

Evaluación de pronósticos del tipo de cambio utilizando redes neuronales y funciones de pérdida asimétricas

Forecast Evaluation of the Exchange Rate Using Artificial Neural Networks and Asymmetric Cost Functions

MUNIR ANDRÉS JALIL^{1,2,a}, MARTHA MISAS^{1,b}

¹BANCO DE LA REPÚBLICA, BOGOTÁ, COLOMBIA

²UNIVERSIDAD NACIONAL DE COLOMBIA, FACULTAD DE ECONOMÍA, BOGOTÁ

Resumen

Se comparan especificaciones lineales y no lineales (estas últimas expresadas en redes neuronales artificiales) ajustadas a la variación porcentual diaria del tipo de cambio utilizando para ello funciones de costo tradicionales (simétricas) y funciones de pérdida asimétricas. Los resultados muestran que las redes neuronales permiten obtener mejores pronósticos con ambos tipos de funciones de costos. Sin embargo, es de anotar que cuando se evalúan los pronósticos con funciones asimétricas, el modelo no lineal supera ampliamente a su contraparte lineal.

Palabras clave: modelos para series de tiempo, modelos no lineales, tipo de cambio extranjero.

Abstract

We compare forecasts obtained via linear vs. non linear specifications. The models are adjusted to the daily percentage change of the exchange rate and the comparison is done using both symmetric and asymmetric cost functions. Results show that the non linear model (which here takes the form of an Artificial Neural Network –ANN) performs better in terms of forecasting ability when evaluated with both types of cost functions. Further more, when using asymmetric costs, the ANN is a much better predictor than its linear counterpart.

Key words: Time series models, Nonlinear models, Foreign exchange.

^aInvestigador y profesor asociado. E-mail: mjalilba@banrep.gov.co

^bEconometrista principal. E-mail: mmisasar@banrep.gov.co

1. Introducción

Tanto para el sector privado como para la autoridad monetaria es útil tener información sobre la evolución futura del tipo de cambio (entendido aquí como cantidad de pesos necesaria para comprar un dólar), ya que con esta es posible establecer la respuesta óptima ante el comportamiento predicho. Por lo anterior, en la literatura internacional se cuenta con un gran número de trabajos en los que se intenta generar *buenos* pronósticos de dicha serie (e. g., Diebold & Nasson (1990), Meese & Rogoff (1983) y Meese & Rose (1991)). Estos trabajos se pueden clasificar en dos grupos: los que usan modelos lineales y los que utilizan modelos no lineales. Una pregunta que surge entonces es ¿por qué se dan distintos modelos para la misma serie?, a la luz de la econometría, toda serie tiene asociado un *proceso generador de datos* y lo que se observa son las realizaciones de este. Sin embargo, dicho proceso es desconocido y lo que los investigadores deben hacer es aproximarlos, de tal manera que el modelo obtenido se ajuste a las realizaciones observadas y pueda ser utilizado para pronosticar de manera adecuada. Históricamente, el dominio lo han tenido los modelos lineales, los cuales han presentado un mayor desarrollo teórico y una mayor difusión. Además, la dificultad computacional de los procesos de estimación no lineal ha hecho que solamente en tiempos recientes se consideren tales modelos que, en principio, pueden representar mejor ciertas características observadas de las series. Es de anotar que, tal como lo menciona Watson (2005), en determinadas ocasiones los modelos no lineales pueden servir para replicar de manera adecuada procesos lineales cuya especificación es desconocida por el investigador.

Otro hecho interesante tiene que ver con la capacidad predictiva de los dos tipos de modelos. Con trabajos como los de Stock (2001), se ha generado la idea que los modelos no lineales no son mucho mejores que los lineales para pronosticar y por ello la relación costo de estimación *vs.* beneficio en pronóstico jugaría en contra de los modelos no lineales. En este sentido, la evaluación tradicional del pronóstico se ha concentrado en medidas simétricas de pérdida, en las que las magnitudes idénticas de errores positivos y negativos tienen el mismo costo asociado. Está por confirmarse, sin embargo, cómo se diferencian estas dos especificaciones a la luz de funciones de costo asimétricas en las que los errores de pronóstico tienen valoraciones distintas dependiendo de si son positivos o negativos.

Con respecto a la determinación del tipo de cambio, antes de los años setenta el modelo dominante para su determinación fue el del flujo de bienes. De acuerdo con este modelo, la demanda por moneda extranjera viene principalmente de compras y ventas de bienes. Por ejemplo, un incremento en las exportaciones aumenta la demanda externa por moneda doméstica con el fin de pagar por los bienes exportados. Entonces de lo anterior se puede concluir que países con superávit comercial (o sea aquellos que exportan más de lo que importan) tendrán flujos de moneda extranjera que al querer ser convertidos en moneda doméstica generarán un incremento en el precio de la moneda doméstica (una apreciación). Sin embargo, cuando se contrasta el resultado anterior con los datos se encuentra que los balances comerciales tienen una correlación muy baja con movimientos del tipo de cambio en los mercados de divisas más importantes. Este resultado negativo es de esperar

cuando se tiene en cuenta que el mercado de bienes y servicios representa una fracción muy pequeña del total de transacciones de moneda.

En los setenta para resolver el anterior inconveniente, surgió el modelo asociado al mercado de activos. Este se construyó tomando como punto de partida las implicaciones del modelo del flujo de bienes pero reconociendo que la demanda de moneda extranjera proviene no solamente de compras y ventas de bienes sino también de compras y ventas de activos. Por ejemplo, con el fin de adquirir TES, un inversionista extranjero localizado en los Estados Unidos debe comprar los pesos. Adicionalmente, el retorno en dólares del inversionista dependerá de los movimientos del peso con respecto al dólar; así su demanda por los bonos depende en parte de su deseo de especular en esos movimientos. Este cambio de perspectiva trajo un cambio en la estrategia de modelación. Los modelos comenzaron a incluir nociones tales como “eficiencia” especulativa: los tipos de cambio comenzaron a ser modelados como eficientes, en el sentido de que ellos incorporaban toda la información públicamente disponible, haciendo la información pública inútil para producir retornos extra. Esta es una característica que el modelo de mercado de bienes no presentaba.

El trabajo empírico no ha confirmado la idea de los mercados de activos. Las variables macroeconómicas que están detrás del mismo no mueven el tipo de cambio de la manera predicha. La referencia clásica en este sentido es Meese & Rogoff (1983), quienes muestran que el modelo de mercado de activos no puede explicar los tipos de cambio más importantes mejor que un modelo tan simple como el de “no cambio” o caminata aleatoria. Peor aún, los modelos de mercados de activos no permiten obtener consistentemente la dirección correcta de los cambios en el tipo de cambio. Lo anterior es recopilado por Meese en su revisión de literatura de 1990, en la que escribe que “la proporción de cambios en el tipo de cambio que pueden ser explicadas por los modelos actuales es esencialmente cero”. (La literatura que documenta el comportamiento “tan pobre” de este modelo es vasta. Para revisiones véase Frankel & Rose (1995), Isard (1995) y Taylor (1995).

Estas observaciones negativas no implican que el modelo de mercado de activos esté completamente equivocado. Por el contrario, en la academia existe un consenso que señala, en términos generales, que el modelo es apropiado. Aparentemente hay algo que hace falta para determinar la forma como el tipo de cambio es determinado, hecho que hace parte de agendas muy importantes de investigación en muchos lugares del mundo.

Por lo anterior, dado que las explicaciones económicas tradicionales necesitan completarse, el trabajo de pronóstico con modelos econométricos de series de tiempo autorregresivos aparece como una primera alternativa para pronosticar el comportamiento del tipo de cambio.

De acuerdo con Kuan & Liu (1995), es ampliamente aceptado que la tasa de cambio es un proceso integrado de orden uno, $I(1)$, y por ende, su cambio se puede describir como un proceso estacionario. Así, los cambios de la tasa de cambio podrían no ser linealmente predecibles¹. Surge la inquietud de si el problema de la no predictibilidad está asociado a limitaciones en los modelos lineales. Es decir,

¹Véase Baillie & McMahon (1989), citados por Kuan & Liu (1995).

si la existencia de no linealidades en el comportamiento del crecimiento de la tasa de cambio conlleva la dificultad que existe para alcanzar pronósticos adecuados al no reconocer tal comportamiento.

Adicionalmente, como lo señala Tenti (1996), la existencia de evidencia que apoya la hipótesis de caminata aleatoria implica que los cambios en la tasa de cambio son independientes e idénticamente distribuidos. Así, la única información relevante de su historia, para la predicción de movimientos futuros, es aquella más reciente.

Por otro lado, tal como lo muestran West et al. (1993), es posible que de maximizar el rendimiento de un portafolio específico que contiene el tipo de cambio, se obtenga que las medidas óptimas de tal procedimiento consistan en analizar los pronósticos de volatilidad a través de una medida asimétrica.

El presente trabajo compara especificaciones lineales y no lineales (expresadas en redes neuronales artificiales) ajustadas a la variación porcentual diaria del tipo de cambio, utilizando para ello funciones de costo tradicionales (simétricas) a la vez que se introduce el análisis asimétrico. Los resultados muestran que las redes neuronales permiten obtener mejores pronósticos con ambos tipos de funciones de costos. Sin embargo, es de anotar que cuando se evalúan los pronósticos con funciones asimétricas, el modelo no lineal supera ampliamente a su contraparte lineal.

La manera como se procede consiste en describir brevemente el método de identificación y estimación de una red neuronal artificial a la vez que se describe la metodología “*rolling*” que se siguió para llevar a cabo la evaluación de pronósticos. Luego se hace una descripción de la teoría general de pronósticos, con el fin de (i) señalar el conjunto de supuestos que tradicionalmente se hacen cuando se trata de evaluar pronósticos, y (ii) mostrar las ventajas que se pueden generar al llevar a cabo cambios en dichos supuestos. Lo anterior desde la perspectiva de la teoría de la decisión. Posteriormente, se lleva a cabo la comparación de los pronósticos obtenidos a través de una red neuronal con un modelo ARIMA y una caminata aleatoria utilizando para ello funciones de pérdida simétricas y asimétricas.

2. Modelos econométricos

La construcción de un modelo que relacione a una variable y_t con su propia historia y/o con la historia de otras variables, X_t , puede llevarse a cabo a través de una variedad de alternativas². Estas dependen de la forma funcional mediante la cual se aproxima la relación, como también de la relación existente entre dichas variables, es decir, de si esta es de carácter lineal o no lineal.

En nuestro caso, se consideran los pronósticos generados a través de modelos paramétricos lineales como son ARIMA y caminata aleatoria. Dichos pronósticos se contrastan con aquellos obtenidos mediante un modelo paramétrico no lineal de redes neuronales.

²Véase al respecto Granger & Teräsvirta (1993).

Dentro del grupo de modelos paramétricos lineales se consideran: (i) el modelo ARIMA, donde el comportamiento de una serie de tiempo, y_t , se explica a través de sus valores pasados y de una suma ponderada de errores, ε_t , pasados y presentes: $\Phi(L)(1-L)^d y_t = \delta + \Theta(L)\varepsilon_t$; con $\{\varepsilon_t\}$ serie de perturbaciones ruido blanco y d número de diferenciaciones requeridas para que $\{y_t\}$ alcance un comportamiento estacionario, y (ii) la caminata aleatoria donde se tiene que $y_t = y_{t-1} + v_t$, con $\{v_t\}$ serie de perturbaciones ruido blanco.

Siguiendo la literatura internacional (Van Dijk et al. 2001), en los últimos años el uso de modelos no lineales de series de tiempo se ha incrementado de manera considerable y, dentro de ellos, los de redes neuronales artificiales (ANN³). En el contexto de análisis de series de tiempo, las ANN se clasifican como modelos entrenados para (i) realizar enlaces entre los valores pasados y presentes de una serie de tiempo, aprendiendo de su error de pronóstico, y (ii) extraer estructuras y relaciones escondidas que gobiernan el sistema de información (Azoff 1996). Su utilización está primordialmente motivada por la capacidad de aproximarse a cualquier función medible de Borel con un muy buen grado de exactitud, como lo señala, entre otros, Rech (2002)⁴.

3. Redes neuronales artificiales

Como lo señalan Swanson & White (1995, 1997a, 1997b), Plasmans et al. (1998), entre otros, los modelos de redes neuronales artificiales se definen como una clase de modelos no lineales flexibles desarrollados por científicos cognitivos. Tales modelos están inspirados en ciertas características asociadas al procesamiento de información en el cerebro humano. El elemento central de este tipo de modelo es la estructura novedosa del sistema de procesamiento de la información, la cual está compuesta por un gran número de elementos interconectados de procesamiento que operan al mismo tiempo para resolver un problema específico. Dichos modelos son capaces de aprender mediante la interacción con su ambiente; tal aprendizaje puede ser entendido como un procedimiento estadístico de estimación recursiva. En particular, una red neuronal artificial se configura para una aplicación específica, de tal forma que el reconocimiento de patrones y la clasificación de información se alcanzan a través de un proceso de aprendizaje. Es de señalar que el aprendizaje tanto en sistemas biológicos como en las redes neuronales artificiales conlleva ajustes en las conexiones sinápticas entre neuronas.

Las redes neuronales artificiales han mostrado, en diferentes campos del conocimiento, una gran capacidad predictiva. Este hecho hace que, en la actualidad, se las considere como una herramienta importante en la elaboración de pronósticos de variables macroeconómicas y financieras. Una posible explicación de tal éxito es su gran habilidad para aproximar cualquier función si la red contiene un número amplio de términos no lineales y una adecuada selección de parámetros.

³Del inglés, *Artificial Neural Network*.

⁴Citando a Hornik et al. (1989).

3.1. Representación

De acuerdo con Kuan & Liu (1995), una red neuronal artificial es un tipo de modelo entrada-salida (*input-output*), que puede ser entendido como una función de regresión no lineal que caracteriza la relación entre una variable dependiente o *output* y_t y un vector de variables explicativas o *inputs* $X_t = (x_{1t}, \dots, x_{pt})$. De tal forma que, sin considerar una función específica no lineal, el modelo se construye combinando muchas funciones no lineales a través de una estructura multicapa (*multilayer structure*). Una clase de ANN, ampliamente estudiada e implementada en este trabajo, es la conocida como de alimentación hacia adelante con una única capa o superficie escondida (*single hidden layer feedforward neural network*).

En este tipo de red, las variables explicativas o *inputs* $\{x_{1t}, x_{2t}, \dots, x_{pt}\}$ activan de manera simultánea a las Q unidades escondidas en la superficie intermedia, a través de una función G , dando como resultado Q unidades escondidas de activación h_{it} , $i = 1, \dots, Q$, de tal forma que, posteriormente, estas unidades se activan a través de una función λ para producir el *output* \hat{y}_t . Simbólicamente, lo anterior se describe a través de las siguientes ecuaciones:

$$h_{it} = G\left(\gamma_{i0} + \sum_{j=1}^p \gamma_{ij}x_{jt}\right), \quad \forall i = 1, \dots, Q \quad (1)$$

$$\hat{y}_t = \lambda(h_{1t}, \dots, h_{Qt}) = \lambda\left(\beta_0 + \sum_{j=1}^Q \beta_j h_{jt}\right) \quad (2)$$

Las funciones de activación son funciones no lineales que pueden ser seleccionadas de manera arbitraria, con una restricción de acotamiento sobre G . Por ejemplo, es usual considerar a G como la función logística: $G(w) = \frac{1}{1+\exp(-w)}$ y a λ como la función idéntica, es decir: $\lambda\left(\beta_0 + \sum_{j=1}^Q \beta_j h_{jt}\right) = \beta_0 + \sum_{j=1}^Q \beta_j h_{jt}$. Como lo señalan Plasmans et al. (1998) y Franses & Van Dijk (2000), es conveniente incluir una conexión directa entre la superficie *input* y la *output* para incorporar de manera explícita el modelo lineal básico. Así,

$$\hat{y}_t = \beta_o + \tilde{X}'_t \Phi + \sum_{j=1}^Q \beta_j G(X'_t \gamma_j) + \varepsilon_t \quad (3)$$

donde $\tilde{X}'_t = (x_{1t}, \dots, x_{pt})$ y $X'_t = (1, x_{1t}, \dots, x_{pt})$. En general, la ecuación (3) se reescribe de la forma presentada en (4).

$$\hat{y}_t = F(X_t; \Theta) = \phi_o + \sum_{r=1}^P \phi_r x_{rt} + \sum_{l=1}^Q \beta_l G(Z'_t \gamma_l) + \varepsilon_t; \quad Z_t \subseteq X_t \quad (4)$$

siendo $\Theta = (\phi_r, r = 0, \dots, p; \beta_l, \gamma_l; l = 1, \dots, Q)$ donde $p \in \{0, 1, \dots, P\}$, de tal forma que cuando $p = P \Rightarrow Z_t = X_t$, de lo contrario Z_t es un subconjunto propio de X_t . Así, cada combinación p y Q determina una arquitectura particular. Es

de señalar que el número total de arquitecturas puede determinarse a través del contador J . Así, $J = 1$ se refiere a la arquitectura correspondiente a ($p = 1, Q = 1$) en tanto que $J = M$ es: ($P =$ máximo número de variables en la componente no lineal, $Q = 4$).

3.2. Aprendizaje

El proceso de aprendizaje de las ANN es de carácter secuencial (Kuan & White 1994). Así, el aprendizaje es un proceso donde la red adquiere conocimiento momento a momento; este se define como la acumulación de experiencias ocurridas. El conocimiento se adquiere a través de los conectores o parámetros de la red. Así, el conocimiento en el momento $t + 1$, $\widehat{\Theta}_{t+1}$ depende del conocimiento en el momento t , $\widehat{\Theta}_t$. Es decir, $\widehat{\Theta}_{t+1} = \widehat{\Theta}_t + \Delta_t$, donde el término Δ_t está asociado a un incremento en el conocimiento o aprendizaje, de tal forma que este depende del conocimiento previo obtenido de experiencias ocurridas $\{(X_1, y_1), \dots, (X_t, y_t)\}$ y de los nuevos valores observados (X_{t+1}, y_{t+1}) a través de una función apropiada: $\Delta_t = \Psi((X_{t+1}, y_{t+1}), \Theta_t)$.

El aprendizaje, en los modelos de redes neuronales, se centra en encontrar aquellos valores del conjunto de parámetros que hace mínima la siguiente diferencia:

$$S(\Theta) = \sum_{i=1}^T (y_t - F(X_t; \Theta))^2 \quad (5)$$

al considerar T observaciones de la forma $\{(X_t, y_t)\}_{t=1}^T$, donde y_t es la variable *output* o variable objetivo que la red neuronal debe generar cuando el t -ésimo vector *input* X_t aparece. Es decir, el aprendizaje puede ser visto como un problema general de minimización. Por consiguiente, este puede ser abordado a través de diferentes métodos de optimización no restringida, los cuales se llevan a cabo mediante algoritmos recursivos (véase Santana (2006)).

3.3. Evaluación fuera de muestra

Como es ampliamente conocido, en el contexto de las ANN es habitual subdividir el período de estudio en dos partes, de tal forma que en la primera se lleva a cabo el proceso de entrenamiento o aprendizaje y en la segunda el de evaluación (figura 1).

Una vez se lleva a cabo la estimación de las diferentes arquitecturas dentro de la muestra de entrenamiento se pasa a la evaluación por fuera de esta bajo un esquema de *rolling*. En dicho esquema se parte del conjunto de parámetros estimados para cada arquitectura en el período de entrenamiento, es decir, $\widehat{\Theta}_{1:T}^J, J = 1, \dots, M$, y se genera un pronóstico de horizonte 12 por arquitectura J , $\widehat{y}_{1:T}^{J,1}, \dots, \widehat{y}_{1:T}^{J,12}$. Posteriormente, se reestiman los parámetros de cada arquitectura considerando (i) como conjunto de información aquella que reúne la inicial o muestra de entrenamiento $\{Obs 1 : Obs T\}$ con la observación inmediatamente siguiente $\{Obs T + 1\}$, es decir, $\{Obs 1 : Obs T\} \cup \{Obs T + 1\} = \{Obs 1 : Obs T + 1\}$, y (ii) como valores

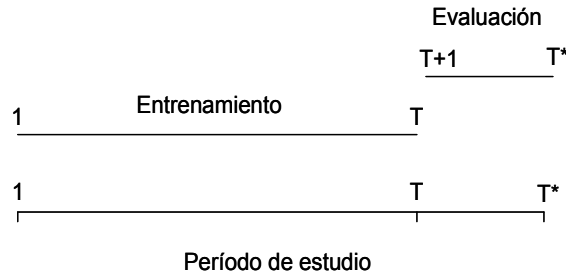


FIGURA 1: Subdivisiones del período de estudio de una red neuronal.

iniciales de los parámetros a aquellos obtenidos en el período de entrenamiento $\hat{\Theta}_{1:T}^J, J = 1, \dots, M$. De esta forma se produce un nuevo conjunto de parámetros: $\hat{\Theta}_{1:T+1}^J, J = 1, \dots, M$ con los cuales se lleva a cabo un pronóstico de horizonte 12 por arquitectura J , $\hat{y}_{1:T+1}^{J,1}, \dots, \hat{y}_{1:T+1}^{J,12}$. Así, en el i -ésimo paso se considera como conjunto de información a $\{Obs\ 1 : Obs\ T\} \cup \{Obs\ T + 1\} \cup \dots \cup \{Obs\ T + i\} = \{Obs\ 1 : Obs\ T + i\}$ para la reestimación y como valores iniciales de los parámetros por arquitectura a aquellos estimados en el paso anterior, es decir, $\hat{\Theta}_{1:T+i-1}^J, J = 1, \dots, M$. Con tales parámetros se generan, por arquitectura J , pronósticos de horizonte 12, $\hat{y}_{1:T+i}^{J,1}, \dots, \hat{y}_{1:T+i}^{J,12}$. Este procedimiento se lleva a cabo hasta el momento $(T^* - 1)$ para tener información observada de la variable y_t y construir las medidas de evaluación simétricas y asimétricas por horizonte $h = 1, 2, \dots, 12$ y por arquitectura ($J = 1, 2, \dots, M$). La figura 2 presenta la metodología anteriormente explicada.

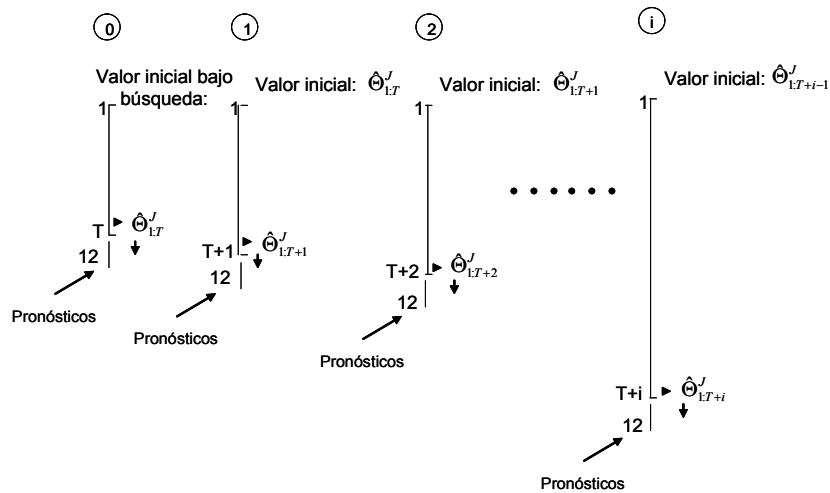


FIGURA 2: Esquema *rolling*.

Una vez se tiene el conjunto de pronósticos para cada arquitectura (J) cubriendo el período de evaluación, se calculan las medidas de comparación tanto

simétricas como asimétricas para cada uno de los horizontes, ($h = 1, 2, \dots, 12$). Así, las medidas de evaluación permiten seleccionar la mejor arquitectura por horizonte de pronóstico.

4. Teoría sobre decisiones y pronóstico

4.1. Caracterización del pronóstico óptimo

Para encontrar el pronóstico óptimo utilizando la teoría de la decisión, la idea es minimizar el riesgo (o maximizar la utilidad esperada).

Es importante mencionar una definición que hace énfasis sobre el rol de la forma funcional del pronóstico:

Definición 1. Un pronóstico óptimo $f_{t+h,t}^*$ es el pronóstico que minimiza la pérdida esperada (riesgo) i.e.

$$f_{t+h,t}^* = \arg \min_{f_{t+h,t}} E[L(y_{t+h}, f_{t+h,t}, z_t) | \Omega_t]$$

s.a. $f_{t+h,t} = g(x_t, \beta)$

donde

$$E[L(y_{t+h}, f_{t+h,t}, z_t) | \Omega_t] = \int_y L(y_{t+h}, f_{t+h,t}, z_t) dF_{Y|\Omega_T} y$$

$x_t \in \Omega_t$, β es un parámetro

Alternativamente, es posible tener un problema más familiar

$$\beta^* = \arg \min_{\beta} E[L(y_{t+h}, g(x_t, \beta), z_t) | \Omega_t]$$

y por consiguiente

$$f_{t+h,t}^* = g(x_t, \beta^*)$$

La definición anterior supone un modelo paramétrico para el pronóstico. Es posible utilizar una forma no paramétrica (y la primera definición así lo permite) pero, por el resto de esta sección, la especificación paramétrica será utilizada. Es importante recalcar que la forma del modelo puede ser lineal o no lineal.

4.2. La función de costos

La función de costos relaciona los resultados y los pronósticos y puede derivarse de la teoría económica. Sin embargo, esta no es la manera en que la literatura ha procedido. Los pronosticadores han utilizado funciones de costos que son matemáticamente convenientes o que tienen otras características llamativas pero que rara vez están relacionadas a los costos económicos o a funciones de utilidad (Pesaran & Skouras 2002).

Las restricciones en la definición son importantes, dado que no cualquier función puede ser una función de costos. En particular una función de costos debería existir si existen errores de pronósticos como tales. La primera restricción $L(x_1, x_2, x_3) = 0$ es solo una normalización. Nótese que los costos están representados como cantidades positivas, así los costos son positivos desde el comienzo o ellos deberían ser multiplicados por -1 .

Un subconjunto de funciones de costos es $L(y_{t+h}, f_{t+h,t}, z_t) = L(f_{t+h,t} - y_{t+h}) = L(e_{t+h,t})$. Este costo es llamado *función de pérdida de error de predicción* por Christoffersen & Diebold (1997). Granger & Newbold (1986) sugieren que mirar los errores de pronóstico es sensible porque tanto los pronósticos como las variables de interés tienen propiedades estadísticas diferentes y compararlas puede prestarse para confusiones.

Granger (1969) discute un conjunto más fuerte de condiciones para las funciones de costos. Las restricciones pueden o no pueden incluirse como parte del ejercicio de pronóstico (véase Granger (1969)):

- $L(0) = 0$,
- $L(e) > 0$ para $e \neq 0$, y
- $L(e)$ monotónicamente no decreciente en $|e|$,
- simetría: $L(-e) = L(e)$ para $p = 1$,
- homogeneidad: $L(\alpha e) = h(\alpha)L(e)$,
- convexidad,
- continuidad, y
- diferenciabilidad.

A continuación se presenta un conjunto de funciones de costos muy usado. Únicamente se muestran funciones convexas en donde el vector aleatorio es tan solo una variable aleatoria $p = 1$.

4.2.1. Error cuadrático medio (ECM)

$$L(e_{t+h,t} : \alpha) = \alpha e_{t+h,t}^2, \quad \alpha > 0$$

α es una constante libre que no es de importancia. Típicamente, $\alpha = \frac{1}{2}$ con el fin de ayudar con las derivadas. Este costo es el más popular en la literatura debido a su tractabilidad matemática. Es monotónicamente creciente, simétrico, homogéneo de grado 2 y diferenciable en todo su rango.

4.2.2. Error absoluto medio (EAM)

$$L(e_{t+h,t} : \alpha) = \alpha |e_{t+h,t}|, \quad \alpha > 0$$

esta función de costos es monotónicamente creciente, simétrica, homogénea y diferenciable en todo su rango con la excepción de $e_{t+h,t} = 0$.

4.2.3. Función linex (Linex)

Introducida por Varian (1974) y estudiada en detalle por Zellner (1986)

$$L(e_{t+h,t} : \alpha_1, \alpha_2) = \alpha_1 [\exp(\alpha_2 e_{t+h,t}) - \alpha_2 e_{t+h,t} - 1], \quad \alpha_1 \geq 0, \quad \alpha_2 \neq 0$$

La función está normalizada de tal manera que $L(0) = 0$. Esta función de costos es asimétrica. Si $\alpha_2 > 0$ es casi lineal a la izquierda del eje y y casi exponencial a la derecha. Esta función se voltea si $\alpha_2 < 0$. La función es diferenciable en todo su dominio. Nótese que si $\alpha_1 = \frac{1}{\alpha_2^2}$ entonces $\lim_{\alpha_2 \rightarrow 0} L(e) = \frac{e^2}{2}$ así para α_2 pequeño, el costo cuadrático está aproximadamente anidado dentro del costo linex.

4.3. La distribución condicional de predicción

La distribución condicional caracteriza completamente la variable aleatoria de interés. Si el interés es en un vector aleatorio (porque existen varias variables de interés o porque la idea es pronosticar varios períodos de tiempo), una distribución conjunta es apropiada. Desde el punto de vista de la teoría de la decisión, la distribución condicional describe la incertidumbre asociada al problema. Sin embargo, la mayor parte de la literatura de pronóstico no considera la distribución dado que únicamente están interesados en pronósticos puntuales, aunque en ocasiones, con el desconocimiento de los autores, existe un supuesto sobre la misma implícito. Desarrollos más recientes estiman el total de la distribución condicional. Esto puede ser hecho, por ejemplo, utilizando regresión por percentiles o análisis no paramétrico. Para una discusión véase Diebold et al. (1998) y Elliott & Timmermann (2002).

4.4. Interacción entre la función de costos y la distribución condicional

La importancia de la interacción entre el costo y la distribución condicional es bien sabida al menos desde el artículo de Granger (1969). Por ejemplo, asimetrías en ambos interactúan para definir el predictor óptimo. Asimetrías en la función de costos indican si existen costos diferentes asignados a sobre o subpredicciones, mientras la asimetría en la distribución indica si la realización de la variable de interés tiende a estar por encima o por debajo de la media.

Elliott & Timmermann (2002) muestran que otra manera de ver la interacción entre la función de costos y la distribución se logra mediante la realización de una expansión de Taylor alrededor de la media condicional del error de pronóstico $\mu_e = E[f_{t+h,t} - y_{t+h} | \Omega_t]$:

$$L(e_{t+h,t}) = L(\mu_e) + L'(\mu_e)(e_{t+h,t} - \mu_e) + \frac{1}{2}L''(\mu_e)(e_{t+h,t} - \mu_e)^2 + \sum_{k=3}^{\infty} \frac{1}{k!}L^k(\mu_e)(e_{t+h,t} - \mu_e)^k$$

y tomando expectativas

$$E[L(e_{t+h,t}) | \Omega_t] = L(\mu_e) + \frac{1}{2}L''(\mu_e)E[(e_{t+h,t} - \mu_e)^2 | \Omega_t] + \sum_{k=3}^{\infty} \frac{1}{k!}L^k(\mu_e)E[(e_{t+h,t} - \mu_e)^k | \Omega_t]$$

La interacción entre la forma de la función de costos (las derivadas) y los momentos (centrales) de la distribución se observa claramente en la expresión. Combinaciones entre los valores de las derivadas y los momentos de la distribución determinarán qué tanto de la una o la otra se necesita para aproximar la función de costos.

4.5. La elección del pronóstico óptimo

La elección del pronóstico óptimo involucra entonces un conjunto de interacciones que tradicionalmente no se tienen en cuenta. La mayor parte de la literatura sobre pronóstico supone una distribución simétrica para la generación de los mismos y evalúan estos con una función simétrica de costos. Lo que hemos visto en los apartados inmediatamente anteriores es que, fuera de este esquema convencional, se pueden presentar situaciones en las que la utilización tanto de funciones condicionales como de funciones de costos asimétricas generan pronósticos óptimos distintos a los obtenidos bajo esquemas simétricos. Es por esto que en la sección siguiente, aparte del análisis tradicional de pronóstico, se añadirá el análisis con funciones de pérdida asimétricas con el fin de ilustrar este punto.

5. Resultados

De las secciones anteriores se observa que un análisis de pronóstico con un modelo no lineal que tenga en cuenta las posibles asimetrías en la función de costos podría brindar información útil para la toma de decisiones por parte de los encargados de las mismas. Por ello se decidió, a manera de ilustración, tomar el cambio de la tasa de cambio nominal y realizar un ejercicio en el que se le ajusta tanto un modelo lineal como uno no lineal, y ambos se evalúan con medidas simétricas y asimétricas.

La evaluación de pronóstico del cambio de la tasa de cambio nominal se lleva a cabo sobre los pronósticos de un modelo lineal ARIMA y de uno no lineal que considera una red neuronal artificial autorregresiva. Este trabajo se lleva a cabo con información diaria correspondiente a la primera diferencia del logaritmo de la tasa de cambio nominal. El estudio abarca el período comprendido entre el 8 de febrero de 2000 y el primero de marzo de 2005. Intervalo de tiempo en el cual, tanto para el modelo ARIMA como para ANN, las últimas 60 observaciones son utilizadas para la evaluación *rolling* fuera de muestra.

5.1. Modelo no lineal

El período de entrenamiento corresponde a la muestra comprendida entre el 8 de febrero de 2000 y el 2 de diciembre de 2004, con un total de 1176 observaciones. La evaluación *rolling* fuera de muestra considera el período entre el 3 de diciembre y el primero de marzo de 2005, es decir, 60 observaciones. Con el propósito de mejorar las propiedades de estimación, como se mencionó anteriormente, la variable crecimiento de la tasa de cambio, $\Delta LTCN_t$, es reescalada en el intervalo $(0, 1)$.

La determinación de las variables *inputs* de la componente lineal o determinación del conjunto X_t , en cada red, se lleva a cabo mediante la estrategia *stepwise*⁵, propuesta por Swanson & White (1995, 1997a). Así, en dicha estrategia se parte de una regresión lineal cuya variable dependiente es $\Delta LTCN_t$ y cuyas posibles variables explicativas se seleccionan dentro de sus primeros 24 rezagos.

Una vez definido el conjunto de variables *input* de la componente lineal, X_t , se realiza el proceso de estimación de la red neuronal mediante el proceso de optimización numérica Quasi-Newton de Broyden, Fletcher, Goldfarb y Shanno⁶, para las diferentes configuraciones del conjunto de información Z_t , $Z_t \subseteq X_t$, de la componente no lineal⁷ y para un número de unidades ocultas Q que varían desde uno hasta cuatro⁸.

Franses & Van Dijk (2000) señalan cómo la convergencia en el proceso de estimación no garantiza la obtención del mínimo global. Por consiguiente, se llevan a cabo 30 estimaciones de cada una de las diferentes arquitecturas utilizando distintos valores iniciales del vector de parámetros γ . Tales valores iniciales se obtienen aleatoriamente a partir de una distribución uniforme entre $[-2, 2]$. Los parámetros del término de *weight decay* en la función objetivo $S(\Theta)$ son $r_\phi = 0.01$, $r_\beta = r_\gamma = 0.0001$.

La selección de las cinco estimaciones óptimas por arquitectura se realiza considerando dos criterios: (i) menor valor de la función objetivo y (ii) vector de gradientes, asociado a los parámetros de la estimación, sin elementos superiores a 1×10^{-3} . Una vez llevada a cabo dicha selección, se realiza el procedimiento de pronósticos fuera de muestra bajo el esquema *rolling*. Finalmente, se procede a calcular las medidas de evaluación simétricas y asimétricas.

⁵A pesar de ser la estrategia *stepwise* de carácter lineal, es frecuentemente utilizada como mecanismo de selección en el contexto de redes neuronales. Como lo expresa Franses frente a una consulta de Arango et al. (2004): “*As nonlinear functions can appear in dozens of formats, it is difficult to make a selection first. Hence, one usually starts with the first order linear approximation*”.

⁶Como lo sugieren Franses & Van Dijk (2000) y Rech (2002), este es uno de los algoritmos más utilizados en el contexto de redes neuronales para solucionar el problema de minimización, planteado en la ecuación (2).

⁷La especificación del conjunto Z_t se lleva a cabo de la siguiente manera: en un primer paso, o $p = 1$, el conjunto Z_t incluye la primera variable del conjunto X_t ; luego, en un segundo paso, o $p = 2$, se adiciona al conjunto Z_t la segunda variable de X_t de tal forma que en el último paso, $p = P$, se tiene la igualdad de los conjuntos, $Z_t = X_t$. Es de resaltar que el conjunto X_t que conforma la componente lineal permanece invariante a través de las diferentes arquitecturas.

⁸La selección de Q desde uno hasta cuatro es una regularidad empírica observada en trabajos similares.

La tabla 1 presenta, por horizonte, las medidas de evaluación RMSPE y MAPE, de carácter simétrico, de los modelos ARIMA, ANN y caminata aleatoria. Es de señalar que en el caso de las ANN se reporta aquella red o arquitectura que es la mejor en el sentido de mínima medida de evaluación. La existencia de evidencia a favor de la caminata aleatoria como proceso generador del nivel de la tasa de cambio lleva a que el mejor pronóstico de $\Delta LTCN$ es cero, valor frente al cual se lleva a cabo la evaluación. Como puede observarse, en el caso del RMSPE a horizontes menores a cinco días el mejor pronóstico se obtendría a través del supuesto de caminata aleatoria. Para los restantes horizontes, claramente se obtienen reducciones del error de pronóstico al considerar las diferentes arquitecturas de la red neuronal.

TABLA 1: Medidas simétricas de evaluación de pronósticos bajo *rolling* $\Delta LTCN$

HORIZONTE	P	Q	RMSPE		
			ANN	ARIMA	RW
1	4.00	3.00	119.80	125.72	100.00
2	2.00	3.00	108.73	118.98	100.00
3	2.00	3.00	102.27	125.05	100.00
4	2.00	3.00	103.40	125.11	100.00
5	4.00	2.00	102.92	124.84	100.00
6	3.00	4.00	96.65	125.11	100.00
7	3.00	4.00	97.82	126.55	100.00
8	5.00	4.00	95.40	130.15	100.00
9	5.00	4.00	92.46	131.22	100.00
10	5.00	4.00	107.90	145.55	100.00
11	9.00	2.00	99.70	102.58	100.00
12	3.00	3.00	100.71	102.66	100.00

HORIZONTE	P	Q	MAPE		
			ANN	ARIMA	RW
1	5.00	4.00	95.99	103.06	100.00
2	2.00	4.00	93.69	103.38	100.00
3	2.00	3.00	90.19	109.18	100.00
4	2.00	2.00	92.92	110.59	100.00
5	4.00	2.00	91.35	110.26	100.00
6	3.00	4.00	89.83	110.43	100.00
7	5.00	4.00	90.60	111.15	100.00
8	4.00	2.00	86.80	112.38	100.00
9	5.00	4.00	84.80	112.84	100.00
10	5.00	4.00	97.97	116.31	100.00
11	1.00	3.00	89.25	100.97	100.00
12	3.00	3.00	87.39	101.05	100.00

En lo referente a la medida MAPE, claramente se observa, para todo horizonte, la ventaja de trabajar con redes neuronales artificiales. Las dos medidas señalan la poca conveniencia de trabajar con modelos lineales como el modelo ARIMA.

La tabla 2 presenta las medidas simétricas de evaluación RMSE y MAE. Se observa, para todo horizonte de pronóstico, un mejor comportamiento de los pronósticos obtenidos a través de las redes neuronales.

TABLA 2: Medidas simétricas de evaluación de pronósticos bajo *rolling* $\Delta LTCN$

HORIZONTE	P	Q	RMSE		
			ANN	ARIMA	RW
1	2.00	2.00	0.0072	0.0076	0.0081
2	3.00	2.00	0.0075	0.0078	0.0081
3	1.00	4.00	0.0079	0.0081	0.0082
4	3.00	4.00	0.0079	0.0081	0.0082
5	7.00	3.00	0.0072	0.0081	0.0083
6	9.00	4.00	0.0078	0.0081	0.0083
7	9.00	4.00	0.0076	0.0080	0.0082
8	9.00	4.00	0.0075	0.0079	0.0081
9	2.00	4.00	0.0077	0.0080	0.0082
10	2.00	4.00	0.0079	0.0082	0.0082
11	7.00	4.00	0.0081	0.0083	0.0083
12	7.00	3.00	0.0080	0.0082	0.0082

HORIZONTE	P	Q	MAE		
			ANN	ARIMA	RW
1	2.00	2.00	0.0047	0.0052	0.0057
2	3.00	4.00	0.0050	0.0055	0.0058
3	5.00	4.00	0.0054	0.0057	0.0058
4	4.00	2.00	0.0055	0.0057	0.0058
5	7.00	3.00	0.0053	0.0058	0.0059
6	7.00	4.00	0.0053	0.0057	0.0058
7	9.00	4.00	0.0053	0.0056	0.0057
8	6.00	2.00	0.0051	0.0055	0.0056
9	2.00	4.00	0.0052	0.0055	0.0056
10	6.00	3.00	0.0054	0.0057	0.0056
11	4.00	4.00	0.0054	0.0057	0.0057
12	9.00	2.00	0.0053	0.0055	0.0056

En la tabla 3 se consignan los resultados de la estrategia asimétrica de evaluación de pronósticos del modelo lineal ARIMA, el modelo no lineal de redes neuronales artificiales, y los resultados de evaluación de los pronósticos de una caminata aleatoria. Como medida asimétrica se eligió una función Lin-Lin en la que se asigna un costo más alto a subpredicciones que a sobrepredicciones⁹. De la tabla se deduce que si existe un interés en un modelo que brinde pronósticos que no arroje demasiadas subpredicciones (dado que estas son costosas), una ANN es mucho más eficiente para este propósito que un modelo ARIMA o una caminata aleatoria a todo horizonte.

⁹La idea detrás de esta valoración proviene del hecho de que subpredicciones (pronosticar por debajo del valor efectivamente observado) significan mucho dinero perdido para el sistema financiero.

TABLA 3: Medida asimétrica de evaluación de pronósticos bajo *rolling* $\Delta LTCN$.

HORIZONTE	P	Q	LINLNP		
			ANN	ARIMA	RW
1	8	1	0.17	0.25	0.27
2	8	1	0.17	0.26	0.27
3	8	1	0.17	0.28	0.28
4	9	1	0.18	0.28	0.28
5	8	1	0.18	0.28	0.28
6	8	1	0.18	0.28	0.28
7	9	1	0.18	0.29	0.28
8	9	1	0.17	0.29	0.29
9	7	1	0.17	0.30	0.29
10	7	1	0.17	0.30	0.30
11	7	1	0.18	0.30	0.30
12	7	1	0.17	0.31	0.31

6. Conclusiones

Este trabajo compara pronósticos provenientes de un modelo no lineal (Red Neuronal) con los de un modelo lineal tradicional (ARIMA). Los pronósticos se obtienen a través de una metodología de *rolling* y su evaluación se lleva a cabo con respecto a medidas tanto simétricas (las cuales asignan la misma valoración a errores de la misma magnitud sin importar su signo) como asimétricas (las cuales permiten diferenciar los errores dependiendo no solamente de su magnitud sino de su signo). La literatura siempre ha tenido la visión que los pronósticos no lineales, si son mejores para pronosticar, no lo son de una manera abrumadora con respecto a sus contrapartes lineales. Este argumento ha sido siempre esbozado utilizando para ello funciones de pérdida simétricas. Por lo anterior este trabajo presenta un esquema de la teoría de decisión y pronóstico en economía, con el fin de ilustrar las distintas posibilidades que existen para evaluar predicciones. De allí se concluye que las funciones basadas en minimización de un error cuadrático medio son tan solo una de las muchas posibilidades existentes para evaluar la bondad de un pronóstico. Por lo anterior, en el presente documento se utilizaron, además de las medidas tradicionales, funciones de pérdida asimétricas con el fin de comparar, bajo este esquema, los pronósticos.

Los resultados obtenidos permiten concluir que, a la luz de las funciones de pérdida asimétricas, los modelos no lineales tienen una mejora considerable en capacidad de pronóstico, con respecto a los modelos lineales. Este resultado es robusto al horizonte de pronóstico, justificando así el uso de técnicas de estimación más complejas si lo que se necesita es solucionar un problema en el que el pronóstico óptimo deba ser evaluado con funciones de pérdida que no son simétricas.

Agradecimientos

Los resultados y opiniones son responsabilidad exclusiva de los autores y su contenido no compromete al Banco de la República ni a su Junta Directiva. Los

autores agradecen los comentarios y sugerencias de Andrés González, de los participantes del Seminario del Banco de la República en Medellín, de la Universidad ICESI en Cali y de los dos árbitros.

Recibido: septiembre de 2006

Aceptado: abril de 2007

Referencias

- Arango, C., Misas, M., López, E. & Hernández, J. N. (2004), 'No-linealidades en la demanda de efectivo en Colombia: las redes neuronales como herramienta de pronóstico', *Ensayos sobre política económica* **45**, 11–57.
- Azoff, E. M. (1996), *Neural Network. Time Series Forecasting of Financial Markets*, in 'Wiley, A Wiley Finance Edition'.
- Christoffersen, P. & Diebold, F. X. (1997), 'Optimal Prediction under Asymmetric Loss', *Econometric Theory* **13**, 806–817.
- Diebold, F. X., Gunther, T. A. & Tay, A. S. (1998), 'Evaluating Density Forecasts with Applications to Financial Risk Management', *International Economic Review* **39**, 863–883.
- Diebold, F. X. & Nasson, J. M. (1990), 'Nonparametric Exchange Rate Prediction', *Journal of International Economics* **28**, 315–332.
- Elliott, G. & Timmermann, A. (2002), 'Optimal Forecast Combinations under General Loss Functions and Forecast Error Distributions', *UCSD Working Paper*.
- Frankel, J. & Rose, A. (1995), *Empirical Research on Nominal Exchange Rates*, Handbook of International Economics, Elsevier Science, Amsterdam.
- Franses, P. H. & Van Dijk, D. (2000), *Non-linear Time Series Models in Empirical Finance*, Cambridge University Press.
- Granger, C. & Teräsvirta, T. (1993), *Modelling Nonlinear Economic Relationships*, Advanced Texts in Econometrics, Oxford University Press.
- Granger, C. W. J. (1969), 'Prediction with a Generalized Cost Function', *Operational Research* **20**, 199–207. Reimpreso en Ghysels, E., Swanson, N. R. & Watson, M. W. (eds.), *Essays in Econometrics: Collected Papers of Clive W. J. Granger*, volume I, 2001. Cambridge: Cambridge University Press.
- Granger, C. W. J. & Newbold, P. (1986), *Forecasting Economic Time Series*, 2nd edn, Academic Press, Orlando.
- Hornik, K., Stinchcombe, M. & White, H. (1989), 'Multi-Layer Feedforward Networks are Universal Approximators', *Neural Networks* **2**, 359–366.

- Isard, P. (1995), *Exchange Rate Economics*, Cambridge University Press, Cambridge.
- Kuan, C. M. & Liu, T. (1995), 'Forecasting Exchange Rates Using Feedforward and Recurrent Neural Networks', *Journal of Applied Econometrics* **10**, 347–364.
- Kuan, C. M. & White, H. (1994), 'Artificial Neural Networks: An Econometric Perspective', *Econometric Reviews* **13**.
- Meese, R. & Rogoff, K. (1983), *Exchange Rate and International Macroeconomics*, University of Chicago Press, Chicago, chapter The Out-of-sample Failure of Empirical Exchange Rate Models.
- Meese, R. & Rose, A. (1991), 'An Empirical Assessment of Non-Linearities in Models of Exchange Rate Determination', *Review of Econometric Studies* **58**.
- Pesaran, M. H. & Skouras, S. (2002), *A Companion to Economic Forecasting*, Blackwell Publishers, Oxford, chapter Decision-Based Methods for Forecast Evaluation.
- Plasmans, J., Verkooijen, W. & Daniels, H. (1998), 'Estimating Structural Exchange Rate Models by Artificial Neural Networks', *Applied Financial Economics* **8**, 541–551.
- Rech, G. (2002), 'Forecasting with Artificial Neural Network Models', *SSE/EFI Working paper Series in Economics and Finance* **491**.
- Santana, J. C. (2006), 'Predicción de series temporales con redes neuronales: una aplicación a la inflación colombiana', *Revista Colombiana de Estadística* **29**(1), 77–92.
- Stock, J. H. (2001), *A Companion to Theoretical Econometrics*, in 'Forecasting Economic Time Series', Blackwell Publishers.
- Swanson, N. R. & White, H. (1995), 'A Model-Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Networks', *Journal of Business & Economic Statistics* **13**(3).
- Swanson, N. R. & White, H. (1997a), 'A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks', *The Review of Economics and Statistics* **79**.
- Swanson, N. R. & White, H. (1997b), 'Forecasting Economic Time Series Using Flexible versus Fixed Specification and Linear versus Nonlinear Econometric Models', *International Journal of Forecasting* **13**.
- Taylor, M. (1995), 'The Economics of Exchange Rates', *Journal of Economic Literature* **83**, 13–47.

- Tenti, P. (1996), 'Forecasting Foreign Exchange Rates Using Recurrent Neural Networks', *Applied Artificial Intelligence* **10**, 567–581.
- Van Dijk, D., Teräsvirta, T. & Franses, P. H. (2001), Smooth Transition Autoregressive Models – A Survey of Recent Developments, Working paper series in economics and finance, Stockholm School of Economics.
- Varian, H. (1974), A Bayesian Approach to Real Estate Assessment, in S. Fienberg & A. Zellner, eds, 'Studies in Bayesian Econometrics and Statistics in Honor of L. F. Savage', North-Holland, Amsterdam, pp. 195–208.
- Watson, M. (2005), Comentario sobre "What's Real about the Business Cycle", Technical report, Federal Reserve Bank of St. Louis Review.
- West, K., Edison, H. & Cho, D. (1993), 'A Utility Based Evaluation of Some Models of Exchange Rate Variability', *Journal of International Economics* **35**, 23–46.
- Zellner, A. (1986), 'Bayesian Estimation and Prediction Using Asymmetric Loss Functions', *Journal of Forecasting* **8**, 446–451.