dos para niveles moderados de volatilidad valida la idea de que el efectivo funciona como un mecanismo de absorción de innovaciones no esperadas en el flujo de ingresos y obligaciones. No obstante la relación parece invertirse cuando dicha volatilidad alcanza niveles anormalmente altos mostrando

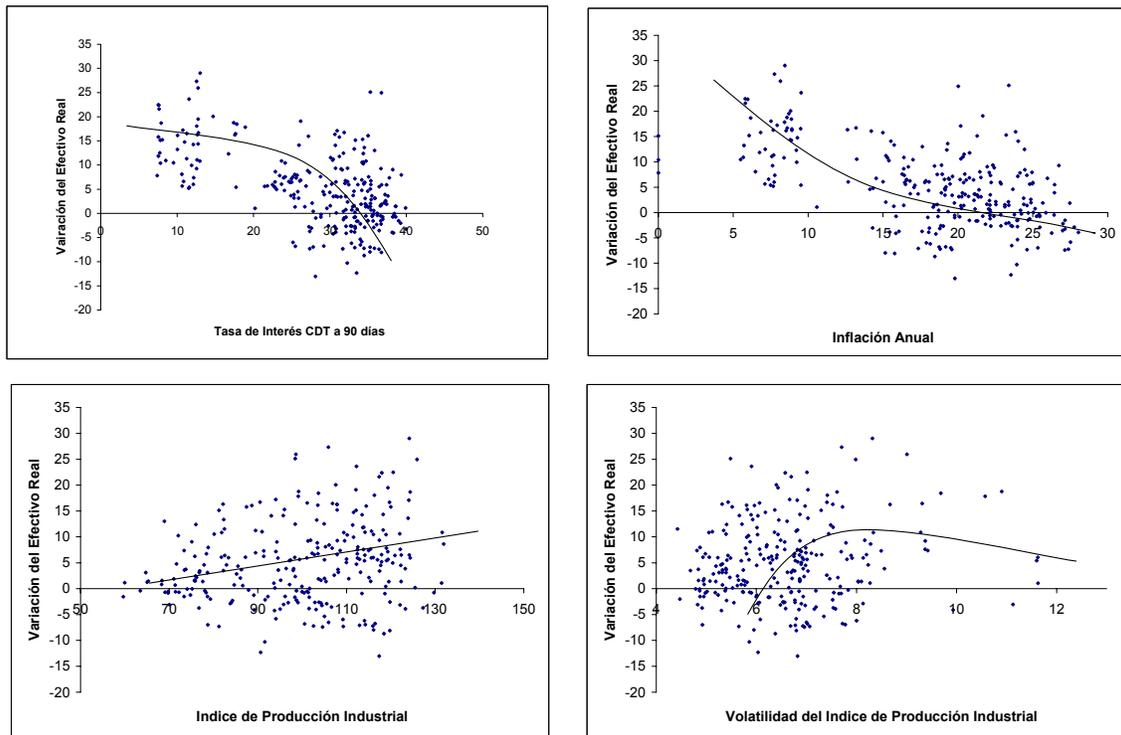
³ Esta tasa ajustada por los costos adicionales de tenencia de depósitos en cuentas de ahorro incluso los costos asociados al impuesto a las transacciones financieras.

⁴ Dada la periodicidad mensual del ejercicio, se selecciona el índice de producción industrial como la variable de escala.

⁵ Por ejemplo en el modelo de Milbourne, Buchholtz y Wasan (1983) que se describe más adelante.

que dicho mecanismo puede resultar muy costoso cuando las innovaciones en el ingreso son considerables.

Gráfico 3: Variación Anual del Efectivo Real vs. Variables Exógenas



3 No-linealidades en la demanda de dinero

En la teoría económica la existencia de no linealidades en la demanda de dinero tiene una larga tradición. El modelo pionero de Baumol (1952), por ejemplo, dio origen a la “regla de la raíz cuadrada” según la cual los saldos en efectivo son proporcionales a la raíz cuadrada del valor de las transacciones por periodo de tiempo.

En este documento, se hace énfasis en aquellos argumentos teóricos que implican una demanda de dinero que responde en forma no lineal a choques exógenos y, que por ende, sus elasticidades varíen en el tiempo en función de los valores de las variables que la determinan. En este sentido, la demanda de dinero puede representarse de la forma:

$$m_t = g(x_t; \theta) \quad (1)$$

donde el efectivo, m_t , es una función de un conjunto de variables exógenas, x_t , y de un conjunto de parámetros, θ , la forma funcional $g(\cdot)$ puede tener cualquier especificación incluyendo la lineal, como caso particular. A este nivel de generalidad la elasticidad de la demanda de dinero con respecto a cualquiera de los argumentos exógenos depende no solo de los parámetros del modelo sino del nivel de las variables explicativas x_t :

$$\varepsilon_{mx} = h(x_t; \theta) \quad (2)$$

3.1 Largo plazo

Aún cuando la mayoría de los trabajos empíricos estiman funciones de demanda de dinero lineales, su derivación teórica, partiendo ya sea de un modelo de optimización del consumidor con dinero entrando en la función de utilidad (Barnett, Fisher y Serletis (1992) y Arrau y Gregorio (1994)) o de un modelo de optimización de portafolio (Buiter y Armstrong (1978)) en general tendrá características no lineales. Solo al imponer restricciones en las preferencias de los individuos y en la tecnología que define los costos de transacción se llega a especificaciones lineales.

Otras fuentes adicionales de no linealidad en la demanda de efectivo se derivan de la agregación de demandas individuales y la modelación explícita de la intermediación financiera. El primer factor generador de no-linealidades en la agregación proviene de sumar demandas individuales de dinero no-lineales. El segundo factor se deriva de que en un mundo estocástico en donde los agentes enfrentan choques idiosincráticos al mismo tiempo que choques macroeconómicos, aún con funciones de demanda lineales y poblaciones homogéneas, el comportamiento agregado puede presentar relaciones no lineales debido a las respuestas diferenciadas de cada individuo (falacia de la composición –Caballero 1992 y 1994-). El tercero se deriva de levantar el supuesto de homogeneidad entre los agentes (aún manteniendo el supuesto de que las demandas individuales de dinero son lineales). En general, el grado de no-linealidad en la respuesta agregada a choques exógenos en los determinantes de la demanda por dinero dependerá de la historia de los choques agregados, del grado de heterogeneidad de los agentes, y del grado de correlación de los choques idiosincráticos.

Por su parte, la introducción explícita de la intermediación financiera provee de nuevos argumentos para la existencia de no-linealidades en la demanda de efectivo. Gertler (1988) divide

en dos las razones para la existencia de intermediarios financieros: costos de información, y reducción de costos transaccionales entre depositantes y ahorradores. Los costos de información inducen respuestas no lineales y asimétricas por parte de los intermediarios financieros frente a individuos heterogéneos. Por su parte, las decisiones de adopción de tecnologías financieras y transaccionales normalmente tienen componentes no lineales relacionados con economías de escala, irreversibilidad de las inversiones y opciones reales las cuales generan una respuesta asimétrica frente al riesgo.

Un ejemplo de cómo el proceso de innovación puede generar no linealidades en la demanda de dinero es el modelo de Soto (1997) en el cual, suponiendo que la innovación tecnológica se comporta como un proceso aleatorio controlado por una distribución de Poisson, la constante de la función de demanda sigue un proceso auto-regresivo que depende no-linealmente del retorno esperado de la innovación. En este caso, un proceso persistente de alta inflación, al inducir cambio técnico, puede producir efectos no lineales sobre la demanda de efectivo.

Finalmente, están los modelos de “inventario” los cuales ven la tenencia de saldos monetarios como un mecanismo de absorción de los choques inesperados en las obligaciones de pago y en los ingresos de los agentes económicos. Estos choques pueden permanecer como excesos de tenencia de efectivo por algún tiempo dados los costos de hacer ajustes continuos en el portafolio. En estos modelos, el flujo de ingresos y obligaciones, la tasa de interés, y sus volatilidades, así como los costos de transacción y eventualmente la tecnología definen la banda dentro de la cual el individuo deja flotar libremente sus saldos y los valores óptimos a los cuales debe restablecer sus saldos en el evento en que estos trasciendan la banda. En este caso la demanda por dinero del individuo es no lineal con respecto a las variables exógenas, ya que su respuesta tiene una forma “quebrada” basada en dicha banda.

3.2 Corto plazo

La aproximación estándar al problema dinámico de corto plazo de la demanda de dinero es el modelo de ajuste parcial (Goldfeld (1973) entre otros) en el cual, en cada período, los agentes ajustan una proporción constante de sus existencias de dinero hacia su nivel deseado, de acuerdo con el siguiente modelo:

$$m_t = m_{t-1} + \theta(m_t^* - m_{t-1}) \quad (3)$$

$$m_t = \theta x_t + (1 - \theta)m_{t-1} \quad (4)$$

donde x_t es el vector de variables relevantes como la tasa de interés, el valor de las transacciones, la tasa de inflación entre otras variables las cuales determinan el stock óptimo de saldos en efectivo. Sin embargo, dicha aproximación no tiene una fundamentación teórica y normalmente se ha justificado basándose en modelos de largo plazo como los ya descritos. En particular, muchos trabajos empíricos mencionan la existencia de costos fijos de ajuste de los saldos reales como sustento para justificar el modelo de ajuste parcial no obstante los modelos de largo plazo predicen que los agentes siguen una regla binaria: ajuste total al nivel óptimo o no ajuste. También se mencionan, de manera heurística, restricciones de liquidez e imperfecciones en el mercado de capital como justificación al modelo de ajuste.

Desde que el modelo de ajuste parcial empezó a presentar deficiencias de pronóstico se empezaron a dar innovaciones tanto econométricas como teóricas con el fin de desentrañar lo que se dio en llamar el "dinero perdido": nombre asociado a la inestabilidad en las estimaciones de la demanda de dinero y al bajo poder predictivo de los modelos de ajuste parcial.

Milbourne, Buchholtz y Wasan (1983) derivan las trayectorias temporales de las tenencias de efectivo partiendo de un modelo de costos de ajuste donde los agentes cambian sus tenencias de efectivo con base en bandas y niveles óptimos de restablecimiento del inventario dados un conjunto de supuestos sobre el comportamiento estocástico de los flujos de ingresos y gastos. En particular, los saldos nominales en efectivo M_t toman valores $0 \leq M_t \leq h$ donde h sigue la regla de optimización de largo plazo de Miller y Orr (1968):

$$h = 3z = -3 \left(\frac{3\gamma\sigma^2}{4r} \right)^2, \quad (5)$$

siendo z el nivel efectivo al cual retornan los saldos cuando estos cruzan los umbrales de la banda; γ los costos de ajuste de portafolio, σ la dispersión de la variación en los saldos de efectivo y r la tasa de interés. En este modelo el nivel óptimo de tenencias de efectivo es:

$$M_t^* = 4 \left(\frac{3\gamma\sigma^2}{4r} \right)^{1/3}, \quad (6)$$

y la trayectoria de los saldos en efectivo del individuo en el corto plazo es:

$$E(M(t) | M_0) = M_0 + (1 - e^{-(\sigma^2\pi^2/2h^2)t})(M^* - M_0) \quad (7)$$

al sustituir (6) en (7), suponiendo que $\sigma = \alpha_1 y$, donde y es el nivel de ingreso del individuo, y $\gamma = \alpha_2 y^\phi$ el costo de ajuste de las tenencias, e integrando sobre todos los individuos se tiene que:

$$\frac{M_t}{P_t} = \frac{M_{t-1}}{P_t} + \theta y_t^\mu r_t^\nu \left(A y_t^\alpha r_t^\beta - \frac{M_{t-1}}{P_t} \right) \quad (8)$$

donde:

$\nu = 2/3$; $\mu = (2/3)(1 - \phi)$; $A = 4(3/4)\alpha_1^2\alpha_2^{1/3}$; $\alpha = (2 + \phi)/3$; $\beta = -1/3$; $\theta = (k/2)(\alpha_1 / \alpha_2)^{2/3}$, y k es una constante.

El modelo de la ecuación (8) sugiere que es el cambio en el stock de saldos reales el que depende del flujo de ingresos y no el nivel del stock. Aun cuando el modelo produce elasticidades de largo plazo constantes con respecto a la tasa de interés, es claro que tanto la elasticidad ingreso como el parámetro de ajuste parcial dependen de los niveles de las variables corrientes y por ende no son constantes a lo largo del tiempo lo cual implica que (8), en su versión paramétrica, debe ser estimada mediante un método que reconozca la no-linealidad de la relación⁶.

4 Modelos lineales y no lineales.

Las dos secciones anteriores han mostrado la posibilidad, tanto empírica como teórica, de existencia de no linealidades en el comportamiento del efectivo y, por ende, la dificultad que existe para alcanzar pronósticos adecuados si no se reconoce dicho comportamiento. Es conveniente, en consecuencia, llevar a cabo un ejercicio que provea evidencia empírica sobre el desempeño en pronóstico que tienen los modelos lineales y los no lineales. En nuestro caso, se consideran modelos lineales de carácter univariado como ARIMA y ARIMA con transferencia y multivariado tipo VAR, en el cual se incorpora la existencia de una relación de cointegración de largo plazo entre las variables que conforman el sistema tradicional de información para funciones de demanda por dinero.

⁶ En un modelo mas general, donde los costos de ajuste del portafolio no son fijos sino que dependen del valor de la transacción, el modelo se vuelve más complejo generando resultados no-lineales en todas las variables independientes.

4.1 Modelos lineales

Dentro del grupo de modelos lineales se consideran los univariados como: (i) los modelos ARIMA, donde el comportamiento de una serie de tiempo, y_t , se explica a través de sus valores pasados y de una suma ponderada de errores, ε_t , pasados y presentes: $\Phi(L)(1-L)^d y_t = \delta + \Theta(L)\varepsilon_t$; con $\{\varepsilon_t\}$ serie de perturbaciones ruido blanco y d número de diferenciaciones requeridas para que $\{y_t\}$ alcance un comportamiento estacionario y (ii) los modelos ARIMA con Transferencia, donde el modelo ARIMA considera información adicional proveniente de una o varias variable(s) exógena(s): $y_t = \frac{W(L)}{\delta(L)} x_{t-1} + \frac{\Theta(L)}{\Phi(L)} \varepsilon_t$; con $\{\varepsilon_t\}$ serie de perturbaciones ruido blanco y $\{y_t\}$ y $\{x_t\}$ series estacionarias⁷; y los multivariados⁸ de series de tiempo VAR bajo existencia de relaciones de cointegración: $\Delta Y_t = \Gamma_1 \Delta Y_{t-1} + \dots + \Gamma_{k-1} Y_{t-k+1} + \Pi Y_{t-1} + \mu + \Psi D_t + \varepsilon_t$; siendo Y_t el sistema de información, D_t el conjunto de elementos determinísticos y $\Pi = \alpha\beta'$ donde α representa las velocidades de ajuste y β los vectores de cointegración.

4.2 Modelos no lineales: redes neuronales artificiales

Como se observa en la literatura internacional (Dijk, Terasvirta y Franses, 2001), en los últimos años el uso de modelos no lineales de series de tiempo se ha ido incrementando de manera considerable y dentro de ellos, los de redes neuronales artificiales⁹. Estos modelos han tenido un desempeño satisfactorio en un gran número de aplicaciones en diversas áreas del conocimiento. Una red neuronal artificial puede definirse como un modelo matemático inspirado en el funcionamiento del cerebro humano y en particular, en la forma como éste procesa información. En el contexto de análisis de series de tiempo, se clasifican como modelos no lineales entrenados para (i) realizar conexiones entre los valores pasados y presentes de una serie de tiempo, aprendiendo de su error de pronóstico y (ii) extraer estructuras y relaciones escondidas que gobiernan el sistema de información (Azoff, 1996). Su utilización está primordialmente motivada por la capacidad de aproximarse a cualquier función medible de Borel con un muy buen grado de exactitud, como lo señala, entre otros, Rech (2002)¹⁰.

⁷ Véanse, Wei (1994), Brockwell and Davis (1987), entre otros.

⁸ Véase, Lutkepohl (1993).

⁹ Artificial Neural Network, ANN.

¹⁰ Citando a Hornik et al. (1989).

4.2.1 Arquitectura

Este artículo se centra en una red neuronal tipo “*feedforward*” multicapa con una única superficie escondida o “*single hidden layer*”. Dicha arquitectura¹¹ se selecciona con base en el concepto de pronóstico de una serie de tiempo donde se desea predecir observaciones futuras utilizando algún tipo de función de las observaciones pasadas (Faraway et al, 1995). En el caso de los modelos de redes neuronales artificiales, dicha función puede ser definida de manera general como en Granger y Terasvirta (1993):

$$y_t = \Phi_0 + X_t' \Phi + \sum_{j=1}^Q \beta_j G(Z_t' \gamma_j) + \varepsilon_t \quad (9)$$

donde $Z_t \subseteq X_t$, $X_t = \{y_{t-1}, \dots, y_{t-l}, w_{1t}, \dots, w_{mt}\}$ con y_{t-j} , $j = 1, \dots, l$ rezagos de la variable dependiente, w_{jt} , $j = 1, \dots, m$ variables exógenas y G la función logística, es decir:

$$G(w) = \frac{1}{1 + \exp(-w)} \quad (10)$$

El término $X_t' \Phi$ puede ser visto como la representación de la parte lineal de la relación entre y_t y X_t , en tanto que, la componente logística caracteriza la parte no lineal, siendo Q el número de componentes logísticas.

El gráfico 4 presenta la arquitectura correspondiente a este modelo. Como se observa, la red está constituida por tres superficies diferentes. La superficie base o superficie-*input* está conformada por las variables explicativas, las cuales pueden presentar una relación lineal con y_t , como también, algunas de ellas, un comportamiento no lineal, conformando los conjuntos X_t y Z_t , respectivamente. Dichos *inputs* son multiplicados por los valores de ponderación de los conectores, ϕ_i $i = 1, \dots, l + m$, en el caso de la relación lineal y $\{\gamma_{ij}, i = 1, \dots, k\} = \gamma_j$ $j = 1, \dots, Q$, en el caso de la relación no lineal, estos últimos atenúan o amplifican las señales originales para posteriormente entrar a la superficie oculta¹², donde existen unidades escondidas asociadas a funciones logísticas. Es importante resaltar que dichas unidades pueden ser entendidas como variables latentes que enriquecen al modelo lineal. En

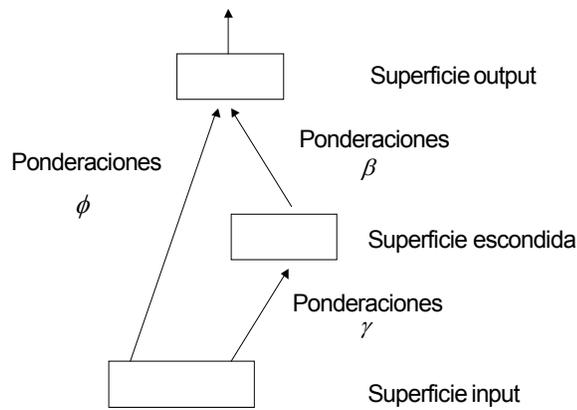
¹¹ En este contexto, el término arquitectura se entiende como el conjunto de *inputs*, P , el número de unidades escondidas, Q , y el número de superficies escondidas, 1 , necesarios en la determinación del componente no lineal.

¹² No relacionada de manera directa con y_t .

esta superficie se forman las combinaciones $Z_t' \gamma_j$ ($j=1, \dots, K$) que son transformadas a valores entre cero y uno mediante las funciones de activación $G(\cdot)$. Finalmente, éstas son multiplicadas por ponderaciones β_j para dar como resultado, al adicionar la componente lineal, el *output* y_t ¹³.

Dado que esta red presenta una única superficie oculta y que la información va en una única dirección entrada salida (*input-output*) se conoce como red neuronal “*feedforward*” multicapa de una única superficie oculta o “*single hidden layer feedforward network*”, (Misas et al, 2002).

Gráfico 4



4.2.2. Modelación bajo redes neuronales artificiales

En una red neuronal artificial, el concepto de aprendizaje se define como un proceso de iteración y ajuste de parámetros (γ, β, ϕ) basado en la minimización de la diferencia entre el *output* producido y el observado para una arquitectura dada. En este contexto, la determinación del modelo comprende las etapas de especificación, estimación de parámetros y evaluación de los modelos estimados.

¹³ Swanson y White (1995) se refieren a los parámetros γ_{ij} como ponderaciones *input to hidden unit*, mientras que los parámetros β_j son ponderaciones *hidden unit to output*. A su vez, los parámetros Φ de la parte lineal en (2) se conocen como ponderaciones *input to output*.

En la primera etapa se seleccionan las variables que conforman los conjuntos X_t y Z_t , como también, el número de unidades escondidas, Q . Existen diferentes estrategias que pueden ser utilizadas para determinar el número de unidades ocultas y el conjunto de variables *input*, es decir, la especificación de la red neuronal. En este trabajo se sigue el esquema de Swanson y White (1995, 1997a), donde la determinación del conjunto de variables “*input*” parte la adopción de un estrategia *step-wise* en la componente lineal, de tal forma que, en cada paso, la variable adicionada sea considerada en el conjunto *input* si su inclusión mejora los criterios de selección¹⁴. Una vez definido el conjunto de variables *input* de la componente lineal, X_t ecuación (9), se conforma el conjunto Z_t , $Z_t \subseteq X_t$. En primera instancia, el conjunto Z_t incluye tan sólo la primera variable del conjunto *input*, luego se adiciona a dicho conjunto la segunda variable *input* de tal forma que en el último paso $Z_t = X_t$. La estimación de la red se lleva a cabo considerando cada uno de estos conjuntos Z_t y diferente número de unidades ocultas, $(Q) (Q = 1, \dots, Q^*)$ ¹⁵.

De acuerdo con Gradojevic y Yang (2000), el número de unidades ocultas Q a elegir presenta una disyuntiva (*trade-off*). Demasiadas unidades pueden llevar a un sobre entrenamiento o sobre ajuste lo cual evita que la red aprenda una solución general; muy pocas unidades ocultas por otro lado, inhiben el aprendizaje del patrón entre el *input* y el *output*. De esta forma, se parte de un amplio número de modelos de redes neuronales, los cuales deben ser estimados y posteriormente evaluados, en lo concerniente a su desempeño muestral.

La segunda etapa o etapa de estimación se lleva a cabo a través de la metodología de mínimos cuadrados ordinarios no lineales (NLS). Así, el vector de parámetros $\Psi = (\phi_0, \phi_1, \dots, \phi_p, \gamma_0^1, \dots, \gamma_k^1, \dots, \gamma_0^Q, \dots, \gamma_k^Q, \beta_1, \dots, \beta_Q)$ de la Ecuación (9) es estimado mediante la minimización de la función $S(\Psi)$ presentada en la Ecuación(11)¹⁶.

$$S(\Psi) = \sum_{t=1}^n [y_t - F(X_t; \Psi)]^2 \quad (11)$$

¹⁴ Criterios tales como AIC y BIC.

¹⁵ Donde Q^* es el máximo número de unidades ocultas.

¹⁶ En este trabajo: (i) las variables son re-escaladas de tal forma que tengan media cero y desviación estándar unitaria y (ii) en la función objetivo del proceso de estimación, ecuación (1), se involucra un término de penalización (*weight decay*), de tal forma que, se evite la presencia de parámetros muy grandes, Franses (2000). De tal forma que la nueva función objetivo es: $S(\Theta) = \sum_{t=1}^n [y_t - F(X_t; \Theta)]^2 + r_\phi \sum_{i=0}^k \phi_i^2 + r_\beta \sum_{j=1}^Q \beta_j^2 + r_\gamma \sum_{j=1}^Q \sum_{i=0}^k \gamma_{ij}^2$

donde:

$$F(X_t, \Psi) = \Phi_0 + X_t' \Phi + \sum_{j=1}^Q \beta_j G(Z_t' \gamma_j) \quad (12)$$

tal minimización se lleva a cabo a través de algoritmos de optimización lineal. En general, los algoritmos de optimización sobre redes neuronales son sensibles a la selección de los valores iniciales como lo presenta Rech (2002), por consiguiente, en la fase de estimación de cada arquitectura se consideran diferentes vectores de valores iniciales¹⁷, Ψ^O , seleccionando aquel que converja al mínimo valor de la función objetivo Ec (11).

Finalmente, todos los modelos estimados son evaluados a través de medidas tradicionales de desempeño de pronóstico dentro y fuera de muestra con el propósito de seleccionar al interior de todas las arquitecturas posibles la mejor de ellas. La relevancia de cada una de las variables que conforman el *input*, en el modelo seleccionado, se lleva a cabo a través de un análisis de sensibilidad, el cual puede realizarse mediante el enfoque de Draisma et al. (1995), donde se compara la variable y_t observada con la variable output estimada a través de la red, \hat{y}_t , al romper la conexión de la variable *input* particular.

5 Resultados

A continuación se presentan los resultados de la estimación de los modelos descritos en la sección anterior. En todos los casos se hace una breve explicación de las características de los modelos y se reportan las principales estadísticas calculadas.

5.1 Modelos lineales.

5.1.1 Modelo univariados: ARIMA con intervención

El modelo ARIMA con intervención para el logaritmo del efectivo real estimado con observaciones mensuales para el período comprendido entre mayo de 1984 y diciembre de 2002 es el siguiente:

$$\left(1 + \frac{0.467L}{(0.06)} + \frac{0.133L^{11}}{(0.06)}\right) \left(1 - \frac{0.161L^{23}}{(0.07)}\right) (1 - L^2)(1 - L) \text{Ln}(Efec_t) = \left(1 - \frac{0.411L^2}{(0.062)} + \frac{0.217L^{26}}{(0.063)}\right) \left(1 - \frac{0.642L^{12}}{(0.058)}\right) e_t$$

$$+ \frac{0.034}{(0.01)} (1 - L^2) SS_t + \frac{0.073}{(0.02)} (1 - L^2)(1 - L) DM_{t-5}$$

$P - value : Ljung - Box(36) : 0.926$

¹⁷ Muestreados de manera aleatoria de una distribución uniforme con límites [-2,2]

donde las intervenciones se refieren a: (i) SS_t : efecto del mes en el cual se celebra la Semana Santa y (ii) efecto del impuesto de dos por mil a las transacciones financieras, DM_{t-5} .

5.1.2 Modelo univariado: ARIMA con intervención y transferencia

El modelo ARIMA con intervención y transferencia para el logaritmo del efectivo real, con la tasa de interés de los CDT a 90 días como variable exógena, estimado con observaciones mensuales para el período comprendido entre mayo de 1981 y diciembre de 2002 es el siguiente:

$$\left(1 + \frac{0.471L}{(0.06)} + \frac{0.143L^{11}}{(0.06)}\right) \left(1 - \frac{0.161L^{23}}{(0.07)}\right) (1 - L^{12})(1 - L) \text{Ln}(Efec_t) = \left(1 - \frac{0.421L^2}{(0.062)} + \frac{0.239L^{26}}{(0.062)}\right) \left(1 - \frac{0.686L^{12}}{(0.056)}\right) e_t \\ + \frac{0.034}{(0.01)} (1 - L^{12}) SS_t + \frac{0.069}{(0.02)} (1 - L^{12})(1 - L) DM_{t-5} - \frac{0.378}{(0.12)} (1 - L) \text{Ln}(TCDT_t)$$

P – value : *Ljung – Box*(36): 0.881

donde las intervenciones se refieren a: (i) SS_t : efecto del mes en el cual se celebra la Semana Santa y (ii) efecto del impuesto de dos por mil a las transacciones financieras, DM_{t-5} y la transferencia se recibe del logaritmo de la tasa de interés de los CDT a 90 días, $\text{Ln}(TCDT_t)$

5.1.3 Modelo multivariado: VEC

El análisis de cointegración, se realizó utilizando para ello la metodología de Johansen (1988). El ejercicio se lleva a cabo para el período comprendido entre octubre de 1980 y mayo de 2003. considerando el sistema de información conformado por el efectivo real, el índice de producción industrial, la tasa de interés de los CDT a 90 días y la inflación, variables transformadas mediante la función logaritmo, con información fin de período.

Se estimó la función de demanda de efectivo de largo plazo que se presenta a continuación y los signos esperados se colocan bajo la variable respectiva:

$$M = f(\Pi, X, R, t) \quad (13)$$

- + - -

donde \mathbf{M} es la demanda por efectivo en términos reales, $\mathbf{\Pi}$ es la inflación, \mathbf{X} es la medida del volumen de transacciones, \mathbf{R} es el vector que recoge los costos alternativos o de uso para los agentes que prefieren mantener efectivo, y \mathbf{t} es la variable de tendencia la cual identifica si hay un cambio tecnológico ahorrador de efectivo en el largo plazo. El signo esperado de los coeficientes, condicionado a la ecuación de demanda, es el que aparece debajo de las variables.

En el caso particular del efectivo, la inflación corresponde a la tasa de retorno de activos reales. Se supone que al incrementarse la inflación los agentes sustituyen tenencias en efectivo por otros activos entre ellos bienes reales (durables) sobre todo aquellos individuos que no tienen acceso o eluden la inversión en activos financieros (sector informal) como sustitutos alternativos (Cagan (1956)). Esta última variable es capturada por la variación anual del IPC. De otro lado, dado que la cointegración se hace con datos mensuales la variable de escala que captura el volumen de transacciones es el índice de producción industrial, mientras que la tasa de interés es capturada por el rendimiento trimestral de los certificados de depósito a término (CDT). La función de demanda de largo plazo por efectivo fue estimada en logaritmos por lo cual la ecuación (13) toma la forma (14), donde las variables en minúscula representan logaritmos naturales:

$$m = \beta_2\pi + \beta_3x + \beta_4r + \beta_5t \quad (14)$$

El análisis se realiza considerando de uno a doce rezagos y los modelos: (i) “drift”, el cual considera una tendencia lineal determinística en los niveles de las variables y (ii) “cidrift”, el cual reconoce la existencia de tendencia lineal en el vector de cointegración. La selección de estos modelos, para el análisis de cointegración, se basa en el hecho que en la mayoría de los casos las variables exhiben tendencia lineal en sus niveles y en consecuencia se hace necesario dirimir entre los dos modelos mencionados, es decir, la selección de componentes determinísticas propuesta por Johansen (1994).

Una vez que se prueba la existencia de cointegración¹⁸, es decir, que el vector de cointegración tiene interpretación económica, la selección de las componentes determinísticas se realiza a través de las pruebas de exclusión, Johansen (1995)¹⁹. El siguiente paso se centra en la evaluación del comportamiento de los residuales. Dichos residuales se definen como adecuados si

¹⁸ Es de señalar que, en el análisis de cointegración se consideran variables “*dummies*” estacionales centradas, como también, de intervención. El uso de variables “*dummies*” de intervención puede verse en Hendry y Doornik (1994) y Arize (1999), entre otros.

¹⁹ En el caso de más de un vector de cointegración por lo menos uno debe tener interpretabilidad económica. Es decir, uno de ellos debe corresponder a una función de demanda por efectivo real.

son ruido blanco y normales multivariados, Lütkepohl (1993). Así, una vez definido como óptimo el comportamiento de los residuales, se llevan a cabo pruebas de exogeneidad débil, estacionariedad y exclusión del vector de cointegración. Es de señalar que, dado el propósito del ejercicio, se requiere que el logaritmo del efectivo real no sea exógeno débil y, adicionalmente, que ninguna de las variables del sistema esté excluida del vector de cointegración.

Los resultados obtenidos, a través del análisis anteriormente mencionado, permiten concluir que existe una relación lineal de largo plazo entre el efectivo real, el índice de producción industrial, la tasa de interés de los CDT a 90 días y la inflación anual del índice de precios al consumidor total nacional.

El cuadro 1 presenta la prueba de cointegración de la traza, ajustada por el tamaño de muestra de acuerdo a Cheung y Lai (1993)²⁰, la estimación del vector de cointegración normalizado, como también, el factor de ajuste. Como se observa, la prueba de la traza señala la existencia de un único vector de cointegración consistente con la teoría. El signo del coeficiente asociado a la desviación del efectivo real de su demanda de largo plazo, en la ecuación dinámica de corto plazo respectiva, es significativo y teóricamente coherente.

CUADRO 1

Sistema / Modelo Longitud del rezago	Prueba de cointegración		Vectores de cointegración $\beta' = [\beta_1 \ \beta_2 \ \beta_3 \ \beta_4 \ \beta_5]$	Velocidad de ajuste $\alpha = [\alpha_1 \ \alpha_2 \ \alpha_3 \ \alpha_4]$ “t” Student			
	Traza	V. Crítico o (90%)		DER_t	$DLIPI_t$	$DLTCDT_t$	$DLINF_t$
$\{LER_t, LIPI_t, LTCDT_t, LINF_t\}$							
Modelo: <i>Cidrift</i> Rezago: 5 Dummies de intervención estacionales	r=1 87.07 r=2 31.56 r=3 15.90 r=4 7.56	58.96 39.07 22.95 10.56	[1 -1.74 0.38 0.26 0.002]	-0.154 (-4.75)	-0.021 (-0.90)	-0.096 (-3.09)	-0.151 (-4.77)

²⁰ En la determinación del rango por el estadístico de la traza se tiene en cuenta que, de acuerdo a Cheung y Lai (1993), dicha prueba muestra más asimetría y exceso de curtosis que la de máximo valor propio, por lo cual se requiere una corrección por tamaño de muestra.

El cuadro 2 presenta las pruebas de exclusión, estacionariedad y exogeneidad débil realizadas sobre las variables que conforman el vector de cointegración. Como se observa, dichas pruebas muestran que ninguna de las variables está excluida de dicho vector y que todas son integradas de orden 1, I(1). Igualmente, existe evidencia para rechazar la hipótesis nula de exogeneidad débil del efectivo real. Es de señalar que, la no exclusión de la tendencia en el vector de cointegración ratifica la selección del modelo *Cidrift*.

CUADRO 2

Exclusión	Estacionariedad	Exogeneidad Débil
$\chi^2_{0.05}(1) = 3.84$	$\chi^2_{0.05}(4) = 9.49$	$\chi^2_{0.05}(1) = 3.84$
$LER_t : 19.92$	$LER_t : 53.52$	$LER_t : 16.38$
$LIP_t : 36.27$	$LIP_t : 45.75$	$LIP_t : 0.69$
$LTCDT_t : 20.82$	$LTCDT_t : 47.23$	$LTCDT_t : 8.14$
$LINF_t : 12.93$	$LINF_t : 51.09$	$LINF_t : 16.44$
$Trend_t : 6.72$		

Los P-values asociados a las pruebas multivariadas de diagnóstico sobre existencia de autocorrelación residual LM(1) , LM(4) y LM(24) son respectivamente 0.24, 0.15 y 0.01. Dichas pruebas permiten concluir comportamiento ruido blanco multivariado en los residuales. Así mismo, el p-value de 0.17 asociado a la prueba de normalidad multivariada de Doornik y Hansen (1994) evidencia un comportamiento normal multivariado en los residuales. Es decir, las pruebas multivariadas de diagnóstico de residuales, Lutkepohl (1993), permiten concluir que éstos presentan un comportamiento óptimo. Finalmente, las pruebas de estabilidad de Hansen y Johansen (1993) muestran, en general, que el vector de cointegración es estable durante los últimos cinco años, Anexo 1.

5.2 Modelos no lineales: redes neuronales artificiales y selección de la mejor red.

La modelación del efectivo real se lleva a cabo a través de dos redes neuronales artificiales que consideran información mensual²¹ para el período comprendido entre junio de 1981 y agosto de 2003. Donde el período entre junio de 1981 y diciembre de 2002 se establece como el período de entrenamiento y evaluación dentro de muestra, en tanto que, el período de evaluación fuera de muestra está conformado por las restantes observaciones.

²¹ Fin de período.

La primera red neuronal es de carácter autorregresivo, es decir, el conjunto X_t está conformado por los doce primeros rezagos de la primera diferencia del logaritmo del efectivo real reescalado²², $X_t = \{\Delta LEFR_{t-1}^*, \dots, \Delta LEFR_{t-12}^*\}$, en tanto que, la segunda o red ampliada, adiciona, a su propia historia, el rezago de orden diez del cambio reescalado del logaritmo de la inflación anual, $X_t = \{\Delta LEFR_{t-1}^*, \dots, \Delta LEFR_{t-12}^*, \Delta LINF_{t-10}^*\}$.

La determinación de las variables *inputs* de la componente lineal o determinación del conjunto X_t , en cada red, se lleva a cabo mediante la estrategia *stepwise*, propuesta por Swanson y White (1995,1997a). Así, se parte de una regresión lineal cuya variable dependiente es $\Delta LEFR_t^*$ y cuyas variables explicativas varían de acuerdo con el modelo considerado. En el primer caso, el conjunto de variables explicativas, o *inputs* de la componente lineal, considera los primeros doce rezagos de $\Delta LEFR_t^*$. En el segundo, se adiciona, a este conjunto de información, los primeros doce rezagos de $\Delta LINF_t$.

Una vez definido, en cada modelo, el conjunto de variables *input* de la componente lineal, X_t , se realiza el proceso de estimación de la red neuronal mediante el proceso de optimización numérica Quasi-Newton de Broyden, Fletcher, Goldfarb y Shano²³, ecuación (9), para las diferentes configuraciones del conjunto de información Z_t , $Z_t \subseteq X_t$, de la componente no lineal y para un número de unidades ocultas o *hidden units* Q que varían desde uno hasta cuatro²⁴.

²²Las variables son reescaladas en el intervalo (0,1). Así, la notación “*” se utiliza para referirse a la transformación de la diferencia del logaritmo de cada serie, es decir:

$$x_t^* = \frac{x_t - \bar{x}_t}{\sigma(x_t)}; \text{ siendo } x_t = \ln(x_t) - \ln(x_{t-1}).$$

²³ Como lo sugieren Franses and Dick V. Dijk (2000) y Rech (2002), este es uno de los algoritmos más utilizados en el contexto de redes neuronales para solucionar el problema de minimización, planteado en la ecuación (2).

²⁴ Como señala Franses y van Dijk (2000), la convergencia en el proceso de estimación no garantiza la obtención del mínimo global. Por consiguiente, se llevan a cabo múltiples estimaciones de cada una de las diferentes redes utilizando distintos valores iniciales del vector de parámetros Ψ . En particular, cada arquitectura se estima para 30 valores iniciales diferentes del vector de parámetros. Tales valores iniciales son obtenidos aleatoriamente a partir de una distribución uniforme entre [-2,2] y los parámetros del término de *weight decay* en la función objetivo $S(\Psi)$ fueron fijados de la siguiente manera: $r_\phi = 0.01$, $r_\beta = r_\gamma = 0.0001$.

La especificación del conjunto Z_t se lleva a cabo de la siguiente manera: en un primer paso, o $P = 1$, el conjunto Z_t incluye la primera variable del conjunto X_t , luego, en un segundo paso, o $P = 2$, se adiciona al conjunto Z_t la segunda variable de X_t de tal forma que en el último paso, $P = l + m$, se tiene la igualdad de los conjuntos, $Z_t = X_t$. Es de resaltar que, el conjunto X_t que conforma la componente lineal permanece invariante a través de las diferentes arquitecturas. El cuadro 3 presenta el esquema de conformación de los conjuntos X_t y Z_t en cada uno de los modelos de redes neuronales.

CUADRO 3

Modelo 1: Red neuronal autorregresiva	Modelo 2: Red neuronal ampliada
Esquema <i>step-wise</i> : $X_t = \{\Delta LEFR_{t-1}, \dots, \Delta LEFR_{t-12}\}$	Esquema <i>step-wise</i> : $X_t = \{\Delta LEFR_{t-1}, \dots, \Delta LEFR_{t-12}, \Delta LINF_{t-10}\}$
$P = 1: Z_t = \{\Delta LEFR_{t-1}\} \subseteq X_t$ $Q = 1,2,3,4$	$P = 1: Z_t = \{\Delta LEFR_{t-1}\} \subseteq X_t$ $Q = 1,2,3,4$
$P = 2: Z_t = \{\Delta LEFR_{t-1}, \Delta LEFR_{t-2}\} \subseteq X_t$ $Q = 1,2,3,4$	$P = 1: Z_t = \{\Delta LEFR_{t-1}\} \subseteq X_t$ $Q = 1,2,3,4$
\vdots	\vdots
$P = 12: Z_t = \{\Delta LEFR_{t-1}, \dots, \Delta LEFR_{t-12}\} = X_t$ $Q = 1,2,3,4$	$P = 12: Z_t = \{\Delta LEFR_{t-1}, \dots, \Delta LEFR_{t-12}\} \subseteq X_t$ $Q = 1,2,3,4$
	$P = 13: Z_t = \{\Delta LEFR_{t-1}, \dots, \Delta LEFR_{t-12}, \Delta LINF_{t-10}\} = X_t$ $Q = 1,2,3,4$

La selección de la estimación óptima, en cada arquitectura, se realiza considerando dos criterios: (i) menor valor de la función objetivo y (ii) vector de gradientes, asociado a los parámetros de la estimación, sin elementos superiores a 1×10^{-3} . Gráficos 5 y 6 presentan, para cada una de las redes neuronales estimadas, las funciones objetivo²⁵, alcanzadas en cada una de las arquitecturas, en el proceso de simulación de los valores iniciales del vector de parámetros Ψ .

²⁵ Dichas funciones objetivo han sido ordenadas de menor a mayor dentro de las 30 iteraciones del proceso de simulación, en cada una de las arquitecturas.

Gráfico 5
Red neuronal artificial autorregresiva

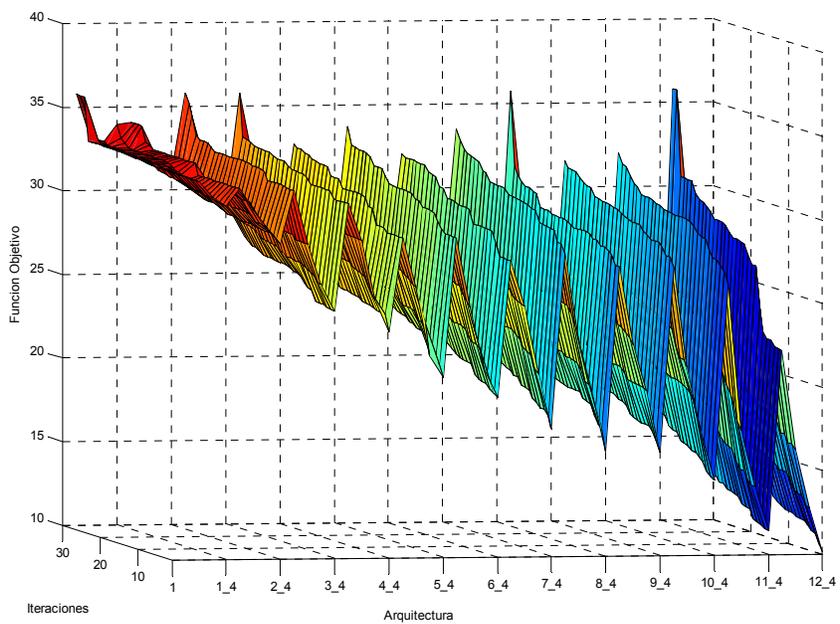
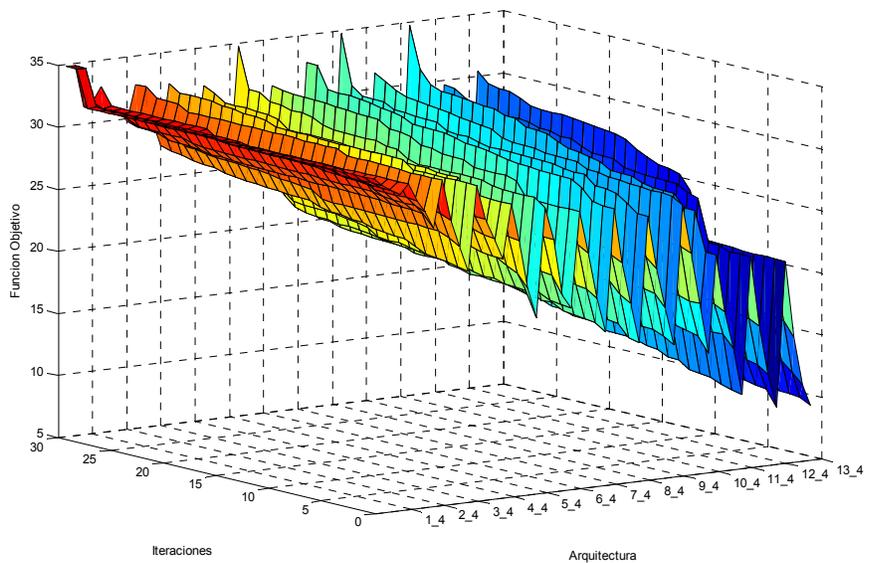


Gráfico 6
Red neuronal artificial ampliada



Una vez seleccionado en los dos modelos, para cada una de las diferentes arquitecturas, el vector de parámetros que produce la menor función objetivo y que cumple con la condición de gradientes, se calculan las medidas de evaluación²⁶ dentro y fuera de muestra, las cuales permiten elegir la mejor combinación en cada modelo. La posible multiplicidad de mínimos locales hace necesario el uso de estas medidas por cuanto la obtención de una función objetivo mínima, para una arquitectura en particular, no garantiza la mejor especificación respecto al pronóstico.

Los cuadros 4 y 5 presentan para el modelo 1 o red neuronal autorregresiva las medidas de evaluación de pronóstico dentro y fuera de muestra, respectivamente. Es de resaltar que, todas las medidas de evaluación de pronóstico presentadas en este documento se llevan a cabo sobre el nivel del efectivo nominal. Es decir, los pronósticos se llevan a cabo sobre $\Delta LEFR_t^*$ y se les aplica la transformación inversa para alcanzar los niveles del efectivo nominal y calcular las medidas de evaluación de pronóstico. En la conversión a nominal se utiliza el valor observado del IPC total nacional.

Del análisis del conjunto de medidas presentadas en dichos cuadros se puede concluir que el mejor desempeño dentro de muestra lo presenta la combinación $P = 8, Q = 4$, en tanto que, el mejor fuera de muestra se tiene con la combinación $P = 12, Q = 1$. Llevando a cabo una elección entre la disyuntiva de desempeño dentro y fuera de muestra, se selecciona como la mejor red autorregresiva a la red cuya arquitectura es $P = 11, Q = 4$. Los gráficos 7 y 8 presentan algunas de las medidas de ajuste de los pronósticos en términos relativos y absolutos, dentro y fuera de muestra, consignadas en los cuadros 4 y 5, respectivamente.

A su vez, en los cuadros 6 y 7 se reportan las medidas de desempeño de pronóstico dentro y fuera de muestra del modelo 2 o red neuronal ampliada, respectivamente. Dichas medidas permiten concluir que el mejor desempeño dentro de muestra se alcanza con la combinación $P = 13, Q = 3$ y fuera de muestra con $P = 3, Q = 4$. La evaluación de los pronósticos dentro y fuera de muestra conduce a la selección de la arquitectura $P = 10, Q = 3$ como la mejor red neuronal ampliada dentro del grupo estudiado. Los gráficos 9 y 10 presentan algunas de las medidas de evaluación dentro y fuera de muestra.

²⁶ Las medidas de evaluación consideradas son las siguientes: AIC, BIC, RMSE, RMSPE, MAE, MAPE y SR dentro de muestra y RMSE, RMSPE, MAE, MAPE, SR y U-THEIL fuera de muestra.

CUADRO 4

Red neuronal autorregresiva: Evaluación dentro de muestra

P	Q	AIC	BIC	RMSE	RMSPE	MAE	MAPE	SR
1	1	28.35	28.39	1414634.2	0.3775	724506.5	0.2572	75.58
1	2	28.10	28.18	1235732.9	0.3347	627752.5	0.2307	74.81
1	3	28.03	28.15	1176907.0	0.3189	607016.2	0.2193	74.81
1	4	28.07	28.24	1191295.3	0.3313	623893.9	0.2363	75.58
2	1	27.43	27.48	889044.0	0.2572	436728.0	0.1830	74.42
2	2	26.68	26.79	603057.6	0.2277	314896.6	0.1750	74.81
2	3	27.18	27.34	760387.7	0.2858	404655.5	0.2313	72.87
2	4	26.43	26.65	516307.3	0.2704	310523.6	0.2250	74.03
3	1	28.26	28.33	1345273.8	0.3448	679855.1	0.2362	74.81
3	2	27.01	27.15	704984.4	0.2969	382415.8	0.2429	74.03
3	3	26.79	26.99	618959.8	0.2687	328318.2	0.2216	77.13
3	4	27.33	27.60	794971.5	0.2272	404641.1	0.1702	77.91
4	1	27.16	27.24	770915.9	0.2411	384555.7	0.1798	75.58
4	2	26.96	27.13	683911.2	0.2415	329481.5	0.1775	77.13
4	3	26.09	26.34	432280.0	0.2877	306288.2	0.2488	77.52
4	4	25.91	26.24	385684.4	0.2711	264380.3	0.2320	77.91
5	1	27.33	27.42	836671.2	0.2866	464987.3	0.2339	76.36
5	2	26.21	26.40	464625.5	0.3197	339674.8	0.2730	77.91
5	3	26.39	26.68	496083.0	0.3016	359442.4	0.2624	79.07
5	4	27.26	27.64	744408.1	0.3121	434106.0	0.2623	81.40
6	1	27.20	27.31	783391.6	0.2906	445791.0	0.2417	74.42
6	2	27.52	27.74	887352.3	0.2528	443803.8	0.1857	79.07
6	3	27.22	27.55	742240.8	0.2734	431959.6	0.2335	81.78
6	4	26.74	27.18	565393.2	0.3557	376588.7	0.3083	84.50
7	1	27.18	27.31	771471.9	0.3295	460204.8	0.2793	76.74
7	2	26.86	27.11	633799.0	0.3198	394619.5	0.2706	77.52
7	3	26.39	26.76	483773.4	0.3420	353098.4	0.2819	81.40
7	4	27.03	27.52	643122.7	0.3082	396290.0	0.2586	82.56
8	1	27.55	27.69	923227.9	0.2641	437883.4	0.1831	79.46
8	2	25.68	25.96	349521.5	0.1573	203372.9	0.1271	76.74
8	3	26.01	26.42	396010.0	0.2525	278360.5	0.2092	81.40
8	4	25.44	25.99	286353.6	0.1176	158322.9	0.0940	84.50
9	1	26.80	26.96	634113.2	0.2986	362119.9	0.2460	76.74
9	2	26.87	27.18	628949.2	0.2589	352998.7	0.2178	79.07
9	3	27.34	27.79	761361.2	0.1984	415953.4	0.1454	82.56
9	4	27.00	27.60	614962.7	0.2816	376867.9	0.2458	87.60
10	1	26.94	27.10	674822.5	0.2999	376129.7	0.2465	77.13
10	2	26.38	26.71	488603.1	0.2002	262849.4	0.1608	80.62
10	3	25.85	26.35	357540.3	0.2045	238021.7	0.1733	84.11
10	4	26.37	27.03	442401.2	0.1733	266407.0	0.1338	83.72
11	1	27.03	27.21	705707.5	0.2418	354971.6	0.1784	75.58
11	2	28.32	28.68	1277264.6	0.3208	731460.7	0.2400	81.40
11	3	27.24	27.78	708042.4	0.2889	457297.9	0.2292	83.33
11	4	26.75	27.47	527360.9	0.1527	290103.1	0.1130	85.66
12	1	30.07	30.26	3203023.1	1.5173	2296049.2	1.2819	77.52
12	2	27.63	28.02	897728.9	0.3176	569787.6	0.2524	81.01
12	3	28.24	28.82	1153155.2	0.5871	805066.1	0.5272	85.27
12	4	27.32	28.09	689580.4	0.2280	435129.6	0.1838	85.27

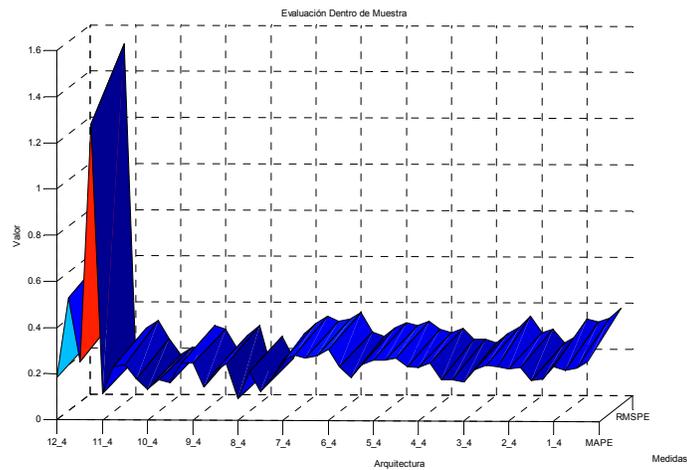
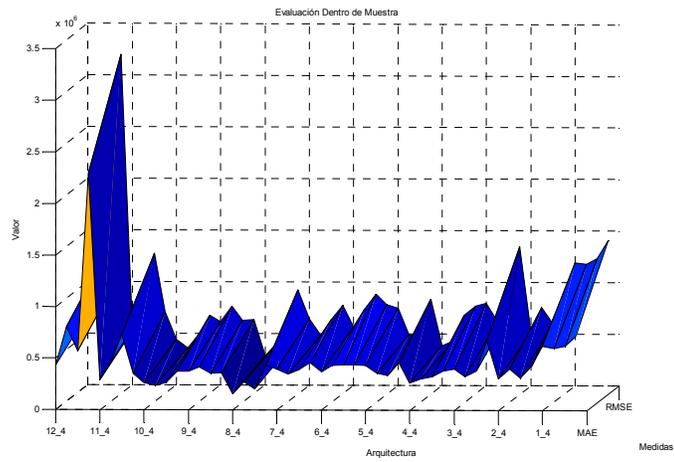
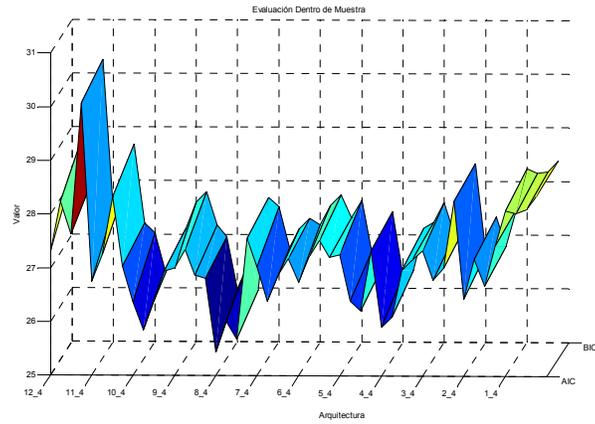
CUADRO 5
Red neuronal autorregresiva: Evaluación fuera de muestra

P	Q	RMSE	RMSPE	MAE	MAPE	UTHEIL
1	1	582843.5	0.0658	501229.7	0.0563	1.60
1	2	655676.8	0.0739	604085.5	0.0676	1.80
1	3	938248.5	0.1054	885059.3	0.0988	2.56
1	4	654035.3	0.0735	614841.8	0.0687	1.79
2	1	819799.1	0.0921	777714.9	0.0868	2.24
2	2	657468.9	0.0739	608176.3	0.0679	1.80
2	3	794632.1	0.0890	760434.8	0.0847	2.17
2	4	781756.5	0.0875	745252.9	0.0830	2.13
3	1	639520.4	0.0717	616880.8	0.0687	1.74
3	2	956749.6	0.1063	930287.4	0.1032	2.59
3	3	488645.0	0.0538	424832.7	0.0469	1.31
3	4	669010.8	0.0750	632937.3	0.0705	1.82
4	1	905948.4	0.1012	882601.3	0.0981	2.46
4	2	796531.4	0.0884	758720.1	0.0841	2.15
4	3	622588.6	0.0697	587449.0	0.0654	1.70
4	4	591241.6	0.0669	488982.3	0.0549	1.63
5	1	1020349.7	0.1137	996163.5	0.1106	2.77
5	2	1374276.7	0.1526	1345607.8	0.1492	3.71
5	3	1015540.3	0.1126	982830.1	0.1089	2.74
5	4	935613.4	0.1039	826552.0	0.0914	2.53
6	1	1074963.6	0.1196	1048303.0	0.1164	2.91
6	2	1252495.4	0.1389	1206405.3	0.1337	3.38
6	3	2073068.8	0.2287	1973106.8	0.2181	5.56
6	4	1254641.0	0.1389	1220815.9	0.1352	3.38
7	1	1122559.9	0.1253	1075869.9	0.1196	3.05
7	2	926801.4	0.1034	869767.5	0.0967	2.52
7	3	1159594.6	0.1287	1124814.9	0.1247	3.13
7	4	1532118.0	0.1689	1409943.3	0.1557	4.11
8	1	1091910.2	0.1218	1056987.6	0.1175	2.96
8	2	872840.0	0.0963	749914.4	0.0827	2.34
8	3	1860783.9	0.2062	1788732.7	0.1982	5.02
8	4	1345621.9	0.1488	1283909.1	0.1421	3.62
9	1	1048503.5	0.1171	1023083.5	0.1138	2.85
9	2	781115.2	0.0876	748922.1	0.0835	2.13
9	3	1086163.1	0.1204	1034352.2	0.1146	2.93
9	4	1640534.5	0.1815	1554243.8	0.1721	4.42
10	1	991682.5	0.1109	965254.5	0.1074	2.70
10	2	964111.3	0.1050	852287.9	0.0936	2.56
10	3	1744732.5	0.1940	1664280.5	0.1847	4.72
10	4	960302.0	0.1039	779079.5	0.0850	2.53
11	1	993009.6	0.1099	944593.4	0.1045	2.67
11	2	396388.2	0.0445	368174.3	0.0411	1.08
11	3	327211.0	0.0368	263104.3	0.0295	0.90
11	4	337475.2	0.0371	246822.1	0.0273	0.90
12	1	299219.4	0.0332	267699.2	0.0297	0.81
12	2	732484.0	0.0822	653859.1	0.0731	2.00
12	3	1239934.2	0.1379	1178596.1	0.1308	3.35
12	4	612236.4	0.0680	577235.9	0.0641	1.65

Gráfico 7

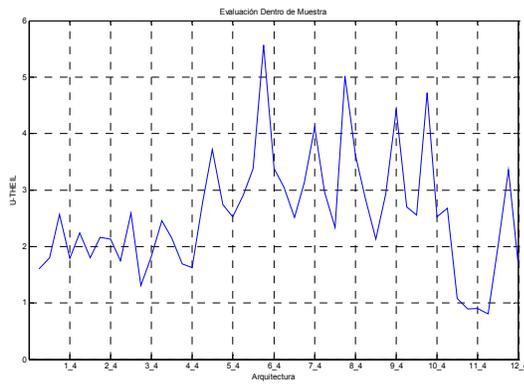
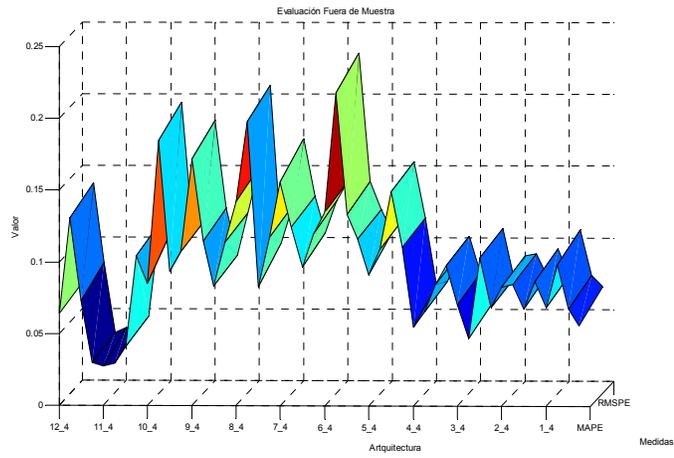
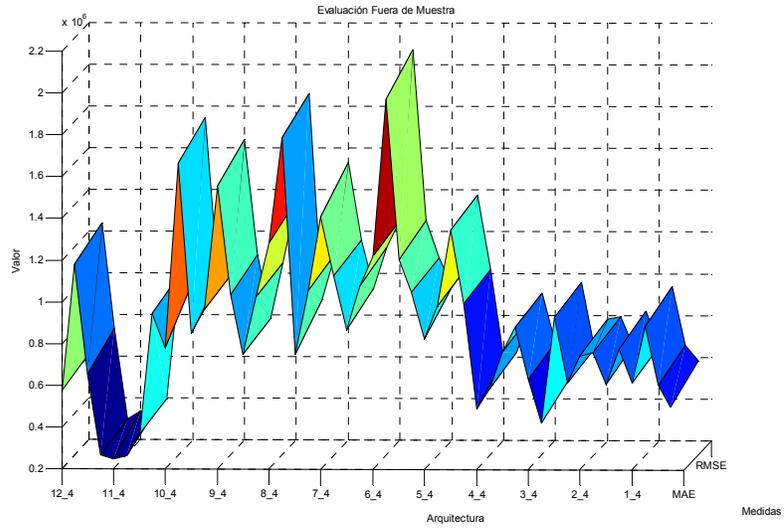
Medidas de evaluación dentro de muestra

Red neuronal artificial autorregresiva *



* Note que el ordenamiento en el eje horizontal cambia en algunas gráficas.

Gráfico 8
Medidas de evaluación fuera de muestra
Red neuronal artificial autorregresiva*



* Note que el ordenamiento en el eje horizontal cambia en algunas gráficas.

CUADRO 6 : Red neuronal ampliada: Evaluación dentro de muestra

P	Q	AIC	BIC	RMSE	RMSPE	MAE	MAPE	SR
1	1	27.67	27.71	1008516.8	0.2796	506423.5	0.1945	75.97
1	2	27.74	27.83	1033711.7	0.2807	520373.6	0.1929	77.13
1	3	27.66	27.78	978426.0	0.2709	496177.1	0.1907	76.36
1	4	27.63	27.79	953811.5	0.2642	485609.8	0.1857	74.81
2	1	26.99	27.05	714772.9	0.2322	371671.2	0.1761	74.42
2	2	26.60	26.71	578199.7	0.2258	323000.4	0.1818	75.58
2	3	26.11	26.27	446200.3	0.2683	298049.4	0.2233	73.64
2	4	26.43	26.65	515827.5	0.2539	311898.2	0.2089	75.19
3	1	27.04	27.11	729121.7	0.2294	376549.5	0.1732	74.42
3	2	27.24	27.38	790536.1	0.2229	411965.5	0.1670	77.13
3	3	26.90	27.10	653829.0	0.2089	357029.4	0.1659	78.29
3	4	26.39	26.66	497560.7	0.1853	260754.4	0.1461	78.29
4	1	26.78	26.86	637821.5	0.2279	341302.5	0.1811	76.74
4	2	26.49	26.66	539775.0	0.3121	348498.6	0.2624	75.58
4	3	26.49	26.73	526595.7	0.3218	338297.0	0.2744	79.07
4	4	26.19	26.52	444411.6	0.3096	306420.6	0.2561	78.68
5	1	27.16	27.26	770905.9	0.2704	441524.1	0.2228	77.52
5	2	26.01	26.20	420586.5	0.2816	286003.3	0.2335	80.23
5	3	26.16	26.45	442336.1	0.2845	318176.4	0.2405	81.01
5	4	26.47	26.85	501629.2	0.2550	300621.1	0.2163	78.29
6	1	27.67	27.78	987311.6	0.2754	501542.2	0.1986	79.84
6	2	27.34	27.56	812413.0	0.2581	422567.2	0.1963	78.68
6	3	27.06	27.39	685037.8	0.3560	452057.9	0.3074	82.56
6	4	26.92	27.36	618605.9	0.3402	403167.6	0.2881	82.17
7	1	27.01	27.14	708752.7	0.3120	430707.3	0.2637	77.91
7	2	27.45	27.70	851966.8	0.2637	448943.8	0.2022	82.56
7	3	26.08	26.45	414816.3	0.2451	274055.2	0.2001	80.62
7	4	26.15	26.65	415811.3	0.2735	300954.0	0.2280	86.05
8	1	26.97	27.11	692856.2	0.3219	426985.1	0.2718	78.29
8	2	27.54	27.81	884605.1	0.2936	514578.9	0.2478	84.50
8	3	26.13	26.54	421102.9	0.2136	250962.0	0.1784	80.23
8	4	27.57	28.12	829573.6	0.2062	438119.7	0.1469	83.72
9	1	26.55	26.70	558310.3	0.2713	333016.1	0.2233	76.36
9	2	27.47	27.78	849530.6	0.2962	478342.1	0.2498	80.62
9	3	27.00	27.45	642143.0	0.2169	331696.7	0.1799	81.40
9	4	26.74	27.34	540267.9	0.1509	254557.1	0.1149	84.88
10	1	26.84	27.00	641474.0	0.2974	370307.3	0.2451	77.13
10	2	26.14	26.47	432054.6	0.3090	310676.1	0.2540	79.46
10	3	25.77	26.26	343149.6	0.1461	206338.8	0.1169	82.17
10	4	25.65	26.31	309213.3	0.1921	208200.2	0.1567	85.66
11	1	29.66	29.84	2618497.8	1.2731	1900878.6	1.0853	78.68
11	2	29.19	29.55	1974702.0	0.8412	1363363.9	0.7123	76.74
11	3	27.50	28.03	804166.4	0.3399	490631.2	0.2692	85.27
11	4	26.45	27.16	452418.9	0.2082	296027.5	0.1631	89.15
12	1	29.62	29.81	2559893.7	1.2099	1838522.6	1.0237	77.91
12	2	27.60	27.99	885591.7	0.2690	521173.1	0.2128	82.95
12	3	28.15	28.73	1103338.6	0.4162	750920.9	0.3405	86.82
12	4	26.56	27.33	471328.5	0.1662	299814.0	0.1352	85.66
13	1	30.17	30.37	3355780.5	1.5685	2392792.4	1.3421	78.29
13	2	28.15	28.56	1153596.6	0.5966	809137.8	0.4921	81.78
13	3	24.86	25.48	210221.3	0.0939	128672.2	0.0775	86.05
13	4	25.80	26.62	317067.8	0.1455	212794.9	0.1209	89.92

CUADRO 7: Red neuronal ampliada: Evaluación fuera de muestra

P	Q	RMSE	RMSPE	MAE	MAPE	SR	UTHEIL
1	1	341747.2	0.0387	268647.9	0.0301	57.14	0.942
1	2	481276.2	0.0545	377166.9	0.0423	57.14	1.325
1	3	415804.9	0.0471	313416.9	0.0352	57.14	1.146
1	4	546057.0	0.0616	470732.5	0.0526	57.14	1.498
2	1	502872.1	0.0567	436617.6	0.0488	42.86	1.380
2	2	639003.8	0.0719	579295.7	0.0647	28.57	1.750
2	3	509351.8	0.0573	461887.8	0.0515	42.86	1.393
2	4	538718.3	0.0608	413201.6	0.0464	57.14	1.480
3	1	492164.8	0.0555	424453.6	0.0474	42.86	1.351
3	2	377317.2	0.0418	352691.9	0.0390	71.43	1.017
3	3	394989.1	0.0438	371827.9	0.0411	85.71	1.065
3	4	246020.4	0.0278	182625.5	0.0205	42.86	0.677
4	1	622152.9	0.0694	575426.2	0.0639	42.86	1.687
4	2	563238.5	0.0627	501329.1	0.0556	42.86	1.524
4	3	519839.7	0.0574	464710.9	0.0514	57.14	1.397
4	4	581754.6	0.0638	519587.3	0.0572	57.14	1.552
5	1	736732.2	0.0820	695143.7	0.0771	42.86	1.994
5	2	791829.8	0.0877	746960.5	0.0827	57.14	2.133
5	3	819910.1	0.0925	704928.5	0.0788	57.14	2.251
5	4	695675.6	0.0777	645257.1	0.0717	57.14	1.890
6	1	645981.4	0.0719	607319.0	0.0674	42.86	1.749
6	2	596094.6	0.0666	521956.9	0.0580	42.86	1.620
6	3	1057150.4	0.1165	1003558.6	0.1109	42.86	2.835
6	4	1177886.6	0.1298	1127897.3	0.1246	42.86	3.158
7	1	883496.0	0.0979	832447.5	0.0922	57.14	2.381
7	2	788856.6	0.0873	727412.0	0.0805	57.14	2.124
7	3	880893.2	0.0976	830563.9	0.0920	42.86	2.375
7	4	1220186.5	0.1356	1151882.9	0.1278	14.29	3.299
8	1	840646.2	0.0932	793113.3	0.0878	57.14	2.267
8	2	1012758.2	0.1107	906564.0	0.0996	57.14	2.693
8	3	2808127.6	0.3103	2704302.1	0.2992	28.57	7.549
8	4	1311181.8	0.1459	1190564.6	0.1320	57.14	3.549
9	1	645262.9	0.0718	586163.3	0.0650	42.86	1.747
9	2	588275.7	0.0661	425590.3	0.0477	85.71	1.607
9	3	1512789.3	0.1653	1302892.1	0.1433	42.86	4.022
9	4	588403.9	0.0656	520440.8	0.0579	57.14	1.596
10	1	476033.0	0.0527	394346.6	0.0436	57.14	1.283
10	2	1153236.9	0.1263	1013331.1	0.1116	28.57	3.074
10	3	246851.6	0.0273	190304.8	0.0211	42.86	0.663
10	4	420882.1	0.0461	276342.6	0.0305	71.43	1.121
11	1	330195.3	0.0359	266765.3	0.0293	71.43	0.873
11	2	286258.1	0.0322	224188.2	0.0251	57.14	0.784
11	3	452072.0	0.0499	363878.3	0.0402	42.86	1.215
11	4	1069580.0	0.1175	965846.4	0.1065	28.57	2.858
12	2	319929.0	0.0358	296473.7	0.0330	57.14	0.870
12	2	319929.0	0.0358	296473.7	0.0330	57.14	0.870
12	3	302504.6	0.0332	258372.9	0.0285	42.86	0.807
12	4	504301.8	0.0558	424859.3	0.0470	28.57	1.358
13	1	322532.8	0.0350	263165.6	0.0288	71.43	0.851
13	2	472199.9	0.0526	413463.1	0.0460	57.14	1.281
13	3	413974.8	0.0461	381294.1	0.0424	57.14	1.122
13	4	432762.9	0.0468	326362.4	0.0357	71.43	1.139

Gráfico 9
Medidas de evaluación dentro de muestra
Red neuronal artificial ampliada

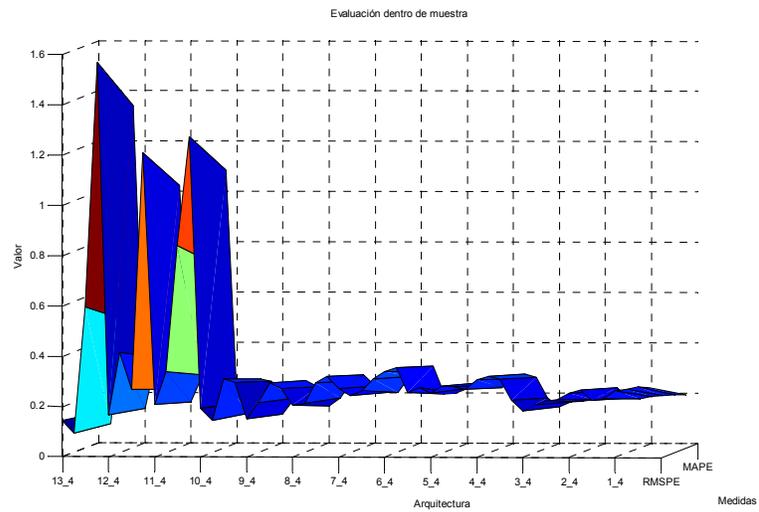
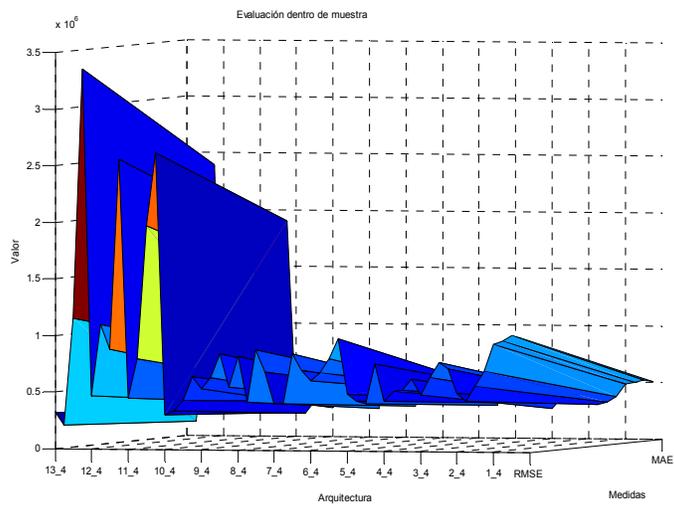
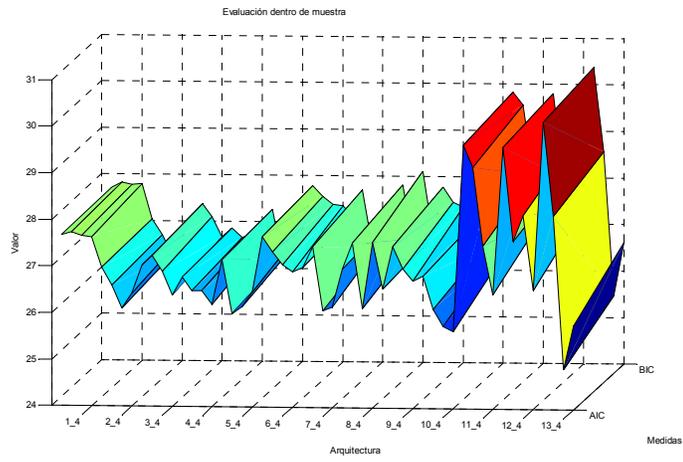
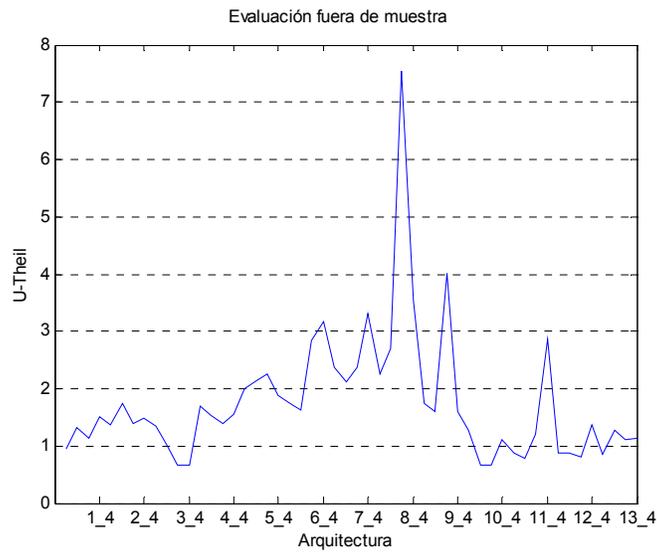
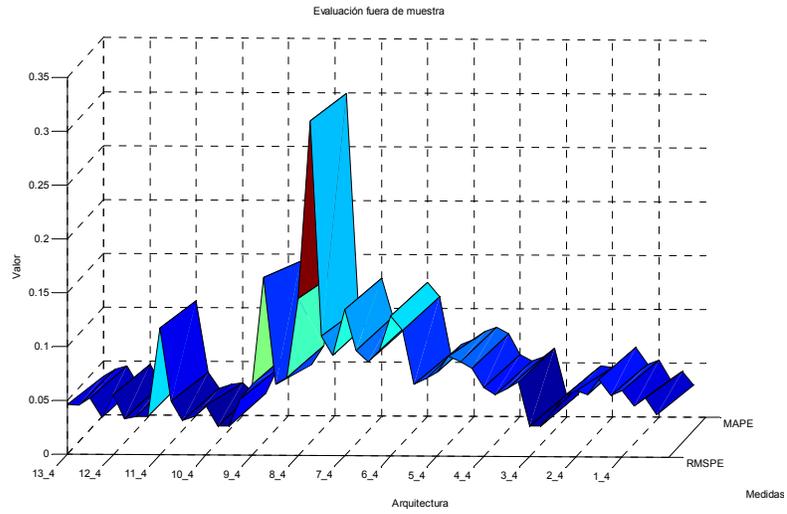
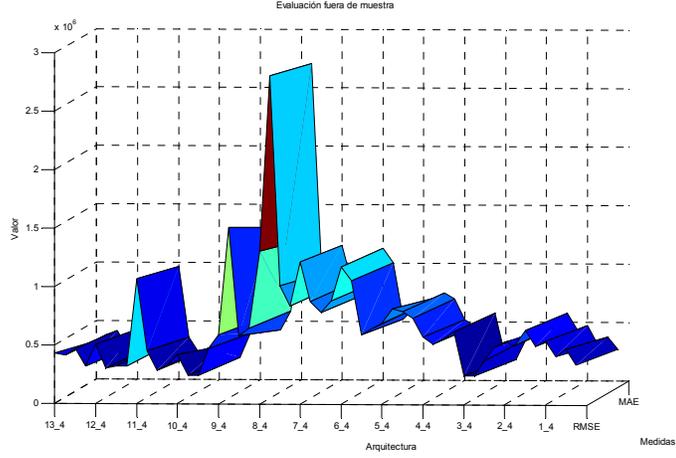


Gráfico 10
Medidas de evaluación fuera de muestra
Red neuronal artificial ampliada



Las redes neuronales seleccionadas pueden representarse a través de la siguiente especificación, en la cual se recuerda para comenzar la forma general de una red (ecuación 9) y posteriormente se presenta cada uno de los casos seleccionados:

Red neuronal artificial

$$y_t = \Phi_0 + X_t' \Phi + \sum_{j=1}^Q \beta_j G(Z_t' \gamma_j) + \varepsilon_t \quad ; \quad G \text{ función logística} \quad (9)$$

- Red neuronal artificial autorregresiva

$$X_t = \{\Delta LEFR_{t-1}, \dots, \Delta LEFR_{t-12}\}; \quad Z_t = \{\Delta LEFR_{t-1}, \dots, \Delta LEFR_{t-11}\}, \quad Q = 3$$

- Red neuronal artificial ampliada

$$X_t = \{\Delta LEFR_{t-1}, \dots, \Delta LEFR_{t-12}, \Delta LINF_{t-10}\}; \quad Z_t = \{\Delta LEFR_{t-1}, \dots, \Delta LEFR_{t-11}\}; \quad Q = 4$$

6. Evaluación de pronósticos

El propósito de esta sección es presentar la evaluación de los pronósticos fuera de muestra de los modelos de redes neuronales con el fin de seleccionar con base en este criterio el mejor de ellos. Se realiza también un análisis de sensibilidad para el modelo de red neuronal ampliada con el fin de determinar la importancia de los *inputs* incorporados en este modelo. Finalmente, se compara el desempeño de pronóstico de los modelos lineales frente a los modelos no lineales.

6.1 Evaluación de pronósticos de los modelos de redes neuronales

La evaluación de los pronósticos fuera de muestra de los dos modelos de redes neuronales, cuadro 8, permite seleccionar a la red neuronal ampliada como el modelo con mejor desempeño de pronóstico dentro de los modelos no lineales.

CUADRO 8

Desempeño fuera de muestra de los modelos no lineales

Medidas	Red Neuronal Autorregresiva	Red Neuronal Ampliada
RMSE	337461.1	246851.6
MAE	246800.0	190304.8
RMSPE	0.0371	0.0272
MAPE	0.0273	0.0211
U-THEIL	0.9028	0.6634

La estimación de un modelo de redes neuronales artificiales exige la existencia de una relación no lineal entre la variable dependiente y las variables que conforman el *input*, en este caso efectivo real e inflación. El cuadro 9 presenta los resultados obtenidos a partir del test de Lee et al (1993), basado en la red neuronal ampliada seleccionada.

Este test, tipo pormanteau, plantea linealidad bajo la hipótesis nula, en tanto que, bajo la alterna, sugiere no linealidad. Los resultados permiten concluir que a un nivel de significancia $\alpha = 5\%$ existe evidencia para rechazar la hipótesis nula de linealidad entre las variables en estudio.

CUADRO 9
Detección de No-linealidad
Test de Lee, White y Granger

Replicaciones	Muestra	Modelo Seleccionado	P-VALUE Test Lee, White y Granger
Replicaciones: 500 Tamaño: 50	Junio 1981: Dic 2002	Red Neuronal Ampliada $P = 10, Q = 3$	Cota de Bonferroni mejorado: 0.00288

Una vez seleccionada la red neuronal y probada la existencia de no linealidad, con base en el modelo seleccionado, se lleva a cabo un análisis de sensibilidad que permita determinar la relevancia de los distintos *inputs* que conforman tanto la parte lineal como la no lineal.

Al desconectar los diferentes *inputs*, las medidas de evaluación de pronóstico dentro de muestra, cuadro 10, permiten concluir que los rezagos más relevantes en la parte no lineal son el 1, el 3 y el 7, su desconexión implica un aumento considerable en todas las medidas²⁷, frente a las obtenidas al considerar el modelo de conexión completa. Es de señalar que, la desconexión de ninguna de las variables del conjunto Z_t produce disminución en las medidas de evaluación frente a las iniciales, hecho que justifica la consideración de $P = 10$.

²⁷ Excepto la medida SR la cual empeora al disminuir.

CUADRO 10

Medidas sobre desempeño dentro de muestra: considerando desconexión

Variable Desconectada Lineal	Variable Desconectada No - Lineal	AIC	BIC	RMSE	RMSPE	MAE	MAPE	SR
Red neuronal aumentada : Modelo seleccionado (completamente conectado)								
0	0	25.77	26.26	343149.6	0.1461	206338.8	0.1169	82.17
Desconexión No-Lineal								
0	1	36.82	37.32	86172354.9	14.2572	42327776.2	9.2867	70.54
0	2	32.49	32.99	9884343.0	2.0111	5605614.8	1.5674	75.58
0	3	36.63	37.13	78484051.7	13.6744	40000170.4	8.9171	73.64
0	4	33.00	33.49	12725993.4	2.5901	7173309.4	2.0839	77.52
0	5	26.36	26.86	461887.0	0.2338	290535.1	0.1941	75.97
0	6	27.29	27.79	734263.2	0.2307	415846.9	0.2033	81.01
0	7	36.42	36.91	70367898.7	11.3934	33929604.3	7.0663	70.54
0	8	29.56	30.05	2281089.5	0.6087	1419005.0	0.5554	77.91
0	9	29.58	30.08	2307086.6	0.6450	1449885.5	0.6160	75.19
0	10	28.11	28.60	1105213.4	0.2991	643587.0	0.2697	79.46
Desconexión Lineal								
1	0	26.44	26.93	479295.3	0.1745	287580.7	0.1466	76.36
2	0	25.66	26.15	324370.6	0.1402	202683.7	0.1153	72.09
3	0	25.90	26.39	365419.2	0.1490	221825.0	0.1218	78.29
4	0	26.03	26.52	390255.0	0.1551	233740.1	0.1271	73.26
5	0	25.96	26.45	376563.9	0.1558	229430.9	0.1276	82.17
6	0	26.25	26.75	436724.0	0.1664	235255.4	0.1271	74.42
7	0	25.68	26.18	328192.0	0.1415	204718.5	0.1163	75.19
8	0	25.68	26.17	327799.6	0.1422	201864.8	0.1154	80.62
9	0	25.74	26.24	338889.6	0.1416	208880.8	0.1170	74.81
10	0	25.86	26.36	359290.2	0.1447	218273.0	0.1187	75.58
11	0	26.50	27.00	495073.0	0.1690	287925.1	0.1398	80.23
12	0	26.82	27.31	579410.3	0.1990	311545.8	0.1510	77.13
13	0	26.21	26.70	427673.5	0.1642	237796.5	0.1261	80.62
Desconexión Lineal y No- Lineal								
1	1	36.55	37.04	75067170.1	11.9854	35810405.9	7.7289	67.44
2	2	32.31	32.81	9049049.4	1.8080	5078981.3	1.4176	69.38
3	3	36.45	36.95	71719490.4	12.2895	36164181.3	8.0085	75.19
4	4	32.74	33.24	11219262.4	2.2420	6251118.3	1.8043	70.93
5	5	25.90	26.39	365999.2	0.2487	251600.6	0.2032	78.29
6	6	27.24	27.74	716496.3	0.2273	418677.9	0.1960	79.46
7	7	36.33	36.83	67424958.2	10.8496	32331427.4	6.7137	69.77
8	8	29.57	30.06	2293389.4	0.6120	1427769.5	0.5586	76.74
9	9	29.62	30.11	2347789.3	0.6561	1478372.6	0.6258	71.32
10	10	28.33	28.82	1232399.7	0.3301	735128.2	0.3001	73.64

La desconexión de las variables que conforman la componente lineal, o conjunto X_t , no produce cambios tan dramáticos como los observados bajo desconexión de inputs no lineales. En algunos casos, como en la desconexión del rezago de orden dos, se presenta una leve disminución en las medidas frente a las iniciales. Finalmente, al desconectar las variables comunes en las dos componentes, los resultados obtenidos son similares a los encontrados bajo desconexión no lineal.

El cuadro 11 presenta las medidas de evaluación de pronóstico bajo desconexión de los diferentes *inputs*. Lo importante de resaltar, en este caso, es que la desconexión en la componente lineal del rezago de orden siete produce una disminución considerable en las medidas, lo cual sugiere mantener la arquitectura seleccionada desconectando dicho rezago de la componente lineal.

La desconexión en la parte lineal del rezago de orden diez del cambio reescalado del logaritmo de la inflación anual (cuadro 10, variable lineal 13) corrobora la importancia de la red neuronal ampliada sobre la autorregresiva en la medida que su exclusión empeora los criterios de selección tanto dentro como fuera de muestra.

CUADRO 11

Medidas sobre desempeño fuera de muestra: considerando desconexión

Variable Desconectada Lineal	Variable Desconectada No - Lineal	RMSE	RMSPE	MAE	MAPE	UTHEIL
Red neuronal aumentada : Modelo seleccionado (completamente conectado)						
0	0	246851.6	0.0273	190304.8	0.0211	0.6634
Desconexión No-Lineal						
0	1	635483.8	0.0716	539184.1	0.0604	1.7429
0	2	797135.8	0.0875	647405.6	0.0713	2.1286
0	3	544364.6	0.0605	435078.5	0.0483	1.4714
0	4	775626.9	0.0851	510369.8	0.0562	2.0699
0	5	374940.1	0.0411	257157.2	0.0283	0.9990
0	6	451348.9	0.0493	347765.3	0.0381	1.1982
0	7	961263.5	0.1045	597263.7	0.0652	2.5417
0	8	321569.2	0.0353	235531.7	0.0259	0.8586
0	9	252451.9	0.0278	185632.3	0.0205	0.6758
0	10	248655.6	0.0275	193064.2	0.0214	0.6680
Desconexión Lineal						
1	0	1317811.6	0.1445	1196066.1	0.1317	3.5153
2	0	548580.9	0.0598	382823.5	0.0420	1.4553
3	0	263500.6	0.0299	193598.2	0.0217	0.7276
4	0	179679.6	0.0195	141849.9	0.0156	0.4734
5	0	288052.3	0.0317	203914.5	0.0225	0.7707
6	0	480474.3	0.0526	317551.7	0.0350	1.2787
7	0	176041.5	0.0196	127259.1	0.0142	0.4780
8	0	249778.8	0.0276	181271.1	0.0201	0.6704
9	0	367313.1	0.0402	288456.2	0.0317	0.9780
10	0	259074.0	0.0285	207568.5	0.0230	0.6940
9	0	367313.1	0.0402	288456.2	0.0317	0.9780
10	0	259074.0	0.0285	207568.5	0.0230	0.6940
11	0	327367.6	0.0359	275557.2	0.0303	0.8730
12	0	498728.5	0.0558	439922.5	0.0491	1.3584
13	0	330182.6	0.0364	251004.9	0.0278	0.8848
Desconexión No-Lineal y Lineal						
1	1	2426162.9	0.2668	2294779.0	0.2531	6.4898
2	2	382591.6	0.0422	293772.4	0.0326	1.0273
3	3	1252478.8	0.1377	1058249.7	0.1166	3.3497
4	4	970515.1	0.1055	598441.6	0.0653	2.5654
5	5	432975.6	0.0473	291767.9	0.0320	1.1508
6	6	401895.8	0.0439	330937.9	0.0363	1.0692
7	7	1010269.6	0.1095	598609.3	0.0652	2.6645
8	8	392886.6	0.0430	291016.5	0.0320	1.0467
9	9	349933.9	0.0382	280963.1	0.0308	0.9299
10	10	257471.2	0.0284	207501.5	0.0230	0.6897

6.2 Comparación entre modelos lineales y no lineales

Una vez estimados los modelos lineales y la red neuronal artificial se comparan sus desempeños en el pronóstico fuera de muestra, cuadro 12. Las redes neuronales, conectada y parcialmente conectada, alcanzan el mejor desempeño de pronóstico fuera de muestra, dentro del conjunto de modelos estudiado.

CUADRO 12
Desempeño de pronósticos fuera de muestra

Medidas	Red neuronal Ampliada	Red neuronal Ampliada con desconexión	ARIMA	ARIMA con Transferencia	VEC
RMSE	246851.6	176041.5	287908.7	370148.1	481327.7
MAE	190304.8	127259.1	250709.7	332697.5	431648.3
RMSPE	0.0273	0.0196	0.0321	0.0412	0.0525
MAPE	0.0210	0.0141	0.0279	0.0369	0.0474
U-THEIL	0.6634	0.4780	0.7813	1.0031	1.2778

Adicionalmente, se lleva a cabo el test de *encompassing* de Fair y Shiller (1990), entre un modelo no lineal y uno lineal, seleccionando para ello a la red neuronal artificial ampliada de conexión completa y al modelo ARIMA. Dicha prueba se lleva a cabo sobre el período comprendido entre junio de 1981 y diciembre de 2002 o período de entrenamiento²⁸.

La significancia estadística de los coeficientes asociados a las dos series de pronóstico, no lineal (f^{NL}) y lineal (f^L), cuadro 13, permiten concluir que los dos modelos contienen información independiente y relevante sobre la variable bajo estudio.

En resumen, tanto la prueba de no linealidad de Lee et al. como la prueba de *encompassing* de Fair y Shiller coinciden en mostrar la existencia de un comportamiento no lineal en el efectivo y soportan una especificación no lineal. Igualmente, el desempeño de los pronósticos fuera de muestra justifica tal especificación.

²⁸ Esta prueba no se lleva a cabo para el período definido como fuera de muestra por estar basada en análisis de regresión, el cual requiere de un número mayor de observaciones.

CUADRO 13

Test de *encompassing* de Fair y Shiller

Dependent Variable: DY				
Method: Least Squares				
Simple(adjusted): Jul 1981 : Dic 2002				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.00146	0.002307	0.636061	0.52532
f^{NL}	0.04698	0.014761	3.182924	0.00164
f^L	0.89427	0.021825	40.9744	0.00000
R-squared	0.92483405			
Adjusted R-squared	0.92392844			
Durbin-Watson stat	1.99596646			
	Test	P-value		
JB	5.18	0.074		
LB(36)	22.45	0.962		
WHITE Cross*	10.24	0.1145		
ARCH(1)*	0.87	0.348		

* Test tipo TR^2 ; $DY = \text{Log}(\text{Efectivo nominal})_t - \text{Log}(\text{Efectivo nominal})_{t-1}$

7. Conclusiones

Existe soporte teórico y evidencia empírica cada vez mayor sobre la existencia de comportamientos no lineales en algunas series macroeconómicas. Hecho que lleva a pensar que los modelos lineales no resultan apropiados para proyectar este tipo de variables. Este documento se ha centrado en el examen empírico de una de ellas: el efectivo real. Se han estimado para la variable dos modelos lineales univariados (ARIMA) y un modelo multivariado (VEC), este último con un mayor contenido teórico, y se ha comparado su desempeño en pronóstico vs. una red neuronal autorregresiva ampliada con la variable inflación, la mejor de las diferentes redes consideradas.

Con el fin de reducir la subjetividad, principal defecto que se ha atribuido a las redes neuronales, la red seleccionada se somete a diversas pruebas. En primer lugar, se realiza un test con el fin de probar la existencia de una relación no lineal entre la variable dependiente y las variables que conforman el *input*, en este caso efectivo real e inflación. Se examina, de otro lado,

la sensibilidad del modelo de red neuronal ampliado con el fin de determinar la relevancia de los distintos *inputs* que conforman tanto la parte lineal como la no lineal. En este caso se muestra que la desconexión de las variables que conforman el componente lineal de la red no produce cambios tan dramáticos como los observados bajo la desconexión de inputs no lineales. Si se desconectan las variables comunes en las dos componentes (lineal y no lineal), los resultados obtenidos son similares a los encontrados bajo desconexión no lineal.

Una vez la red neuronal ampliada ha sido sometida a estos procedimientos, se compara su desempeño en pronóstico con los modelos lineales. De esta comparación surge el resultado central de este trabajo: la superioridad manifiesta que en el pronóstico tiene la red neuronal ampliada sobre todos los modelos lineales considerados.

Bibliografía

Arango y Gonzalez (2000), "A nonlinear specification of demand for cash in Colombia" *Money Affaire*, Vol. XIII, No. 2.

Arize, A. (1999), "The Demand for LDC Exports: Estimates from Singapore", *The International Trade Journal*, Vol. XIII, No.4.

Arrau y Gregorio (1994) "Financial Innovation and Money Demand: Application to Chile and Mexico" *Review of Economics and Statistics*, Vol. 75.

Azoff, E. M. (1996), *Neural Network. Time Series Forecasting of Financial Markets*, Wiley, A Wiley Finance Edition.

Barnett, Fisher y Serletis(1992)"Consumer Theory and the Demand for Money," *Journal of Economic Literature*, Vol. 30 (4)

Baumol J.(1952) "The transaction demand for cash: An inventory-theoretic approach" *Quarterly Journal of Economics*, 66, 545-556.

Buitre y Armstrong (1978)), "A Didactic Note on the Transactions Demand for Money and Behavior Towards Risk" *Journal of Money, Credit, and Banking*, Vol. 10, no. 4.

Caballero (1994) "On the dynamics of aggregate investment" in A. Solimano and L. Serven (eds.), *Striving for Growth after Adjustment*, The World Bank.

Caballero (1992) "A fallacy of Composition" *American Economic Review*, Vol. 82.

Cagan, P (1956), "The Money Dynamics of Hyperinflation" En: Friedman, M n(ed.) *Studies in the Quantity Theory of Money*, University of Chicago Press, Chicago.

Cheung, Y. W. and K. S. Lai (1993), "Finite-Sample Sizes of Johansen's Likelihood Ratio Tests for Cointegration", *Oxford Bulletin of Economics and Statistics*, No. 55 Vol.3.

Dowd K. (1990) "Monetary policy in the 21st. century: An impossible task" *Cato Journal*,17

Draisma, G., J. Kaashoek and H. K. van Dijk (1995), "A neural network applied to economic time series", *Econometric Institute, Erasmus University, Rotterdam and Tinbergen Institute Rotterdam*. S.F.

Fair R.C. and R. Shiller (1990), "Comparing Information in Forecast from Econometric Models", *The American Economic Review*, Vol. 80, No. 3.

Faraway J. and C. Chatfield (1995), "Time Series Forecasting with Neural Networks: A Case of Study", *Research Report No. 95-06 of the Statistics Group, University of Bath*.

Franses P.H. and D. van Dijk (2000), *Non-linear time series models in empirical finance*, Cambridge University Press.

Friedman B. (1999) "The future of monetary policy: The central bank as an army with only a signal corps?" *International Finance*, 2, 231-338.

Gertler (1988) "Financial Structure and Aggregate Economic Activity: An Overview" *Journal of Money, Credit and Banking*, Vol. 20.

Goldfeld (1973) "The Case of Missing Money" *Brooking Papers on Economic Activity*, Vol. 3.

Gradojevic, N. and J. Yang (2000), "The Application of Artificial Neural Networks to Exchange Rate Forecasting: The Role of Market Microstructure Variables", Working paper 2000-23, Bank of Canada.

Goodhart, C.A.E. (2000) "Can central banking survive the IT revolution?" London School of Economics Financial Market Group Special Discussion Paper No: SP0125.

Goodhart, C.A.E. y M. Krueger (2001) "The impact of technology on cash usage" London School of Economics Financial Market Group Discussion Paper No:DP0374.

Hansen, H. and S. Johansen (1993), "Recursive estimation in cointegrated VAR models". Referenciado en el manual de CATS for RATS.

Hendry, D.F. and A. Doornik (1994), "Modelling Linear Dynamic Econometric Systems" *Scottish Journal of Political Economy*, Vol. 45. Referenciado por Arize (1999).

Jansenn N. (1998) "The Demand for M0 in the United Kingdom Reconsidered: Some Specification Issues" Mimeo, Bank of England

Johansen, S. (1988), "Statistical analysis of cointegration vectors", *Journal of Economic Dynamics and Control*, No. 12, pages 231-54.

----- (1994), "The role of the constant and linear terms in cointegration analysis of non-stationary variables", *Econometric Reviews*, No. 13, pages 205-29.

----- (1995), *Likelihood Based Inference in Cointegrated Vector Auto-regressive Models*, *Advanced Text in Econometrics*, Oxford University Press.

King M. (1999) "Challenges for monetary policy: New and old" Bank of England Quarterly Bulletin, 397-415.

Lee, T.H., H. White and C. W.J. Granger (1993), "Testing for neglected nonlinearity in time series models", *Journal of Econometrics*, No. 56.

Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, Spriger-Verlag, Second Edition.

Markose S. y Y. Loke (2002) "Can cash hold its own ? International comparisons: theory and evidence" Mimeo, University of Essex.

Milbourne, Buchholtz y Wasan (1983)

Misas M, E. López y P. Querubín (2002), "La inflación en Colombia: una aproximación desde las redes neuronales", *Ensayos sobre política económica*, Banco de la República, No. 41-42, Junio – Diciembre 2002.

Soto(1997) "Nonlinearities in the Demand for Money: A Neural Network Approach" ILADES, Documento de Investigación # 107, diciembre.

Swanson, N. R. and H. White (1995), "A Model-Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Networks", *Journal of Business & Economic Statistics*, Vol. 13, No.3.

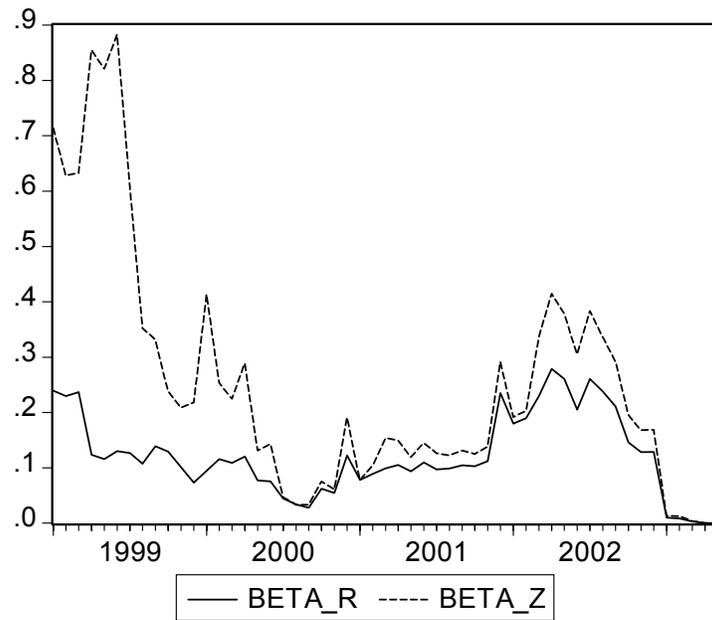
----- (1997a), "A model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks", *The Review of Economics and Statistics*, No. 79.

Reh G. (2002), "Forecasting with artificial neural network models", *SSE/EFI Working paper Series in Economics and Finance*, No. 491.

Van Dijk, D., T Terasvirta and P. H. Franses (2001), "Smooth Transition Autoregressive Models – A Survey of Recent Developments", *Econometric Reviews*. Xxxxxxx

Anexo 1

Test de estabilidad* del vector de cointegración



* 1 es el nivel de significancia al 5%