

Evaluación de pronóstico de una red neuronal sobre el PIB en Colombia

Por: José Mauricio Salazar Sáenz

Núm. 575
2009

Borradores de ECONOMÍA



tá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Col



Evaluación de pronóstico de una red neuronal sobre el PIB en Colombia

José Mauricio Salazar Sáenz*

jsalazsa@banrep.gov.co

Las redes neuronales artificiales han mostrado ser modelos robustos para dar cuenta del comportamiento de diferentes variables. En el presente trabajo se emplean para modelar la relación no lineal del crecimiento del PIB. Tres modelos son considerados: dos autoregresivos (especificación lineal y no lineal) y una red neuronal que usa la tasa de interés. Evaluando el desempeño de los modelos dentro y fuera de muestra, los pronósticos realizados por las redes neuronales artificiales superan ampliamente a los modelos lineales, siendo esta evidencia de relaciones asimétricas en el comportamiento del PIB en Colombia.

Palabras claves: Red neuronal artificial, no linealidad, PIB, *Rolling* de pronóstico, evaluación de pronóstico.

Clasificación JEL: C45, C53, E17, E23

* El autor es profesional del Departamento técnico y de información económica. Trabajo realizado durante la pasantía en el Banco de la República. Se agradece especialmente a Martha Misas A. por su valiosa y oportuna dirección y por el código de SAS sobre redes neuronales artificiales. Igualmente, se agradece a Munir Jalil por su apoyo y por el código de SAS sobre evaluación de pronóstico. Las opiniones acá expresadas son responsabilidad del autor y no comprometen al Banco de la República ni a su Junta Directiva.

1. Introducción

Desde la antigüedad el poder anticipar el futuro y contar con información sobre éste, se ha convertido en una necesidad. Los seres humanos en cada momento del tiempo deciden la mejor opción posible de acuerdo con la información disponible para tal efecto. Así, la generación de pronósticos adecuados genera mejores decisiones. Para los agentes económicos anticiparse a movimientos futuros del mercado supone anticipar choques, posibilitando la generación de mecanismos que permitan aprovechar de la mejor forma el futuro desde el presente.

Una de las variables macroeconómicas más importantes es el PIB. Para los agentes económicos conocer el posible comportamiento futuro de esta variable es fundamental. Por ejemplo, al Gobierno le beneficia saber cómo se comportará el PIB para planear su presupuesto de ingresos y gastos del próximo año; para el Banco de la República conocer su comportamiento le permitirá manejar de manera más eficiente la política monetaria, fijando de manera óptima la tasa de intervención en el mercado de acuerdo al clima económico; para los bancos comerciales, una posible estrategia sería anticipar la cantidad de recursos extras que la economía necesitará para su desarrollo; e incluso para los sindicatos conocer como se comportará el PIB les permite negociar de mejor manera el aumento salarial que se fija cada fin de año de acuerdo con la inflación y el crecimiento esperados.

Las variables económicas observadas en cada momento del tiempo pueden ser entendidas como realización de un proceso generador de datos. La modelación econométrica empieza con la selección de las variables independientes que expliquen el comportamiento de las variables dependientes que son de interés; posteriormente, se decidirá cuál es el proceso generador de datos, que en el caso de los modelos paramétricos supone asignar una forma funcional a dicho proceso generador. Dentro de éstos, el modelo lineal ha sido ampliamente usado. Siguiendo la hipótesis de linealidad se tendrá que la magnitud de un cambio en una variable explicativa afecta en igual magnitud a las variables explicadas y que en valor absoluto los cambios positivos y negativos de una variable explicativa afecta en la misma magnitud a la variable explicada. Sin embargo, los modelos lineales omiten relaciones presentes en el proceso generador de datos de algunas variables económicas; estas omisiones han abierto el paso a modelos no lineales que permiten asimetrías ante diferentes movimientos de variables independientes.

Dentro de los modelos paramétricos no lineales encontramos a las redes neuronales artificiales (RNA), las cuales se desarrollaron para replicar el comportamiento del cerebro en lo concerniente al reconocimiento de patrones, a las interconexiones entre diferentes neuronas y a la posibilidad de aprender de los errores. Ante un evento particular, al cerebro le resulta mucho más fácil la interpretación de dicho evento en la medida que haya recibido conocimiento previo o relacionado con este suceso.

Del mismo modo, las redes neuronales artificiales (RNA) pueden reconocer patrones y relaciones no observados. Así, en la medida que se cuente con un amplio número de observaciones las RNA pueden llegar a reproducir de manera exacta la variable dependiente que se está explicando, logrando una bondad de ajuste cercana a uno ($R^2 \approx 1$ ver Frances y van Dyk (2000)). Haciendo un símil, una red con pocas observaciones o unidades escondidas se comportará como un humano que se enfrenta a eventos poco conocidos.

Esta interesante propiedad de reconocimiento neuronal tiene sus costos; el primero es el posible sobreajuste de la red llegando a explicar datos atípicos, el segundo es que los parámetros de las RNA carecen de interpretación económica. Por ello, las RNA son conocidas como “modelos de cajas negras” (Misas, López y Querubín (2002)).

Las RNA han mostrado ser modelos muy versátiles y han sido usadas en diversas ramas del conocimiento, desde pronósticos de sismos, pasando por el reconocimiento de cáncer de seno hasta el trabajo de Swanson y White (1995) en donde se modelan 14 series macroeconómicas.

Uno de los trabajos pioneros que usan las RNA en la econometría fue realizada por Kuan y White (1994) y desde entonces han sido empleadas para pronosticar diferentes variables como la inflación, el PIB, la tasa de cambio, la demanda de efectivo, etc. El presente trabajo busca explorar las fortalezas de las RNA para explicar y pronosticar el comportamiento del PIB sin desestacionalizar en Colombia. El periodo de estudio comprende desde principios de la década de los 80 hasta el segundo trimestre de 2008, explorando la existencia de comportamientos no lineales presentes en el producto. Esto se realiza a través de una RNA autoregresiva y de una RNA aumentada que usa la política monetaria como variable explicativa, comparando estos dos modelos con el mejor ARIMA lineal encontrado.

Este trabajo esta ordenado en 5 secciones. La primera incluye esta introducción. En la segunda sección se presentan algunas variables modeladas a través de RNA y se discuten elementos teóricos que expliquen la existencia de relaciones no lineales del PIB, justificando la modelación de éste a través de RNA. En la tercera sección se presenta la metodología y se introduce al lector en las redes neuronales. En la cuarta se discuten los resultados de la aplicación de las RNA al crecimiento del PIB, presentando en esta misma sección la comparación con el modelo lineal. En la quinta, se concluye.

2. No linealidad

El uso cada vez mayor de las RNA en economía, finanzas y otras disciplinas como la medicina o el reconocimiento caligráfico se debe a la presencia de comportamientos no lineales en las relaciones entre diferentes variables. En el caso de la economía y finanzas, donde generalmente se han usado modelos lineales, el uso de herramientas no lineales como las RNA han permitido explicar de mejor manera el comportamiento de las interrelaciones entre diferentes variables económicas.

2.1 Modelación no lineal de variables económicas

El comportamiento de las variables macroeconómicas durante las diferentes fases del ciclo económico ha despertado el interés de los investigadores, encontrándose evidencia de comportamientos asimétricos de algunas variables durante periodos expansionistas y contraccionistas. Por esta razón en los últimos años, tanto a nivel mundial como en Colombia, se ha venido incrementando el número de trabajos que emplean esquemas no lineales para la modelación de las relaciones económicas.

En esta sección se explican brevemente algunos trabajos realizados con RNA en Colombia.

La inflación en Colombia ha sido abordada en varios trabajos. Uno de éstos, el desarrollado por Misas, López y Querubín (2002) utiliza las RNA para explicar la posible relación no lineal entre el dinero y la inflación, la cual puede ser explicada en razón a: i.) los efectos asimétricos de la política monetaria¹ y ii.) las rigideces de precios. En el trabajo mencionado se concluye el gran potencial que tienen las RNA para explicar el comportamiento entre el dinero y la inflación, superando a las siguientes especificaciones lineales: el modelo P* de la inflación, el modelo escandinavo, curva de Phillips, precios relativos y ARIMA.

¹ En el trabajo se cita a Soto (1997) quien estudió las propiedades no lineales de la demanda de dinero utilizando RNA.

En el trabajo de Torres (2005) se explica el comportamiento de la inflación básica de bienes transables y no transables, mediante la estimación de distintas especificaciones de la curva de Phillips. Diferentes teorías muestran la existencia de no linealidades en la curva de Phillips; por esta razón se utilizan cuatro modelos para estimar la curva de Phillips: dos lineales (Mínimos cuadrados ordinarios y ARIMA) y dos no lineales (RNA y mínimos cuadrados flexibles). Los resultados en términos de error de pronóstico resultaron muy favorables los modelos no lineales.

Aristizábal (2006) utiliza RNA para explicar la inflación en Colombia. La red neuronal empleada usa el dinero y rezagos de la inflación como variables independientes; se encuentra que las redes neuronales explican mejor el comportamiento de la inflación que un modelo ARIMA, mostrando evidencia de no linealidades en la inflación. El comportamiento de las RNA mejora aún más usando funciones de pérdida asimétrica² que ponderan de forma diferente los errores de pronóstico positivos y negativos.

Para evaluar la demanda de especies monetarias tanto a nivel agregado como por denominación en Colombia³, Arango, Misas y Hernández (2004) justifican el uso de modelos no lineales debido a: *i.*) los cambios en los ritmos de la inflación presentes en la desaceleración a finales de los 90, *ii.*) el creciente uso de tarjetas de crédito y dinero electrónico, *iii.*) la introducción de impuestos a las transacciones financieras como el 4xmil. Para pronosticar el agregado de las especies monetarias en este trabajo, se desarrollan diferentes modelos (RNA, mínimos cuadrados flexibles, VEC y ARIMA); al evaluarlos, las especificaciones no lineales muestran un mejor desempeño.

Para pronosticar las especies monetarias por denominación, además de las razones ya expuestas, tanto el tiempo de circulación como la inflación alteran de forma no lineal la curva de participaciones. En esta sección del trabajo la modelación se realiza con RNA, con mínimos cuadrados flexibles y usando un modelo conocido como la métrica D, que en ese momento era usado por muchas tesorerías de Bancos Centrales. Los resultados de la modelación muestran que los modelos no lineales superan ampliamente a la métrica D y que las redes neuronales pronostican mejor las altas denominaciones, mientras que los mínimos cuadrados flexibles pronostican mejor las bajas denominaciones.

El comportamiento de la tasa de cambio ha sido explicado por Jalil y Misas (2005) con una red neuronal autoregresiva, comparando su poder explicativo contra un modelo ARIMA. Se utilizan funciones de pérdida simétricas y asimétricas para evaluar los pronósticos de los dos modelos; las dos funciones encuentran evidencia a favor de las RNA, siendo mayor la evidencia utilizando funciones de pérdida asimétrica.

Estos trabajos sugieren que a través de la modelación de redes neuronales artificiales se obtienen mejores resultados que aquellos alcanzados mediante modelos lineales. Así, las RNA son una herramienta atractiva para la modelación y elaboración de pronósticos de variables macroeconómicas.

2.2 Evidencias de comportamientos no lineales del PIB.

² Como se explica en este trabajo, el uso de funciones de pérdida asimétrica dependerá de la ponderación social que se le asigna a los pronósticos. Cuando se ponderan de forma diferente los pronósticos por arriba o por abajo respecto al valor observado, lo más conveniente será usar funciones de pérdida asimétrica. Esto sucede con la inflación, dado que a principio de cada año el Banco de la República anuncia la meta de inflación. Si este pronóstico estuvo por arriba de la inflación efectiva la sociedad estará tranquila sobre el manejo que se le ha dado a la política monetaria; sin embargo, la sociedad perderá la credibilidad en el banco si el pronóstico está por debajo. Así los pronósticos por abajo deberían ser penalizados con mayor severidad para reflejar la forma como la sociedad los percibe.

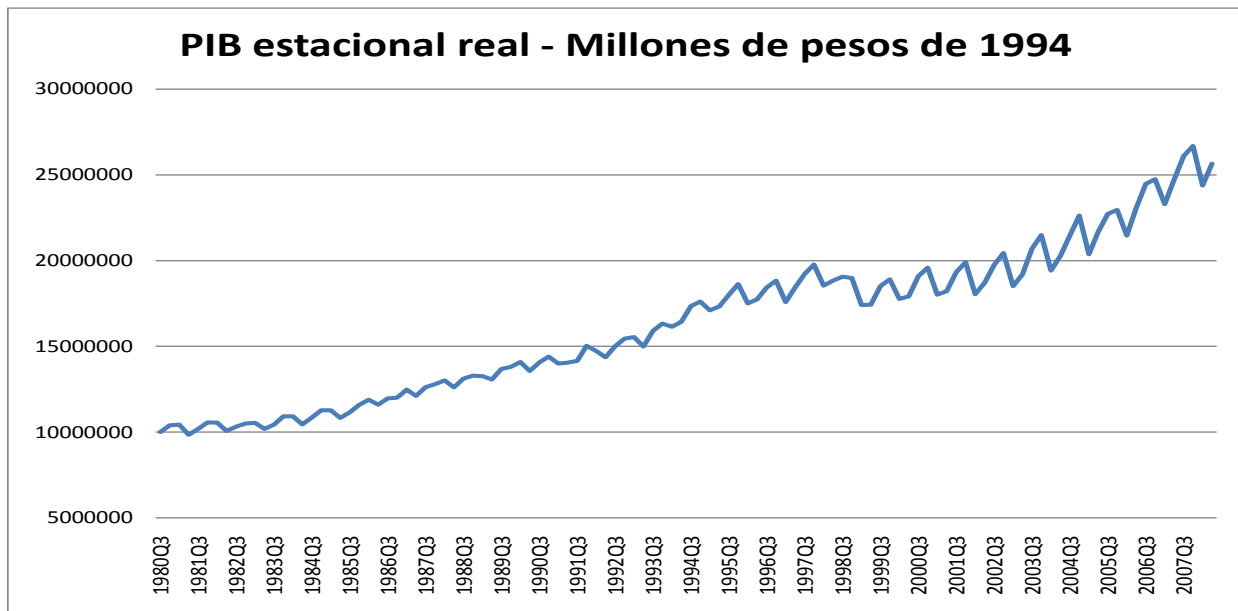
³ Para el caso Colombiano se tienen las siguientes denominaciones de billetes: 1.000, 2.000, 5.000, 10.000, 20.000 y 50.000. Las denominaciones monetarias son los diferentes valores en circulación.

Como se había mencionado anteriormente, en una relación lineal se tendrá que la magnitud de un cambio en una variable explicativa afecta en igual magnitud a las variables explicadas y que, en valor absoluto, los cambios positivos y negativos de una variable explicativa afectan en la misma magnitud a la variable explicada. Si contemplamos el valor del PIB presente como función de su pasado, estas dos propiedades indican que la magnitud de las recesiones o las expansiones afectan en la misma proporción al PIB actual. Por otro lado, se tendrá que, en valor absoluto, el PIB será afectado en la misma magnitud por momentos pasados de recesión y expansión.

El comportamiento del PIB se aleja de estos supuestos, dado que se ha considerado que los ciclos económicos presentan comportamientos no lineales. La evidencia empírica muestra la diferencia en los tiempos asumidos desde una expansión económica hasta una recesión y viceversa.

A continuación se enunciarán propiedades teóricas y empíricas que sustenten el supuesto de no linealidad.

Gráfico 2.1



Durante la gran depresión los intentos fallidos de estimular la producción a través de la política monetaria hicieron que los economistas de la época creyeran que la política monetaria expansionista era inefectiva. Posteriormente, y hasta nuestros días, existe consenso sobre las asimetrías de la política monetaria. Friedman (1969) mostró cómo la política monetaria contraccionista afectaba más al producto que la política expansionista; ello se debe al comportamiento de la economía durante el ciclo económico. Mientras la política monetaria expansionista se aplica en periodos de recesión, la contraccionista se efectúa en periodos de alto crecimiento.

Morgan (1993) expone tres argumentos sobre el comportamiento de la economía que explican el poco impacto de la política monetaria expansionista:

1. Cambios en las expectativas: durante las recesiones los inversionistas y los consumidores se vuelven pesimistas, esperando que la economía no se recupere rápidamente y generando cierta inercia en el ciclo económico. Ante la política monetaria expansionista los inversionistas no ahorrarán ni invertirán

más pues sus expectativas sobre el ciclo económico son bajas; los consumidores tampoco consumirán más por la misma razón.

Sin embargo, el solo pesimismo no explica la asimetría de la política monetaria, pues si los agentes económicos se vuelven pesimistas durante las recesiones, también se vuelven más optimistas durante las expansiones, haciendo que la política monetaria contraccionista reduzca su impacto sobre el PIB por el optimismo, de igual forma que el pesimismo debilita a la política expansiva.

Para que la asimetría se alcance, el pesimismo debe ser mayor durante las recesiones que el optimismo durante las expansiones del PIB; ésto sólo se cumple si las expectativas cambian asimétricamente. A nivel de las firmas, las expectativas cambian asimétricamente, dado que es más costoso entrar al mercado que salir del mismo y si una firma se encuentra produciendo con toda su capacidad instalada, aumentar la producción resulta mucho más costoso que disminuirla. Es por ello que la política monetaria expansionista es menos efectiva que la contraccionista, pues aumentar la producción es mucho más difícil que contraerla y esto se refleja en las expectativas de las firmas.

2. Restricciones de crédito: en periodos expansivos de la economía, con la ejecución de políticas monetarias contraccionistas que reflejen una alta tasa de interés, los bancos enfrentarán mayores costos financieros ante los mayores riesgos que esto supone. Una subida en las tasas de interés hará que las obligaciones de los deudores sean mayores, aumentando la posibilidad de bancarrota de los bancos, ante el evento de que los deudores no paguen. Debido a esto los bancos generan restricción del crédito, otorgando el mismo sólo a deudores con mayores probabilidades de pago. Dado que la economía se encuentra en expansión, se presenta una gran demanda de crédito simultáneamente con la restricción del mismo impuesta por los bancos; ello hace que los efectos reales de la política monetaria contraccionista sean mayores.

En una economía en recesión, por otro lado, una caída en las tasa de interés no necesariamente hará que los consumidores y los inversionistas aumenten su nivel de consumo e inversión respectivamente: este efecto, que es conocido como trampa de liquidez y que fue descrito por Keynes, consiste precisamente en la imposibilidad de estimular la economía cuando la tasa de interés está en niveles bajos y al mismo tiempo se realiza política monetaria expansionista.

3. Rigidez de los precios a la baja: La política monetaria tendrá efectos asimétricos si los precios se ajustan de manera asimétrica. Para las firmas recolectar información sobre los diferentes choques en la economía para cambiar los precios y discriminar cuáles son permanentes y cuáles son transitorios es bastante costoso; por esta razón la meta de inflación y la inflación esperada constituyen referentes fundamentales y de bajo costo que ayudan a las firmas a determinar la tendencia de los precios. Por ello, los cambios de los precios son más flexibles al alza que a la baja, pues la inflación de largo plazo siempre será positiva y genera una presión positiva sobre los precios.

Otra posible causa de relaciones no lineales en el PIB se generaría a través de los precios. Diferentes teorías se han desarrollado para explicar por qué los precios pueden generar distorsiones sobre las variables reales y sobre el PIB; estas teorías son conocidas como rigideces nominales (Romer (2005)). Las firmas no solo deben recolectar información sobre el devenir económico para decidir un ajuste en precios, sino que, una vez se ha decidido cambiar éstos, se debe incurrir en costos extras para difundir a la sociedad los cambios en precios; estos costos son conocidos como “costos de menú”.

Suponiendo que las firmas se encuentran produciendo en un punto óptimo, ante perturbaciones en la economía, éstas sólo cambiarán sus precios si los costos de no ajustar los mismos son mayores que los costos de menú de la firma, obteniendo situaciones en que no se cambien los precios, pese a que las perturbaciones generen un precio óptimo diferente al actual; de esta forma se sacrifica el producto en lugar del ajuste de precios. Ajustar o no los precios dependerá del tamaño de la perturbación, haciendo que el primer supuesto sobre las relaciones lineales no se cumpla, pues el tamaño de la perturbación sí modifica el efecto sobre la variable dependiente, pudiendo ser este nulo en el caso de los precios, debido a los costos de menú. Así mismo, se ha encontrado evidencia empírica sobre la asimetría de los precios cuando deben ser ajustados a la baja y al alza ante diferentes choques. Fabiani et al (2005) muestran cómo en la zona Euro los precios son más rígidos al alza frente a choques positivos de demanda y cómo son más rígidos a la baja ante choques negativos de costos.

Todos estos argumentos dan cuenta de comportamientos no lineales de otras variables respecto al PIB y del PIB en sí mismo; de esta forma modelar el PIB a través de modelos no lineales seguramente tendrá un mejor desempeño que la modelación lineal. Zukime y Junoh (2004) utilizan diferentes variables que miden el grado de avance tecnológico y de acceso al conocimiento en Malasia para predecir el crecimiento del PIB; en este trabajo, comparando con un modelo lineal las RNA explican de mejor manera la tasa de crecimiento del PIB.

Con el mismo propósito, Tkacz y Hu (1999), del banco de Canadá, utilizan RNA y modelos lineales usando como variables explicativas para la tasa de crecimiento del PIB canadiense la diferencia entre la tasa de interés de largo plazo de los bonos y los papeles comerciales a 90 días, la tasa de crecimiento de M1 y M2 y la tasa de crecimiento del índice 300 TSE. Este trabajo encuentra que las redes neuronales aumentan su poder de predicción respecto a los modelos lineales cuando el horizonte de pronóstico es más lejano, que cuando se pronostica un periodo consecutivo.

Finalmente, Araújo y Gama (2004) emplean las redes neuronales artificiales para explicar la duración y la cantidad de los ciclos económicos del PIB en los países del G-7, Méjico y Brasil. La mayoría de las características de los ciclos del PIB son explicados de mejor manera por RNA, con un mejor desempeño en los países del G-7. En resumen, se encuentra clara evidencia de las ventajas de emplear estos modelos para la explicación de los ciclos frente a los modelos lineales ARIMA.

La definición del conjunto de variables explicativas es el primer paso en cualquier esquema de modelación. El presente trabajo considera los rezagos de las tasas de crecimiento del PIB y de la tasa de interés de los CDT a 90 días como variables explicativas de la tasa de crecimiento del PIB, desde el segundo trimestre de 1980 hasta el segundo trimestre de 2008.

Para identificar la variable que mida la política monetaria se pueden emplear los agregados monetarios o las tasas de interés. Diferentes trabajos empíricos han mostrado la conveniencia de usar la tasa de interés. Sims (1980) muestra que el poder explicativo de los agregados monetarios sobre el comportamiento del PIB se diluye al introducir la tasa de interés en la regresión. El trabajo de Reinhart y Reinhart (1991) para el caso colombiano, utiliza un VAR estructural sobre las siguientes variables: dinero, tasa de interés, PIB, salarios, precios y tasa de cambio. En este trabajo se encuentra que la tasa de interés es la única variable que influye significativamente al PIB, siendo coherente este resultado con Sims (1980). En el trabajo de Morgan (1993) se usa la tasa de interés de los bonos federales para medir el estado de la política; el autor afirma que los agregados monetarios pueden estar reflejando movimientos de las inversiones o del crecimiento económico, y, por lo tanto, no necesariamente reflejan el estado de la política monetaria. En dicho trabajo se encuentra

evidencia econométrica de la asimetría de la política monetaria para Estados Unidos con un modelo lineal en dos etapas.

El trabajo de Saito y Takeda (2000), de hecho, va un paso más adelante y usa la diferencia de las tasas de interés de largo y corto plazo para capturar el estado de la política monetaria. Diferencias positivas muestran que la tasa de interés de largo plazo es mayor que la de corto plazo. Es decir, que se está en un momento económico donde la política monetaria es expansionista; por otra parte, si la diferencia es negativa, la política monetaria será contraccionista. Sin embargo, dado que en Colombia no se cuenta con medidas de la tasa de interés de largo plazo desde los años 80, se usa la tasa de interés de los CDT a 90 días como aproximación de la política monetaria; para trabajos futuros sería conveniente utilizar la diferencia de la tasa de interés buscando mejores resultados.

Otro argumento a favor de la implementación de RNA para realizar pronósticos de la tasa de crecimiento del PIB es la existencia de variables intermedias que afectan las relaciones entre la tasa de interés y el PIB. Por ejemplo, una subida en la tasa de interés hace que la inversión caiga, de tal forma que ante la caída de la inversión el producto caerá, y por lo tanto, entre la tasa de interés y el producto existe una variable intermedia que las relaciona. Tkacz y Hu (1999) recomiendan usar RNA para modelar la existencia de variables intermedias, sin necesidad de tenerlas plenamente identificadas, a través de unidades escondidas⁴, las cuales se encargarán de realizar conexiones indirectas y no lineales entre las variables independientes y la variable dependiente.

Este trabajo utiliza una red neuronal autoregresiva, es decir con rezagos del PIB como variables explicativas, y comparará el desempeño de dicha red con un modelo ARIMA construido con la metodología Box-Jenkins. Por otra parte, se introduce la tasa de interés de los CDT a 90 días como variable explicativa en la red neuronal, para indagar si los pronósticos mejoran con la introducción de variables que midan el estado de la política monetaria.

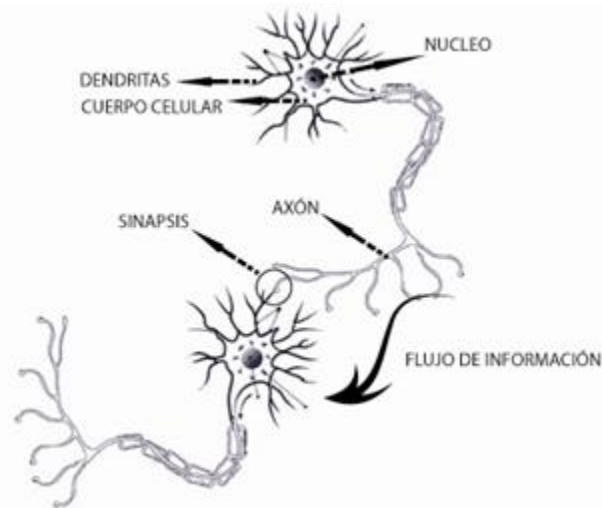
3. Redes Neuronales Artificiales

3.1 Relación entre Redes Neuronales Biológicas y Redes Neuronales Artificiales.

El cerebro, y en general todo el sistema nervioso humano, ya sean órganos de los sentidos o sistemas autónomos centrales y periféricos, están compuestos en su mayoría por células especializadas denominadas neuronas, consideradas como unidad funcional y estructural de dicho sistema nervioso, encargadas de percibir el entorno por medio de diferentes estímulos externos (temperatura o luz, por ejemplo) y decodificar la información para así crear una visión entendible y manejable de esta información. Las neuronas están compuestas por diferentes partes: las dendritas, el cuerpo o soma y el axón. La información viaja de neurona a neurona a través de impulsos eléctricos entrando por las dendritas, que son las conexiones de entrada, pasando por el soma, que es la encargada del procesamiento de la información, y saliendo por el axón, que es la conexión de salida, transmitiendo la información a otras neuronas. La transmisión de impulsos entre neuronas se llama sinapsis.

⁴ Este concepto será explicado más adelante.

Gráfico 3.1



Existen muchos tipos de neuronas; sin embargo, hay tres grandes categorías de acuerdo a su función (véase Snell (2007)):

Neuronas sensoriales: encargadas de la conducción de los diferentes impulsos, ya sean propioceptivos (dolor, movimiento gastrointestinal) o impulsos externos, hasta el sistema nervioso central. Esta conducción es llevada a cabo mediante los cordones espinales posteriores.

Neuronas motoras: encargadas de la interacción con el músculo y otras células efectoras en una estructura denominada placa neuromuscular; son las responsables de transmitir órdenes ya procesadas para su ejecución. La conducción se da por medio de los cordones espinales anteriores.

Interneuronas: también llamadas neuronas de asociación, se encuentran únicamente en el sistema nervioso central, más exactamente en el cerebro. Su función es precisamente la conexión entre las neuronas receptoras y las efectoras con el fin de transmitir la información requerida. Cumplen un papel fundamental ya que en ellas está el éxito de la sincronización de las órdenes dadas por el sistema nervioso central.

Dado un estímulo específico, la transmisión de la información de una neurona a otra puede ser inhibitoria o excitadora; si la carga eléctrica de la señal alcanza los umbrales para los cuales la neurona está capacitada, la señal será transmitida, de lo contrario la señal será interrumpida.

Las redes neuronales artificiales son modelos computacionales que han sido utilizadas en diferentes áreas del conocimiento y desarrollados para replicar el comportamiento de las redes neuronales biológicas. Las RNA (Redes Neuronales Artificiales) son una colección de neuronas interconectadas en diferentes capas, que buscan reproducir la forma como las neuronas biológicas procesan la información. En las RNA las variables explicativas *-inputs-* se conectan con las variables explicadas *-outputs-* a través de ponderadores que miden la importancia relativa; así se transmite la información de los *-inputs-* a los *-outputs-*. Siguiendo el modelo propuesto por Frances y van Dyjk (2000) representaremos esta relación a través de dos canales: el lineal y el no lineal.

$$\text{Ec. 3.1} \quad y = \varphi_0 + X\varphi + \sum_{i=1}^q \beta_i * G(X*\gamma) + \varepsilon$$

La parte no lineal está compuesta por funciones de activación que buscan activar o desactivar comportamientos de la variable de salida y sus variables de entrada no observables en el mundo lineal. Es a

través de estas funciones de activación que las RNA consiguen reproducir el comportamiento de las neuronas biológicas, logrando mecanismos mediante los cuales la serie responde de manera diferente ante umbrales de información determinados por la estimación. La función de activación más usada es la función logística.

$$\text{Ec. 3.2} \quad \text{Función logística : } \frac{1}{1+e^{-(\gamma_1x_1+\gamma_2x_2-\dots+\gamma_ix_i-\gamma_0)}}$$

Siendo $z = \gamma_1x_1 + \gamma_2x_2 - \dots + \gamma_ix_i - \gamma_0$, la función logística posee las siguientes propiedades:

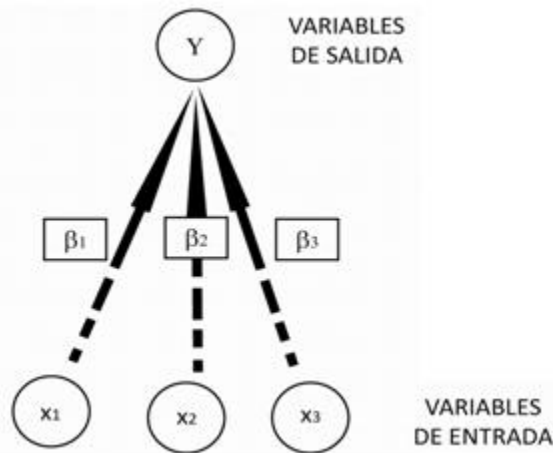
$$f(z = 0) = 0.5, \text{ donde necesariamente } \gamma_0 = \gamma_1x_1 - \gamma_2x_2 - \dots - \gamma_ix_i$$

$$\lim_{z \rightarrow -\infty} f(z) = 0, \text{ donde } \gamma_0 > \gamma_1x_1 - \gamma_2x_2 - \dots - \gamma_ix_i$$

$$\lim_{z \rightarrow \infty} f(z) = 1, \text{ donde } \gamma_0 < \gamma_1x_1 - \gamma_2x_2 - \dots - \gamma_ix_i$$

Se observa cómo los valores que tome la función logística dependen de la diferencia entre la combinación lineal de las variables x_i y el parámetro γ_0 . Así, las RNA responden de manera distinta ante diferentes umbrales de información tal como las neuronas biológicas responden a diferentes impulsos eléctricos; del mismo modo como una neurona motora inhibe estímulos sensoriales, una función de activación inhibirá información menor al umbral γ_0 . Cuando no existe evidencia de asimetrías en la respuesta a la información o evidencia de comportamientos no lineales de la serie la función de activación será la idéntica, dando como resultado el modelo lineal simple, donde la red envía la información de las variables de entrada *-input-* a las variables de salida *-output-*, a través de ponderadores que son los parámetros del modelo lineal. Esta red contiene dos capas, la de inputs y la de outputs.

Gráfico 3.2



La importancia de cada variable viene dada por el parámetro β_i ; de esta forma en cada momento del tiempo la red colecciona los valores de cada una de las variables de entrada, los multiplica por sus parámetros y logra identificar el valor esperado del *-output-*. Así, el valor de y será la suma ponderada de las variables x , donde la transmisión de los estímulos se realiza a través de las ponderaciones β_i , como lo muestra la ecuación:

$$\text{Ec. 3.3} \quad y = \beta_0 + \beta_1X_1 + \dots + \beta_kX_k = \beta_0 + \sum_{i=1}^k \beta_iX_i = \beta_0 + X\beta$$

Sin embargo, este modelo lineal, el primero y más simple utilizado, no logra ajustarse bien al comportamiento de las variables ante la existencia de comportamientos no lineales, discontinuidades o variables intermedias

entre los *inputs*- y los *outputs*-. Para capturar estos comportamientos las redes neuronales usan funciones de activación no lineales que permitan identificar umbrales de la suma ponderada de los *inputs*-. Estos cambios generalmente son modelados con funciones de transición suave, que conformarán la parte no lineal de la relación, donde la función más usada para este fin es la función logística. De esta forma, la RNA con su componente lineal, no lineal y el término de perturbación queda especificada de la siguiente manera:

$$\text{Ec. 3.4 } y = \varphi_0 + X\varphi + \sum_{i=1}^q \beta_i * \frac{1}{1+e^{-(\gamma_1 x_1 + \gamma_2 x_2 + \dots + \gamma_p x_p - \gamma_0 i)}} + \varepsilon$$

Esta ecuación generalmente se reescribe de la siguiente forma y así será trabajada:

$$\text{Ec. 3.5 } y = \varphi_0 + X\varphi + \sum_{i=1}^q \beta_i * G_i(Z\gamma) + \varepsilon \quad Z \subseteq X$$

No obstante, trabajos como el de Tkacz y Hu (1999) usan otra forma de representar las redes neuronales omitiendo la parte lineal, pero conservando el intercepto:

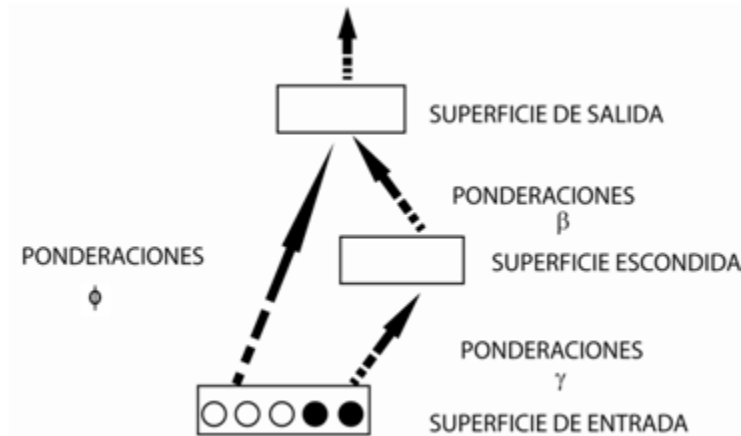
$$\text{Ec. 3.6 } y = \varphi_0 + \sum_{i=1}^q \beta_i * G_i(Z\gamma) + \varepsilon$$

3.2 Arquitectura y nomenclatura de una RNA

Siguiendo la propuesta de Frances y van Dyjk (2000) las variables a introducir en las funciones de activación serán un subconjunto propio de las variables introducidas en la parte lineal; de este modo, si se cuenta con k variables en la parte lineal se tendrá p variables en la parte no lineal, donde $p \leq k$. Se puede introducir un número q de funciones de activación. La arquitectura de una red neuronal es una combinación particular de p y q . Es decir, una arquitectura es una combinación de un número de variables en la parte no lineal con un número de funciones de activación. Dado un número máximo de funciones de activación a introducir en un ejercicio particular tendremos un total de $k*q$ arquitecturas.

Se ha desarrollado una nomenclatura especial para describir la forma funcional de la red de acuerdo a la cantidad de capas y la forma como fluye la información. Una clase de RNA, ampliamente estudiada e implementa, es la conocida como *single hidden layer feedforward network*. Ésta es una red con una única capa o superficie escondida y de alimentación hacia adelante. A la capa compuesta por las funciones de activación se le conoce como capa escondida, pues representa una conexión no directa entre *inputs*- y *outputs*- y será explicada en detalle más adelante. La alimentación hacia adelante hace referencia a la forma como fluye la información en la RNA, que va de las variables explicativas (*input*) a las variables explicadas (*output*):

Gráfico 3.3: single hidden layer feedforward network⁵



En la base de la red se encuentran las variables *inputs*- conformadas por el conjunto de variables explicativas X . Estas k variables pueden ser l rezagos de la variable explicada y m variables exógenas, $X = \{y_{t-1}, \dots, y_{t-l}, z_{1,t}, \dots, z_{m,t}\}$; a esta superficie se le conoce como la superficie o capa *input*. La información viaja hacia los *outputs* a través de dos caminos: el lineal y el no lineal. En la parte lineal cada una de las señales emitidas por cada variable es ponderada por su coeficiente; estos efectos son transmitidos directamente a los *outputs*.

$$\text{Ec. 3.7 } y_{\text{lineal}} = \widehat{\varphi}_0 + \widehat{\varphi}_1 X_1 + \widehat{\varphi}_2 X_2 + \dots + \widehat{\varphi}_k X_k$$

En la parte no lineal encontramos que los parámetros $\gamma_{i,j}$ con $i=1 \dots p$ y $j=1 \dots q$ son los encargados de amplificar o disminuir el efecto de las señales introducidas en cada función de activación. Las variables introducidas en la parte no lineal deben ser un subconjunto propio de las variables contenidas en la superficie *input*. Los efectos conjuntos de las variables, que es la combinación lineal de éstas y sus ponderadores, son enviadas al intervalo $[0,1]$ a través de la función logística⁶. Estos valores son multiplicados por los β_j para transmitir el efecto final sobre el *output*. Es por esta transferencia indirecta de las señales que a la capa compuesta por funciones de activación se le conoce como superficie escondida.

$$\text{Ec. 3.8 } y_{\text{no lineal}} = \widehat{\beta}_1 * G_1(Z\hat{\gamma}) + \widehat{\beta}_2 * G_2(Z\hat{\gamma}) + \dots + \widehat{\beta}_q * G_q(Z\hat{\gamma})$$

Para obtener el valor pronosticado de y por la RNA se suman la parte lineal y no lineal:

$$\text{Ec. 3.9 } y_{\text{estimado}} = f(X_t, \hat{\gamma}, \hat{\beta}, \hat{\varphi}) = \widehat{\varphi}_0 + X\hat{\varphi} + \sum_{i=1}^q \widehat{\beta}_i * G_i(Z\hat{\gamma}) \quad Z \subseteq X$$

Los parámetros β_j y φ_i son estimados a través de mínimos cuadrados ordinarios y los parámetros gamas son encontrados a través de muestreo partiendo de valores iniciales. El proceso de estimación de la RNA será explicado más en detalle en la próxima sección.

⁵ Este gráfico representa una RNA con 5 variables en la parte lineal; en este ejemplo las dos variables subrayadas transmitirán su información a través del camino no lineal, siendo ésta una representación para una arquitectura con $p=2$.

⁶ Otra función que ha sido ampliamente usada en las RNA es la tangente hiperbólica.

3.3 Estimación de las RNA

El proceso de estimación consistirá en encontrar los valores de $\hat{\varphi}$, $\hat{\beta}$ y $\hat{\gamma}$ de tal manera que describan de la mejor forma posible el comportamiento de las variables de salida. Se llama $\hat{\theta}$ al vector conformado por todos los parámetros de la RNA, en donde $\hat{\theta} = (\hat{\gamma}, \hat{\beta}, \hat{\varphi})$; este vector contiene $(p+1)*q+q+k+1$ elementos, donde los primeros $(p+1)*q$ corresponden a todos los parámetros introducidos en las funciones de activación, los siguientes q elementos corresponden a los ponderadores de las funciones de activación y los últimos $k+1$ elementos corresponden a los parámetros de la parte lineal. Swanson y White (1995) denominan estos parámetros de la siguiente manera: los parámetros $\hat{\gamma}$ son llamados pesos de los *inputs* a las unidades escondidas, los parámetros $\hat{\beta}$ son llamados pesos de las unidades escondidas a los *outputs* y los parámetros $\hat{\varphi}$ son llamados pesos de *inputs* a *outputs*.

La estimación consistirá en encontrar el vector theta que optimice una función objetivo, que podría ser la función de máxima verosimilitud o la suma de los errores al cuadrado. Dado que difícilmente se conoce la distribución de los errores en las RNA, no es posible usar la función máxima verosimilitud; por consiguiente, se usará el método de mínimos cuadrados no lineales (MCNL) para minimizar la función objetivo que será la suma de los errores al cuadrado:

$$\text{Ec. 3.10} \quad S(\hat{\theta}) = \sum_{t=1}^n [y_t - f(X_t, \hat{\theta})]^2$$

Donde $f(X_t, \hat{\theta}) = \hat{\gamma} = \hat{\varphi}_0 + X\hat{\varphi} + \sum_{i=1}^q \hat{\beta}_i * G_i(Z\hat{\gamma})$ y por definición $\hat{\varepsilon}_t = y_t - f(X_t, \hat{\theta})$

Para ver cómo las RNA's simulan el comportamiento de las redes neuronales biológicas, se observan dos interesantes propiedades: una RNA con un número apropiado de unidades escondidas puede aproximar cualquier función continua (Cybenko (1989)); esta propiedad ampliamente estudiada se conoce como la propiedad universal de aproximación de las redes neuronales. Sin la existencia de las unidades escondidas, tal aproximación no es posible. Por otro lado, las derivadas parciales de cada uno de los parámetros de la red contienen los errores del modelo. Esta interesante propiedad junto con la posibilidad de las RNA de aproximar cualquier función, permiten simular la interconexión entre neuronas en el cuerpo humano y el aprendizaje del cerebro basado en ensayo y error. El aprendizaje en las RNA puede ser entendido como un proceso de estimación recursiva que modifica los parámetros ante nueva información, buscando reproducir de la mejor manera posible el comportamiento de la variable de salida para minimizar la función objetivo. Las derivadas parciales de las RNA son:

$$\text{Ec. 3.11} \quad \frac{dS(\theta)}{d\beta_j} = -2 \sum_{t=1}^n [y_t - f(X_t, \hat{\theta})] * G_j(Z\hat{\gamma}) = -2 \sum_{t=1}^n [\hat{\varepsilon}_t] * G_j(Z\hat{\gamma})$$

$$\begin{aligned} \text{Ec. 3.12} \quad \frac{dS(\theta)}{d\gamma_{ij}} &= -2 \sum_{t=1}^n [y_t - f(X_t, \hat{\theta})] * \left[\hat{\beta}_j G_j(Z\hat{\gamma}) (1 - G_j(Z\hat{\gamma})) \right] * x_{i,t} \\ &= -2 \sum_{t=1}^n [\hat{\varepsilon}_t] * \left[\hat{\beta}_j G_j(Z\hat{\gamma}) (1 - G_j(Z\hat{\gamma})) \right] * x_{i,t} \end{aligned}$$

La derivada con respecto a φ es exactamente igual a la derivada de MCO lineales:

$$\text{Ec. 3.13} \quad \frac{dS(\theta)}{d\varphi_i} = -2 \sum_{t=1}^n [y_t - f(X_t, \hat{\theta})] * x_{i,t} = -2 \sum_{t=1}^n [\hat{\varepsilon}_t] * x_{i,t}$$

Para optimizar estas derivadas y dado que las RNA son modelos paramétricos no lineales, no pueden ser optimizadas a través de métodos convencionales; para la optimización de funciones no lineales existen gran cantidad de métodos numéricos de optimización que usan las condiciones de primer orden, ayudando a resolver este tipo de problemas. Dentro de estos algoritmos encontramos Gauss-Newton, Newton-Rapson, entre muchos otros. Todos estos algoritmos tienen la siguiente forma:

$$\text{Ec. 3.14} \quad \hat{\theta}^{r+1} = \hat{\theta}^r - \lambda[A(\hat{\theta}^r)]^{-1}\nabla S(\hat{\theta}^r)$$

Estos métodos numéricos funcionan optimizando la función objetivo a través de la derivada y en algunos casos de la concavidad, construyendo una sucesión de parámetros que converjan a algún mínimo local (en algunas ocasiones global). Estos algoritmos son procesos iterativos que dependen de la estimación inmediatamente anterior. Con base en la información aportada por los parámetros ya estimados se decide hacia qué dirección se deben mover los nuevos parámetros, de tal forma que $S(\hat{\theta}^{r+1}, X_t) < S(\hat{\theta}^r, X_t)$; el proceso iterativo terminará cuando $\hat{\theta}^{r+1} - \hat{\theta}^r = 0$. Sin embargo, esta convergencia perfecta no siempre es posible en problemas multivariados debido a la naturaleza de las funciones. Alternativamente se buscará que esta diferencia tienda a cero, $\hat{\theta}^{r+1} - \hat{\theta}^r \approx 0$; por consiguiente, $S(\hat{\theta}^{r+1}, X_t) \approx S(\hat{\theta}^r, X_t)$. Si hemos llegado a un vector de parámetros que cumple estas condiciones, la función habrá sido minimizada. Esta convergencia implica que $\lambda[A(\hat{\theta}^r)]^{-1}\nabla S(\hat{\theta}^r) \approx 0$, es decir, que la derivada de $\hat{\theta}^r$ tienda a cero, que es de hecho una condición para alcanzar los puntos mínimos.

La diferencia entre los algoritmos reside en el valor asignado a la matriz $A(\hat{\theta}^r)$. Para Gauss-Newton $A(\hat{\theta}^r) = 0.5 * \nabla S(\hat{\theta}^r) \nabla S(\hat{\theta}^r)$; para Newton-Rapson $A(\hat{\theta}^r) = H(\hat{\theta}^r)$, donde $H(\hat{\theta}^r)$ es la matriz Hessiana. Dicha matriz permite observar la concavidad de la suma de los errores al cuadrado evaluada en el vector $\hat{\theta}^r$. Para entender el comportamiento de las RNA se explicará el método conocido como *steepest descent* o máximo descenso que le asigna a la matriz $A(\hat{\theta}^r)$ la matriz idéntica. Este algoritmo es representado de la siguiente manera:

$$\text{Ec. 3.15} \quad \hat{\theta}^{r+1} = \hat{\theta}^r - \lambda \nabla S(\hat{\theta}^r)$$

En este caso, la dirección hacia la cual se deben mover los parámetros para optimizar la función estará determinada por el negativo del gradiente; el tamaño del paso estará dado por el tamaño de la derivada y por el *step length*. Este parámetro notado con la letra griega λ repondera el impacto del gradiente sobre los nuevos parámetros. Dado un valor inicial, el algoritmo funciona de la siguiente manera: con el valor inicial se calcula el gradiente de la suma de los errores al cuadrado y se calcula la diferencia descrita en la ecuación 3.15, encontrando un nuevo vector de parámetros; este proceso se repite un número apropiado de veces hasta encontrar convergencia en los parámetros, en donde $\nabla S(\hat{\theta}) \approx 0$. Para encontrar el valor óptimo de λ se emplea la siguiente estimación (Hendry (1995)):

$$\text{Ec. 3.16} \quad \lambda^* = \min S(\hat{\theta}^r - \lambda \nabla S(\hat{\theta}^r))$$

De esta forma, el tamaño del paso variará de acuerdo al punto de la función en que nos encontremos, siendo el paso cada vez más pequeño en la medida en que los parámetros estén en la vecindad de un mínimo.

En el contexto de las RNA los algoritmos pueden ser entendidos como la forma en que la red es capaz de “aprender” el comportamiento de las variables de salida. Esto se obtiene a través de una regla de aprendizaje que encuentre el valor óptimo de los parámetros.

$$\text{Ec. 3.17} \quad \hat{\theta}^{r+1} = \hat{\theta}^r + \Delta^r$$

Δ^r será la regla de aprendizaje.

La ecuación anterior indica que la RNA optimiza el vector de parámetros mediante un proceso iterativo que encuentra convergencia cuando la red no tiene posibilidad de incrementar su conocimiento sobre el comportamiento de las series a explicar, haciendo mínimos los errores. Sin embargo, al encontrar que $\Delta^r \neq 0$, se quiere decir con ello que aún existen posibilidades de aprender y los movimientos deberán tener la magnitud y el signo que la regla indique. Si se tiene

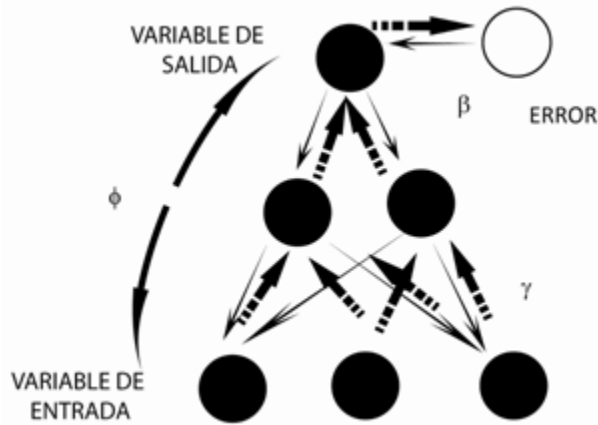
$$\text{Ec. 3.18} \quad \Delta^r = -\lambda \nabla S(\hat{\theta}^r)$$

la regla de aprendizaje de la RNA será el *steepest descent*. Así, de todas las posibles direcciones que lleven a posibles aprendizajes, la trayectoria escogida estará determinada por el gradiente y λ . Un gradiente diferente de cero nos indica que todavía es posible obtener un descenso, logrando a través de una sucesión de parámetros un menor valor de la función objetivo hasta obtener algún mínimo. Es conveniente recordar que el vector gradiente está conformado por todas las derivadas parciales de la función respecto a los parámetros. Estas derivadas parciales miden el efecto marginal que cada uno de los parámetros tiene sobre la suma de los residuales al cuadrado $S(\hat{\theta})$ y ya fueron descritas en las ecuaciones 3.11, 3.12 y 3.13.

En econometría se posee la ventaja de poseer los valores observados. Estos son usados para realizar un entrenamiento supervisado, el cual se entiende como la comparación de los valores efectivamente observados contra los valores estimados en el modelo construyendo los errores. De esta forma, cuando se aplica la regla de aprendizaje lo que la red está logrando es replicar de la mejor forma posible el comportamiento de la variable efectivamente observada; esta propiedad es deseable para cualquier modelo. Sin embargo, dado que teóricamente las RNA pueden aproximar cualquier función continua, son estos modelos los que logran realizar un determinado “aprendizaje” sobre el comportamiento de las variables.

Este aprendizaje funciona de la siguiente manera: teniendo un vector de parámetros en la iteración r $\hat{\theta}^r$, se calcula el valor estimado de la variable de salida; con este valor y el valor observado se calcula el error. La información aportada por este error viaja de vuelta a través de las derivadas y a través de la regla de aprendizaje se calcula $\hat{\theta}^{r+1}$, y así sucesivamente, hasta que no sea posible aumentar el aprendizaje y $\hat{\theta}^{r+1} \approx \hat{\theta}^r$. Es por esta razón que este tipo de métodos se conoce como “*backward propagation*”, dado que la información viaja de regreso haciendo que la red aprenda de sus propios errores; es a través de esta metodología que las RNA logran, reproduciendo el comportamiento del cerebro (ver Montenegro (2001)), simular la interconexión de las neuronas biológicas, dado que todos los parámetros responden ante los errores cometidos, obteniendo la mejor interpretación de la realidad que en el universo de las RNA son las variables de salida.

Gráfico 3.4



Para encontrar los parámetros óptimos, otro enfoque utilizado es la estimación recursiva. Este tipo de metodología introduce paso a paso nuevas observaciones; en este escenario el aprendizaje se da en dos momentos: a través de la regla de aprendizaje en cada momento del tiempo y con la introducción de nuevas observaciones en diferentes momentos del tiempo. Así, la nueva regla supone que en cada momento del tiempo se encuentran los parámetros óptimos $\hat{\theta}_t^*$, siendo su representación de la siguiente manera:

$$\text{Ec. 3.19} \quad \hat{\theta}_{t+1}^* = \hat{\theta}_t^* + \Delta_t^r \quad \text{donde } t \in (1, n)$$

Cada $\hat{\theta}_t^*$ denota los parámetros óptimos basados en las primeras t observaciones. Cada nuevo dato supone la posibilidad de aprender mejor el comportamiento de la variable de salida, ajustando el aprendizaje a través de los parámetros, reduciendo de esta forma cada vez más los errores.

El aprendizaje a través de *steepest descent* ha mostrado ser muy ineficiente relativo a otros métodos numéricos de optimización no lineales. Los métodos de cuasi-Newton son bastante usados en RNA. Todos los algoritmos de cuasi-Newton son aproximaciones del algoritmo Newton-Rapson, pues usan aproximaciones numéricas de la matriz Hessiana para extraer información sobre la concavidad, dado que estimar la matriz Hessiana suele ser computacionalmente costoso y, además, se requiere que la matriz sea invertible. Estos métodos numéricos poseen la siguiente estructura:

$$\text{Ec. 3.20} \quad \hat{\theta}^{r+1} = \hat{\theta}^r - \lambda K_r \nabla S(\hat{\theta}^r)$$

Donde K_r es una matriz definida positiva que aproxime a $[H(\hat{\theta}^r)]^{-1}$. K_r también es estimada de forma recursiva: $K_{r+1} = K_r + D_r$. En efecto, la diferencia entre los distintos algoritmos radica en el método empleado para calcular la matriz K_r . Uno de los primeros métodos fue propuesto por Davidon y desarrollado por Fletcher y Powell (DFP). Sin embargo, el método de Broyden-Fletcher-Goldfarb-Shanno (BFGS)⁷ ha mostrado un mejor desempeño; este algoritmo es calculado a partir de la siguiente ecuación:

$$\text{Ec. 3.21} \quad K_{r+1} = K_r + \left(1 + \frac{p_r' K_r p_r}{p_r' d_r}\right) * \frac{d_r d_r'}{p_r' d_r} - \frac{d_r p_r' K_r + K_r p_r d_r'}{p_r' d_r}$$

⁷ Para mayor información sobre este y otros algoritmos de optimización numérica ver Hendry (1995)

Donde

$$p_r = \nabla S(\hat{\theta}^{r+1}) - \nabla S(\hat{\theta}^r)$$

y

$$d_r = \hat{\theta}^{r+1} - \hat{\theta}^r$$

La concavidad mide el cambio de la primera derivada, es decir, la curvatura de la función original $S(\hat{\theta}^r)$. Esta información adicional de la función en los parámetros $\hat{\theta}^r$ ayuda a los algoritmos a obtener el tamaño de los movimientos en $\hat{\theta}^r$ para obtener un mínimo local, haciendo que los pasos sean grandes donde la función tenga una gran curvatura y dando pasos pequeños en las regiones planas de la función, pues seguramente se está cerca de un mínimo. En estos métodos la concavidad de la función dada por K_r modifica el tamaño del paso indicado por el gradiente $\nabla S(\hat{\theta}^r)$, logrando una mayor velocidad de convergencia ante apropiados valores iniciales (Hendry(1995)). Para la implementación en este trabajo de la metodología BFGS se usa la subrutina NLPQN⁸ en SAS versión 9.1.

Dado que los procesos de optimización no lineales son procesos iterativos, dependen de un vector inicial $\hat{\theta}^0$. Uno de los problemas adicionales de los modelos no lineales es la existencia de muchos mínimos y máximos locales; de hecho, la posibilidad de alcanzar el mínimo global disminuye en la medida en que se tienen más variables involucradas en el proceso de optimización. De esta forma, para aumentar la posibilidad de obtener un mínimo global o un mínimo local con menor $S(\hat{\theta}^*)$ se recomienda escoger varios vectores $\hat{\theta}^0$. En el presente trabajo se escogieron 30 vectores de parámetros iniciales.

Frances y van Dyk (2000) proponen dos transformaciones a los datos para mejorar las propiedades numéricas de los estimadores. La primera consiste en transformar todos los valores de la red y llevarlos al intervalo [0,1]:

$$\text{Ec. 3.22} \quad z_i = \frac{y_i - \min(y_i)}{\max(y_i) - \min(y_i)}$$

La segunda transformación normalizará todos los datos para que tengan una distribución normal estándar, de tal forma que tengan una media equivalente a cero y una desviación estándar igual a uno:

$$\text{Ec. 3.23} \quad z_i = \frac{y_i - \text{media}(y_i)}{\text{desviación estandar}(y_i)}$$

En este trabajo se utilizan las variables normalizadas. Por otro lado, para evitar que los parámetros de la red tengan valores muy grandes se propone usar una función objetivo que penalice valores muy altos; esta función objetivo es conocida como *Weight Decay* y será:

$$\text{Ec. 3.24} \quad S(\theta) = \sum_{t=1}^n [y_t - f(X_t, \theta)]^2 + r_\varphi \sum_{i=0}^k \varphi_i^2 + r_\beta \sum_{j=1}^q \beta_j^2 + r_\gamma \sum_{j=1}^q \sum_{i=0}^p \gamma_{ij}^2$$

⁸ Non Linear Optimization by Quasi-Newton Method

Frances y van Dijk (2000) recomiendan usar $r_\varphi=0.01$ y $r_\beta = r_\gamma = 0.0001$. En este trabajo se usan estos valores para encontrar los parámetros óptimos, controlando el valor de éstos y dejando con gran valor únicamente los parámetros que aumentan el poder explicativo de la RNA al reducir la suma de los errores al cuadrado.

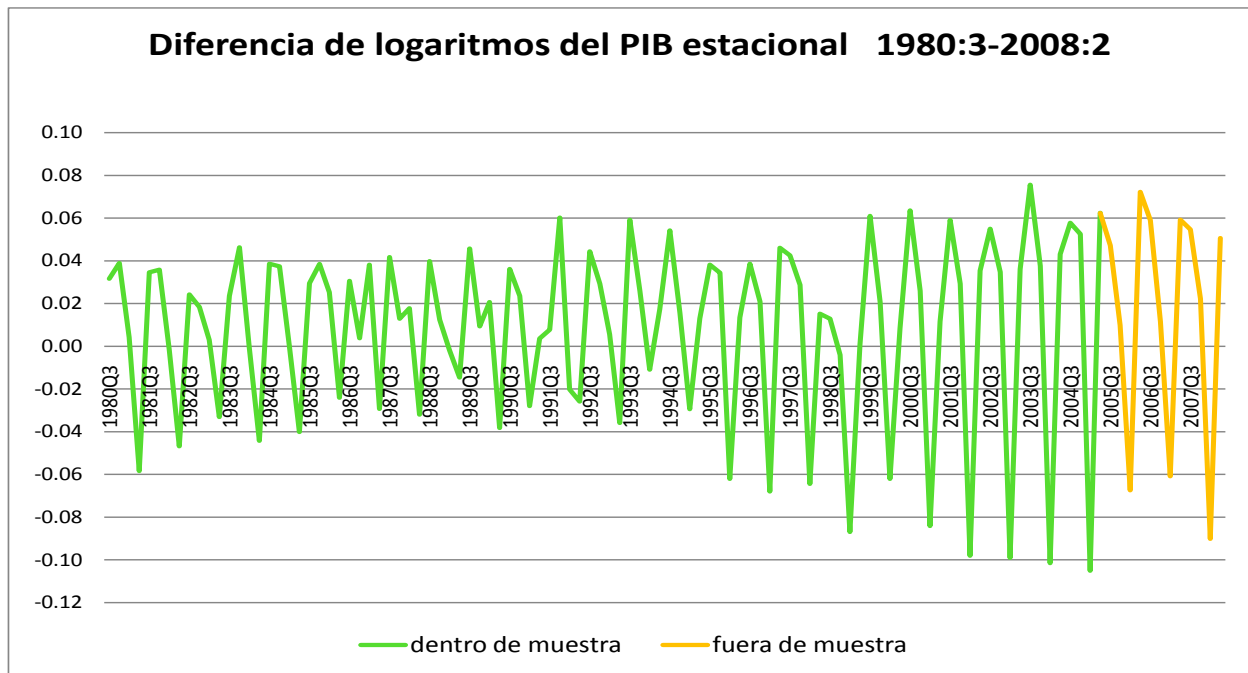
Una de las propiedades más interesantes de las RNA puede suponer un gran riesgo en la estimación: la propiedad universal de aproximación de las redes neuronales. Cuando la red captura todo el comportamiento de la variable de salida se dice que la red esta sobre entrenada pues los cambios de regímenes capturados a través de las funciones de activación pueden llegar a explicar comportamientos atípicos, quitándole el poder predictivo a la RNA. Para evitar el sobre entrenamiento, la estimación se realiza dividiendo la muestra en dos periodos mutuamente excluyentes: la muestra de entrenamiento o dentro de muestra y la muestra de evaluación de pronóstico o fuera de muestra; en inglés se llaman *in-sample* y *out-sample*. Esta división permite hacer una validación cruzada sobre el comportamiento de una RNA particular y su capacidad para hacer pronósticos consistentes con la realidad.

3.3.1 Evaluación por dentro de muestra

Para la elección de las variables de entrada (*inputs*) se usa la metodología *step-wise* propuesta por Swanson y White (1995). En la implementación de esta metodología se usan diferentes enfoques que trae programados SAS en su versión 9.1: *stepwise selection*, *forward entry* y *backward removal*. Estos hacen uso de criterios de decisión como el R^2 , el R^2 ajustado, el estadístico de Mallow y la significancia de las variables para seleccionar de un conjunto de candidatos las variables de entrada que tengan un mayor explicativo sobre las variables de salida.

Teniendo un conjunto de variables candidatas, la metodología *forward* (hacia adelante) adiciona una variable por paso. Esta variable es escogida teniendo en cuenta los criterios de decisión, que midan el poder explicativo de las variables; se incluye la variable que genere mejores criterios. Este proceso adiciona al modelo variables hasta que la introducción de una nueva variable no mejore significativamente los criterios de decisión de acuerdo con un valor crítico especificado. La metodología *backward* (hacia atrás) parte de un modelo con todas las variables candidatas y paso a paso remueve la variable que no muestre tener un impacto significativo sobre los criterios de decisión; si se llega a un punto donde la eliminación de una variable reduce los criterios significativamente el proceso concluye. Para la elección de las variables de entrada en el presente trabajo se escogieron las variables seleccionadas a través de diferentes métodos. Cabe aclarar que la búsqueda de las variables a través de estos métodos se hace con modelos lineales.

Gráfico 3.5



Con las variables de entrada seleccionadas y una vez se ha escogido el tamaño de las observaciones dentro de muestra, se estimarán los parámetros de cada una de las diferentes arquitecturas. El tamaño escogido para este trabajo son 92 observaciones dentro de muestra y 12 por fuera de muestra. El complejo proceso de estimación inicia una vez se han transformado las variables a través de la estandarización, haciendo un muestreo aleatorio de los parámetros $\hat{\gamma}_{ij}$. Para este propósito Frances y van Dijk (2000) recomiendan usar una distribución uniforme con valores entre -2 y 2; así, los parámetros iniciales de los *inputs* a las unidades escondidas se distribuirán $\hat{\gamma}_{ij}^0 \sim U(-2,2)$. Como la función objetivo puede presentar diversos mínimos locales, la convergencia de los parámetros no significa que se haya encontrado un mínimo global. Por esta razón, se recomienda utilizar diferentes vectores iniciales de parámetros $\hat{\gamma}_{ij}^0$; partiendo de cada vector y a través de mínimos cuadrados ordinarios (MCO) se estiman los valores iniciales $\hat{\beta}_j^0$ y $\hat{\varphi}_i^0$, conformando de esta manera el vector inicial $\hat{\theta}^0$, para iniciar el proceso iterativo de optimización de la función *weight decay*.

Una vez que cada sucesión de parámetros haya confluído a algún valor, se escogen aquellos $\hat{\theta}^*$ que tienen un gradiente cercano a cero; cada vector de parámetros se organizan de menor a mayor de acuerdo a la función objetivo, encontrando los mínimos con menor valor de la función objetivo. En este trabajo se escogen solo aquellos parámetros con gradiente menor a 0.1 en todos sus componentes $\nabla S(\hat{\theta}^*) < 0.1$ y se conservan solo los cinco vectores de parámetros con menor función objetivo.

Con el conjunto de parámetros escogidos por arquitectura, se evaluará el comportamiento de cada uno de éstos dentro de muestra evaluando el poder explicativo y la bondad de ajuste que cada modelo obtiene a través de criterios de información como el de AIC y el BIC, con criterios de bondad de ajuste como el R-cuadrado y el SR, y a través de medidas de bondad de pronóstico como el MAE, MAPE, RMSPE y el RMSE.

De esta manera, se puede obtener la arquitectura con el número óptimo de variables escondidas y de variables de entrada. Sin embargo, esta red con el mejor comportamiento dentro de muestra no necesariamente es la mejor fuera de muestra, pues una red sobre entrenada puede perder su poder predictivo.

3.3.2 Evaluación por fuera de muestra

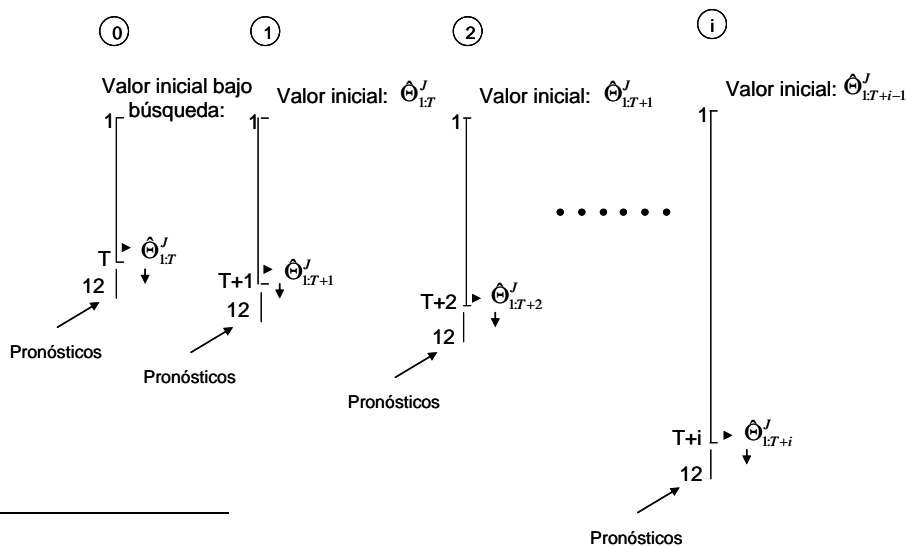
En este trabajo se cuenta con 12 observaciones por fuera de muestra, para evaluar el comportamiento de cada arquitectura. Frances y van Dijk (2000) utilizan los parámetros escogidos dentro de muestra y con la información disponible por fuera de muestra se realizan pronósticos de la variable de salida. Con los valores efectivamente observados se realiza la evaluación de pronóstico a través de los criterios ya mencionados para encontrar las arquitecturas que presenten el mejor comportamiento por fuera de muestra o, lo que es igual, aquellas que tengan un mayor poder predictivo.

Como señalan los autores, no es muy claro cuál debe ser el balance entre el comportamiento de una arquitectura dentro de muestra y por fuera de muestra para escoger la red que tenga un mejor comportamiento en la muestra completa y, por otro lado, este tipo de evaluación no mide el grado de versatilidad de cada red para ajustarse a nueva información presentada en el periodo por fuera de muestra.

Teniendo en cuenta estos problemas se seguirá la metodología usada por Misas y Jalil (2005), en la cual se utiliza estimación recursiva de los parámetros por fuera de muestra. En dicha metodología se mide el “aprendizaje” de cada vector de parámetros por arquitectura ante la introducción de información contenida por fuera de muestra. Esta metodología conocida como *Rolling*⁹, funciona usando como vector de parámetros iniciales los parámetros encontrados por dentro de muestra; partiendo de éstos, se identifican nuevos modelos reestimando los parámetros un número de veces, y con cada modelo estimado se generan pronósticos otro número de periodos hacia adelante.

La metodología *Rolling* adoptada en este trabajo funciona de la siguiente manera: con el vector de parámetros estimado dentro de muestra $\hat{\Theta}_{92}^*$ se realizan b pronósticos hacia adelante desde la observación 93 hasta la observación $92+b$. Véase gráfico 3.6 tomado del trabajo de Jalil y Misas (2005)¹⁰:

Gráfico 3.6



⁹ Metodología también usada por Torres (2005) y Aristizábal (2006).

¹⁰ En el trabajo de Jalil y Misas se usa $h = 12$.

Posteriormente, se ingresa la observación 93, se reestima el vector de parámetros minimizando el *weight decay* con la observación adicional, tomando como valor inicial $\hat{\theta}_{92}^*$ y obteniendo como resultado $\hat{\theta}_{93}^*$; con este vector de parámetros se realizan b pronósticos hacia adelante desde la observación 94 hasta la observación $93+b$.

Este proceso continúa reestimando un vector de parámetros cada vez que una observación adicional es introducida y proyectando pronósticos b periodos hacia adelante a partir de cada observación introducida. El proceso termina con la penúltima observación disponible por fuera de muestra; con ésta se obtiene el vector $\hat{\theta}_{103}^*$ y se realizan pronósticos desde la observación 104 hasta la observación $103+b$. Dado que el último dato observado efectivamente es la observación 104, no tiene sentido realizar pronósticos desde esta observación, pues no sería posible evaluar la calidad de estos pronósticos al no existir más datos disponibles contra los cuales ser evaluados.

En este trabajo con cada vector de parámetros se realizan pronósticos ocho periodos adelante; de esta manera $b=8$. Al finalizar el *Rolling* se cuenta con cinco matrices de 8×12 por cada arquitectura. Son cinco matrices porque en la estimación dentro de muestra se escogieron los cinco vectores de parámetros que optimizaban con menor valor la función objetivo. Cada una de estas matrices está compuesta por los pronósticos realizados con cada vector de parámetros obtenidos en el *Rolling*; las filas de esta matriz son los 8 periodos pronosticados, y las 12 columnas indican cada una de las reestimaciones realizadas fuera de muestra.

Tabla 3.1: Matriz de pronósticos

	$\hat{\theta}_{92}^* = f(X_{92})$	$\hat{\theta}_{93}^* = f(X_{93})$...	$\hat{\theta}_{103}^* = f(X_{103})$
1 periodo hacia adelante	$y_{est}^{93} = f(X_{92}, \hat{\theta}_{92}^*)$	$y_{est}^{94} = f(X_{93}, \hat{\theta}_{93}^*)$...	$y_{est}^{104} = f(X_{103}, \hat{\theta}_{103}^*)$
2 periodos hacia adelante	$y_{est}^{94} = f(X_{92}, \hat{\theta}_{92}^*, y_{est}^{93})$	$y_{est}^{95} = f(X_{93}, \hat{\theta}_{93}^*, y_{est}^{94})$...	$y_{est}^{105} = f(X_{103}, \hat{\theta}_{103}^*, y_{est}^{104})$
⋮	⋮	⋮	⋮	⋮
8 periodos hacia adelante	$y_{est}^{100} = f(X_{92}, \hat{\theta}_{92}^*, y_{est}^{93}, \dots, y_{est}^{99})$	$y_{est}^{101} = f(X_{93}, \hat{\theta}_{93}^*, y_{est}^{94}, \dots, y_{est}^{100})$...	$y_{est}^{111} = f(X_{103}, \hat{\theta}_{103}^*, y_{est}^{104}, \dots, y_{est}^{109})$

En la tabla se observa cómo cada vector de parámetros es función de las observaciones disponibles para su estimación. En la columna 2 se encuentran los 8 pronósticos realizados desde la reestimación del vector de parámetros con 93 observaciones. Puede observarse que en cada periodo del *Rolling* los pronósticos dependen de los valores predichos anteriormente.

Esta metodología busca capturar cuáles vectores parámetros, correspondientes a diferentes arquitecturas y a diferentes repeticiones de cada arquitectura, responden mejor ante la nueva información y, adicionalmente, poder discriminar sobre los modelos que generen mejores pronósticos a diferentes periodos adelante. Por

ejemplo, puede ser que la tercera replicación de $\hat{\theta}_{2,3}^*$ genere los mejores pronósticos tres periodos hacia adelante, pero la cuarta replicación de $\hat{\theta}_{5,4}^*$ sea la que genere mejores pronósticos ocho periodos adelante.

Para la evaluación de pronósticos se pueden emplear las medidas usadas dentro de muestra o como lo hacen Misas y Jalil (2005) y Aristizábal (2006) funciones de pérdida asimétrica. En el presente trabajo se evaluarán los pronósticos con funciones de pérdida simétrica.

3.4. Análisis de sensibilidad de una RNA

Una pregunta usual en los modelos econométricos versa sobre la significancia estadística de las variables, esto es, indagar sobre la relevancia de cada una de las variables explicativas sobre la variable explicada; este proceso, bastante frecuente en los modelos lineales, debe usar análisis de sensibilidad en las RNA. Para tal efecto, Misas, López y Querubín (2002) desconectan las variables de entrada de las variables de salida, haciendo los coeficientes de conexión iguales a cero. De esta forma, es posible evaluar la contribución de cada variable sobre el poder explicativo del modelo.

Por otro lado, si se tienen en cuenta todos los posibles ordenamientos para introducir paso a paso las variables en las unidades escondidas, se tienen $k!$ posibles ordenamientos. De estos $k!$ posibles ordenamientos, se trabaja con el orden determinado por la metodología *stepwise* reduciendo las posibilidades a una sola. Utilizar el análisis de sensibilidad, garantiza considerar modelos con configuraciones de variables no posibles con el orden ya establecido.

Partiendo del ordenamiento inicial, la desconexión se puede hacer de tres maneras: i.) desconectar la variable de la parte lineal y no lineal, ii.) desconectar la variable de la parte lineal y iii.) desconectar la variable de la parte no lineal. De esta forma, la conexión y desconexión de los parámetros nos permite identificar las variables, que según el ordenamiento del *stepwise* y en una arquitectura determinada, son irrelevantes en la estructura de un modelo dado.

Así, partiendo del ordenamiento dado por el *stepwise*, el análisis de sensibilidad permite identificar la importancia de las variables en los modelos y considerar arquitecturas con variables diferentes a las iniciales dentro y fuera de muestra.

4. Aplicación sobre el PIB en Colombia a través de las Redes Neuronales Artificiales

En este punto se ha decidido usar las Redes Neuronales Artificiales como proceso generador de datos del crecimiento trimestral del PIB en Colombia. Para elegir las variables independientes se usan las metodologías propuestas por Swanson y White (1995) ya explicadas con anterioridad. Los resultados de este proceso se encuentran en el anexo 1.

La información empleada en este trabajo corresponde a la primera diferencia de logaritmo del PIB (DLPIB) y la primera diferencia de la tasa de interés de los CDT a 90 días (DLCDT). Siguiendo a Frances y Dijk (1999), las dos series son normalizadas para que tengan media igual a cero y varianza igual a uno (NDLPIB y NDLCDT). Esto se hace para mejorar las propiedades numéricas de los estimadores haciendo que la sucesión de parámetros obtenida en el proceso de optimización numérica tiendan a algún mínimo. Para la red neuronal autoregresiva se consideran 8 rezagos de NDLPIB como posibles variables explicativas; a través de los tres

¹¹ k factorial, donde k es el número de variables presentes en la parte lineal.

métodos empleados (step-wise, forward y backward) se encontró convergencia sobre las cinco variables a introducir en el modelo. Estas fueron: los rezagos 1, 3, 4, 6 y 8 del NDLPIB. Dada la naturaleza de la metodología forward, que introduce paso a paso las variables con mayor poder explicativo, se siguió el orden de entrada de las variables considerados en esta técnica. El orden es: primero entra el rezago 4, y, sucesivamente, 8, 6, 1 y 3. Para la red neuronal con CDT se consideran 8 rezagos de NDLPIB y de NDLCDT como posibles variables explicativas. Los tres métodos empleados convergen sobre las variables que debían ingresar en el modelo; sin embargo, forward y step-wise encuentran que los rezagos 4, 8 del NDLPIB, y, 1 y 3 de NDLCDT deben ingresar en el modelo; la estrategia backward encuentra estas variables y además los rezagos 1 y 5 del NDLPIB. Debido a la compleja forma funcional de las redes y que éstas buscan identificar relaciones no lineales, se usan las variables presentes en los tres métodos y se sigue el orden de inserción en el modelo arrojado por la metodología forward: rezago 4 del NDLPIB, 1 de NDLCDT, 3 de NDLCDT y 8 del NDLPIB.

En la búsqueda de la variable que refleje el estado de la política monetaria se practicaron las tres metodologías de selección de variables a un modelo que tendría 8 rezagos del NDLPIB, NDLCDT y NDLM1 como variables explicativas. Como se ve en el anexo 1 la metodología forward encuentra exactamente las mismas variables que fueron encontradas en la especificación que contenía a rezagos del NDLPIB y NDLCDT; este resultado confirma la evidencia encontrada para Colombia por Reinhart y Reinhart (1991) y para Estados Unidos por Sims (1980) sobre la pérdida de poder explicativo de los agregados monetarios sobre el PIB al incluir algún tipo de tasa de interés en el modelo¹².

Debido a que en los dos modelos se tiene como máximo 8 rezagos del NDLPIB, en la estimación de los modelos se pierden estas observaciones. Es por esta razón que de las 112 observaciones iniciales para la estimación de las redes sólo se cuenta, finalmente, con 104 observaciones.

Para la red autoregresiva el conjunto X que contiene a todas las variables es $X = \{NDLPIB_{t-1}, NDLPIB_{t-3}, NDLPIB_{t-4}, NDLPIB_{t-6}, NDLPIB_{t-8}\}$. Las variables escogidas en cada red conforman las k variables que estarán presentes en la especificación lineal de todas las arquitecturas (cinco en el modelo autoregresivo). Sucesivamente se introduce una variable en la parte no lineal hasta que $p=k$. Por esta razón, establecer el orden de la sucesión es importante y generar criterios de elección sobre este orden también ahorra tiempo. Por ello, en este trabajo se asegura que el orden mediante el cual las variables ingresan en la superficie escondida responden al poder explicativo de éstas en el mundo lineal.

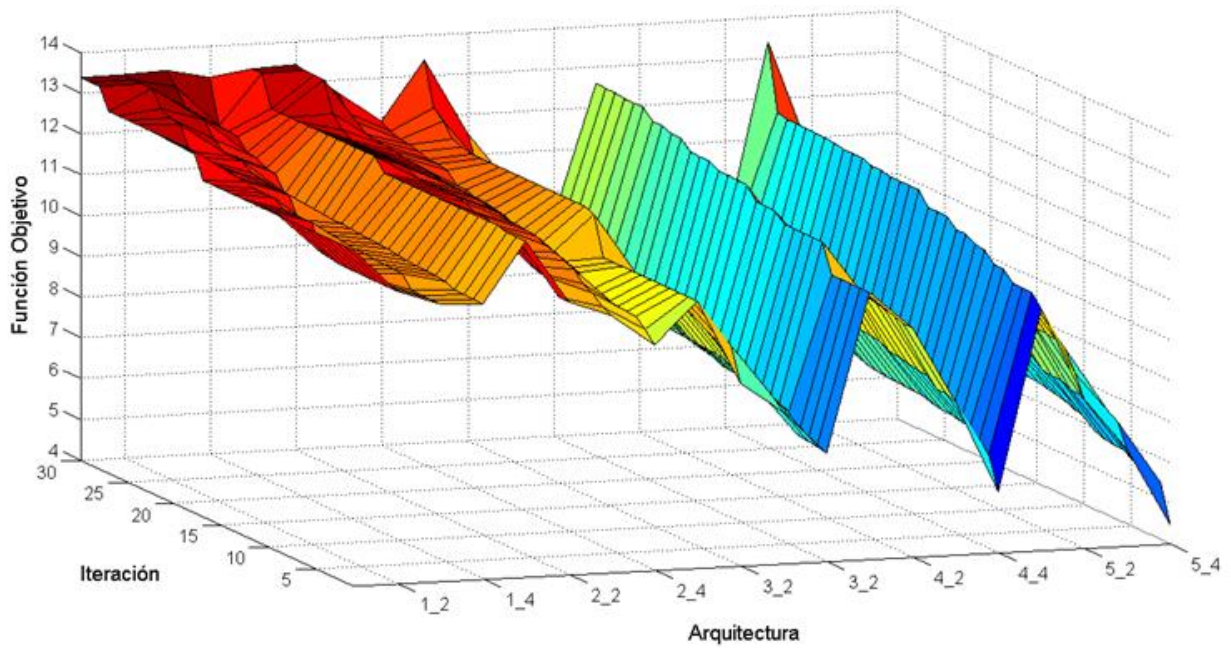
4.1 Evaluación dentro de muestra de las Redes Neuronales Artificiales aplicadas al PIB

Debido a que las RNA son funciones no lineales para aumentar la probabilidad de encontrar un mínimo global, se hicieron 30 repeticiones por arquitectura a través de diferentes valores iniciales de los parámetros γ . Una vez terminada la minimización de la función objetivo (*Weight Decay*), y de depurar aquellos parámetros que tengan un gradiente en alguna componente mayor a 0.1, se tienen 600 modelos para la red autoregresiva que cuenta con 5 variables y 480¹³ modelos para la red con CDT que cuenta con 4 variables.

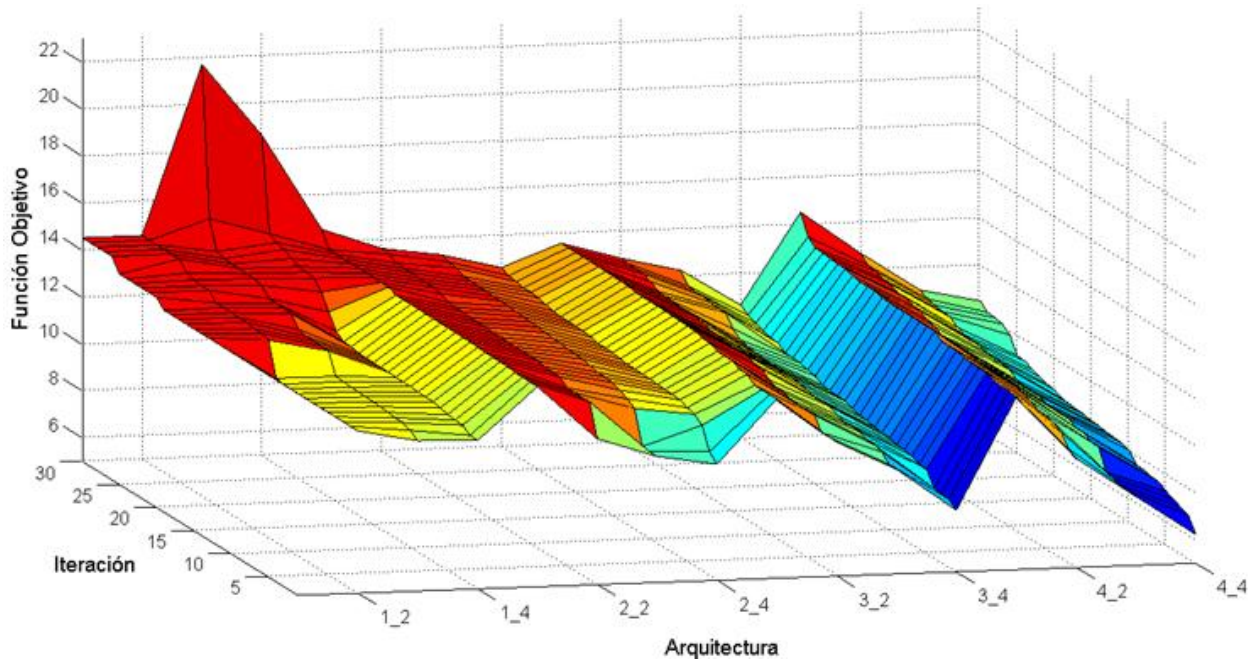
¹² La estrategia Step-wise confirma estos resultados; dichos resultados no son presentados en este trabajo por no ser considerados de relevancia para el desarrollo del mismo.

¹³ En este trabajo y en la mayoría de las aplicaciones económicas se usan máximo cuatro unidades escondidas. Estos números surgen debido a que el número de arquitecturas en una red viene dado por $k*q$; por consiguiente, en el caso autoregresivo tendremos 20

Grafica 4.1: Función Objetivo Red Neuronal Autoregresiva



Grafica 4.2: Función Objetivo Red Neuronal Aumentada (con CDT)



En las gráficas 4.1 y 4.2 se observa claramente que el hecho de que los parámetros converjan no implica que los parámetros encontrados correspondan a un mínimo global. Para continuar con el proceso de estimación y

arquitecturas y, como para cada arquitectura se realizaron 30 iteraciones, entonces surge el número 600 como el total de modelos estimados.

elección de la mejor red para realizar pronósticos del PIB, se escogen por arquitectura los cinco parámetros con menor función objetivo, reduciendo la búsqueda de 600 modelos en la especificación autoregresiva a 100, y de 480 en la especificación con CDT a 80 modelos. Gráficamente el proceso continúa con las cinco iteraciones que se ubican en la parte delantera de las dos graficas.

Como el interés de este trabajo es encontrar el modelo que mejor aproxime el comportamiento de la tasa de crecimiento del PIB, para la evaluación dentro de muestra los datos deben ser destransformados para contrastar el valor observado con el valor dado por el modelo en las mismas unidades. En el segundo anexo se presentan las formas funcionales de las medidas usadas en este trabajo; en el tercer anexo se observan los resultados numéricos de las estimaciones dentro de muestra.

Se escogen los cuatro modelos que presentan un mejor desempeño dentro de muestra de acuerdo a cada medida; para el modelo autoregresivo la arquitectura 5_4¹⁴ está presente en cinco medidas, siendo la primera bajo AIC, R2, RMSE, MAE y la cuarta en SR. La arquitectura 4_4 está presente en cuatro medidas: AIC, R2, RMSE y SR; se ve claramente que las arquitecturas con tres variables en las unidades escondidas aparecen en tres de los siete criterios (BIC, RMSPE y MAPE). La cantidad de unidades escondidas en este caso está muy repartido; la arquitectura 3_1 domina completamente el criterio BIC, mientras que las arquitecturas con dos, tres y cuatro unidades escondidas dominan el RMSPE y MAPE. Otras dos arquitecturas que están presentes en las medidas son 5_3 en AIC y 4_3 en SR.

Para el modelo aumentado, la arquitectura 4_4 aparece en cinco medidas, domina el AIC, R2, RMSE, MAE y aparece en SR; las arquitecturas 4_3 y 4_2 aparecen cada uno en dos medidas AIC, SR y RMSPE, MAPE respectivamente. La arquitectura 3_4 domina RMSPE y MAPE.

De los anteriores resultados se puede concluir que los modelos más complejos presentan un mejor comportamiento dentro de muestra; esto se cumple excepto en el criterio BIC del modelo aumentado donde domina completamente la arquitectura más simple 1_1. Se observa que las arquitecturas más complejas en ambos modelos (5_4 en el modelo autoregresivo y 4_4 en el modelo con CDT) están presentes en cinco de las siete medidas. Así, un mayor número de unidades escondidas y variables en la parte no lineal permite capturar con mayor exactitud el comportamiento del crecimiento del PIB. Sin embargo, como se explicó anteriormente y como se verá más adelante, a menudo un excelente ajuste dentro de muestra implica una sobre especificación que hace que los modelos logren aprenderse el comportamiento de una serie, generando pronósticos errados.

4.2 Especificación de un modelo ARIMA para el PIB en Colombia

Los trabajos que usan redes neuronales para explicar otras variables referenciados en la segunda sección, estiman modelos lineales dado que la hipótesis de no linealidad debe ser contrastada con algún modelo lineal. Para la aplicación sobre el PIB estacional en Colombia se estimó un modelo ARIMA siguiendo la metodología de Box-Jenkins¹⁵. En los modelos ARIMA las variables no deben ser transformadas, pues las propiedades matemáticas son más simples en relación a las redes neuronales. De esta forma, el modelo se especifica mediante DLPIB. Para comparar los pronósticos entre los tres modelos, se dejan los mismos doce periodos

¹⁴ 5_4 indica que $p=5$, $q=4$; cuando en el presente trabajo se encuentren más adelante referencias a arquitecturas como 5_4_3 se estará refiriendo a $p=5$, $q=4$ en la tercera replicación.

¹⁵ Para la estimación del modelo ARIMA se siguió a Guerrero (2003) y a Peña (2005).

fuera de muestra. Se emplea la metodología *Rolling* para la realización de pronósticos fuera de muestra y de esta forma se logra que los tres modelos compitan en las mismas condiciones.

Los resultados del modelo ARIMA se encuentran en el anexo cinco. El modelo especificado cuenta con la cuarta diferencia estacional, la parte autoregresiva cuenta con los rezagos uno y seis y la parte MA tiene el cuarto rezago. En el anexo cinco se muestran tres periodos del *Rolling* constatando la estabilidad del modelo y verificando que los errores son ruido blanco a lo largo de los periodos fuera de muestra.

4.3 Resultados de la evaluación fuera de muestra

El periodo fuera de muestra corresponde a los periodos comprendidos entre el tercer semestre de 2005 y el segundo trimestre de 2008. Para evaluar los pronósticos se usaron cuatro medidas de evaluación de pronóstico RMSE, MAE, RMSPE y MAPE. Dentro de los modelos escogidos por diferentes criterios se observan arquitecturas que en la evaluación dentro de muestra no eran contempladas; para el modelo aumentado, arquitecturas como 1_4_5 tienen el mejor desempeño en los periodos 4, 5 y 6 y para la red autoregresiva bajo los criterios MAE y RMSE. El problema de la sobre especificación de las redes neuronales se hace evidente en la evaluación fuera de muestra. En la tabla 4.1 se muestran los cuatro modelos que obtuvieron un menor RMSE; en esta tabla no aparecen las especificaciones más complejas (5_4 y 4_4). En el cuarto anexo están las tablas de la evaluación fuera de muestra organizadas por MAE, RMSPE y MAPE. Se observa que en la evaluación medida a través de RMSPE para el modelo aumentado, la arquitectura 4_4 está presente en los periodos cinco y seis; en las otras dos medidas las arquitecturas complejas no son favorecidas. Mientras en la evaluación dentro de muestra estas especificaciones estaban presentes en cinco de las siete medidas, en la evaluación fuera de muestra aparecen en un criterio de cuatro y en dos periodos de ocho para el modelo aumentado.

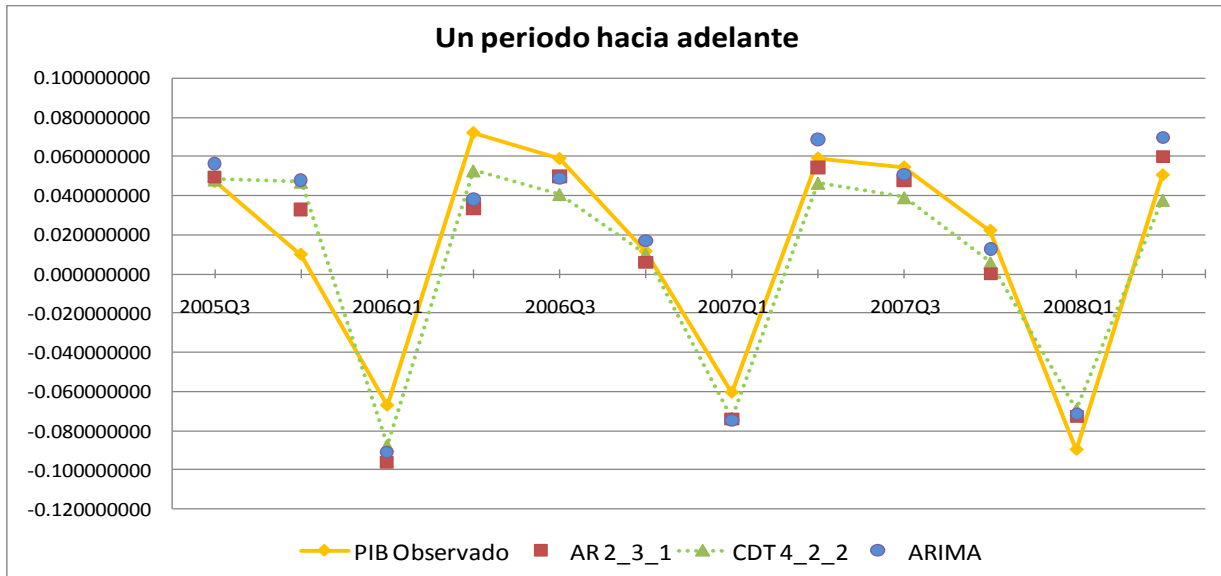
Esto comprueba la necesidad de generar dos muestras diferentes, partiendo la muestra completa en dos partes para hacer una validación cruzada sobre la flexibilidad de las diferentes redes ante diferente información. Una vez que se escoge una red neuronal para realizar pronósticos se debe garantizar que dicha red sea la que mejor responda ante cambios en el entorno económico. El fin de utilizar en este caso la metodología *rolling* consiste en generar un ambiente controlado que permita al investigador simular las condiciones a las cuales la red se enfrentará por fuera de la muestra estudiada, permitiendo escoger una red que no esté sobre entrenada y que haya mostrado el mejor desempeño para explicar el comportamiento de la serie en estudio.

Recuérdese que en las redes neuronales las variables han sido transformadas para que tengan media igual a cero y varianza igual a uno. Antes de realizar las medidas los pronósticos de la red deben ser destransformados; una vez destransformados, los pronósticos de las redes se pueden comparar con los pronósticos del modelo ARIMA que genera los datos en diferencia de logaritmos.

Las tablas de las diferentes medidas tienen subrayadas los dos mejores modelos por horizonte de pronóstico comparando los tres modelos: ARIMA, red neuronal autoregresiva y red neuronal ampliada. Dado que la metodología *rolling* permite identificar qué modelo pronostica mejor a diferentes horizontes, para realizar pronósticos un periodo hacia adelante el RMSE muestra que la red con CDT 4_4_2 es la que logra el mejor desempeño, seguida de la red autoregresiva 2_3_1. El modelo ARIMA solo supera a las especificaciones no lineales en dos periodos y en dos medidas: el periodo 3 bajo RMSPE y el periodo 8 bajo MAPE. El ARIMA es el segundo mejor modelo en el periodo 8 bajo la especificación RMSPE. Sin embargo, si consideramos el total de los posibles primeros lugares tenemos 32 escenarios, en donde en solo dos de ellos la especificación

ARIMA supera a las redes neuronales artificiales; esta clara evidencia muestra que las redes neuronales superan ampliamente a la especificación lineal para explicar el crecimiento del PIB.

Gráfica 4.3



En el desarrollo del trabajo se encontró que las medidas de evaluación de pronóstico porcentuales al ser realizadas sobre crecimientos son inadecuadas. En el dominio de los crecimientos se encuentran los números reales, siendo los valores extremadamente grandes muy extraños o casi inexistentes; en efecto, el cero está contenido en este dominio. Las medidas como RMSPE y MAPE parten de:

$$\text{Ec. 4.1} \quad \frac{\hat{y}_t - y_t}{y_t} \quad 16$$

Si en algún periodo la tasa de crecimiento es cero, estas medidas no existirán; por otro lado, si tenemos tasas de crecimiento cercanas a cero en algunos momentos, las desviaciones de los pronósticos en estos periodos en particular serán penalizados más que las desviaciones en otros periodos. Por estas razones no se escogieron estas medidas para presentar los resultados gráficos. Para futuros trabajos este comportamiento de las medidas podría ser estudiado con mayor detalle.

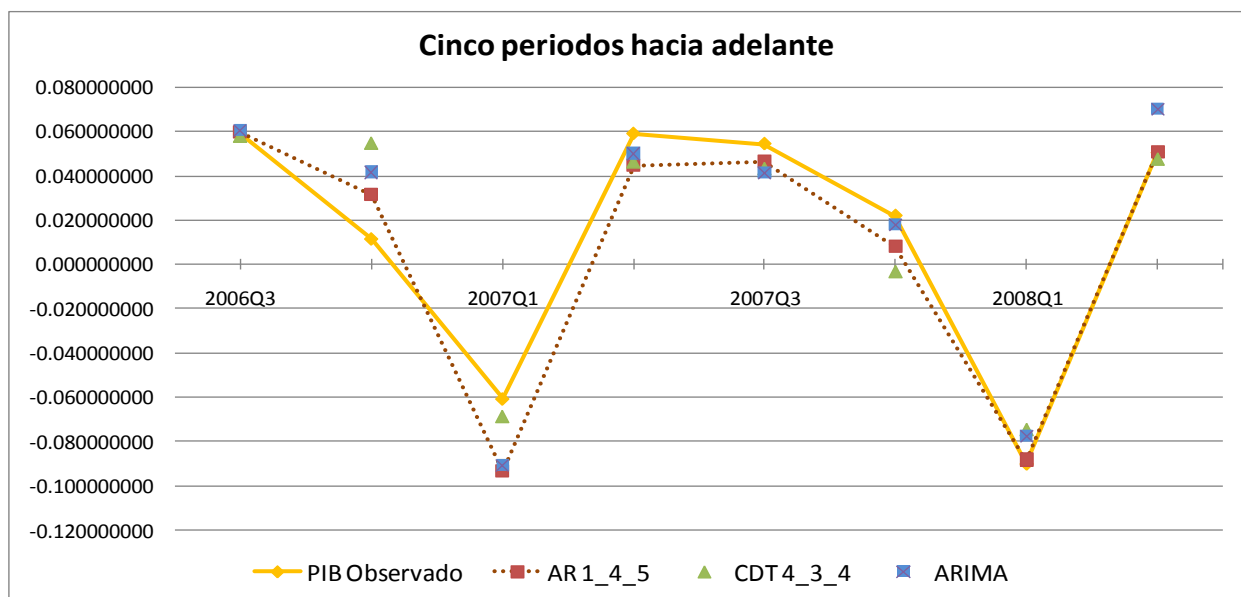
¹⁶ Ver anexo 2.

Tabla 4.1

RMSE	Horizonte	ARIMA	REDES NEURONALES ARTIFICIALES							
			Autoregresivo				Con CDT			
	1	0.01929	2_3_1 0.01855	1_4_3 0.01944	2_2_4 0.01947	2_2_3 0.01948	4_2_2 0.0181	2_1_2 0.01868	2_1_4 0.01868	2_1_1 0.01868
	2	0.01983	2_4_3 0.01831	2_3_1 0.01858	5_4_2 0.01881	4_3_2 0.01929	4_1_4 0.01888	4_2_2 0.01897	4_3_4 0.01915	2_1_3 0.0195
	3	0.01675	2_3_4 0.01695	2_3_3 0.01768	4_3_4 0.0179	5_1_5 0.01807	4_3_4 0.01318	4_1_4 0.01564	4_2_2 0.0157	4_1_1 0.0159
	4	0.01391	2_3_3 0.01204	2_4_4 0.01369	2_4_3 0.01369	2_4_2 0.01477	3_3_3 0.01158	4_2_3 0.01254	4_3_4 0.01363	2_1_5 0.01388
	5	0.01816	1_4_5 0.01568	1_4_1 0.01569	1_4_2 0.01572	2_3_5 0.01578	4_3_4 0.01978	3_1_5 0.02103	1_4_4 0.02152	1_3_1 0.02161
	6	0.02008	1_4_5 0.01589	1_4_2 0.01648	2_3_5 0.01654	1_4_3 0.01669	4_3_4 0.02138	3_1_5 0.02265	3_4_2 0.02301	4_2_2 0.02331
	7	0.0189	1_4_5 0.01627	2_4_2 0.01642	2_3_5 0.01664	1_4_3 0.01711	4_3_4 0.01438	1_4_4 0.01884	4_4_3 0.01886	4_1_4 0.01935
	8	0.01015	1_4_2 0.00696	1_4_5 0.00726	2_3_3 0.00772	2_4_3 0.00916	1_4_4 0.00758	3_1_5 0.00924	1_4_3 0.00992	1_3_4 0.01031

La metodología *Rolling* busca identificar el modelo que genere mejores pronósticos cuando se recibe nueva información en diferentes periodos de tiempo. Las gráficas 4.3 y 4.4 muestran la serie observada y los pronósticos de los tres modelos para cada uno de los periodos en cuestión; la línea punteada indica aquel modelo que tuvo el mejor desempeño según el RMSE. Este criterio fue escogido, teniendo en cuenta que penaliza duramente las desviaciones más grandes de los pronósticos respecto al valor observado; empero, cualquier otro criterio puede ser empleado. En el anexo seis se presentan las graficas para los demás horizontes de pronóstico.

Gráfica 4.4



Para los demás periodos, las otras medidas indican que el modelo autoregresivo supera ampliamente a los otros dos modelos. Dado que se seleccionaron los dos mejores modelos por cada medida en cada horizonte del tiempo, resultan 64 posibles selecciones. El modelo ARIMA es seleccionado en solo 3 selecciones, la red aumentada en 13 ocasiones y la red neuronal autoregresiva en 48 ocasiones. Esta evidencia nos muestra que la red neuronal autoregresiva supera ampliamente a los demás modelos empleados, en la mayoría de los periodos.

En la tabla 4.1 se observa que para los periodos 1, 3, 4 y 7 la red aumentada es mejor que la red autoregresiva; el método MAE favorece al modelo con CDT en los periodos 1, 2 y 3; RMSPE muestra que el modelo aumentado supera a la red autoregresiva en el periodo 3. Se podría concluir que la información aportada al modelo por la injerencia de la tasa de interés de los CDT a 90 días mejora el poder predictivo del mismo para el periodo 1 y sobre todo para el periodo 3.

4.4. Análisis de Sensibilidad

Para el análisis de sensibilidad se usan únicamente las arquitecturas expuestas en la tabla 4.1 y que fueron escogidas por su buen desempeño fuera de muestra siguiendo el criterio RMSE. En el anexo 7 se muestran los resultados de la desconexión dentro de muestra de las tres formas posibles (parte lineal y no lineal, lineal, y no lineal). Para la desconexión debe tenerse en cuenta el orden dado por el *stepwise*. Así, en la RNA autoregresiva el conjunto X en orden es $X = \{NDLPIB_{t-4}, NDLPIB_{t-8}, NDLPIB_{t-6}, NDLPIB_{t-1}, NDLPIB_{t-3}\}$; entonces, cuando la columna *hide* sea igual a 1, quiere decir que los parámetros asociados al rezago cuatro del NDLPIB es igual a cero, y así sucesivamente. Para la RNA aumentado el orden de las variables es $X = \{NDLPIB_{t-4}, NDLCDT_{t-1}, NDLCDT_{t-3}, NDLPIB_{t-8}\}$; de esta forma, cuando la columna *hide* sea igual a 3 quiere decir que los parámetros asociados al tercer rezago del NDLCDT es igual a cero.

4.4.1. Análisis de Sensibilidad dentro de muestra

Para la desconexión, en una arquitectura particular con p variables en las unidades escondidas, se define la variable *hide* que toma valores desde 1 hasta p , y que desconecta de la red los parámetros que están en la posición que *hide* toma, bien sea en la parte lineal o no lineal de la RNA.

Los resultados de la desconexión dentro de muestra se observan en el anexo 7; las tablas presentan los cinco mejores modelos de acuerdo a cada una de las medidas, considerando las diferentes formas de realizar la desconexión. Pese a que se encuentran algunos modelos competitivos, se observa que la capacidad de la red neuronal para explicar el comportamiento del PIB no mejora en el modelo autoregresivo, ni en el modelo aumentado, pues bajo ninguna medida, los modelos con desconexión llegan a ser los mejores.

En las tablas del anexo 7, se subrayan los modelos que según diferentes medidas se encuentran entre los cinco mejores, comparándolos con los modelos originales. La desconexión lineal y no lineal no logra tener ningún modelo entre los cinco mejores bajo ninguna medida. La desconexión no lineal solo logra que al desconectar el rezago uno de NDLPIB en el modelo autoregresivo 4_4_4 compita bajo el criterio SR.

El análisis de sensibilidad evidencia que debido al sobre entrenamiento que se da en la estimación dentro de muestra, la desconexión no lineal genera una pérdida de poder explicativo, ya que cada parámetro fue optimizado para hacer mínimos los errores de una muestra que no cambia. Aún más, las medidas muestran la importancia que tienen las superficies escondidas para las RNA, en el propósito de explicar cualquier serie en estudio.

En la desconexión lineal de la RNA aumentada, el modelo 4_4_2 se ubica entre los cinco mejores bajo los criterios RMSE, MAE y el R2, al desconectar el tercer rezago de NDLCDT. Así mismo, ubica otros dos modelos bajo los criterios MAE y SR. Por otro lado, en la RNA autoregresiva logra tener 2 modelos bajo MAE y 3 bajo SR. La arquitectura 5_4_1 logra estar entre los cinco mejores modelos bajo AIC, R2 y RMSE, al desconectar linealmente el rezago ocho de NDLPB.

4.4.2. Análisis de Sensibilidad fuera de muestra

El análisis fuera de muestra revela otros resultados, logrando mejoras significativas de diferentes RNA para realizar pronósticos. La sensibilidad fue evaluada sobre el esquema *rolling*, lo cual supone un gran avance en la aplicación de las RNA, siendo ésta la primera vez en el país que se realiza un análisis semejante. En el anexo 8, se muestran dos tablas que evidencian la forma como se realizó el análisis de sensibilidad. En la tabla 4.2 se muestran las arquitecturas que, finalmente, fueron seleccionadas.

El análisis fuera de muestra debe ser evaluado en cada uno de los ocho periodos considerados para todas las arquitecturas y con las diferentes desconexiones. De esta forma, al desconectar una variable, el nuevo modelo se evalúa para los ocho periodos y se comparan los valores de RMSE y MAE, de cada uno, con el mínimo valor encontrado para cada medida en los modelos sin desconexión, expuestos en el anexo 4 para MAE y en la tabla 4.1 para RMSE. Este proceso se presenta en el anexo 8, donde se asigna el valor de 1 para los modelos que exhiben medidas menores a las logradas por los modelos sin desconexión.

Para construir la tabla 4.2, se seleccionan únicamente los modelos que tienen menores valores para RMSE y MAE conjuntamente. De nuevo, se usa el criterio RMSE para escoger la mejor RNA para cada horizonte de pronóstico. En ésta tabla se observa que para los periodos 1, 2, 3, 4, 7 y 8 la RNA aumentada con desconexión de variables genera mejores pronósticos que la RNA autoregresiva, la cual solo es el mejor modelo en los periodos 5 y 6.

Estos resultados muestran la conveniencia de usar análisis de sensibilidad para encontrar nuevos y mejores modelos; de los mejores modelos escogidos solo uno resultó ser el modelo sin desconexión, a saber, la RNA autoregresiva 1_4_5 en el periodo cinco. Por otro lado, debido a la desconexión el balance entre la RNA autoregresiva y la RNA aumentada cambia significativamente; los segundos superan a los primeros en seis de ocho periodos. Finalmente, esta mejora en las RNA sustenta la conveniencia de usar modelos no lineales.

Tabla 4.2: Análisis de Sensibilidad fuera de muestra

		AUTOREGRESIVO						CON CDT							
		P	Q	ITER	HIDE	RMSE	MAE	MODELO	P	Q	ITER	HIDE	RMSE	MAE	MODELO
HORIZONTE	1	2	3	1	1	0.01694	0.01357	3	4	1	4	2	0.01621	0.01287	1
		4	2	2	2	0.01746	0.01417	1	4	2	2	2	0.01736	0.01392	2
	4	1	4	2	0.01628	0.01302	3	4	1	4	2	0.01628	0.01302	3	
	4	2	2	0	0.01810	0.01568	0	4	2	2	0	0.01810	0.01568	0	
	2	2	3	1	1	0.01719	0.01330	3	4	1	4	2	0.01630	0.01268	1
		4	1	4	2	0.01810	0.01413	2	4	1	4	3	0.01831	0.01424	2
	4	1	4	2	0.01633	0.01278	3	4	1	4	2	0.01633	0.01278	3	
	4	1	4	0	0.01888	0.01483	0	4	1	4	0	0.01888	0.01483	0	
	3	2	3	1	1	0.01637	0.01211	3	4	1	4	2	0.01222	0.01014	1
		4	1	4	2	0.01217	0.01025	3	4	1	4	2	0.01217	0.01025	3
	4	3	4	2	0.01278	0.01072	3	4	3	4	2	0.01278	0.01072	3	
	4	2	4	3	0	0.01695	0.01382	0	4	3	4	0	0.01318	0.01091	0
	4	3	3	3	2	0.01001	0.00804	1	4	1	4	2	0.01061	0.00884	1
		4	2	3	2	0.01081	0.00899	1	4	2	3	2	0.01081	0.00899	1
		4	3	4	2	0.01117	0.00847	1	3	3	3	2	0.00925	0.00638	2
		4	2	3	2	0.01045	0.00873	2	4	2	3	2	0.01045	0.00873	2
		4	2	3	4	0.01123	0.00966	2	4	1	4	2	0.01123	0.00966	2
		4	1	4	2	0.01081	0.00910	3	4	1	4	2	0.01081	0.00910	3
		4	3	4	2	0.01148	0.00965	3	4	3	4	2	0.01148	0.00965	3
		2	3	3	0	0.01204	0.00992	0	3	3	3	0	0.01158	0.01025	0
	5	4	1	4	3	0.01860	0.01428	1	4	1	4	3	0.01859	0.01420	3
		4	3	4	0	0.01978	0.0149	0	4	3	4	0	0.01978	0.0149	0
	6	2	3	1	1	0.01444	0.01128	3	4	1	4	2	0.02128	0.01610	1
		4	1	4	3	0.02030	0.01570	1	4	1	4	3	0.02030	0.01570	1
4	1	4	3	0.02021	0.01552	3	4	1	4	3	0.02021	0.01552	3		
4	3	4	0	0.02138	0.01726	0	4	3	4	0	0.02138	0.01726	0		
7	2	3	5	2	0.01436	0.01061	2	4	1	4	2	0.01409	0.01107	1	
	2	3	1	1	0.01541	0.01107	3	4	1	4	3	0.01387	0.01102	1	
4	1	4	3	0.01347	0.01059	3	4	1	4	3	0.01347	0.01059	3		
4	3	4	0	0.01438	0.01249	0	4	3	4	0	0.01438	0.01249	0		
8	3	3	3	2	0.00696	0.00669	2	3	3	3	2	0.00696	0.00669	2	
	3	3	3	2	0.00368	0.00310	3	3	3	3	2	0.00368	0.00310	3	
4	1	4	2	0	0.00696	0.00557	0	1	4	4	0	0.00758	0.00673	0	

0 = MODELO ORIGINAL; 1 = LINEAL Y NO LINEAL; 2 = LINEAL; 3 = NO LINEAL

Mínimos valores MAE obtenidos: 1 periodo: 0.01519; 2 periodos: 0.01472, 3 periodos: 0.01382; 4 periodos: 0.00992; 5 periodos: 0.01155; 6 periodos: 0.01169; 7 periodos: 0.0114; 8 periodos: 0.00557.	Mínimos valores MAE obtenidos: 1 periodo: 0.01455; 2 periodos: 0.01435, 3 periodos: 0.01091; 4 periodos: 0.01025; 5 periodos: 0.0149; 6 periodos: 0.01726; 7 periodos: 0.01249; 8 periodos: 0.00673.
--	--

5. Conclusiones

Las redes neuronales artificiales son modelos que intentan reproducir el proceso mediante el cual el cerebro reconoce la realidad a través de complejos procedimientos que permiten identificar patrones, para generar interpretaciones de la realidad. Debido a estas cualidades, que las redes neuronales artificiales tratan de reproducir, estos modelos son capaces de reconocer patrones y relaciones inherentes a las series que otras especificaciones no logran capturar.

La aplicación de las redes neuronales artificiales sobre el PIB en Colombia muestra que esta variable contiene relaciones no lineales en su proceso generador de datos y que, además, las redes logran capturar estas relaciones de manera exitosa. Se concluye esto debido a que las redes neuronales superan ampliamente al modelo lineal ARIMA. Sin embargo, dado que las redes neuronales son conocidas como modelos de “cajas negras”, no es posible identificar la fuente de las relaciones no lineales presentes en el PIB como las expuestas en la sección dos de este trabajo. Para evaluar si la explicación de los comportamientos no lineales mejora con la introducción de la política monetaria, se usó una aproximación de ésta en una red neuronal con la tasa de interés de los CDT a 90 días; los resultados muestran que el modelo ampliado mejora los pronósticos en por lo menos dos periodos, mostrando que esta variable resulta ser relevante para el comportamiento del PIB.

Los resultados sobre la evaluación dentro de muestra y fuera de muestra señalan que las mejores especificaciones dentro de muestra usualmente son inadecuadas pronosticando; esto se debe a que la sobre especificación de una red le quita la flexibilidad necesaria para ajustarse a condiciones económicas diferentes a las encontradas dentro de muestra. Para la evaluación de pronósticos fuera de muestra se usó la metodología *rolling*, permitiendo escoger los mejores modelos ante la introducción de nueva información periodo a periodo y a diferentes horizontes de pronósticos. El mayor horizonte contemplado en este trabajo es de dos años; de esta manera, la elección sobre el mejor modelo fue hecha discriminando diferentes horizontes de pronóstico.

El análisis de sensibilidad fue realizado sobre las arquitecturas seleccionadas utilizando la metodología *rolling* para indagar sobre la relevancia de cada una de las variables en los diferentes modelos. Por un lado, debido al mejor comportamiento de algunos modelos con desconexión, se revela la ventaja de considerar modelos sin las variables irrelevantes. Por otro lado, se evidencia la superioridad del modelo aumentado, el cual responde a la desconexión, generando mejores pronósticos que el modelo autoregresivo; los resultados señalan que la red neuronal desconectada con CDT genera mejores pronósticos en seis periodos de ocho.

Sin embargo, para realizar pronósticos con la RNA aumentada se necesitará construir un modelo para la tasa de interés de los CDT a 90 días, debido a que esta variable es exógena en el modelo; de lo contrario, no es posible realizar pronósticos del PIB con la red neuronal aumentada, pues no contamos con información sobre la tasa de interés en el futuro.

Para trabajos venideros, con el fin de mejorar la capacidad de pronóstico de las redes neuronales sobre el PIB, se deberán introducir más observaciones. Tzacz y Hu (1999) señalan que el número de observaciones son un factor determinante para el desempeño de las redes neuronales debido a la gran cantidad de parámetros a estimar y, como se mostró en la sección tres, en las redes neuronales contar con mayores observaciones supone la posibilidad de “aprender” más fácilmente el comportamiento de la serie en estudio. Por otro lado, el debate sobre la tasa de interés que refleje el estado de la política es bastante amplio, los mecanismos de transmisión de la política monetaria son complejos y desarrollos teóricos posteriores podrían sugerir el uso de una tasa diferente a la de los CDT a 90 días. Otra opción para medir la política monetaria sería aplicar el trabajo de Saito y Takeda (2000), el cual usa la diferencia entre la tasa de interés de largo y corto plazo. Para

implementar dicho trabajo, se deberá esperar a tener una medida de la tasa de interés de largo plazo suficientemente larga o reproducir este ejercicio en un futuro cuando la información haya logrado un tamaño suficientemente grande.

Bibliografía

- Araújo, E y Gama, C (2004) “Replicando características de ciclos económicos: um estudo comparativo entre Redes Neurais Artificiais e modelos ARIMA” IBMEC.
- Aristizábal, M (2006) “Evaluación asimétrica de una red neuronal artificial: aplicación al caso de la inflación en Colombia”. *Lecturas de Economía*, Núm 65, Universidad de Antioquia.
- Cybenko, G(1989) “Approximation by Superpositions of a sigmoidal function”. *Mathematics of Control Signals and Systems* 2, 303-314.
- Fabiani,S., Druant,M., Hernando,I., Kwapil,C., Landau,B., Loupias,C., Martins,F., Mathä,T., Sabbatini,R., Stahl,H., y Stokman,C.J. (2005) “The pricing behaviour of firms in the euro area: new survey evidence.” Working papers, Núm. 535, European Central Bank.
- Franses, P.,H. y van Dijk, D. (2000). “Non-linear Time Series Models in Empirical Finance”. Cambridge University Press.
- Friedman, M. (1968), “The Role of Monetary Policy”. *American Economic Review* 58: 1-17.
- Guerrero, V (2003) “Análisis estadístico de series de tiempo económicas”. Thomson editores, segunda edición.
- Hendry, David (1995) “Dynamic Econometrics”. Oxford University Press.
- Jalil, M. A. y Misas, M. (2005). “Evaluación de pronósticos del tipo de cambio utilizando redes neuronales y funciones de pérdida asimétrica”. *Borradores de Economía*, Núm 376, Banco de la República.
- Kuan C. M. and H. White (1994), “Artificial Neural Networks: An Econometric Perspective”. *Econometric Reviews* 13.
- Misas, M. A., E., Arango, C. A. y Hernández, N. (2003). “La demanda de efectivo en Colombia: una caja negra a la luz de las redes neuronales”. *Borradores de Economía*, Núm.268, Banco de la República.
- Misas, M. A., López, E. y Querubín, P. (2002). “La Inflación en Colombia: una aproximación desde las redes neuronales”. *Borradores de Economía*, Núm.199, Banco de la República.
- Montenegro, Álvaro (2001) “Redes Neurales Artificiales”. *Documentos de Economía*, 2001-11. Pontificia Universidad Javeriana.
- Morgan, D. (1993) “Asymmetric effects of monetary policy”. *Federal Reserve Bank of Kansas City Economic Review* 78.
- Peña, D (2005) “Análisis de series temporales”. Alianza Editorial, Madrid.

- Reinhart, C y Reinhart, V (1991) “Fluctuaciones del producto y choques monetarios: evidencia Colombiana”.
Revista ESPE, Núm. 20, Art. 02.
- Saito, Yoshihito y Takeda, Yoko (2000) “Predicting the US Real GDP Growth Using Yield Spreads of
Corporate Bonds”. International Department, Bank of Japan.
- Sims, C.A.(1980) “Comparison of Interwar and Postwar Business Cycles”. American Economic, Review 70.
- Snell, Richard (2007) “Neuroanatomía clínica de Snell”. Editorial Panamericana, sexta edición.
- Soto R. (1997), “Non-linearities in the Demand for Money: A Neural Network Approach” Documento de
Investigación, Núm.107, Ilades/Georgetown University.
- Stock, J y Watson, M (2003) “Forecasting Output and Inflation: The Role of Asset”. Journal of Economic
Literature, Vol. 41.
- Swanson, N. R. y White, H. (1995). “A Model Selection Approach to Real Time Macroeconomic Forecasting
using Linear Models and Artificial Neural Networks”. The Review of Economics and Statistics,
Núm.79.
- Tkacz, G. y Hu, S. (1999). “Forecasting GDP growth using Artificial Neural Networks”. Working Paper 99-3,
Bank of Canada.
- Torres, José Luis (2006) “Modelos para la inflación básica de bienes transables y no transables en Colombia”.
Borradores de Economía, Núm 365, Banco de la República.
- Zukime, Mohd y Junoh, Mat (2001) “Predicting macroeconomic time series in Malaysia : using neural
networks approaches”. Sunway College Journal, Universiti Utara, Malaysia.

ANEXO 1

Resultados de la Metodología Stepwise

Selección de variables para RNA autoregresiva

Resultados de la metodología Stepwise

Variable	Parámetro Estimado	Error estándar	Type II SS	F-Valor	Pr > F
Intercept	0.0027	0.0394	0.0007	0.0000	0.9465
Y1	-0.2437	0.0636	2.3663	14.6900	0.0002
Y3	-0.1773	0.0645	1.2179	7.5600	0.0071
Y4	0.5126	0.0999	4.2394	26.3300	<.0001
Y6	-0.2591	0.0657	2.5062	15.5600	0.0002
Y8	0.2208	0.0936	0.8954	5.5600	0.0204

Resultados de la metodología Forward

Variable	Parámetro Estimado	Error estándar	Type II SS	F-Valor	Pr > F
Intercept	0.0027	0.0394	0.0007	0.0000	0.9465
Y1	-0.2437	0.0636	2.3663	14.6900	0.0002
Y3	-0.1773	0.0645	1.2179	7.5600	0.0071
Y4	0.5126	0.0999	4.2394	26.3300	<.0001
Y6	-0.2591	0.0657	2.5062	15.5600	0.0002
Y8	0.2208	0.0936	0.8954	5.5600	0.0204

Resumen de Selección hacia delante

Paso	Variable introducida	Número vars en	R-cuadrado parcial	R-cuadrado del modelo	C(p)	F-Valor	Pr > F
1	Y4	1	0.8084	0.8084	22.4481	430.2900	<.0001
2	Y8	2	0.0123	0.8207	16.5629	6.9500	0.0097
3	Y6	3	0.0070	0.8277	14.1090	4.0400	0.0470
4	Y1	4	0.0110	0.8387	9.0970	6.7300	0.0109
5	Y3	5	0.0116	0.8502	3.7105	7.5600	0.0071

Resultados de la metodología Backward

Variable	Parámetro Estimado	Error estándar	Type II SS	F-Valor	Pr > F
Intercept	0.0027	0.0394	0.0007	0.0000	0.9465
Y1	-0.2437	0.0636	2.3663	14.6900	0.0002
Y3	-0.1773	0.0645	1.2179	7.5600	0.0071
Y4	0.5126	0.0999	4.2394	26.3300	<.0001
Y6	-0.2591	0.0657	2.5062	15.5600	0.0002
Y8	0.2208	0.0936	0.8954	5.5600	0.0204

Selección de variables para RNA aumentada (con CDT)

Resultados de la metodología Stepwise

Variable	Parámetro Estimado	Error estándar	Type II SS	F-Valor	Pr > F
Intercept	-0.0016	0.0404	0.0003	0.0000	0.9694
Y4	0.7622	0.0959	10.6742	63.2400	<.0001
Y8	0.2021	0.0963	0.7435	4.4100	0.0384
C1	-0.1019	0.0405	1.0667	6.3200	0.0136
C3	-0.1006	0.0399	1.0737	6.3600	0.0133

Resultados de la metodología Forward

Variable	Parámetro Estimado	Error estándar	Type II SS	F-Valor	Pr > F
Intercept	-0.0016	0.0404	0.0003	0.0000	0.9694
Y4	0.7622	0.0959	10.6742	63.2400	<.0001
Y8	0.2021	0.0963	0.7435	4.4100	0.0384
C1	-0.1019	0.0405	1.0667	6.3200	0.0136
C3	-0.1006	0.0399	1.0737	6.3600	0.0133

Resumen de Selección hacia delante

Paso	Variable introducida	Número vars en	R-cuadrado parcial	R-cuadrado del modelo	C(p)	F-Valor	Pr > F
1	Y4	1	0.8084	0.8084	30.5151	430.2900	<.0001
2	C1	2	0.0153	0.8237	22.0752	8.7800	0.0038
3	C3	3	0.0107	0.8344	16.8206	6.4300	0.0128
4	Y8	4	0.0071	0.8414	14.0143	4.4100	0.0384

Resultados de la metodología Backward

Variable	Parámetro Estimado	Error estándar	Type II SS	F-Valor	Pr > F
Intercept	-0.0014	0.0377	0.0002	0.0000	0.9715
Y1	-0.3599	0.0891	2.4059	16.3300	0.0001
Y4	0.7577	0.0917	10.0634	68.2900	<.0001
Y5	0.3228	0.0917	1.8269	12.4000	0.0007
Y8	0.2224	0.0910	0.8813	5.9800	0.0163
C1	-0.1119	0.0393	1.1966	8.1200	0.0053
C3	-0.14128	0.03866	1.96829	13.36	0.0004

Selección de variables para RNA aumentada (con CDT y M1)

Resultados de la metodología Forward

Variable	Parámetro Estimado	Error estándar	Type II SS	F-Valor	Pr > F
Intercept	-0.0016	0.0404	0.0003	0.0000	0.9694
Y4	0.7622	0.0959	10.6742	63.2400	<.0001
Y8	0.2021	0.0963	0.7435	4.4100	0.0384
C1	-0.1019	0.0405	1.0667	6.3200	0.0136
C3	-0.1006	0.0399	1.0737	6.3600	0.0133

Resumen de Selección hacia delante

Paso	Variable introducida	Número vars en	R-cuadrado parcial	R-cuadrado del modelo	C(p)	F-Valor	Pr > F
1	Y4	1	0.8084	0.8084	35.9152	430.2900	<.0001
2	C1	2	0.0153	0.8237	27.0433	8.7800	0.0038
3	C3	3	0.0107	0.8344	21.4886	6.4300	0.0128
4	Y8	4	0.0071	0.8414	18.4834	4.4100	0.0384

ANEXO 2

Medidas de evaluación de pronóstico

$$AIC^{17} = n * \ln(\hat{\sigma}^2) + \frac{2k}{n}$$

$$BIC^{18} = n * \ln(\hat{\sigma}^2) + \frac{k * \ln(n)}{n}$$

$$RMSE^{19} = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2}$$

$$RMSPE^{20} = 100 * \sqrt{\frac{1}{n} \sum_{t=1}^n \left(\frac{\hat{y}_t - y_t}{y_t} \right)^2}$$

$$MAE^{21} = \frac{1}{n} \sum_{t=1}^n |\hat{y}_t - y_t|$$

¹⁷ Akaike Information Criterion.

¹⁸ Bayesian Information Criterion.

¹⁹ Root Mean Squared Error.

²⁰ Root Mean Squared Percentage Error.

²¹ Mean Absolut Error.

$$MAPE^{22} = 100 * \frac{1}{n} \sum_{t=1}^n \left| \frac{\hat{y}_t - y_t}{y_t} \right|$$

$$SR^{23} = \frac{1}{n} \sum_{t=1}^n I[\{(y_t - y_{t-1}) * (\hat{y}_t - \hat{y}_{t-1})\} > 0]$$

Donde

n=número de observaciones

\hat{y}_t = valor estimado por el modelo en el periodo t.

y_t = valor observado de la variable en el periodo t.

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^n \hat{\epsilon}_t^2}{n}$$

$\hat{\sigma}^2$ es la varianza estimada del error y $\hat{\epsilon}_t = \hat{y}_t - y_t$

k=número de parámetros estimados por modelo.

La medida SR mide la capacidad de los modelos para reproducir los movimientos de la serie observada. Estos movimientos se miden en términos del signo de los movimientos de la serie y de los valores estimados por el modelo.

²² Mean Absolut Percentage Error.

²³ Success Ratio.

ANEXO 3

Resultados de la evaluación dentro de muestra

RNA Autoregresiva

CRITERIO	P	Q	ITERACIÓN	AIC	BIC	R2	RMSE	RMSPE	MAE	MAPE	SR
AIC	5	4	1	-8.81016	-8.04266	0.95158	0.00901	22.41153	0.00722	2.97729	93.40659
	4	4	1	-8.73461	-8.07675	0.94304	0.00977	29.16944	0.00814	3.85344	94.50549
	5	4	2	-8.63023	-7.86273	0.94204	0.00986	23.48248	0.00734	3.11059	91.20879
	5	3	1	-8.59600	-8.02038	0.93016	0.01082	18.45378	0.00828	2.51236	93.40659
BIC	3	1	4	-8.41341	-8.27636	0.88129	0.01411	16.44714	0.01129	2.91183	86.81319
	3	1	2	-8.41341	-8.27636	0.88129	0.01411	16.44709	0.01129	2.91183	86.81319
	3	1	1	-8.41341	-8.27636	0.88129	0.01411	16.44706	0.01129	2.91182	86.81319
	3	1	3	-8.41341	-8.27636	0.88129	0.01411	16.44710	0.01129	2.91183	86.81319
R2	5	4	1	-8.81016	-8.04266	0.95158	0.00901	22.41153	0.00722	2.97729	93.40659
	4	4	1	-8.73461	-8.07675	0.94304	0.00977	29.16944	0.00814	3.85344	94.50549
	5	4	2	-8.63023	-7.86273	0.94204	0.00986	23.48248	0.00734	3.11059	91.20879
	5	4	3	-8.58900	-7.82150	0.93960	0.01006	48.78655	0.00725	5.72340	89.01099
RMSE	5	4	1	-8.81016	-8.04266	0.95158	0.00901	22.41153	0.00722	2.97729	93.40659
	4	4	1	-8.73461	-8.07675	0.94304	0.00977	29.16944	0.00814	3.85344	94.50549
	5	4	2	-8.63023	-7.86273	0.94204	0.00986	23.48248	0.00734	3.11059	91.20879
	5	4	3	-8.58900	-7.82150	0.93960	0.01006	48.78655	0.00725	5.72340	89.01099
RMSPE	3	2	1	-8.50675	-8.23265	0.90301	0.01275	2.07162	0.01036	0.91511	89.01099
	3	4	4	-8.52709	-7.97887	0.92353	0.01132	3.28574	0.00948	1.04965	92.30769
	3	4	2	-8.53903	-7.99082	0.92444	0.01126	4.39728	0.00921	1.16162	90.10989
	3	3	2	-8.47595	-8.06479	0.91028	0.01226	4.59904	0.00962	1.18146	92.30769
MAE	5	4	4	-8.55703	-7.78953	0.93763	0.01023	13.64057	0.00704	2.13397	92.30769
	5	4	1	-8.81016	-8.04266	0.95158	0.00901	22.41153	0.00722	2.97729	93.40659
	5	4	3	-8.58900	-7.82150	0.93960	0.01006	48.78655	0.00725	5.72340	89.01099
	5	4	2	-8.63023	-7.86273	0.94204	0.00986	23.48248	0.00734	3.11059	91.20879
MAPE	3	2	1	-8.50675	-8.23265	0.90301	0.01275	2.07162	0.01036	0.91511	89.01099
	3	4	4	-8.52709	-7.97887	0.92353	0.01132	3.28574	0.00948	1.04965	92.30769
	3	4	2	-8.53903	-7.99082	0.92444	0.01126	4.39728	0.00921	1.16162	90.10989
	3	3	2	-8.47595	-8.06479	0.91028	0.01226	4.59904	0.00962	1.18146	92.30769
SR	4	3	1	-8.42677	-7.93338	0.91170	0.01217	23.76118	0.00966	3.47012	94.50549
	4	4	1	-8.73461	-8.07675	0.94304	0.00977	29.16944	0.00814	3.85344	94.50549
	5	4	5	-8.54907	-7.78157	0.93713	0.01027	13.70290	0.00773	1.97486	94.50549
	4	4	2	-8.53443	-7.87657	0.93041	0.01080	12.79152	0.00841	2.15213	93.40659

RNA con CDT a 90 días

CRITERIO	P	Q	ITERACIÓN	AIC	BIC	R2	RMSE	RMSPE	MAE	MAPE	SR
AIC	4	4	1	-8.53977	-7.88192	0.93078	0.01077	27.18495	0.00843	3.86843	91.20879
	4	3	4	-8.44146	-7.94806	0.91299	0.01208	18.73212	0.00970	2.98670	87.91209
	4	4	3	-8.43702	-7.77916	0.92329	0.01134	14.10809	0.00867	2.44210	89.01099
	4	4	2	-8.43286	-7.77501	0.92297	0.01136	7.32888	0.00833	1.37141	91.20879
BIC	1	1	2	-8.18244	-8.10021	0.84381	0.01618	6.77680	0.01246	1.85635	83.51648
	1	1	1	-8.18244	-8.10021	0.84381	0.01618	6.77680	0.01246	1.85635	83.51648
	1	1	3	-8.18244	-8.10021	0.84381	0.01618	6.77680	0.01246	1.85635	83.51648
	1	1	5	-8.18244	-8.10021	0.84381	0.01618	6.77669	0.01246	1.85634	83.51648
R2	4	4	1	-8.53977	-7.88192	0.93078	0.01077	27.18495	0.00843	3.86843	91.20879
	4	4	3	-8.43702	-7.77916	0.92329	0.01134	14.10809	0.00867	2.44210	89.01099
	4	4	2	-8.43286	-7.77501	0.92297	0.01136	7.32888	0.00833	1.37141	91.20879
	4	4	5	-8.38363	-7.72577	0.91909	0.01165	23.63604	0.00891	3.38522	91.20879
RMSE	4	4	1	-8.53977	-7.88192	0.93078	0.01077	27.18495	0.00843	3.86843	91.20879
	4	4	3	-8.43702	-7.77916	0.92329	0.01134	14.10809	0.00867	2.44210	89.01099
	4	4	2	-8.43286	-7.77501	0.92297	0.01136	7.32888	0.00833	1.37141	91.20879
	4	4	5	-8.38363	-7.72577	0.91909	0.01165	23.63604	0.00891	3.38522	91.20879
RMSPE	3	4	2	-8.39135	-7.84313	0.91241	0.01212	3.06536	0.00904	1.05963	85.71429
	4	2	3	-8.25780	-7.92888	0.88089	0.01413	3.31091	0.01056	1.11794	84.61538
	3	4	3	-8.36623	-7.81802	0.91019	0.01227	4.44029	0.00950	1.24469	89.01099
	1	3	2	-8.21629	-7.96959	0.86747	0.01491	4.97161	0.01172	1.45910	83.51648
MAE	4	4	2	-8.43286	-7.77501	0.92297	0.01136	7.32888	0.00833	1.37141	91.20879
	4	4	1	-8.53977	-7.88192	0.93078	0.01077	27.18495	0.00843	3.86843	91.20879
	4	4	3	-8.43702	-7.77916	0.92329	0.01134	14.10809	0.00867	2.44210	89.01099
	4	4	5	-8.38363	-7.72577	0.91909	0.01165	23.63604	0.00891	3.38522	91.20879
MAPE	3	4	2	-8.39135	-7.84313	0.91241	0.01212	3.06536	0.00904	1.05963	85.71429
	4	2	3	-8.25780	-7.92888	0.88089	0.01413	3.31091	0.01056	1.11794	84.61538
	3	4	3	-8.36623	-7.81802	0.91019	0.01227	4.44029	0.00950	1.24469	89.01099
	4	4	2	-8.43286	-7.77501	0.92297	0.01136	7.32888	0.00833	1.37141	91.20879
SR	3	3	1	-8.30931	-7.89815	0.89401	0.01333	27.09923	0.01051	3.68259	91.20879
	4	3	1	-8.42273	-7.92934	0.91135	0.01219	21.80254	0.00937	3.13343	91.20879
	4	4	1	-8.53977	-7.88192	0.93078	0.01077	27.18495	0.00843	3.86843	91.20879
	4	4	2	-8.43286	-7.77501	0.92297	0.01136	7.32888	0.00833	1.37141	91.20879

ANEXO 4

Resultados de la evaluación fuera de muestra

MAE	ARIMA	REDES NEURONALES ARTIFICIALES							
		Autoregresivo				Con CDT			
Horizonte									
1	0.01622	2_3_1 0.01519	5_1_1 0.01527	1_4_3 0.01571	2_3_2 0.01579	4_3_4 0.01455	4_4_2 0.01513	3_3_3 0.01528	4_1_4 0.01534
2	0.01658	2_3_1 0.01472	2_4_3 0.01506	1_4_2 0.01524	2_4_1 0.01527	4_3_4 0.01435	4_1_4 0.01483	3_3_3 0.01596	4_1_5 0.01634
3	0.01427	2_4_3 0.01382	1_4_2 0.01401	1_4_5 0.01412	2_3_1 0.01412	4_3_4 0.01091	4_1_4 0.01247	3_3_3 0.01331	4_2_3 0.01367
4	0.01242	2_3_3 0.00992	2_3_1 0.01137	2_4_1 0.01146	2_4_3 0.01167	3_3_3 0.01025	4_2_3 0.0111	4_3_4 0.0115	4_1_5 0.0117
5	0.01506	1_4_5 0.01155	1_4_2 0.01179	1_4_1 0.01212	1_4_3 0.01294	4_3_4 0.0149	1_4_4 0.01553	3_1_5 0.01619	1_3_1 0.0169
6	0.01608	1_4_5 0.01169	2_4_3 0.01244	1_4_2 0.01269	2_4_1 0.0129	4_3_4 0.01726	3_1_5 0.01782	1_4_4 0.01795	1_3_1 0.01884
7	0.01481	1_4_5 0.0114	2_4_2 0.01215	2_3_5 0.01221	1_4_2 0.01231	4_3_4 0.01249	1_4_4 0.01343	3_1_5 0.01482	1_3_1 0.01612
8	0.00761	1_4_2 0.00557	1_4_5 0.00561	2_3_3 0.00722	2_3_5 0.00745	1_4_4 0.00673	3_1_5 0.00839	1_4_1 0.00918	1_4_3 0.00925

RMSPE	ARIMA	REDES NEURONALES ARTIFICIALES							
		Autoregresivo				Con CDT			
Horizonte									
1	114.52	5_1_1 50.9141	5_1_5 63.7537	5_3_5 71.4087	4_2_4 76.3065	3_4_5 101.731	1_4_5 107.978	1_3_1 108.218	1_3_2 108.276
2	112.582	4_3_4 48.3389	5_1_1 52.2614	5_2_4 66.6975	5_1_5 70.5744	3_4_5 108.973	1_4_5 113.826	1_3_1 113.969	1_3_2 114.062
3	31.6716	1_4_2 38.7801	1_4_5 38.873	2_3_1 41.3452	1_4_1 41.4185	4_2_2 33.8612	2_3_2 38.3593	2_3_1 38.3834	4_1_5 38.5571
4	30.5919	2_3_3 27.8297	2_4_3 29.8777	2_4_2 33.3955	2_3_5 33.4715	4_2_2 34.2171	4_2_3 36.3987	4_1_5 36.6381	3_3_1 37.3162
5	96.0601	5_3_5 45.9869	2_4_1 49.2832	5_1_3 50.953	2_3_5 51.4275	4_4_5 69.6954	3_4_2 70.9775	4_2_2 87.2101	3_4_5 88.4749
6	97.8152	2_4_1 48.3595	2_3_5 50.9706	2_3_1 52.6535	2_4_2 57.3709	3_4_2 73.2118	4_4_5 79.7688	3_4_5 92.3215	4_2_2 94.3076
7	32.5657	2_3_5 29.671	1_4_5 30.9628	2_4_2 31.1647	1_4_2 31.4781	1_4_4 34.6583	3_1_5 36.2296	1_3_1 38.3231	1_3_2 38.6222
8	18.8064	2_3_3 16.6598	1_4_2 19.4276	1_4_5 20.2213	2_4_3 21.7488	3_1_5 19.9206	1_4_4 20.4396	1_4_1 22.7225	3_4_2 22.924

MAPE	ARIMA	REDES NEURONALES ARTIFICIALES							
		Autoregresivo				Con CDT			
1	58.042	5_1_1	5_1_5	2_3_1	2_4_1	4_2_2	4_1_5	2_3_2	2_3_1
		38.4186	44.1974	48.3842	48.7571	55.7552	56.7761	58.6607	58.7571
2	60.1356	4_3_4	5_1_1	5_1_5	2_3_1	4_1_4	4_1_5	4_2_2	2_1_2
		36.3861	40.7057	46.0372	46.9543	60.7166	60.7281	62.0294	62.7745
3	28.4178	1_4_2	1_4_5	1_4_3	1_4_1	4_1_4	4_1_5	2_3_4	4_2_2
		27.3971	28.1085	30.2692	30.7454	28.5786	28.6827	29.7793	29.8472
4	27.0962	2_3_3	2_3_1	2_4_1	2_4_3	4_1_5	4_2_3	4_4_2	4_2_2
		20.3934	23.7534	24.182	24.2609	25.6833	26.1014	28.9416	29.0376
5	53.1551	5_3_5	3_5_5	3_3_4	2_4_1	3_4_2	4_4_5	3_1_5	1_3_1
		34.6014	36.861	36.9154	37.2662	50.442	57.226	57.9352	58.5149
6	56.5111	2_4_1	2_3_5	2_3_1	2_4_2	3_4_2	3_1_5	1_3_1	1_3_2
		37.3512	38.6873	40.1846	42.2789	53.8798	65.7138	66.8815	67.1461
7	26.5022	2_3_5	1_4_5	2_4_2	1_4_2	1_4_4	3_1_5	1_4_1	1_3_1
		22.8811	23.3739	23.9466	24.3153	26.8229	28.2402	30.9622	31.0397
8	13.7249	1_4_2	1_4_5	2_3_3	2_4_2	1_4_4	3_1_5	1_4_1	1_3_4
		14.1087	14.5593	14.885	16.7051	16.5084	17.6727	20.146	20.5333

ANEXO 5

Resultados del modelo ARIMA seleccionado en el Rolling

Maximum Likelihood Estimation 96 obs

Parámetro	Estimador	Error				Retardo
		estándar	Valor t	Pr > t		
MA1,1	0.36771	0.10011	3.6700	0.0002	4	
AR1,1	-0.27202	0.09814	-2.7700	0.0056	1	
AR1,2	-0.19221	0.09942	-1.9300	0.0532	6	

Correlations of Parameter Estimates

Parámetro	MA1,1	AR1,1	AR1,2
MA1,1	1	-0.075	0.104
AR1,1	-0.075	1	-0.033
AR1,2	0.104	-0.033	1

Autocorrelation Check of Residuals				
To Lag	Chi-Square	DF	Pr > ChiSq	
6	1.72	3	0.633	
12	4.58	9	0.8695	
18	10.02	15	0.8187	
24	16.26	21	0.7551	

Maximum Likelihood Estimation 102 obs

Parámetro	Estimador	Error estándar	Valor t	Pr > t	Retardo
MA1,1	0.36222	0.09895	3.66	0.0003	4
AR1,1	-0.24867	0.09491	-2.62	0.0088	1
AR1,2	-0.20163	0.10026	-2.01	0.0443	6

Correlations of Parameter Estimates

Parámetro	MA1,1	AR1,1	AR1,2
MA1,1	1	-0.061	0.111
AR1,1	-0.061	1	-0.055
AR1,2	0.111	-0.055	1

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	
6	0.98	3	0.8059	
12	3.77	9	0.9261	
18	9.53	15	0.8484	
24	14.99	21	0.8233	

Maximum Likelihood Estimation 107 obs

Parámetro	Estimador	Error estándar	Valor t	Pr > t	Retardo
MA1,1	0.38256	0.09123	4.19	<.0001	4
AR1,1	-0.26402	0.09268	-2.85	0.0044	1
AR1,2	-0.19158	0.09361	-2.05	0.0407	6

Correlations of Parameter Estimates

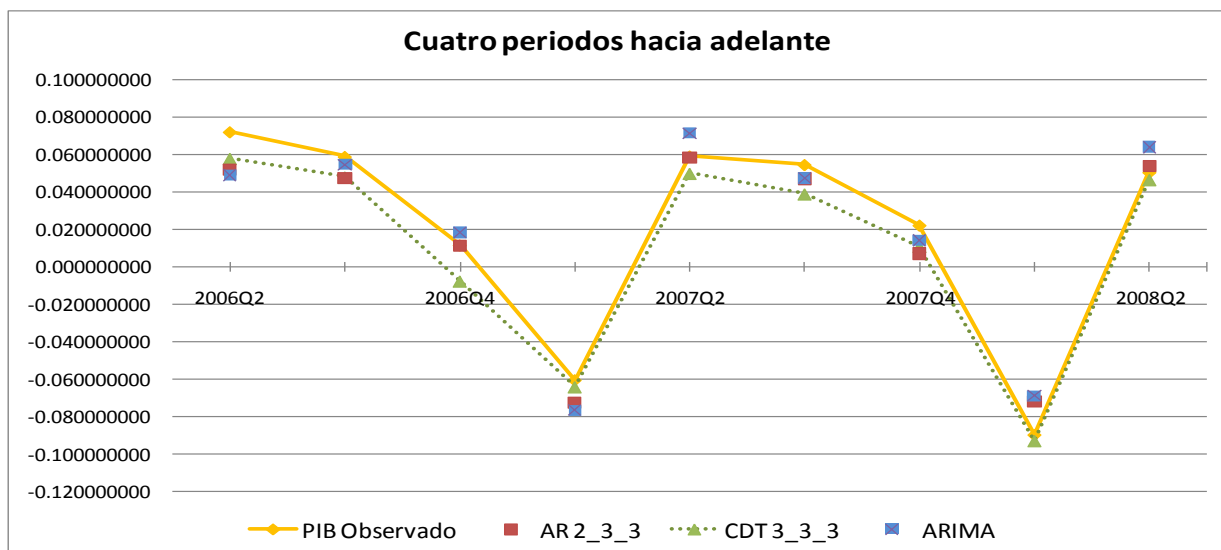
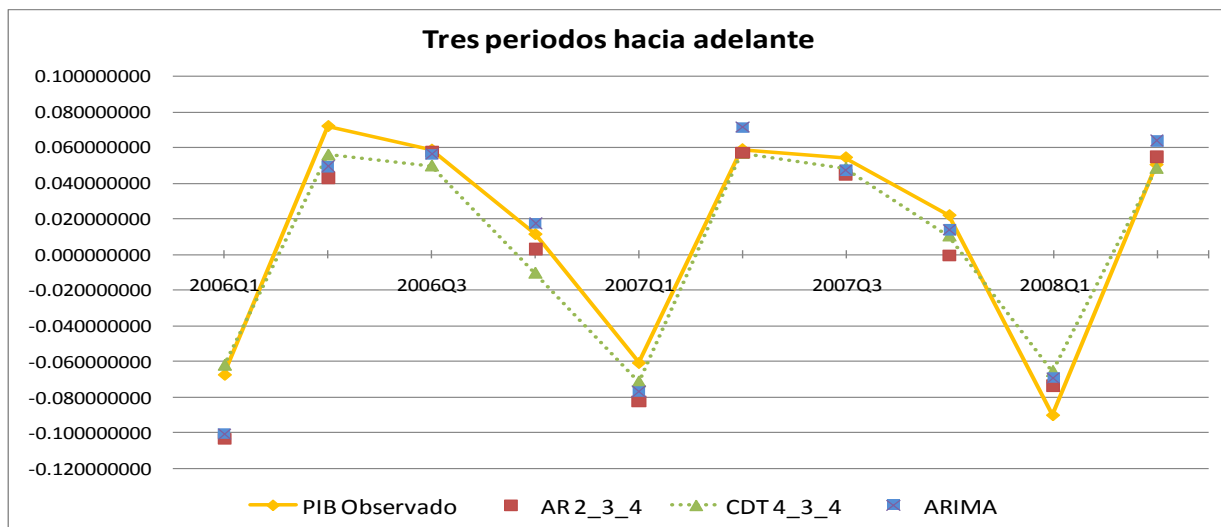
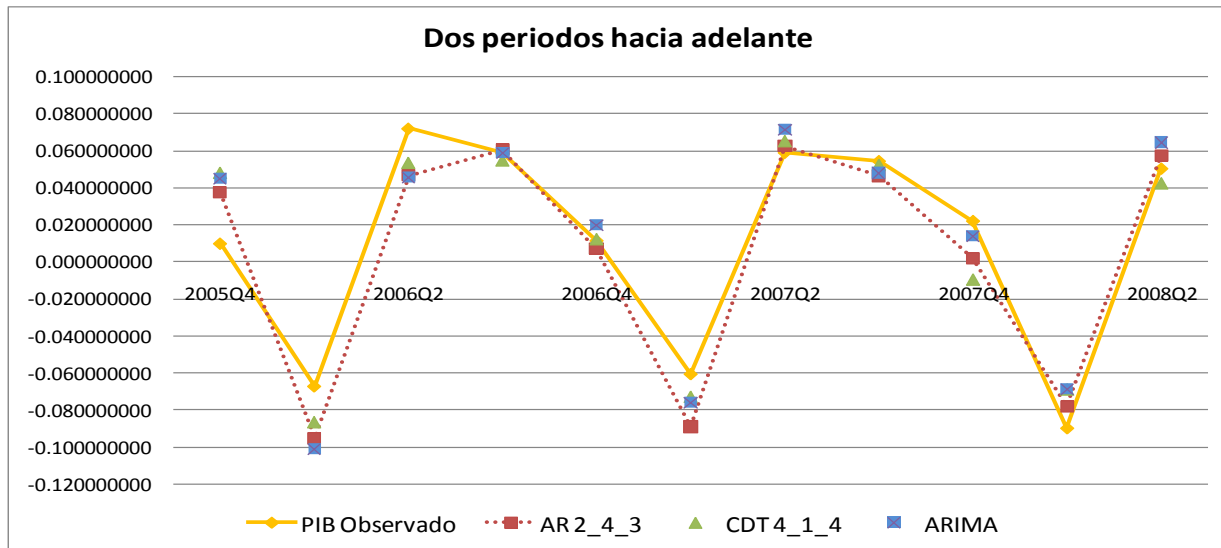
Parámetro	MA1,1	AR1,1	AR1,2
MA1,1	1	-0.03	0.039
AR1,1	-0.03	1	-0.021
AR1,2	0.039	-0.021	1

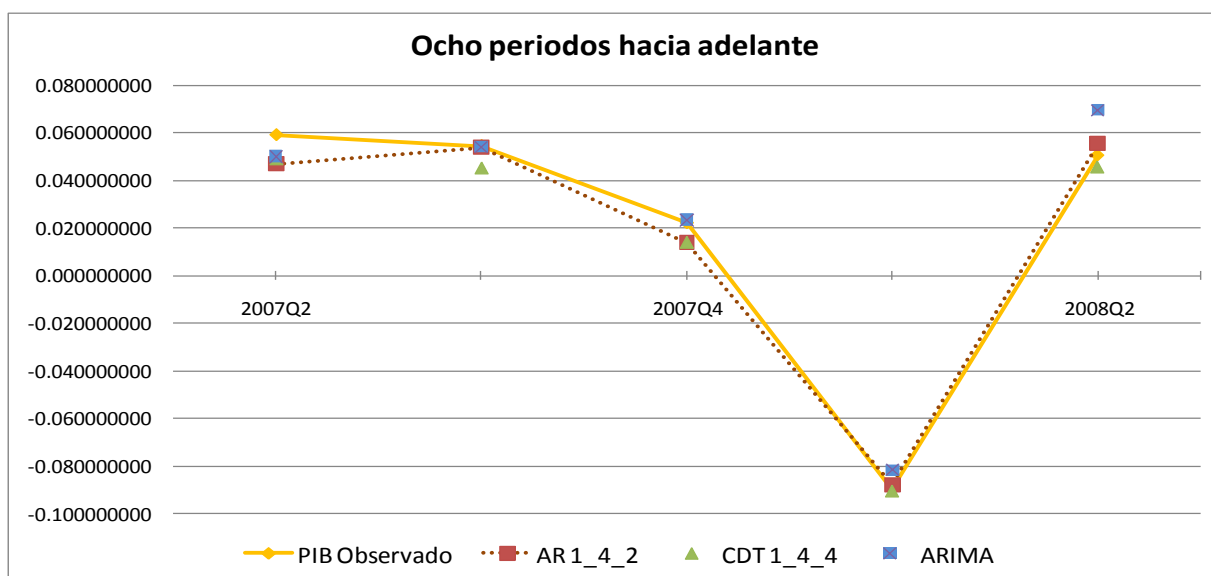
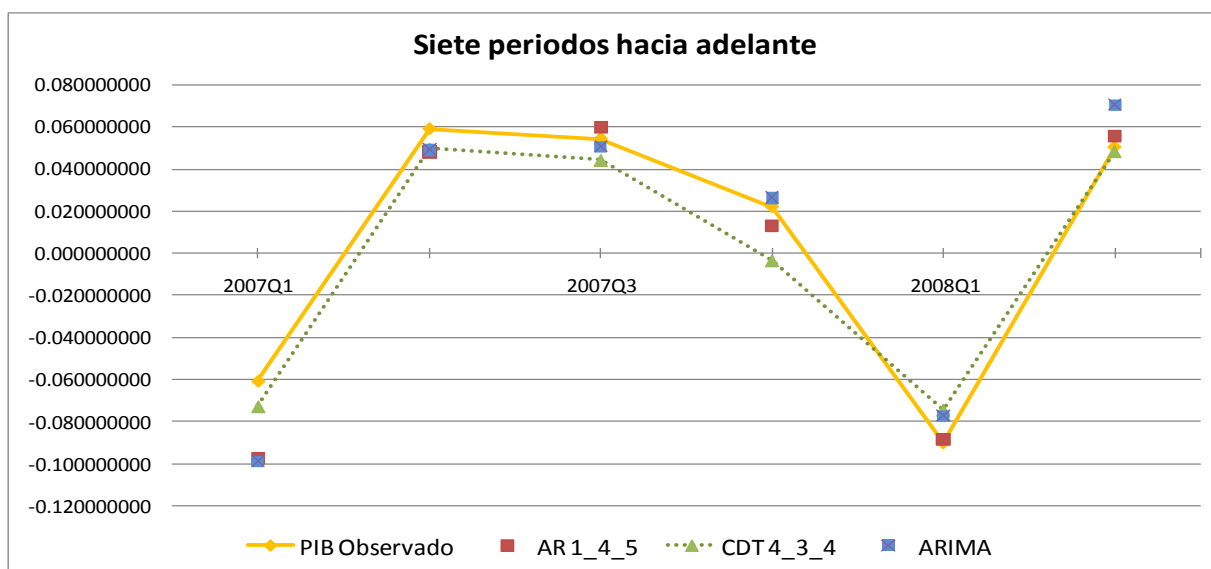
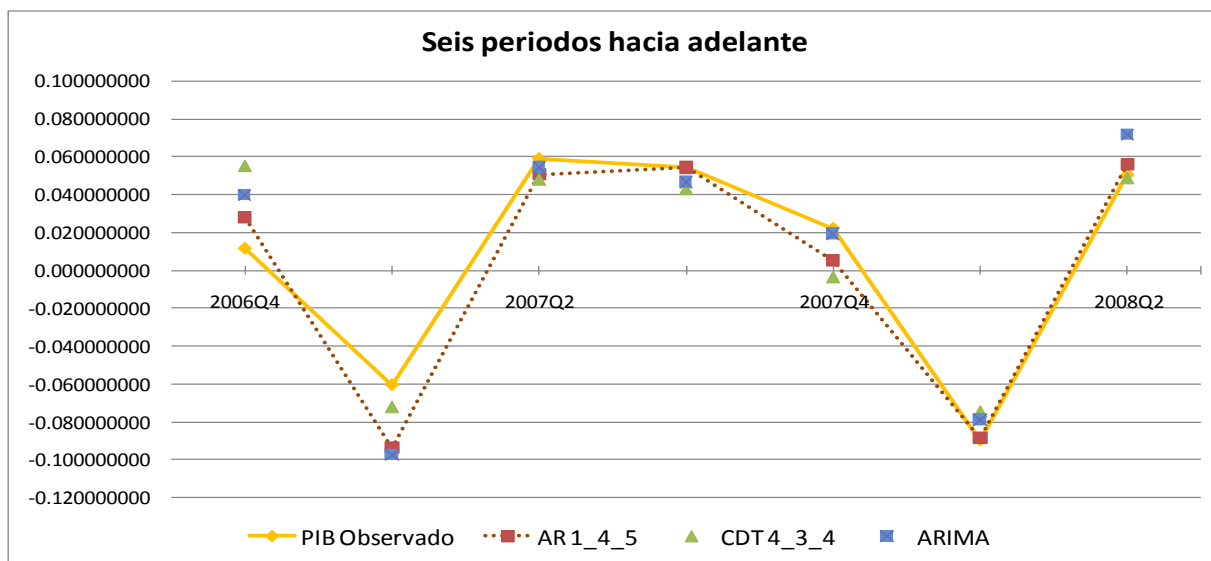
Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr >ChiSq
6	0.63	3	0.8905
12	4.57	9	0.8701
18	10.16	15	0.8097
24	16.14	21	0.7619

ANEXO 6

Gráficas de los tres modelos a diferentes horizontes de pronósticos





ANEXO 7

Análisis de Sensibilidad – RNA Autoregresiva – Evaluación dentro de muestra

Desconexión lineal y no lineal

$X = \{NDLPIB_{t-4}, NDLPIB_{t-8}, NDLPIB_{t-6}, NDLPIB_{t-1}, NDLPIB_{t-3}\}$; $\hat{Y}_{HIDE,j} = \hat{\phi}_{HIDE} = 0$, donde $HIDE \in [1, p]$

CRITERIO	P	Q	ITERACIÓN	HIDE	AIC	BIC	R2	RMSE	MAE	SR
AIC	4	3	2	4	-8.183604	-7.690211	0.887399	0.013740	0.010438	85.714286
	4	1	2	4	-8.174368	-8.009904	0.852481	0.015727	0.012595	86.813187
	4	1	1	4	-8.174367	-8.009903	0.852481	0.015727	0.012595	86.813187
	4	1	3	4	-8.174367	-8.009902	0.852481	0.015727	0.012595	86.813187
	4	1	4	4	-8.174119	-8.009655	0.852444	0.015729	0.012598	86.813187
BIC	4	1	2	4	-8.174368	-8.009904	0.852481	0.015727	0.012595	86.813187
	4	1	1	4	-8.174367	-8.009903	0.852481	0.015727	0.012595	86.813187
	4	1	3	4	-8.174367	-8.009902	0.852481	0.015727	0.012595	86.813187
	4	1	4	4	-8.174119	-8.009655	0.852444	0.015729	0.012598	86.813187
	4	1	5	4	-8.174119	-8.009654	0.852444	0.015729	0.012598	86.813187
R2	4	4	1	4	-8.125700	-7.467842	0.895277	0.013251	0.010422	92.307692
	4	4	4	4	-8.085525	-7.427668	0.890984	0.013519	0.011302	86.813187
	4	3	2	4	-8.183604	-7.690211	0.887399	0.013740	0.010438	85.714286
	5	4	2	2	-7.852538	-7.085038	0.873843	0.014543	0.010619	87.912088
	4	2	5	4	-8.104268	-7.775339	0.861118	0.015259	0.012329	83.516484
RMSE	4	4	1	4	-8.125700	-7.467842	0.895277	0.013251	0.010422	92.307692
	4	4	4	4	-8.085525	-7.427668	0.890984	0.013519	0.011302	86.813187
	4	3	2	4	-8.183604	-7.690211	0.887399	0.013740	0.010438	85.714286
	5	4	2	2	-7.852538	-7.085038	0.873843	0.014543	0.010619	87.912088
	4	2	5	4	-8.104268	-7.775339	0.861118	0.015259	0.012329	83.516484
MAE	4	4	1	4	-8.125700	-7.467842	0.895277	0.013251	0.010422	92.307692
	4	3	2	4	-8.183604	-7.690211	0.887399	0.013740	0.010438	85.714286
	5	4	2	2	-7.852538	-7.085038	0.873843	0.014543	0.010619	87.912088
	4	4	4	4	-8.085525	-7.427668	0.890984	0.013519	0.011302	86.813187
	5	4	4	5	-7.642089	-6.874588	0.844293	0.016157	0.011357	89.010989
SR	5	3	5	5	-7.398129	-6.822504	0.768609	0.019696	0.014586	93.406593
	5	3	1	5	-7.144519	-6.568894	0.701814	0.022359	0.016574	92.307692
	5	3	3	1	-7.090868	-6.515242	0.685379	0.022967	0.017089	92.307692
	4	4	1	4	-8.125700	-7.467842	0.895277	0.013251	0.010422	92.307692
	2	3	5	1	-7.415591	-7.086662	0.723476	0.021532	0.017964	90.109890

Desconexión lineal

$X = \{NDLPIB_{t-4}, NDLPIB_{t-8}, NDLPIB_{t-6}, NDLPIB_{t-1}, NDLPIB_{t-3}\}; \hat{\varphi}_{HIDE} = 0$, donde $HIDE \in [1, p]$

CRITERIO	P	Q	ITERACIÓN	HIDE	AIC	BIC	R2	RMSE	MAE	SR
AIC	5	4	1	2	-8.7321663	-7.9646654	0.9476529	0.0093682	0.0076797	93.4065934
	5	4	3	5	-8.5859095	-7.8184086	0.9394085	0.0100790	0.0072864	89.0109890
	5	4	3	2	-8.5846880	-7.8171871	0.9393345	0.0100851	0.0072127	89.0109890
	3	4	3	2	-8.5577918	-8.0095769	0.9258430	0.0111503	0.0092198	89.0109890
	5	4	4	1	-8.5342197	-7.7667189	0.9361942	0.0103429	0.0073825	93.4065934
BIC	3	2	2	1	-8.4319040	-8.1577965	0.8954711	0.0132382	0.0108798	89.0109890
	3	2	4	1	-8.4319030	-8.1577955	0.8954710	0.0132382	0.0108798	89.0109890
	3	2	3	1	-8.4319029	-8.1577955	0.8954709	0.0132382	0.0108798	89.0109890
	3	1	2	3	-8.2693490	-8.1322953	0.8628995	0.0151611	0.0120290	84.6153846
	3	1	5	3	-8.2693488	-8.1322951	0.8628994	0.0151611	0.0120290	84.6153846
R2	5	4	1	2	-8.7321663	-7.9646654	0.9476529	0.0093682	0.0076797	93.4065934
	5	4	3	5	-8.5859095	-7.8184086	0.9394085	0.0100790	0.0072864	89.0109890
	5	4	3	2	-8.5846880	-7.8171871	0.9393345	0.0100851	0.0072127	89.0109890
	5	4	4	1	-8.5342197	-7.7667189	0.9361942	0.0103429	0.0073825	93.4065934
	5	4	2	2	-8.4988049	-7.7313040	0.9338941	0.0105276	0.0080045	91.2087912
RMSE	5	4	1	2	-8.7321663	-7.9646654	0.9476529	0.0093682	0.0076797	93.4065934
	5	4	3	5	-8.5859095	-7.8184086	0.9394085	0.0100790	0.0072864	89.0109890
	5	4	3	2	-8.5846880	-7.8171871	0.9393345	0.0100851	0.0072127	89.0109890
	5	4	4	1	-8.5342197	-7.7667189	0.9361942	0.0103429	0.0073825	93.4065934
	5	4	2	2	-8.4988049	-7.7313040	0.9338941	0.0105276	0.0080045	91.2087912
MAE	5	4	3	2	-8.5846880	-7.8171871	0.9393345	0.0100851	0.0072127	89.0109890
	5	4	3	5	-8.5859095	-7.8184086	0.9394085	0.0100790	0.0072864	89.0109890
	5	4	4	1	-8.5342197	-7.7667189	0.9361942	0.0103429	0.0073825	93.4065934
	5	4	1	2	-8.7321663	-7.9646654	0.9476529	0.0093682	0.0076797	93.4065934
	5	4	2	5	-8.4964446	-7.7289438	0.9337378	0.0105401	0.0079033	91.2087912
SR	5	3	1	5	-8.1129550	-7.5373293	0.8867857	0.0137772	0.0104347	96.7032967
	5	2	4	4	-8.2967777	-7.9130272	0.8903127	0.0135609	0.0105942	95.6043956
	5	3	5	4	-8.4348882	-7.8592626	0.9179483	0.0117288	0.0093612	95.6043956
	2	4	4	1	-8.0203952	-7.5818233	0.8615460	0.0152358	0.0123413	94.5054945
	4	4	1	2	-7.9998560	-7.3419981	0.8812329	0.0141110	0.0115490	94.5054945

Desconexión no lineal

$X = \{NDLPIB_{t-4}, NDLPIB_{t-8}, NDLPIB_{t-6}, NDLPIB_{t-1}, NDLPIB_{t-3}\}; \hat{Y}_{HIDE,j} = 0$, donde $HIDE \in [1, p]$

CRITERIO	P	Q	ITERACIÓN	HIDE	AIC	BIC	R2	RMSE	MAE	SR
AIC	5	2	2	3	-8.4206575	-8.0369071	0.9030928	0.0127464	0.0105895	89.0109890
	5	1	1	2	-8.4133087	-8.2214335	0.8863326	0.0138048	0.0113711	89.0109890
	4	1	5	4	-8.3921705	-8.2277060	0.8813528	0.0141039	0.0112921	86.8131868
	4	1	4	4	-8.3921704	-8.2277060	0.8813528	0.0141039	0.0112921	86.8131868
	4	1	1	4	-8.3921650	-8.2277005	0.8813522	0.0141040	0.0112921	86.8131868
BIC	4	1	5	4	-8.3921705	-8.2277060	0.8813528	0.0141039	0.0112921	86.8131868
	4	1	4	4	-8.3921704	-8.2277060	0.8813528	0.0141039	0.0112921	86.8131868
	4	1	1	4	-8.3921650	-8.2277005	0.8813522	0.0141040	0.0112921	86.8131868
	4	1	3	4	-8.3921650	-8.2277005	0.8813522	0.0141040	0.0112921	86.8131868
	4	1	2	4	-8.3921648	-8.2277003	0.8813522	0.0141040	0.0112921	86.8131868
R2	4	4	4	4	-8.3734060	-7.7155481	0.9182542	0.0117070	0.0096075	94.5054945
	4	4	1	4	-8.3327434	-7.6748855	0.9148617	0.0119474	0.0095782	93.4065934
	5	2	2	3	-8.4206575	-8.0369071	0.9030928	0.0127464	0.0105895	89.0109890
	4	3	2	4	-8.3017845	-7.8083911	0.8999500	0.0129515	0.0102274	92.3076923
	5	2	1	2	-8.3708655	-7.9871151	0.8981454	0.0130678	0.0103440	87.9120879
RMSE	4	4	4	4	-8.3734060	-7.7155481	0.9182542	0.0117070	0.0096075	94.5054945
	4	4	1	4	-8.3327434	-7.6748855	0.9148617	0.0119474	0.0095782	93.4065934
	5	2	2	3	-8.4206575	-8.0369071	0.9030928	0.0127464	0.0105895	89.0109890
	4	3	2	4	-8.3017845	-7.8083911	0.8999500	0.0129515	0.0102274	92.3076923
	5	2	1	2	-8.3708655	-7.9871151	0.8981454	0.0130678	0.0103440	87.9120879
MAE	4	4	1	4	-8.3327434	-7.6748855	0.9148617	0.0119474	0.0095782	93.4065934
	4	4	4	4	-8.3734060	-7.7155481	0.9182542	0.0117070	0.0096075	94.5054945
	5	4	2	2	-7.9775794	-7.2100785	0.8886716	0.0136620	0.0099521	90.1098901
	4	3	2	4	-8.3017845	-7.8083911	0.8999500	0.0129515	0.0102274	92.3076923
	5	2	1	2	-8.3708655	-7.9871151	0.8981454	0.0130678	0.0103440	87.9120879
SR	4	4	4	4	-8.3734060	-7.7155481	0.9182542	0.0117070	0.0096075	94.5054945
	4	4	1	4	-8.3327434	-7.6748855	0.9148617	0.0119474	0.0095782	93.4065934
	4	2	2	4	-8.0387484	-7.7098194	0.8517142	0.0157674	0.0126250	92.3076923
	4	3	2	4	-8.3017845	-7.8083911	0.8999500	0.0129515	0.0102274	92.3076923
	5	4	4	1	-6.9062659	-6.1387650	0.6750089	0.0233425	0.0167901	92.3076923

Análisis de Sensibilidad – RNA Aumentada – Evaluación dentro de muestra

Desconexión lineal y no lineal

$X = \{NDLPIB_{t-4}, NDLCDT_{t-1}, NDLCDT_{t-3}, NDLPIB_{t-8}\}; \hat{Y}_{HIDE,j} = \hat{\phi}_{HIDE} = 0$, donde $HIDE \in [1, p]$

CRITERIO	P	Q	ITERACIÓN	HIDE	AIC	BIC	R2	RMSE	MAE	SR
AIC	4	1	2	2	-8.1112082	-7.9467438	0.8428633	0.0162312	0.0125374	83.5164835
	4	1	3	2	-8.1112081	-7.9467437	0.8428633	0.0162312	0.0125374	83.5164835
	4	2	1	2	-8.0948191	-7.7658901	0.8597999	0.0153315	0.0118362	84.6153846
	2	1	5	2	-8.0897301	-7.9800871	0.8323174	0.0167670	0.0128239	83.5164835
	4	1	1	2	-8.0846787	-7.9202142	0.8386387	0.0164479	0.0128181	84.6153846
BIC	2	1	5	2	-8.0897301	-7.9800871	0.8323174	0.0167670	0.0128239	83.5164835
	2	1	2	2	-8.0812587	-7.9716157	0.8308909	0.0168382	0.0128561	83.5164835
	2	1	1	2	-8.0812586	-7.9716157	0.8308909	0.0168382	0.0128561	83.5164835
	2	1	3	2	-8.0812586	-7.9716156	0.8308909	0.0168382	0.0128561	83.5164835
	2	1	4	2	-8.0812584	-7.9716155	0.8308909	0.0168382	0.0128561	83.5164835
R2	3	4	4	2	-7.9558435	-7.4076286	0.8646136	0.0150660	0.0118869	86.8131868
	3	3	2	2	-8.0324252	-7.6212640	0.8601952	0.0153099	0.0119467	85.7142857
	4	2	1	2	-8.0948191	-7.7658901	0.8597999	0.0153315	0.0118362	84.6153846
	4	3	1	2	-7.9248394	-7.4314459	0.8541446	0.0156377	0.0121348	87.9120879
	4	2	4	2	-8.0506974	-7.7217684	0.8534755	0.0156735	0.0123042	83.5164835
RMSE	3	4	4	2	-7.9558435	-7.4076286	0.8646136	0.0150660	0.0118869	86.8131868
	3	3	2	2	-8.0324252	-7.6212640	0.8601952	0.0153099	0.0119467	85.7142857
	4	2	1	2	-8.0948191	-7.7658901	0.8597999	0.0153315	0.0118362	84.6153846
	4	3	1	2	-7.9248394	-7.4314459	0.8541446	0.0156377	0.0121348	87.9120879
	4	2	4	2	-8.0506974	-7.7217684	0.8534755	0.0156735	0.0123042	83.5164835
MAE	4	2	1	2	-8.0948191	-7.7658901	0.8597999	0.0153315	0.0118362	84.6153846
	3	4	4	2	-7.9558435	-7.4076286	0.8646136	0.0150660	0.0118869	86.8131868
	3	3	2	2	-8.0324252	-7.6212640	0.8601952	0.0153099	0.0119467	85.7142857
	4	3	3	2	-7.9171485	-7.4237551	0.8530185	0.0156979	0.0120615	87.9120879
	4	3	1	2	-7.9248394	-7.4314459	0.8541446	0.0156377	0.0121348	87.9120879
SR	4	4	1	2	-7.6683114	-7.0104536	0.8345434	0.0166553	0.0124540	89.0109890
	4	4	4	2	-7.5759773	-6.9181194	0.8185386	0.0174423	0.0131923	89.0109890
	3	3	1	2	-7.9260993	-7.5149381	0.8445113	0.0161458	0.0125563	87.9120879
	3	4	5	2	-7.5262261	-6.9780112	0.7919558	0.0186762	0.0149163	87.9120879
	4	3	1	2	-7.9248394	-7.4314459	0.8541446	0.0156377	0.0121348	87.9120879

Desconexión lineal

$X = \{NDLPIB_{t-4}, NDLCDT_{t-1}, NDLCDT_{t-3}, NDLPIB_{t-8}\}; \hat{\phi}_{HIDE} = 0$, donde $HIDE \in [1, p]$

CRITERIO	P	Q	ITERACIÓN	HIDE	AIC	BIC	R2	RMSE	MAE	SR
AIC	4	4	2	3	-8.4328243	-7.7749664	0.9229699	0.0113643	0.0083266	91.208791
	3	4	1	3	-8.4140286	-7.8658137	0.9143775	0.0119813	0.0095218	85.714286
	4	3	5	3	-8.3953691	-7.9019757	0.9088883	0.0123594	0.0098494	90.109890
	4	4	5	2	-8.3834860	-7.7256281	0.9190741	0.0116481	0.0089118	91.208791
	3	4	2	3	-8.3803324	-7.8321175	0.9114432	0.0121849	0.0092398	85.714286
BIC	3	1	3	3	-8.2145396	-8.0774859	0.8551753	0.0155823	0.0125157	85.714286
	3	1	4	3	-8.2145395	-8.0774858	0.8551753	0.0155823	0.0125157	85.714286
	3	1	2	3	-8.2145363	-8.0774826	0.8551748	0.0155824	0.0125157	85.714286
	3	1	1	3	-8.2145361	-8.0774823	0.8551748	0.0155824	0.0125157	85.714286
	4	1	1	3	-8.2238440	-8.0593795	0.8596021	0.0153423	0.0119037	84.615385
R2	4	4	2	3	-8.4328243	-7.7749664	0.922970	0.0113643	0.0083266	91.208791
	4	4	5	2	-8.3834860	-7.7256281	0.9190741	0.0116481	0.0089118	91.208791
	4	4	5	3	-8.3603398	-7.7024819	0.9171791	0.0117837	0.0087710	91.208791
	3	4	1	3	-8.4140286	-7.8658137	0.9143775	0.0119813	0.0095218	85.714286
	4	4	2	2	-8.3158458	-7.6579879	0.9134109	0.0120488	0.0091033	90.109890
RMSE	4	4	2	3	-8.4328243	-7.7749664	0.9229699	0.011364	0.0083266	91.208791
	4	4	5	2	-8.3834860	-7.7256281	0.9190741	0.0116481	0.0089118	91.208791
	4	4	5	3	-8.3603398	-7.7024819	0.9171791	0.0117837	0.0087710	91.208791
	3	4	1	3	-8.4140286	-7.8658137	0.9143775	0.0119813	0.0095218	85.714286
	4	4	2	2	-8.3158458	-7.6579879	0.9134109	0.0120488	0.0091033	90.109890
MAE	4	4	2	3	-8.4328243	-7.7749664	0.9229699	0.0113643	0.008327	91.208791
	4	4	5	3	-8.3603398	-7.7024819	0.9171791	0.0117837	0.0087710	91.208791
	4	4	5	2	-8.3834860	-7.7256281	0.9190741	0.0116481	0.0089118	91.208791
	4	4	2	2	-8.3158458	-7.6579879	0.9134109	0.0120488	0.0091033	90.109890
	3	4	2	2	-8.3794895	-7.8312746	0.9113685	0.0121900	0.0091689	84.615385
SR	3	3	1	2	-8.0603398	-7.6491787	0.8640438	0.0150977	0.0118956	92.30769
	3	3	1	3	-8.2991530	-7.8879919	0.8929261	0.0133984	0.0107196	91.208791
	4	3	4	3	-8.2590184	-7.7656250	0.8955784	0.0132314	0.0104138	91.208791
	4	3	5	2	-8.3597345	-7.8663411	0.9055831	0.0125816	0.0097995	91.208791
	4	4	2	3	-8.4328243	-7.7749664	0.9229699	0.0113643	0.0083266	91.208791

Desconexión no lineal

$X = \{NDLPIB_{t-4}, NDLCDT_{t-1}, NDLCDT_{t-3}, NDLPIB_{t-8}\}; \hat{Y}_{HIDE,j} = 0$, donde $HIDE \in [1, p]$

CRITERIO	P	Q	ITERACIÓN	HIDE	AIC	BIC	R2	RMSE	MAE	SR
AIC	4	1	2	2	-8.1112082	-7.9467438	0.8428633	0.0162312	0.0125374	83.5164835
	4	1	3	2	-8.1112081	-7.9467437	0.8428633	0.0162312	0.0125374	83.5164835
	4	2	1	2	-8.0948191	-7.7658901	0.8597999	0.0153315	0.0118362	84.6153846
	2	1	5	2	-8.0897301	-7.9800871	0.8323174	0.0167670	0.0128239	83.5164835
	4	1	1	2	-8.0846787	-7.9202142	0.8386387	0.0164479	0.0128181	84.6153846
BIC	2	1	5	2	-8.0897301	-7.9800871	0.8323174	0.0167670	0.0128239	83.5164835
	2	1	2	2	-8.0812587	-7.9716157	0.8308909	0.0168382	0.0128561	83.5164835
	2	1	1	2	-8.0812586	-7.9716157	0.8308909	0.0168382	0.0128561	83.5164835
	2	1	3	2	-8.0812586	-7.9716156	0.8308909	0.0168382	0.0128561	83.5164835
	2	1	4	2	-8.0812584	-7.9716155	0.8308909	0.0168382	0.0128561	83.5164835
R2	3	4	4	2	-7.9558435	-7.4076286	0.8646136	0.0150660	0.0118869	86.8131868
	3	3	2	2	-8.0324252	-7.6212640	0.8601952	0.0153099	0.0119467	85.7142857
	4	2	1	2	-8.0948191	-7.7658901	0.8597999	0.0153315	0.0118362	84.6153846
	4	3	1	2	-7.9248394	-7.4314459	0.8541446	0.0156377	0.0121348	87.9120879
	4	2	4	2	-8.0506974	-7.7217684	0.8534755	0.0156735	0.0123042	83.5164835
RMSE	3	4	4	2	-7.9558435	-7.4076286	0.8646136	0.0150660	0.0118869	86.8131868
	3	3	2	2	-8.0324252	-7.6212640	0.8601952	0.0153099	0.0119467	85.7142857
	4	2	1	2	-8.0948191	-7.7658901	0.8597999	0.0153315	0.0118362	84.6153846
	4	3	1	2	-7.9248394	-7.4314459	0.8541446	0.0156377	0.0121348	87.9120879
	4	2	4	2	-8.0506974	-7.7217684	0.8534755	0.0156735	0.0123042	83.5164835
MAE	4	2	1	2	-8.0948191	-7.7658901	0.8597999	0.0153315	0.0118362	84.6153846
	3	4	4	2	-7.9558435	-7.4076286	0.8646136	0.0150660	0.0118869	86.8131868
	3	3	2	2	-8.0324252	-7.6212640	0.8601952	0.0153099	0.0119467	85.7142857
	4	3	3	2	-7.9171485	-7.4237551	0.8530185	0.0156979	0.0120615	87.9120879
	4	3	1	2	-7.9248394	-7.4314459	0.8541446	0.0156377	0.0121348	87.9120879
SR	4	4	1	2	-7.6683114	-7.0104536	0.8345434	0.0166553	0.0124540	89.0109890
	4	4	4	2	-7.5759773	-6.9181194	0.8185386	0.0174423	0.0131923	89.0109890
	3	3	1	2	-7.9260993	-7.5149381	0.8445113	0.0161458	0.0125563	87.9120879
	3	4	5	2	-7.5262261	-6.9780112	0.7919558	0.0186762	0.0149163	87.9120879
	4	3	1	2	-7.9248394	-7.4314459	0.8541446	0.0156377	0.0121348	87.9120879

ANEXO 8

Análisis de Sensibilidad – RNA Autoregresiva – Evaluación fuera de muestra – Desconexión lineal – 7 periodos hacia adelante.

P	Q	ITERACION	HIDE	RMSE	MAE	PRUEBA si RMSE < 0.01627	PRUEBA si MAE < 0.0114
1	4	1	1	0.067375	0.056788	FALSE	FALSE
1	4	2	1	0.032843	0.023965	FALSE	FALSE
1	4	3	1	0.024741	0.019877	FALSE	FALSE
1	4	5	1	0.070039	0.057590	FALSE	FALSE
2	2	3	1	0.034033	0.030117	FALSE	FALSE
2	2	3	2	0.019258	0.016471	FALSE	FALSE
2	2	4	1	0.033984	0.030064	FALSE	FALSE
2	2	4	2	0.019283	0.016489	FALSE	FALSE
2	3	1	1	0.035875	0.028585	FALSE	FALSE
2	3	1	2	0.019652	0.016214	FALSE	FALSE
2	3	3	1	0.035664	0.030315	FALSE	FALSE
2	3	3	2	0.042781	0.030198	FALSE	FALSE
2	3	4	1	0.030782	0.026970	FALSE	FALSE
2	3	4	2	0.019947	0.017532	FALSE	FALSE
2	3	5	1	0.041751	0.035475	FALSE	FALSE
2	3	5	2	0.014364	0.010608	1	1
2	4	2	1	0.027595	0.024535	FALSE	FALSE
2	4	2	2	0.016945	0.014430	FALSE	FALSE
2	4	3	1	0.036522	0.032624	FALSE	FALSE
2	4	3	2	0.025844	0.022107	FALSE	FALSE
2	4	4	1	0.028336	0.024177	FALSE	FALSE
2	4	4	2	0.017249	0.011296	FALSE	1
4	3	2	1	0.028600	0.025510	FALSE	FALSE
4	3	2	2	0.042820	0.034884	FALSE	FALSE
4	3	2	3	0.038961	0.032669	FALSE	FALSE
4	3	2	4	0.038677	0.030901	FALSE	FALSE
4	3	4	1	0.038830	0.028057	FALSE	FALSE
4	3	4	2	0.090622	0.077331	FALSE	FALSE
4	3	4	3	0.036001	0.032262	FALSE	FALSE
4	3	4	4	0.048541	0.038965	FALSE	FALSE
5	1	5	1	0.035528	0.027496	FALSE	FALSE
5	1	5	2	0.027564	0.022603	FALSE	FALSE
5	1	5	3	0.024234	0.018458	FALSE	FALSE
5	1	5	4	0.033037	0.028722	FALSE	FALSE
5	1	5	5	0.032583	0.028313	FALSE	FALSE
5	4	2	1	0.029113	0.022801	FALSE	FALSE
5	4	2	2	0.028351	0.024919	FALSE	FALSE
5	4	2	3	0.032396	0.025751	FALSE	FALSE
5	4	2	4	0.041365	0.032929	FALSE	FALSE
5	4	2	5	0.022849	0.020929	FALSE	FALSE

Análisis de Sensibilidad – RNA Aumentada – Evaluación fuera de muestra – Desconexión no lineal – 1 periodo hacia adelante

P	Q	ITERACION	HIDE	RMSE	MAE	PRUEBA si RMSE < 0.0181	PRUEBA si MAE < 0.01455
1	3	1	1	0.020036	0.016026	FALSE	FALSE
1	3	4	1	0.019450	0.014725	FALSE	FALSE
1	4	3	1	0.021085	0.015893	FALSE	FALSE
1	4	4	1	0.032137	0.028430	FALSE	FALSE
2	1	1	1	0.018394	0.014376	FALSE	1
2	1	1	2	0.018685	0.016137	FALSE	FALSE
2	1	2	1	0.018394	0.014376	FALSE	1
2	1	2	2	0.018685	0.016137	FALSE	FALSE
2	1	3	1	0.018394	0.014376	FALSE	1
2	1	3	2	0.018685	0.016137	FALSE	FALSE
2	1	4	1	0.018394	0.014376	FALSE	1
2	1	4	2	0.018685	0.016137	FALSE	FALSE
2	1	5	1	0.018571	0.014690	FALSE	FALSE
2	1	5	2	0.018947	0.016319	FALSE	FALSE
3	1	5	1	0.020659	0.015770	FALSE	FALSE
3	1	5	2	0.018483	0.015575	FALSE	FALSE
3	1	5	3	0.018486	0.015580	FALSE	FALSE
3	3	3	1	0.027222	0.023279	FALSE	FALSE
3	3	3	2	0.019284	0.017141	FALSE	FALSE
3	3	3	3	0.023393	0.018974	FALSE	FALSE
3	4	2	1	0.068119	0.033526	FALSE	FALSE
3	4	2	2	0.021197	0.017708	FALSE	FALSE
3	4	2	3	0.022886	0.018439	FALSE	FALSE
4	1	1	1	0.022391	0.017046	FALSE	FALSE
4	1	1	2	0.020655	0.016800	FALSE	FALSE
4	1	1	3	0.020945	0.017372	FALSE	FALSE
4	1	1	4	0.019423	0.015219	FALSE	FALSE
4	1	4	1	0.024282	0.020361	FALSE	FALSE
4	1	4	2	0.016283	0.013018	1	1
4	1	4	3	0.019172	0.014603	FALSE	FALSE
4	1	4	4	0.025023	0.021457	FALSE	FALSE
4	2	2	1	0.018241	0.015745	FALSE	FALSE
4	2	2	2	0.018133	0.015720	FALSE	FALSE
4	2	2	3	0.017916	0.015483	1	FALSE
4	2	2	4	0.060875	0.033989	FALSE	FALSE
4	2	3	1	0.040925	0.026968	FALSE	FALSE
4	2	3	2	0.019180	0.015571	FALSE	FALSE
4	2	3	3	0.020899	0.016716	FALSE	FALSE
4	2	3	4	0.043762	0.027152	FALSE	FALSE
4	3	4	1	0.019517	0.015915	FALSE	FALSE
4	3	4	2	0.022615	0.014624	FALSE	FALSE
4	3	4	3	0.027694	0.023798	FALSE	FALSE
4	3	4	4	0.032660	0.027400	FALSE	FALSE
4	4	3	1	0.024221	0.018999	FALSE	FALSE
4	4	3	2	0.023925	0.017145	FALSE	FALSE
4	4	3	3	0.029235	0.025568	FALSE	FALSE
4	4	3	4	0.028973	0.025267	FALSE	FALSE