
Predicción de series de tiempo usando un modelo híbrido basado en la descomposición wavelet

Time series prediction using a hybrid model based on wavelet decomposition

Michael Vásquez^a
mvasquezl@unal.edu.co

Resumen

El pronóstico de series de tiempo que exhiben una estructura de segundo orden que varía en función del tiempo ha recibido especial atención debido a la dificultad de obtener buenos pronósticos, especialmente cuando existe una estructura poco homogénea al final de los datos. En este trabajo, se usa una metodología adecuada para pronosticar series de tiempo, con un alto nivel de ruido que evidencien no estacionariedad. Específicamente, se combina la transformación wavelet discreta de máximo traslape (MODWT) con el modelo ARFIMA-HYGARCH y redes neuronales. Ambos modelos se aplican para pronosticar la tasa de cambio USD/COP. Los resultados sugieren que la metodología basada en wavelets y redes neuronales, proveen pronósticos más precisos para pronosticar una apreciación/depreciación del tipo de cambio.

Palabras clave: pronóstico, modelos híbridos,, ARFIMA-HYGARCH, Descomposición Wavelet, tipos de cambio.

Abstract

The forecast of time series that exhibit a time varying second order structure has received special attention due to the difficulty to get accurate forecast. In this research, we used a methodology to forecast nonstationary time series contaminated with high noise levels. In particular, we combine the maximal overlap discrete wavelet transform (MODWT) with the model ARFIMA-HYGARCH and neural networks (NN). In the present study, both methodologies are applied to forecast USD/COP exchange rate. The results suggest that the hybrid methodology, based on wavelets and neural networks, provided more accurate forecast of an appreciation or a depreciation of the exchange rates.

Keywords: Forecast, Hybrid models, ARFIMA-HYGARCH, wavelet decomposition, exchange rates..

^aEscuela de Estadística, Universidad Nacional de Colombia, Medellín.

1. Introducción

Una gran cantidad de series de tiempo en las ciencias aplicadas muestran una estructura de segundo orden que varía en función del tiempo. Lo anterior, hace necesario no solo modelar la media condicional del proceso, sino también la varianza condicional. Una de las especificaciones más empleadas en la literatura ha sido el modelo ARFIMA-GARCH. Sin embargo, un hecho común surge en la mayoría de trabajos. El modelo se ajusta a la serie temporal de los datos que está contaminada, usualmente, por un alto nivel de ruido, lo cual deteriora el pronóstico de corto plazo.

Recientemente se ha propuesto la estimación a través de wavelets que en lugar de considerar la serie temporal de los datos tiene en cuenta el dominio tiempo escala. En este sentido, las wavelets permiten descomponer la serie de tiempo en diferentes niveles de resolución. En cada nivel se obtiene una señal de aproximación y una señal de detalle. La señal de aproximación toma en cuenta las frecuencias bajas, mientras que la señal de detalle corresponde a los componentes de alta frecuencia.

El análisis de wavelets, en el contexto de series de tiempo univariadas, se utiliza conjuntamente con modelos de pronóstico tradicionales de tipo paramétricos y semiparamétricos. Por lo tanto, el análisis de wavelets híbrido busca descomponer una serie de tiempo, que puede ser no estacionaria, en un conjunto de componentes más simples que pueden ser modeladas independientemente.

A continuación se mencionan algunos trabajos que usan la metodología wavelet para pronosticar series de tiempo. Conejo et al. (2005) proponen un modelo híbrido para el pronóstico de los precios de la electricidad en el mercado español. Específicamente, emplean un modelo wavelet-ARMA. La metodología híbrida es superior, en comparación con el modelo ARMA, para todas las estaciones del año. Karthikeyan & Nagesh Kumar (2013) emplean las wavelets en combinación con modelos ARMA para la predicción de series hidrológicas no estacionarias. Particularmente, emplean series mensuales relativas a caudales. A pesar de la no estacionariedad, el modelo wavelets-ARMA funciona de gran manera para predecir eventos de sequía a largo plazo. Zhu et al. (2014) proponen una metodología híbrida que combina una transformación wavelet no diezmada y el modelo ARMA. Específicamente, emplean la Transformada Wavelet Discreta de máximo Traslape (MODWT). El modelo propuesto es evaluado para una serie de precipitaciones diarias y la serie MacKey-Glass. En términos del pronóstico, el modelo híbrido MODWT-ARMA resulta superior que su contraparte híbrida basada en la Transformada Wavelet Discreta (DWT).

Algunos trabajos han empleado las wavelets con el modelo GARCH. Tan et al. (2010) proponen un modelo mixto que combina la metodología wavelet con el modelo ARMA GARCH. El modelo se aplica para la serie de precios de la elec-

tricidad del mercado español. Los resultados sugieren una evidencia clara a favor de la superioridad de la metodología híbrida respecto al modelo tradicional simple. Ismail et al. (2016) comparan el modelo GARCH y el MODWT-GARCH para pronosticar la volatilidad del índice de precios diarios para cuatro países africanos. La metodología híbrida arroja pronósticos más precisos.

En cuanto al uso de metodologías no paramétricas en combinación con la descomposición wavelet es común el uso de redes neuronales y Máquinas de Vectores de Soporte (SVM). Soltani (2002) proponen el uso del análisis multiescala en combinación con redes neuronales para la predicción de la serie MacKey-Glass. La metodología resulta superior comparada con otras técnicas semiparamétricas como las SVM y Funciones de Base Radial. Fernandez (2007) usa distintas series de tiempo mensuales de la industria manufacturera estadounidense para comparar la capacidad predictiva de varios modelos tradicionales y una metodología híbrida basada en wavelets. Particularmente compara un modelo ARIMA estacional multiplicativo, SVM, un modelo de componentes no observables y un modelo híbrido ARIMA-wavelets. Además, realiza una combinación lineal de los pronósticos para capturar las distintas características de cada modelo. Los resultados sugieren, en general, la superioridad de las combinaciones basadas en wavelets. Ben Mabrouk et al. (2008) proponen el uso de un modelo autorregresivo (AR) en combinación con wavelets y estimadores Kernel. Los datos analizados corresponden a la serie de manchas solares y a una serie generada por la ecuación MacKey-Glass. La metodología propuesta resulta superior, en términos del pronóstico, cuando se compara con modelos no paramétricos, el modelo autorregresivo de umbral y el modelo propuesto por Soltani (2002). Guo et al. (2008) proponen el uso de wavelets con SVM para pronosticar series de producción de diferentes industrias en Japón. Empleando distintas medidas de error de pronóstico, encuentran que el modelo híbrido es superior a las SVM tradicionales. Seo et al. (2017) usan la MODWT en combinación con SVM y Algoritmos Genéticos para modelar la corriente diaria del agua en una cuenca hidrográfica de Korea. Los resultados sugieren que la metodología híbrida mejora el rendimiento y la precisión de los modelos individuales.

En este trabajo se comparará la capacidad predictiva de la implementación híbrida de wavelets, tanto paramétrica como semiparamétrica, con otros modelos tradicionales. La principal motivación para el uso de wavelets es el análisis de fenómenos multiescala. Muchos autores observan que varios factores operan en distintas escalas de tiempo. Esta característica es usual en retornos logarítmicos financieros. En este trabajo se emplean los retornos diarios del tipo de cambio para el Peso Colombiano (COP) Dólar Estadounidense (USD).

2. Algunos modelos para series de tiempo

2.1. Modelo ARIMA Y ARFIMA

Un modelo empleado tradicionalmente en el pronóstico de series de tiempo ha sido el propuesto por Box & Jenkins (1970) conocido como Modelo Autorregresivo de Medias Móviles Integrado (ARIMA). Para una serie de tiempo $\{X_t, t = 1, \dots, T\}$, el modelo $ARIMA(p, d, q)$ se define como

$$\phi(L)(1-L)^d X_t = \theta(L)\epsilon_t, \quad (1)$$

donde L denota el operador de retardo, d el parámetro de diferenciación, $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ y $\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$ son los polinomios autorregresivos y de medias móviles, respectivamente. Si $\{X_t\}$ es estacionario e invertible, las raíces de ambos polinomios igualados a cero se encuentran fuera del círculo unidad, además los errores aleatorios ϵ_t son independientes e idénticamente distribuidos normales, con media cero y varianza σ^2 .

El parámetro de diferenciación es un número entero positivo, generalmente $|d| \leq 2$. Este parámetro surge como consecuencia de la existencia de series que no satisfacen la condición de estacionariedad en media. Sin embargo, algunas series pueden volverse estacionarias al ser diferenciadas una o más veces. Es decir, si se diferencia una vez la serie no estacionaria X_t , se puede obtener una nueva serie que cumpla esta propiedad $\nabla X_t = X_t - X_{t-1}$.

Si $d = 1$, el proceso puede ser persistente. La persistencia es entendida como el efecto permanente de un cambio unitario en la innovación sobre el nivel futuro (pronóstico) de la serie ¹. En este caso, se dice que el proceso tiene memoria infinita. Si $d = 0$, el modelo se reduce a un $ARMA(p, q)$ con memoria corta. Para determinar los órdenes de los polinomios autorregresivos y de medias móviles se sigue el procedimiento de modelación proporcionado por Box & Jenkins (1970).

Como lo señalan algunos autores (Granger (1980) y Castaño et al. (2008)), para algunas series de tiempo, la diferenciación entera puede generar que la componente asociada a la baja frecuencia desaparezca de la serie original, deteriorando el pronóstico de largo plazo.

Granger & Joyeux (1980) y Hosking (1981) introducen los modelos de memoria larga $ARFIMA$. Este proceso sigue la misma forma de (1) pero el parámetro de diferenciación toma un valor fraccional en un intervalo continuo de números reales.

¹ Como lo señala Cochrane (1991), la existencia de una raíz unitaria no es sinónimo de persistencia. En este sentido, de manera informal sería más adecuado decir que una serie que tenga una probabilidad mayor de rechazar la prueba de raíz unitaria generalmente puede tener menos persistencia.

El operador de diferencia fraccional se especifica como

$$(1 - L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k = \sum_{k=0}^{\infty} \frac{\Gamma(-d + k)}{\Gamma(k + 1) \Gamma(-d)} L^k, \quad (2)$$

donde $\Gamma(\cdot)$ denota la función Gamma. Si $-\frac{1}{2} < d < \frac{1}{2}$ el proceso es estacionario e invertible. Si $0 < d < \frac{1}{2}$, el proceso es estacionario con memoria larga, y si $-\frac{1}{2} < d < 0$ el proceso será estacionario con memoria corta y tendrá una fuerte reversión a la media.

2.2. Modelos GARCH

A menudo el supuesto de varianza constante en las dos especificaciones anteriores es bastante restrictivo, los retornos de los activos financieros tienen una naturaleza heteroscedástica. Lo anterior, hace necesario no solo modelar la media condicional del proceso, sino también la varianza condicional. El modelo de heteroscedasticidad más común en la modelización de series de tiempo es el Modelo de Heteroscedasticidad Condicional Autorregresiva Generalizada (GARCH) propuesto por Bollerslev (1986) el cual es una generalización del modelo ARCH introducido por Engle (1982). Dada una serie corregida por la media, $a_t = X_t - u_t$, donde u_t sigue una especificación ARMA o ARFIMA, el modelo $GARCH(m, s)$ se define como

$$a_t = \sigma_t \epsilon_t, \quad (3)$$

$$\sigma_t^2 = \alpha_0 + \alpha(L) a_t^2 + \beta(L) \sigma_t^2,$$

donde ϵ_t es una variable aleatoria independiente e idénticamente distribuida, $\alpha_0 > 0$, $\beta(L) = \beta_1 L + \dots + \beta_s L^s$ y $\alpha(L) = \alpha_1 L + \dots + \alpha_m L^m$. Cuando $s = 0$, se obtiene el modelo ARCH. Una condición necesaria y suficiente para que el proceso sea estacionario es $\sum_{i=1}^m \alpha_i + \sum_{i=0}^s \beta_i < 1$. Esta condición implica que la varianza incondicional es finita.

En muchos trabajos empíricos se ha detectado que no se cumple la condición de estacionariedad. Para resolver esta limitación, Engle & Bollerslev (1986) introducen el modelo GARCH Integrado (IGARCH). Asumiendo que $n_t = a_t^2 - \sigma_t^2$, la expresión (3) puede reescribirse como

$$(1 - \alpha(L) - \beta(L)) a_t^2 = \alpha_0 + (1 - \beta(L)) n_t, \quad (4)$$

donde n_t tiene media cero y no está serialmente correlacionada. Si el polinomio autorregresivo, $1 - \alpha(L) - \beta(L)$, tiene una raíz unitaria, se dice que el proceso es integrado. El proceso $IGARCH(p, q)$ es

$$\phi(L) (1 - L) a_t^2 = \alpha_0 + (1 - \beta(L)) n_t, \quad (5)$$

donde $\phi(L) = (1 - \alpha(L) - \beta(L))(1 - L)^{-1}$. Todas las raíces de $\phi(L)$ y $(1 - \beta(L))$ están por fuera del círculo unidad. Una extensión natural de este proceso, es reemplazar el operador de diferencia en (5) por el operador de diferencia fraccional definido en (2). Este modelo es conocido en la literatura como modelo GARCH fraccionalmente integrado (FIGARCH) y fue propuesto originalmente por Baillie et al. (1996). En la literatura se reporta evidencia de memoria larga en volatilidad, Ding & Granger (1996) para el índice bursátil de Japón (Nikkei) y Liu (2000) para el S&P 500.

La varianza condicional del FIGARCH (m, d, s) está dada por

$$\sigma_t^2 = \alpha_0 [1 - B(L)]^{-1} \left\{ 1 - [1 - B(L)]^{-1} \phi(L) (1 - L)^d \right\} \varepsilon_t^2 \quad (6)$$

ó $\sigma_t^2 = \omega^* + \sum_{i=1}^{\infty} \lambda_i L^i \varepsilon_t^2 = \omega^* + \lambda(L) \varepsilon_t^2$, con $0 \leq d \leq 1$. Las condiciones suficientes para asegurar que la varianza condicional del FIGARCH ($1, d, 1$) es estacionaria son: $\alpha_0 > 0$, $\beta_1 - d \leq \phi_1 \leq \frac{2-d}{2}$ y $d(\phi_1 - \frac{1-d}{2}) \leq \beta_1(\phi_1 - \beta_1 + d)$, ver Bollerslev & Mikkelsen (1996).

La amplitud del proceso FIGARCH es $S = \lambda(1) = 1$. La amplitud indica la dimensión de un shock en la varianza condicional. Consecuentemente, el segundo momento de la distribución incondicional de ε_t es infinito y el proceso no será estacionario en covarianza, ver Baillie et al. (1996) para más detalles. Davidson (2004) argumenta que la longitud de la memoria se incrementa cuando d se aproxima a cero. Esto contrasta con la interpretación convencional que sugiere que la memoria se incrementa conforme d aumenta. Davidson (2004) demuestra que este comportamiento inusual es producto de la restricción en la amplitud igual a uno y propone una generalización del proceso conocido como GARCH Hiperbólico HYGARCH. En la especificación anterior $\lambda(1)$ es reemplazado por $1 - [1 - B(L)]^{-1} \phi(L) \left\{ 1 + \alpha \left[(1 - L)^d \right] \right\}$.

La amplitud del proceso HYGARCH será $S = 1 - \frac{\phi}{B}(1 - \alpha)$. El modelo soluciona algunas limitaciones de la especificación anterior. El modelo hiperbólico es estacionario en covarianza, modela la amplitud y la memoria de forma separada. Finalmente, la memoria aumenta cuando d aumenta. El proceso será estacionario en covarianza, si satisface las restricciones de estacionariedad para la componente GARCH y $\ln(\alpha) < 0$, ver Davidson (2004) para más detalles. Kwan et al. (2012) muestran propiedades asintóticas de la estimación y derivan procedimientos de diagnóstico para el modelo general ARFIMA-HYGARCH.

Un hecho estilizado en los retornos de activos financieros es que su distribución incondicional tiene colas pesadas. La distribución Gaussiana es incapaz de capturar este fenómeno. Por lo tanto, los procesos de la clase GARCH han sido combinados con distribuciones que capturan este fenómeno. En este trabajo, además de la distribución normal, se emplea la Distribución de Error Generalizado (GED). Los

modelos son estimados usando máxima verosimilitud.

2.3. Generalidades sobre redes neuronales

Los modelos que se han presentado hasta ahora no capturan las dinámicas no lineales que presentan algunas series de tiempo en la media condicional. Un modelo no lineal usado ampliamente en la predicción de series de tiempo es la red neuronal perceptrón multicapa (MLP) con una sola capa oculta. Las redes neuronales artificiales son modelos semiparamétricos que, dada su estructura, imitan el funcionamiento del cerebro humano. están constituidas por unidades básicas de procesamiento de información denominadas neuronas que se interconectan entre sí y se organizan en varias capas: de entrada, ocultas y de salida.

La red neuronal opera de la siguiente manera: la capa de entrada recibe un conjunto de variables del mundo externo $x_i, i = 1, \dots, k$. Cada variable es ponderada por un factor de intensidad denotado γ_{ji} de forma que el input total que entra a la unidad oculta esta dado por $u_j = \sum_i \gamma_{ji} x_i$ (o $\mathbf{x}'\boldsymbol{\gamma}_j$, donde $\mathbf{x} = (1, x_1, \dots, x_k)'$ y $\boldsymbol{\gamma}_j = (\gamma_{j0}, \gamma_{j1}, \dots, \gamma_{jk})'$). Cada neurona de la capa oculta produce como output $h_j = \psi(\sum_i \gamma_{ji} x_i)$, es decir, una transformación no lineal del input total mediante el empleo de una función de activación diferenciable, acotada y no lineal. Las neuronas de la capa de salida reciben el input de la capa previa y, repitiendo el proceso antes descrito, producen el resultado final:

$$f(\mathbf{x}, \boldsymbol{\theta}) = \varphi(\beta_0 + \sum_{j=1}^q \beta_j \psi(\gamma_{j0} + \sum_i \gamma_{ji} x_i)), \tag{7}$$

donde φ es la función de activación de la capa de salida, ψ es la función de activación de la capa oculta, de tipo logístico $\psi(u) = \frac{1}{(1+e^{-u})}$, γ_{ji} es el peso de la neurona i de la capa de entrada, a la neurona j de la capa oculta. β_j representa la fuerza de la conexión de la neurona j , de la capa oculta, a la capa de salida. n es el número de inputs, q es el número de unidades ocultas y β_0, γ_{j0} corresponden a los sesgos de un modelo tradicional. Generalmente se usa la función identidad para φ , en este caso (7) se puede representar como

$$f(\mathbf{x}, \boldsymbol{\theta}) = \beta_0 + \sum_{i=1}^q \frac{\beta_j}{1 + e^{-(\gamma_{j0} + \sum_i \gamma_{ji} x_i)}} \tag{8}$$

Si se asume una conexión directa entre la capa de entrada y la de salida, (7) se puede definir como

$$f(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{x}'\boldsymbol{\delta} + \sum_{j=1}^q \beta_j \psi(\mathbf{x}'\boldsymbol{\gamma}_j), \tag{9}$$

donde $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_k)$. Tomando como variable de entrada retardos de la serie $\{X_t\}$, de modo que $\mathbf{x} = (X_{t-1}, X_{t-2}, \dots, X_{t-p})'$ y el número de unidades ocultas

$q = 0$, tendremos un modelo autorregresivo $AR(p)$ de la forma $f(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{x}'\boldsymbol{\delta} + \epsilon_t$.

La red que se acaba de describir, se conoce como red *feedforward* dado que las unidades de cada capa se conectan con las unidades de la capa siguiente. más detalles sobre este tipo de redes se pueden encontrar en Haykin (1994).

La red neuronal es estimada minimizando la siguiente función de costo

$$S(\boldsymbol{\theta}) = \sum_{t=1}^n |x_t - f(\mathbf{x}, \boldsymbol{\theta})|^2. \quad (10)$$

De modo que los parámetros $\boldsymbol{\theta} = \{\beta_0, \dots, \beta_q, \delta_1, \dots, \delta_k, \gamma'_1, \dots, \gamma'_q\}$ se ajustan recursivamente hasta minimizar la suma de residuales al cuadrado y obtener el mejor modelo que capture el comportamiento de la serie bajo estudio. Los pesos iniciales se asignan de forma aleatoria. En cada iteración el vector es actualizado de acuerdo a una regla de aprendizaje

$$\widehat{\boldsymbol{\theta}}^{(r+1)} = \widehat{\boldsymbol{\theta}}^{(r)} + \Delta^{(r)}. \quad (11)$$

La definición de esta regla conduce a diversas formas de optimización. Un método común es el de gradiente descendente que tiene la forma

$$\widehat{\boldsymbol{\theta}}^{(r+1)} = \widehat{\boldsymbol{\theta}}^{(r)} - \lambda A(\boldsymbol{\theta}^{(r)})^{-1} \nabla S(\boldsymbol{\theta}^{(r)}), \quad (12)$$

donde el parámetro λ es la tasa de aprendizaje y $A(\boldsymbol{\theta}^{(r)})$ es la matriz Hessiana de derivadas parciales de la función de costo respecto a los parámetros. En este trabajo se emplea el algoritmo BFGS sugerido por Franses & van Dijk (2000) para calcular dicha matriz. Además, para aumentar la capacidad de generalización de la red se emplea una componente que penalice la función de costo original. Esta penalización garantiza que cada conexión tiende al valor de cero a menos que sea importante. Para ello, se asigna una componente de regularización correspondiente al cuadrado de los pesos.

Como lo señala Soltani (2002), el desafío en la predicción de series de tiempo consiste en que el modelo sea capaz de capturar las dinámicas rápidas, correspondientes a las altas frecuencias, y cancelar el ruido simultáneamente. El sobreentrenamiento está relacionado directamente con el aprendizaje de ruido.

3. Metodología híbrida basada en wavelets

La DWT para una serie de tiempo, $X = \{x_t, t = 1, \dots, N\}$, se define como la transformación ortogonal

$$\omega = WX \tag{13}$$

$$X = W^T \omega \tag{14}$$

donde $\omega = \{\omega_n, n = 0, 1, \dots, N - 1\}$ es un vector columna de coeficientes wavelet de dimensión $N = 2^J$ y $W \in \mathbb{R}^{N \times N}$ es una matriz ortogonal. En términos de filtros, la DWT es un filtro $\{(h_n, n \in \mathbb{Z})\}$ con a lo más L valores no nulos. Las filas de la matriz W se componen de diferentes h_j y ceros.

Escribiendo $\omega = (\omega_1, \dots, \omega_J, v_j)'$, donde los primeros $\frac{N}{2}$ elementos se denotan como $\omega_1 = \{\omega_{1,t} = 0, 1, \dots, \frac{N}{2} - 1\}$ se define

$$\omega_{1,t} = \sum_{j=0}^{L-1} h_j x_{2t+1-j} \tag{15}$$

y

$$2t + 1 = j(\text{mod}N). \tag{16}$$

Los coeficientes de detalle se pueden definir como $d_j = W^T \omega_j, j = 1, 2, \dots, J$ y la serie x_t se puede expresar en términos de una combinación lineal de los coeficientes de detalle como

$$x_t = \sum_{j=1}^{J+1} d_{j,t}, \quad t = 0, \dots, N - 1. \tag{17}$$

Zhu et al. (2014), descomponen la expresión (17) para una serie de tiempo $X = \{x_t, t = 0, \dots, T - 1\}$, como

$$x_t = A_{J,t} + \sum_{j=1}^J D_{j,t}, \quad t = 0, \dots, T - 1, \tag{18}$$

donde $A_{J,t}$ es una subserie que representa la tendencia del proceso, mientras que $D_{j,t}$ presenta los detalles locales de la serie. Para realizar la proyección \hat{X}_{N+h} es necesario evaluar $\hat{A}_{J,N+h}$ y $\hat{D}_{j,N+h}$, así que se definen los siguientes modelos para ambas componentes

$$\hat{A}_{J,N+h} = f_0(A_{J,N}, A_{J,N-1}, \dots, A_{J,N-p_0}), \tag{19}$$

$$\hat{D}_{j,N+h} = f_j(D_{j,N}, D_{j,N-1}, \dots, D_{j,N-p_j}), \quad j = 1, \dots, J, \tag{20}$$

donde cada modelo $f_j(j = 0, \dots, J)$ tiene su propio orden p_j . La elección de f_j está relacionada con el comportamiento dinámico de cada subserie.

En este trabajo se propone usar el modelo general ARFIMA-HYGARCH y, debido al uso frecuente en la literatura de modelos no paramétricos, se usa una red neuronal. Para la descomposición de la serie, se emplea la transformación wavelet de máximo traslape (MODWT). Como se mencionó antes, es una transformación no diezmada y a diferencia de la DWT no requiere que $N = 2^J$. Además Zhu et al. (2014), usando la serie Mackey-Glass y una serie de precipitaciones diarias, muestran que la transformación MODWT es superior en términos de pronóstico a la transformación DWT. Con la transformación MODWT para cada nivel de descomposición, hay T coeficientes wavelet y de escalamiento. Dados los filtros en espejo $h_{j,n}$ y $g_{j,n}$ para cada escala $j = -1, -2, \dots, -J$, el filtro wavelet MODWT, $\tilde{h}_{j,n}$, y el filtro escalar $\tilde{g}_{j,n}$ se definen como

$$\tilde{h}_{j,n} = \frac{h_{j,n}}{2^{j/2}}, \quad \tilde{g}_{j,n} = \frac{g_{j,n}}{2^{j/2}}, \tag{21}$$

donde $n = 1, 2, \dots, N_h$ es la longitud del filtro. Los coeficientes wavelet en cada nivel j , serán la convolución de la serie de tiempo, $X = \{x_t, t = 0, \dots, T - 1\}$, y los filtros MODWT

$$\tilde{W}_{j,t} = \sum_{n=0}^{N_j-1} \tilde{h}_{j,n} X_{t-n \bmod T}, \quad \tilde{V}_{j,t} = \sum_{n=0}^{N_j-1} \tilde{g}_{j,n} X_{t-n \bmod T} \tag{22}$$

donde $N_j = (2^j - 1)(N_h - 1) + 1$. En notación matricial se pueden definir como

$$\tilde{W}_j = \tilde{w}_j X, \quad \tilde{V}_j = \tilde{v}_j X, \tag{23}$$

donde,

$$\tilde{w}_j = \frac{1}{2^k} \begin{bmatrix} \tilde{h}_{j,0} & \tilde{h}_{j,T-1} & \tilde{h}_{j,T-2} & \dots & \tilde{h}_{j,2} & \tilde{h}_{j,1} \\ \tilde{h}_{j,1} & \tilde{h}_{j,0} & \tilde{h}_{j,T-1} & \dots & \tilde{h}_{j,3} & \tilde{h}_{j,2} \\ \tilde{h}_{j,2} & \tilde{h}_{j,1} & \tilde{h}_{j,0} & \dots & \tilde{h}_{j,4} & \tilde{h}_{j,3} \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ \tilde{h}_{j,T-2} & \tilde{h}_{j,T-3} & \tilde{h}_{j,T-4} & \dots & \tilde{h}_{j,0} & \tilde{h}_{j,T} \\ \tilde{h}_{j,T-1} & \tilde{h}_{j,T-2} & \tilde{h}_{j,T-3} & \dots & \tilde{h}_{j,1} & \tilde{h}_{j,0} \end{bmatrix}.$$

\tilde{v}_j se define de forma similar reemplazando $\tilde{h}_{j,n}$ por $\tilde{g}_{j,n}$. La serie original puede reconstruirse como

$$X_t = \tilde{v}_J^T \tilde{V}_J + \sum_{j=1}^J \tilde{w}_j^T \tilde{W}_j = \tilde{S}_{J,t} + \sum_{j=1}^J \tilde{D}_{j,t}. \tag{24}$$

Zhu et al. (2014) demuestran que $\tilde{W}_{j,t} = \sum_{n=0}^{N_j-1} \tilde{h}_{j,n} X_{t-n}$, basado en un filtro wavelet Daubechies, será un proceso estacionario y de memoria corta. Para la elección de la base wavelet se recomienda un número de momentos de desvanecimiento que

sea razonable. Si se elige un número pequeño se puede originar variaciones bruscas artificiales en el momento de reconstruir la serie. Por el contrario, si el número de coeficientes es muy alto, estarán influenciados por las condiciones de borde y el esfuerzo computacional podría ser mayor. Fernandez (2007) señala que una elección común en la literatura es el filtro Daubechies de mínima asimetría de orden $N = 8$.

Los modelos híbridos, pueden definirse como sigue. La subserie de tendencia $\tilde{S}_J = \{\tilde{S}_{J,t}, t = 1, \dots, T\}$, seguirá el proceso ARFIMA-HYGARCH

$$\phi_{p_J}(L)(1-L)^{d_m}\tilde{S}_J = \theta_{q_J}(L)\epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2) \tag{25}$$

$$\sigma_t^2 = \omega [1 - B(L)]^{-1} \left\{ 1 - [1 - B(L)]^{-1} \phi_v(L) \left\{ 1 + \alpha [(1-L)]^{d_v} \right\} \right\} \varepsilon_t^2. \tag{26}$$

Los polinomios autorregresivos, $\phi_J(L)$, y de medias móviles, $\theta_J(L)$, se definen como en (1). Los parámetros de diferenciación fraccional para la media condicional, d_m , y la varianza condicional, d_v , se definen como en (2).

Las subseries de detalles, $\tilde{D}_j = \{\tilde{D}_{j,t}, t = 1, \dots, N\}$ seguirán el proceso ARMA-HYGARCH

$$\phi_{p_j}(L)\tilde{D}_j = \theta_{q_j}(L)\epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2) \tag{27}$$

y σ_t^2 seguirá un proceso HYGARCH definido como antes.

La descomposición en sucesivas subseries puede generar dependencias no lineales en la estructura temporal de los datos por lo que algunos autores han combinado la metodología wavelet con modelos no lineales. En este trabajo se emplea una red neuronal con una sola capa oculta. Así, para la serie de tendencia el modelo de redes neuronales se puede definir como

$$f(\tilde{Y}_i, \theta) = \beta_0^J + \sum_{k=1}^{q^J} \beta_k^J \psi(\gamma_{k0}^J + \sum_i \gamma_{ki}^J \tilde{Y}_i), \tag{28}$$

donde $\psi(\cdot)$ representa la función de activación logística, q^J es el número de unidades en la capa oculta y $\tilde{Y}_i = (\tilde{S}_{J,t-1}, \tilde{S}_{J,t-2}, \dots, \tilde{S}_{J,t-p})'$ son las variables explicativas. De forma análoga, para las subseries de detalles los modelos considerados son

$$f(\tilde{Z}_i, \theta) = \beta_0^j + \sum_{k=1}^{q^j} \beta_k^j \psi(\gamma_{k0}^j + \sum_i \gamma_{ki}^j \tilde{Z}_i), \tag{29}$$

donde $\tilde{Z}_i = (\tilde{D}_{j,t-1}, \tilde{D}_{j,t-2}, \dots, \tilde{D}_{j,t-p})'$. $\psi(\cdot)$ y q^j se define como antes. Soltani (2002) señala que las subseries correspondientes a los detalles, en los niveles más bajos, están contaminadas con un alto nivel de ruido y, por lo tanto, pueden ser omitidas (pronóstico cero) en el momento de realizar la predicción agregada. Esto puede verse como una medida adicional para evitar sobreentrenamiento.

Para la evaluación del pronóstico fuera de la muestra se compara la capacidad predictiva de la metodología ARFIMA-HYGARCH y el modelo híbrido basado en wavelets. Asimismo, estas metodologías se comparan con el modelo referente en la literatura, el paseo aleatorio. El pronóstico se evalúa usando el RMSE (Raíz del error cuadrático medio). Además, se usa el test de Diebold & Mariano (1995) para determinar si los errores de predicción de los modelos propuestos son menores a los del paseo aleatorio.

El estadístico de contraste de Diebold & Mariano (1995) examina la hipótesis nula de que la función de pérdida esperada entre los pronósticos sea la misma. Si e_{1t} y e_{2t} representan los errores de predicción correspondientes al paseo aleatorio y al modelo propuesto, respectivamente, el estadístico se define como

$$DM = \frac{\bar{d}}{\sqrt{\frac{LRV_{\bar{d}}}{T}}}, \quad (30)$$

donde $\bar{d} = \frac{1}{T-t_0} \sum_{t=t_0}^T (e_{1t}^2 - e_{2t}^2)$ es el diferencial medio de la función de pérdida y $LRV_{\bar{d}}$ es un estimador consistente del error estándar. Diebold & Mariano (1995) muestran que el estadístico se distribuye asintóticamente normal estándar bajo la hipótesis nula. El test es robusto al incumplimiento de supuestos estándar sobre los errores de previsión como normalidad y no autocorrelación. Un valor significativo y positivo de DM muestra una mayor precisión del predictor propuesto con relación al paseo aleatorio.

Una aproximación adicional a la evaluación puntual, consiste en predecir si el tipo de cambio sufrirá una apreciación o depreciación en el futuro, ver Walzack (2001) y Álvarez & Álvarez (2004). Este enfoque es adecuado particularmente en el análisis del mercado cambiario porque la existencia de errores predictivos muy pequeños, pero en la dirección equivocada, puede originar importantes pérdidas de capital. Para lo anterior, se emplea el test de acierto direccional (DA) de Pesaran & Timmermann (1992) que prueba la hipótesis nula de independencia entre el retorno observado X_t y el pronosticado \hat{X}_t . Si $P_x = P_r(X_t > 0)$, $P_{\hat{x}} = P_r(\hat{X}_t > 0)$ y \hat{P} es la tasa de acierto de signo, el estadístico DA se define como

$$DA = \frac{\hat{P} - \hat{P}_*}{\sqrt{\text{var}(\hat{P}) - \text{var}(\hat{P}_*)}}, \quad (31)$$

donde $\hat{P} = \sum_{t=t_0}^T \theta(X_t, \hat{X}_t)$ y $\hat{P}_* = \hat{P}_x \hat{P}_{\hat{x}} + (1 - \hat{P}_x)(1 - \hat{P}_{\hat{x}})$. $\theta(X)$ es la función heaviside definida como

$$\theta(X) = \begin{cases} 1 & \text{si } X_t \hat{X}_t > 0 \\ 0 & \text{si } X_t \hat{X}_t < 0. \end{cases} \quad (32)$$

Bajo la hipótesis nula el estadístico DA se distribuye normal estándar.

4. Resultados empíricos

Los datos empleados en este estudio corresponden a los retornos logarítmicos de la TRM (tasa de cambio representativa del mercado) del Peso Colombiano (COP) respecto al Dólar Estadounidense (USD) ². Sólo se consideraron días comerciales. Los datos tienen una longitud de 1074 observaciones comenzando desde el 4 de junio de 2012 hasta el 14 de diciembre de 2016. El período de estudio se divide en dos partes. Un conjunto de 1024 observaciones destinadas para la estimación y 50 observaciones por fuera de la muestra para llevar a cabo la predicción. En la Figura 1 se observa la evolución de los retornos y sus cuadrados a lo largo del período bajo estudio. La inspección visual parece indicar estacionariedad en media. Se observa claramente que existe un período de mayor volatilidad. Este hecho sugiere un proceso de segundo orden no estacionario.

4.1. Estadísticas preliminares

Las estadísticas básicas de la serie se muestran en la Tabla 1. El estadístico de Ljung-Box muestra evidencia de correlación serial significativa. La prueba de Jarque-Bera rechaza la hipótesis de normalidad, se observa exceso de curtosis característico de funciones de distribución leptocúrticas.

4.2. Ajuste modelo ARFIMA-HYGARCH

Una vez realizado el análisis inicial de la serie bajo estudio, se procede a modelar la media condicional. Para la estimación del modelo $ARFIMA(p, d, q)$ ³ se emplea una distribución GED. Los resultados del modelo se muestran en la Tabla 2. El modelo que mejor se ajusta a los datos es el $ARFIMA(1, d, 2)$. A pesar de que en la estimación conjunta el parámetro de diferenciación fraccional no resulta significativo, cuando se compara el criterio de información de Shibata del modelo

²Fuente: Los datos fueron tomados de www.superfinanciera.gov.co.

³La estimación y pronóstico de los modelos ARFIMA y de la familia GARCH es llevada a cabo usando la librería rugarch. Además, para la estimación de los modelos FIGARCH e HYGARCH se emplea Ox y se enlaza con la consola R Project.

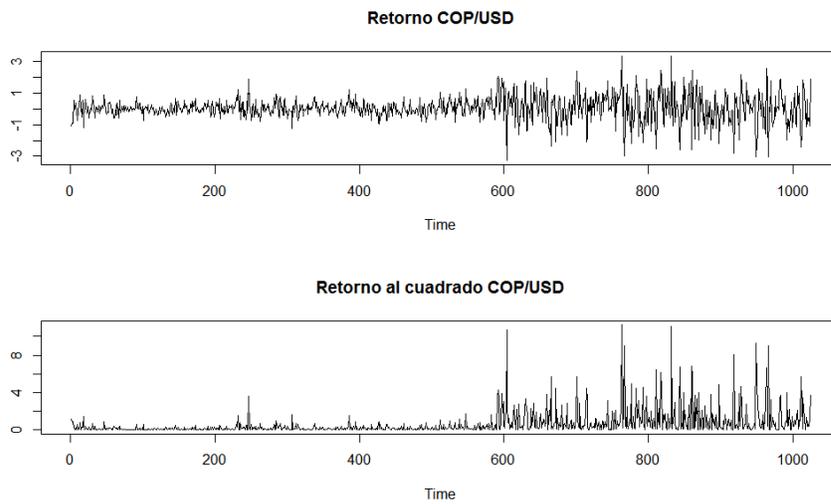


Figura 1: Retornos del tipo de cambio

Tabla 1: Estadísticas básicas para los retornos

Estadísticas	USDCOP
Media	0,045
Mediana	0,037
máximo	3,352
Mínimo	-3,278
Desviación Estandar	0,774
Asimetría	-0,176
Exceso de curtosis	2,441
Primer Cuartil	-0,318
Tercer Cuartil	0,421
Jarque-Bera	259,49 (0,000)

Fuente: elaboración propia.

USDCOP representa los retornos logarítmicos de la TRM del Peso Colombiano respecto al Dólar Estadounidense.

$Q(20)$ y $Q2(20)$ representan los estadísticos del contraste de Pormanteau de Ljung-Box hasta el rezago 20 sobre los retornos y sus cuadrados, respectivamente. Entre paréntesis se encuentran los p – valores.

Tabla 2: Estimación de la media condicional para los retornos

Retornos	USDCOP
Especificación	ARFIMA (1,d, 2)
Media condicional	
u	-
d	0,150 (1,829) {0,067}
ϕ_1	0,653 (20,217) {0,000}
θ_1	-0,592 (-19,465) {0,000}
θ_2	-0,1659 (-5,004) {0,000}
σ	0,748 (29,414) {0,000}
GED	1,023 (16,687) {0,000}
Log-L	-1087,933
Akaike	2,136
Bayes	2,1655
Shibata	2,137
Q(5)	4,405 (0,493)
Q(10)	8,173 (0,611)
Q(15)	16,178 (0,370)
Q2(5)	153,12 (0,000)
Q2(10)	287,38 (0,000)
Q2(15)	481,06 (0,000)
ARCHLM(5)	95,36 (0,000)

Fuente: Elaboración propia.

Q() y Q2() hacen referencia a la prueba de Ljung Box para los retornos y sus cuadrados, respectivamente. ARCHLM hace referencia a la prueba de Multiplicadores de Lagrange de Engle.

ARFIMA, respecto al modelo ARMA, resulta levemente superior. Como puede observarse, los residuales no están autocorrelacionados. Sin embargo, el test de Ljung-Box sobre los residuales al cuadrado rechaza hipótesis nula de no autocorrelación. Además, la prueba de multiplicadores de Lagrange de Engle rechaza hipótesis nula de no existencia de efectos ARCH.

Para modelar la varianza condicional se ajusta el modelo general ARFIMA-HYGARCH. Los resultados, para la varianza condicional, se presentan en la Tabla 3. Como puede observarse, los parámetros no resultan significativos, así que se procede a modelar el proceso de memoria larga FIGARCH(1,d,1). De nuevo, la no significancia de los parámetros no respalda esta especificación. El paso siguiente es modelar el proceso GARCH(1,1), sin embargo, los parámetros estimados no satisfacen la restricción $\alpha_1 + \beta_1 < 1$ para garantizar estacionariedad y que la varianza incondicional exista.

Se procede entonces a estimar el modelo integrado ARFIMA-IGARCH(1,1). Los resultados se presentan en la Tabla 4. Esta última especificación presenta parámetros estimados estadísticamente significativos y, ni los residuales estandarizados, ni

Tabla 3: Estimación de la varianza condicional para los retornos

Retornos	USDCOP		
Especificación	HYGARCH(1,1)	FIGARCH(1,1)	GARCH(1,1)
Varianza condicional			
α_0	0,003 {0,256}	0,001 {0,241}	0,000 {0,286}
d	0,451 {0,160}	1,000 {0,000}	-
α_1	0,372 {0,028}	0,032 {0,763}	0,0678 {0,000}
β_1	0,755 {0,000}	0,937 {0,000}	0,934 {0,000}
GED	1,668 {0,000}	1,678 {0,000}	1,671 {0,000}
HY	0,089 {0,618}	-	-

Fuente: elaboración propia.

Los p – valores para probar la significancia estadística de los parámetros están entre llaves.

Tabla 4: Estimación modelo ARFIMA-IGARCH para los retornos

Retornos	USDCOP
Especificación	ARFIMA (1, d ,2)- IGARCH(1,1)
Media condicional	
u	-
d	0,166 (1,374) {0,169}
ϕ_1	0,669(6,424) {0,000}
θ_1	-0,603 (-4,868) {0,000}
θ_2	-0,144 (-5,004) {0,000}
Varianza condicional	
α_0	0,001 (1,754) {0,079}
α_1	0,066 (5,310) {0,000}
β_1	0,933
GED	1,683 (16,056) 0,000
Log-L	-916,257
Akaike	1,803
Bayes	1,836
Shibata	1,803
Q(5)	1,875 {0,866}
Q(10)	5,366 {0,865}
Q(15)	9,984 {0,821}
Q2(5)	3,817 {0,576}
Q2(10)	5,488 {0,856}
Q2(15)	12,107 {0,670}
ARCHLM(5)	2,149 {0,439}
Nyblom	1,639 (1,9)

Fuente: Elaboración propia.

Q() y Q2() hacen referencia a la prueba de Ljung Box para los retornos y sus cuadrados, respectivamente.

sus cuadrados, exhiben algún tipo de autocorrelación. A pesar de que el parámetro de diferenciación fraccional en la media condicional no es significativo, esta especificación cuando se compara con un modelo ARMA en la media condicional tiene mejor criterio de información. Por lo tanto, se conserva dicho parámetro dentro del modelo. En cuanto a la estabilidad de los parámetros, el test de Nyblom sugiere que todos los parámetros son estables. El proceso IGARCH es de memoria corta, por lo tanto, un choque sobre el proceso de volatilidad desaparece en forma exponencial. Por lo tanto el modelo es adecuado para pronosticar por fuera de la muestra.

4.3. Metodología híbrida

Finalmente se lleva a cabo la metodología híbrida. La serie se descompone en distintos niveles ⁴. Se emplea como base wavelet la función Daubechies de mínima Asimetría de orden $N = 8$. Esta familia fue propuesta por Daubechies (1992) para reducir la asimetría de las wavelets originales. Esta familia tiene las mismas características de la familia original en cuanto a que posee p momentos de desvanecimiento, soporte en $[-p + 1, p]$ y es ortogonal.

A cada subserie de detalle se le aplica la metodología ARMA-HYGARCH (tener en cuenta que, desde el punto de vista teórico, las subseries de detalle son de memoria corta en la media condicional). La subserie de tendencia se modela como un proceso fraccionalmente integrado ARFIMA-HYGARCH.

En la Tabla 5, columna uno, se muestran las especificaciones resultantes. Sin embargo, un hecho común surge en todos los modelos, ninguna especificación logra cumplir que sus residuales sean incorrelacionados. La prueba de Ljung-Box rechaza la hipótesis nula de no autocorrelación para distintos rezagos. Por lo tanto, se concluye que la metodología ARFIMA-IGARCH no es adecuada para modelar las subseries de detalle.

En cuanto a la subserie de tendencia, aplicando la prueba KPSS, el estadístico para probar hipótesis nula de estacionariedad alrededor de una tendencia es igual a 0.22 con un p -valor de 0.01, se concluye entonces que el proceso es integrado. Para la especificación final estimada, se encuentra que los residuales están correlacionados.

Por lo anterior, se modela cada subserie mediante el uso de redes neuronales. Para determinar el orden de rezagos de la variable input, se realiza un procedimiento stepwise de dos vías y se selecciona el número de rezagos óptimos de acuerdo al criterio de información AIC. Además, dado que las redes neuronales son modelos

⁴Para la descomposición de la serie original se emplea la librería wavelets. En el anexo se muestra el código implementado.

Tabla 5: Prueba de Ljung Box sobre los residuales

Subserie	Especificación	Q(5)	Q(10)	Q(15)	Q(20)
D1	ARMA(2,4)-IGARCH	28,385 (0,000)	31,477 (0,000)	34,221 (0,003)	37,792 (0,009)
D2	ARMA(1,0)-IGARCH	614,13 (0,000)	626,33 (0,000)	629,47 (0,000)	634,37 (0,000)
D3	ARMA(2,2)*-IGARCH	502,99 (0,000)	669,36 (0,000)	672,15 (0,000)	676,27 (0,000)
D4	ARMA(1,0)-IGARCH	891,02 (0,000)	2231,3 (0,000)	2352,7 (0,000)	2568,7 (0,000)
D5	ARMA(3,0)*-IGARCH	211,46 (0,000)	246,67 (0,000)	329,42 (0,000)	539,23 (0,000)
D6	ARMA(0,0)-GARCH	4528,1 (0,000)	7009 (0,000)	7695,9 (0,000)	7740,2 (0,000)
S6	ARIMA(0,1,0)-GARCH	4480,7 (0,000)	7812,1 (0,000)	10134 (0,000)	11601 (0,000)

Fuente: Elaboración propia.

$Q()$ representa el estadístico de Ljung-Box sobre los retornos, entre paréntesis se encuentran los p – valores. (*) Indica que existen raíces dentro del círculo unidad.

semiparamétricos, se realiza un proceso de validación cruzada para determinar el número de unidades ocultas y el valor del parámetro de regularización. Para ello, se fija un rango de valores posibles para ambos parámetros y se dividen los datos en cinco submuestras del mismo tamaño. Por turno, una de las submuestras se usa como conjunto de prueba y las cuatro restantes como conjunto de entrenamiento. Posteriormente se promedian los errores obtenidos en los cinco turnos para cada valor del parámetro y se selecciona el que tenga el mejor desempeño. Este proceso se repite cinco veces.⁵

Una vez determinada la estructura de la red neuronal se procede a la estimación. Para evitar los mínimos locales en el proceso de optimización, se ajustan múltiples redes cambiando el valor de los pesos iniciales. Estos pesos se eligen de forma aleatoria y la red es evaluada en el segmento de entrenamiento de acuerdo al RMSE⁶.

4.4. Evaluación de los pronósticos

Para todos los modelos, se realiza el pronóstico un paso hacia adelante en un horizonte de 50 observaciones. El pronóstico sigue un esquema rolling en el sentido de que el modelo inicial se usa para generar el pronóstico un paso hacia adelante, posteriormente se reestiman los parámetros de cada modelo considerando como conjunto de información aquella que reúne la muestra inicial con la observación inmediatamente siguiente⁷. Este proceso se repite hasta recorrer todo el horizonte de pronóstico.

⁵La validación cruzada de la red neuronal es llevada a cabo usando la librería caret de R Project. En el anexo se muestra el código implementado.

⁶Para la estimación y pronóstico de la red neuronal se emplea la librería rminer. En el anexo se muestra el código implementado.

⁷Para conservar el tamaño muestral de 1024 observaciones se van eliminando de las muestra las primeras observaciones

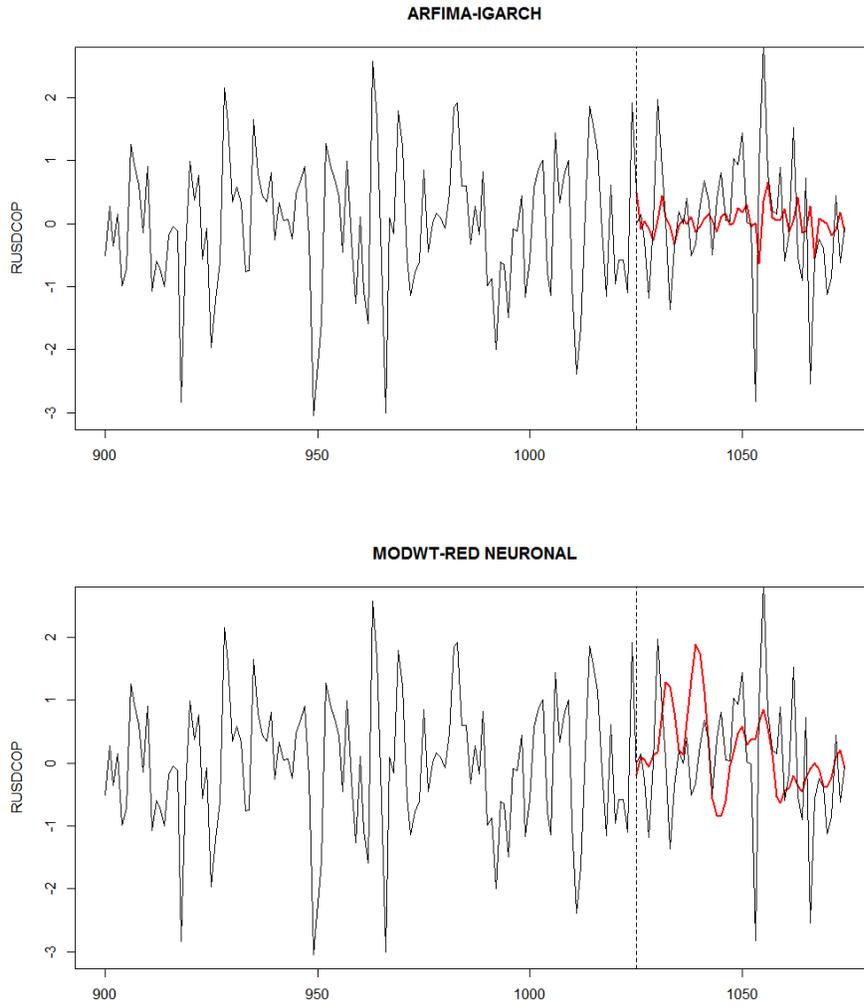


Figura 2: pronóstico modelo ARFIMA-IGARCH y modelo híbrido MODWT-RED. La línea negra corresponde a la serie original y la línea roja al pronóstico, por fuera de la muestra, de cada especificación.

Tabla 6: Resultados predicción RMSE, DA y DM

Especificación	RMSE	DA	DM
ARFIMA-IGARCH	0,961	56 % (0,778){0,218}	-0,2092 (0,835)
MODWT-RED	1,107	64 % (1,940){0,026}	1,434 (0,158)
RW	0,969	0 %	-

Fuente: elaboración propia.

En la Figura 2 se muestra el pronóstico un paso hacia adelante para los modelos propuestos. Además, en la Tabla 6 se muestran las medidas de evaluación de los pronósticos. En términos del RMSE, los modelos ARFIMA-IGARCH y paseo aleatorio (RW) tienen un desempeño similar. Para el modelo híbrido MODWT-RED, en términos del RMSE, tiene el peor desempeño. Sin embargo, de acuerdo al test de Diebold y Mariano (DM), la diferencia entre los errores de predicción respecto a los errores del paseo aleatorio de todos los métodos no son significativos. La hipótesis nula de que los dos métodos tienen la misma exactitud de pronóstico no puede ser rechazada en ningún caso. Por otro lado, cuando se evalúa el porcentaje de aciertos de signo y el contraste estadístico de Pesaran & Timmermann (1992), el único modelo que rechaza la hipótesis nula de independencia entre los valores pronosticados y observados en un nivel de significancia del 5 %, es el modelo MODWT-RED. El porcentaje de aciertos de signo es significativamente superior al que se obtendría bajo el modelo de paseo aleatorio. Por lo tanto, la metodología híbrida que combina wavelets y redes neuronales es capaz de predecir una eventual apreciación o depreciación del tipo de cambio.

5. Conclusiones

El uso de modelos no lineales en series de tiempo económicas representa un complemento al modelamiento macroeconómico tradicional, adecuado para trabajar con series procedentes de sistemas lineales simples y con principios fundamentales que subyacen en poco nivel de ruido. En la literatura, se sugiere que el proceso generador de datos de la tasa de cambio en un proceso integrado de orden uno (Kuan & Liu (1995)); por lo tanto, su cambio no estaría correlacionado y no sería linealmente predecible. Lo anterior, ha generado que se use un paseo aleatorio para pronosticar su valor. Sin embargo, este análisis no incorpora algunas características empíricas detectadas para esta serie como memoria larga, cluster de volatilidad y dependencias no lineales.

En este trabajo se ha comparado el pronóstico de dos modelos basados en wavelets respecto al modelo tradicional estacionario ARFIMA-HYGARCH y el modelo referente en la literatura, el paseo aleatorio. La serie bajo estudio fue el tipo de cambio USD/COP. El uso de la red neuronal se debe a que las subseries resultantes de la descomposición wavelet no se ajustaron bien a especificaciones lineales

y parecen exhibir dependencias no lineales. En términos del pronóstico, el modelo híbrido, basado en wavelets y redes neuronales, resulta superior que las demás especificaciones para pronosticar una apreciación o depreciación del tipo de cambio. Como trabajo futuro debe investigarse el desempeño de los modelos wavelet para otras series de tiempo.

Como trabajo futuro debe investigarse el desempeño de los modelos wavelet para otras series de tiempo. Especialmente para series de retornos financieros la evidencia empírica sugiere la existencia de no linealidad en la media condicional y una estructura de segundo orden (en varianza o covarianza) que varía en función del tiempo y, por lo tanto, su estructura espectral es variable. En la literatura comúnmente se usa el modelo ARMA-GARCH para pronosticar este tipo de procesos. Sin embargo, esta metodología asume el supuesto de estacionariedad. Los wavelets, empleados conjuntamente con modelos no lineales de pronóstico, permiten incorporar la mayoría de hechos estilizados reportados para series financieras sin asumir este último supuesto. La mejora en el desempeño del pronóstico será significativa, en la medida de que existan fuertes estructuras no lineales en la dinámica temporal de los precios.

Recibido: 2018-11-20
Aceptado: 2018-12-11

Referencias

- Álvarez, M. & Álvarez, A. (2004), 'Predicción no lineal de los tipos de cambio. aplicación de un algoritmo genético', *Revista de Economía Aplicada* **12**(35), 55–64.
- Baillie, R., Bollerslev, T. & Mikkelsen, H. (1996), 'Fractionally integrated generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **74**, 3–30.
- Ben Mabrouk, A., Ben Abdallah, N. & Z., D. (2008), 'Wavelet decomposition and autoregressive model for time series prediction', *Applied Mathematics and Computation* **199**, 334–40.
- Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**, 307–327.
- Bollerslev, T. & Mikkelsen, H. (1996), 'Modelling and pricing long memory in stock market volatility', **73**, 151–184.
- Box, G. & Jenkins, G. (1970), *Time series analysis: Forecasting and Control*, Holden-Day, San Francisco.
- Castaño, E., Gómez, K. & Gallón, S. (2008), 'Una nueva prueba para el parámetro de diferenciación fraccional', *Revista Colombiana de Estadística* **31**(1), 67–84.

- Cochrane, J. H. (1991), 'Pitfalls and opportunities: What macroeconomists should know about unit roots: Comment', *6*, 201–210.
- Conejo, A. J., Plazas, M. A., Espínola, R. & Molina, A. B. (2005), 'Day-ahead electricity price forecasting using the wavelet transform and arima models', *IEEE TRANSACTIONS ON POWER SYSTEMS* **20**, 1035–1042.
- Daubechies, I. (1992), *Ten Lectures on Wavelets*, SIAM, Philadelphia.
- Davidson, J. (2004), 'Moment and memory properties of linear conditional heteroscedasticity models, and a new model', *Journal of Business and Economics Statistics* **22**, 16–29.
- Diebold, F. X. & Mariano, R. S. (1995), 'Comparing predictive accuracy', *Journal of Business & Economic Statistics* **20**(1), 134–144.
- Ding, Z. & Granger, C. (1996), 'Modelling volatility persistence of speculative returns: A new approach', pp. 185–215.
- Engle, E. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation', *Econometrica* **50**, 987–1008.
- Engle, R. & Bollerslev, T. (1986), 'Modelling the persistence of conditional variances', *Econometric Reviews* **5**, 1–50.
- Fernandez, V. (2007), 'Wavelet-and svm-based forecasts: An analysis of the u.s. metal and materials manufacturing industry', *Resources Policy* **32**, 80–89.
- Franses, P. & van Dijk, D. (2000), *Non-linear time series models in empirical finance*, Cambridge University Press.
- Granger, C. W. J. & Joyeux, R. (1980), 'An introduction to long-memory time series models and fractional differencing', *Journal of Time Series Analysis* **1**, 15–39.
- Granger, W. (1980), 'Long memory relationships and the aggregation of dynamic models', *Journal of Econometrics* **14**, 227–238.
- Guo, X., Sun, L., Li, G. & Wang, S. (2008), 'A hybrid wavelet analysis and support vector machines in forecasting development of manufacturing', *Expert Systems with Applications* **35**, 415–422.
- Haykin, S. (1994), *Neuronal Networks: A Comprehensive Foundation*, New York: Mcmillan College Publishing Company.
- Hosking, J. R. M. (1981), 'Fractional differencing', *Biometrika* **68**, 165–176.
- Ismail, M. T., B., A. & M., T. M. (2016), 'Volatility forecasting with the wavelet transformation algorithm garch model: Evidence from african stock markets', pp. 125–135.

- Karthikeyan, L. & Nagesh Kumar, D. (2013), 'Predictability of nonstationary time series using wavelet and emd based arma models', *Journal of Hydrology* **502**, 103–119.
- Kuan, C. M. & Liu, T. (1995), 'Forecasting exchange rates using feedforward and recurrent neural networks', **10**, 347–364.
- Kwan, W., Keung, W. & Li, G. (2012), 'On the estimation and diagnostic checking of the arfima- h garch model', **56**, 3632–3644.
- Liu, M. (2000), 'Modelling long memory in stock market volatility', pp. 139–171.
- Pesaran, M. & Timmermann, A. (1992), 'A simple nonparametric test of predictive performance', *Journal of Business and Economic Statistics* **4**, 461–465.
- Seo, Y., Choi, Y. & Choi, J. (2017), 'River stage modeling by combining maximal overlap discrete wavelet transform, support vector machines and genetic algorithm', **9**(525).
- Soltani, S. (2002), 'On the use of the wavelet decomposition for time series prediction', *Neurocomputing* **48**, 267–277.
- Tan, Z., Zhang, J., Wang, J. & Xu, k. (2010), 'Day-ahead electricity price forecasting using wavelet transform combined with arima and garch models', *Applied Energy* **87**, 3606–3610.
- Walzack, S. (2001), 'An empirical analysis of data requirements for financial forecasting with neural networks', *Journal of management Information Systems* **4**, 203–222.
- Zhu, L., Wang, Y. & Fan, Q. (2014), 'Modwt-arma model for time series prediction', *Applied Mathematical Modelling* **38**, 1859–1865.

Anexo

6. Código: metodología híbrida para modelo ARFIMA-HYGARCH

La serie de retornos se descompone en 6 niveles de descomposición. Para ello, se emplea el comando para la Transformación Wavelet de Máximo Traslapo *modwt()* de la librería *wavelets*. Para cada subserie resultante, se emplea la misma estrategia de identificación empleada para la serie original.

```
library(wavelets)
mraout<-modwt(RUSDCOP[1:1024], filter="la8",
n.levels=floor(log(length(RUSDCOP[1:1024]))), fast =TRUE)
WaveletSeries<-cbind(do.call(cbind, mraout@W),
mraout@V[[floor(log(length(RUSDCOP[1:1024])))]])
ts_1<-as.numeric(WaveletSeries[,1])
ts_2<-as.numeric(WaveletSeries[,2])
ts_3<-as.numeric(WaveletSeries[,3])
ts_4<-as.numeric(WaveletSeries[,4])
ts_5<-as.numeric(WaveletSeries[,5])
ts_6<-as.numeric(WaveletSeries[,6])
ts_7<-as.numeric(WaveletSeries[,7])
```

7. Código: metodología híbrida para redes neuronales

7.1. Selección de variables

Se emplea la librería *MASS* para seleccionar las variables rezagadas que sean más importantes. Para ello, se emplea el comando *stepAIC* para aplicar la metodología stepwise. Sólo se muestran los comandos para la modelación de una subserie.

```
#ts1
variablesUSDCOPts_1<-CasesSeries(ts_1[1:1024],c(1,2,3,4,5,6,7,8))
fit=lm(y~.,variablesUSDCOPts_1[1:1024,])
stepts_1 <- stepAIC(fit, direction="both")
stepts_1$anova
```

7.2. Código: validación cruzada redes neuronales

Se realiza un análisis de validación cruzada para determinar cuáles son los parámetros adecuados para ajustar el modelo. El procedimiento se lleva a cabo usando la librería *caret*. Para entrenar las redes neuronales se emplea la librería *nnet*. Para la validación cruzada se emplean K-iteraciones. Los datos se dividen en k subconjuntos, uno de los subconjuntos se utiliza como datos de prueba y el resto como datos de entrenamiento. En nuestro caso, se emplean 5 folds y se repite durante 5 veces para cada uno de los posibles subconjuntos de prueba.

```
library(caret)
d_RUSDCOPts_1<-CasesSeries(ts_1[1:1024],c(1,2,3,4,5,6,7,8))
fitControl<-trainControl(method="repeatedcv", number=5, repeats=5)
nnetGrid<-expand.grid(size=seq(from=1, to=10, by=1),
  decay = seq(from=0.01, to=0.2, by=0.02))

set.seed(45567)
netFit_ts_1<-train(x=d_RUSDCOPts_1[,1:8],y=d_RUSDCOPts_1[,9],
  method="nnet", metric="RMSE", trControl=fitControl,
  tuneGrid=nnetGrid, verbose=FALSE)
plot(netFit_ts_1)
plot(varImp(object=netFit_ts_1),main="GBM - Variable Importance")

MFOR_RUSDCOPts_1=lforecast(netFit_ts_1,d_RUSDCOPts_1,1017,50)
```

7.3. Código: pesos iniciales redes neuronales y pronóstico

Se entrenan diferentes redes eligiendo distintos pesos iniciales para evitar caer en mínimos locales. Las redes son entrenadas con el 80% de las observaciones, el restante 20% se usa para la validación. Se usa el comando *fit* para estimar la red y se emplean los parámetros obtenidos en la fase de validación cruzada.

```
nstep<-100
model_semilla_ts_1<-rep(0,nstep)
rmse_ts_1<-rep(0,nstep)
searchs<-seq(1,nstep,1)
for(i in searchs){
  set.seed(c(i))
  M_RUSDCOPts_1<-fit(y~.,d_RUSDCOPts_1[1:792,],model="mlpe",size=1,
    decay=0.01, scale="all", rang = 0.7,abstol = 1.0e-3,
    reltol = 1.0e-3, maxit = 400, MaxNWts = 500)

  model_semilla_ts_1[i]=i
  rmse_ts_1[i]=mmetric(lforecast(M_RUSDCOPts_1,d_RUSDCOPts_1[1:792,],
```

```

      793,224),d_RUSDCOPts_1$y[793:1016],"RMSE")
}
pesos_ts_1<-cbind(model_semilla_ts_1,rmse_ts_1)
pesos_ts_1_f<-pesos_ts_1[which.min(pesos_ts_1[,2]), ]
d_RUSDCOPts_1_2<-CasesSeries(ts_1_2[1:1024],c(1,2,3,4,5,6,7,8))
set.seed(pesos_ts_1_f)
M_RUSDCOPts_1_2<-fit(y~.,d_RUSDCOPts_1_2[1:1016,],model="mlpe",
                    size=1, decay=0.01, scale="all", rang = 0.7,
                    abstol = 1.0e-3, reltol = 1.0e-3,maxit = 400,
                    MaxNWts = 500)
MFOR_RUSDCOPts_1_2<-lforecast(M_RUSDCOPts_1,d_RUSDCOPts_1_2,1017,1)

```

7.4. Medidas de evaluación de pronóstico

Para el pronóstico agregado de todas las subseries z se emplean las distintas medidas de evaluación.

```

rmse(z,RUSDCOP[1025:1074])
DACTest(z, RUSDCOP[1025:1074], test = "PT", conf.level = 0.99)
error_forred=z-RUSDCOP[1025:1074]
dm.test(error_forred,error_RW)

```