

**REDES NEURONALES ARTIFICIALES:  
PREDICCIÓN DE LA VOLATILIDAD DEL  
TIPO DE CAMBIO DE LA PESETA**

**María Bonilla, Paulina Marco e Ignacio Olmeda\***

WP-EC 2002-08

Correspondencia a: P. Marco. Universitat de València. Dpto. Economía Financiera y Matemática, Campus de los Naranjos. 46071 Valencia (Spain). E-mail: Paulina.Marco@uv.es.

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.

Primera Edición Abril 2002

Depósito Legal: V-1485-2002

Los documentos de trabajo del IVIE ofrecen un avance de los resultados de las investigaciones económicas en curso, con objeto de generar un proceso de discusión previo a su remisión a las revistas científicas.

---

\* M. Bonilla y P. Marco: Dpto. Economía Financiera y Matemática, Universitat de València. I. Olmeda: Dpto. Economía e Historia Económica, Universidad de Alcalá de Henares.

# **REDES NEURONALES ARTIFICIALES: PREDICCIÓN DE LA VOLATILIDAD DEL TIPO DE CAMBIO DE LA PESETA**

**María Bonilla, Paulina Marco e Ignacio Olmeda**

## **RESUMEN**

El presente trabajo propone el empleo de las Redes Neuronales Artificiales (RNA) al objeto de predecir la volatilidad del tipo de cambio de la peseta. En primer lugar, realizamos una comparación exhaustiva de la capacidad predictiva de las RNA en relación con otros modelos de la clase ARCH. Los resultados sugieren que, en media, las RNA se comportan mejor que los modelos tipo ARCH. Finalmente, también proponemos nuevos modelos híbridos para predecir la volatilidad que, basados en la técnica de las RNA, utilizan las predicciones de diferentes modelos paramétricos. Nuestros resultados muestran que el modelo híbrido que proponemos, en media, por lo general se comporta mejor que los otros modelos paramétricos y que la agregación lineal de las predicciones.

**PALABRAS CLAVE:** Volatilidad, predicción, no linealidad, modelos no paramétricos, redes neuronales artificiales (RNA).

## **ABSTRACT**

In this work, we propose the use of Artificial Neural Networks (ANNs), with the objective of predicting the volatility of peseta exchange rate. Firstly, we perform an exhaustive analysis of the forecasting ability of ANNs by comparing them against other ARCH-type models. The results suggest that ANN are, on average, better than ARCH models. Finally, we also propose new hybrid prediction models of volatility, based on ANNs, which use the forecasts of different parametric models. Our results show that the model is generally better, in mean, than other parametric models as well as a linear aggregation of forecasts.

**KEYWORDS:** Volatility, prediction, no linearity, non parametric models, artificial neural networks (ANNs)

# 1. INTRODUCCIÓN

El objetivo de numerosos estudios lo constituye la posibilidad de predecir las variaciones de los tipos de cambio, partiendo, para ello, de que la dinámica subyacente en los mismos es lineal. No obstante, los modelos lineales son incapaces de explicar numerosos fenómenos económicos y financieros. Además del fracaso empírico de dichos modelos en múltiples ámbitos, como el de los mercados financieros, sus fundamentos teóricos también son cuestionables. Los modelos lineales no permiten generar ciclos económicos aperiódicos, ni incorporar shocks como algo intrínseco en la evolución económica. Por ello, nos vemos obligados a extender nuestro campo de estudio a los modelos dinámicos no lineales que, a la hora de modelizar determinados fenómenos resultan mucho más potentes que el modelo lineal, aunque exigen elevados requisitos computacionales, así como un tratamiento analítico mucho más complejo.

En ocasiones, las series de variables financieras se presentan incorrelacionadas mostrando, por el contrario, dependencia no lineal. Además, existe una evidencia empírica suficiente a favor de que los tipos de cambio exhiben una dependencia no lineal<sup>1</sup>.

Así, recientemente, aparecen nuevas herramientas econométricas que, mediante la construcción de modelos teóricos que introducen elementos no lineales, ofrecen la posibilidad de modelizar y predecir la evolución de diversos fenómenos económicos y financieros, como el de la formación de los tipos de cambio y determinación de la volatilidad de los mismos.

Este trabajo se inserta en estas líneas de investigación y estudia la capacidad de las Redes Neuronales Artificiales (RNA), y en especial de un modelo específico y novedoso que introducimos de RNA híbrida, para explicar y predecir la volatilidad del tipo de cambio.

Las RNA se han aplicado con frecuencia en la predicción de rendimientos<sup>2</sup>, pero no en volatilidades. Pues bien, en este trabajo, no sólo empleamos las RNA para predecir series de volatilidades, sino que además proponemos un modelo híbrido,

---

<sup>1</sup> Véase Mañas (1986) y Hsieh (1988), entre otros.

<sup>2</sup> Pueden encontrarse múltiples ejemplos en Marco y Puertas (1999).

basado en la técnica de las RNA, comparando sus resultados con los de otros procedimientos competitivos de predicción.

Es interesante destacar que planteamos el empleo de las RNA, no sólo como alternativa de predicción a partir de observaciones históricas, sino también como modelo de agregación de otros modelos individuales de volatilidad, con el fin de explotar las posibles ventajas comparativas de cada uno de dichos modelos. Además, consideramos que la combinación lineal de la variable a predecir sobre las predicciones de los diferentes modelos, efectuada entre otros por Granger y Ramanathan (1984), no es necesariamente óptima, por lo que es posible que otras agregaciones no lineales permitan combinar los diferentes modelos más eficientemente. En este contexto, abordamos el problema de la agregación no lineal de modelos de predicción a través de métodos no paramétricos, en concreto, mediante RNA. Pensamos, por tanto, que nuestra propuesta puede resultar atractiva, ampliando, aún más, las herramientas disponibles de predicción.

Las RNA se aplican con éxito a numerosos campos de las ciencias, y en particular en la economía y las finanzas donde, su capacidad de ofrecer muy buenos resultados en la estimación no paramétrica y en la modelización y predicción de series temporales y sistemas dinámicos, facilita la toma de decisiones de los inversores y analistas financieros, así como, la resolución de múltiples tipos de problemas en el ámbito empresarial. Destaquemos la importancia que en este contexto cobran las RNA, como consecuencia de su capacidad de predecir variables financieras como rendimientos y volatilidades, tal y como muestran los numerosos ejemplos que pueden encontrarse en Marco y Puertas; 1999.

El artículo se estructura del siguiente modo. La sección 2 se dedica al estudio de las RNA, al constituir una técnica no paramétrica de modelización con gran capacidad para modelizar y predecir dinámicas no lineales. En la sección 3, se propone el empleo de las RNA para la predicción de la volatilidad del tipo de cambio, así como la utilización de modelos híbridos, que basados en una sencilla combinación de diferentes modelos, permiten construir modelizaciones más complejas que aprovechan las ventajas de cada uno de los modelos individuales. Igualmente hacemos una breve introducción de los métodos paramétricos que también empleamos en nuestro trabajo con fines predictivos, en concreto modelos de la clase ARCH. En la sección 4, efectuamos una aplicación de las RNA a la predicción de la volatilidad del tipo de cambio de la peseta frente a las seis monedas contempladas en nuestro análisis, exponiendo la base de datos,

metodología empleada en nuestro análisis y los principales resultados del estudio que realizamos. Las conclusiones y referencias citadas cierran el trabajo.

## **2. LAS REDES NEURONALES ARTIFICIALES**

Las RNA tratan de emular el sistema nervioso, de forma que son capaces de reproducir algunas de las principales tareas que desarrolla el cerebro humano, al reflejar las características fundamentales de comportamiento del mismo. Lo que realmente intentan modelizar las redes neuronales es una de las estructuras fisiológicas de soporte del cerebro, la neurona y los grupos estructurados e interconectados de varias de ellas, conocidos como redes de neuronas. De este modo, construyen sistemas que presentan un cierto grado de inteligencia. No obstante, debemos insistir en el hecho de que los Sistemas Neuronales Artificiales, como cualquier otra herramienta construida por el hombre, tienen limitaciones y sólo poseen un parecido superficial con sus contrapartidas biológicas.

Como afirman Kuan y White (1994), las redes neuronales artificiales, en relación con el procesamiento de información, heredan tres características básicas de las redes de neuronas biológicas: paralelismo masivo, respuesta no lineal de las neuronas frente a las entradas recibidas y procesamiento de información a través de múltiples capas de neuronas.

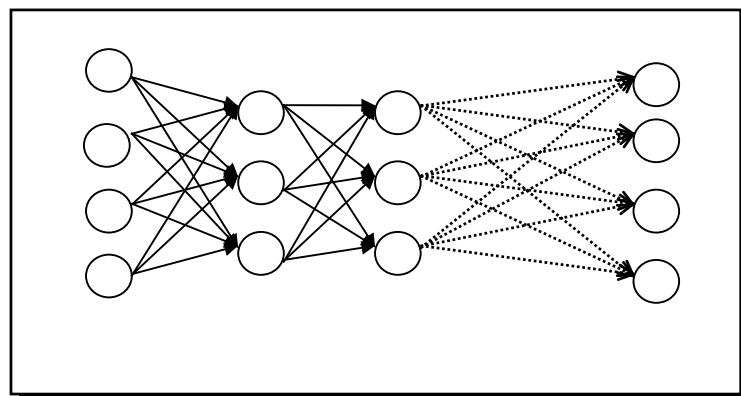
Una de las principales propiedades de este sistema es la capacidad de aprender y generalizar a partir de ejemplos reales. Es decir, la red aprende a reconocer la relación (que no deja de ser equivalente a estimar una dependencia funcional) que existe entre el conjunto de entradas proporcionadas como ejemplos y sus correspondientes salidas, de modo que, finalizado el aprendizaje, cuando a la red se le presenta una nueva entrada (aunque esté incompleta o posea algún error), en base a la relación funcional establecida en el mismo, es capaz de generalizarla ofreciendo una salida.

En consecuencia, podemos definir una red neuronal artificial como un sistema inteligente capaz, no sólo de aprender, sino también de generalizar. Una RNA se compone de un conjunto de procesadores simples altamente interconectados, denominados *nodos*, *neuronas* y *unidades* o *elementos de proceso*, los cuáles se organizan en grupos, llamados *capas*, que permiten la conexión con el exterior y el procesamiento de información. Pueden distinguirse tres tipos de capas:

- 1) La *capa de entrada*, mediante la cuál se presentan los datos a la red. Está formada por nodos de entrada que reciben la información directamente del exterior.
- 2) Las *capas ocultas o intermedias*, que se encargan de procesar la información, se interponen entre las capas de entrada y salida. Son las únicas que no tienen conexión con el exterior.
- 3) La *capa de salida*, que representa la respuesta de la red a una entrada dada, siendo esta información transferida al exterior.

La estructura de red más conocida es la que representamos en la figura 1, denominada *red alimentada hacia delante* o “*feedforward*”, puesto que las conexiones entre neuronas se establecen en un único sentido, por el siguiente orden: capa de entrada, capa(s) oculta(s) y capa de salida. Por ejemplo, en la figura se muestra una red con dos capas ocultas. No obstante, existen también *redes retroalimentadas* o “*feedback*”, que pueden tener conexiones hacia atrás, es decir, de nodos de una capa a elementos de proceso de capas anteriores, así como *redes recurrentes*, que pueden poseer conexiones, tanto entre neuronas de una misma capa, como de un nodo a sí mismo. La figura 2 ilustra un modelo de red en que coexisten los distintos tipos de conexiones que hemos comentado, es decir, hacia delante, hacia atrás y recurrentes, mostrando una interconexión total.

**Figura 1: Red alimentada hacia delante o “feedforward”**

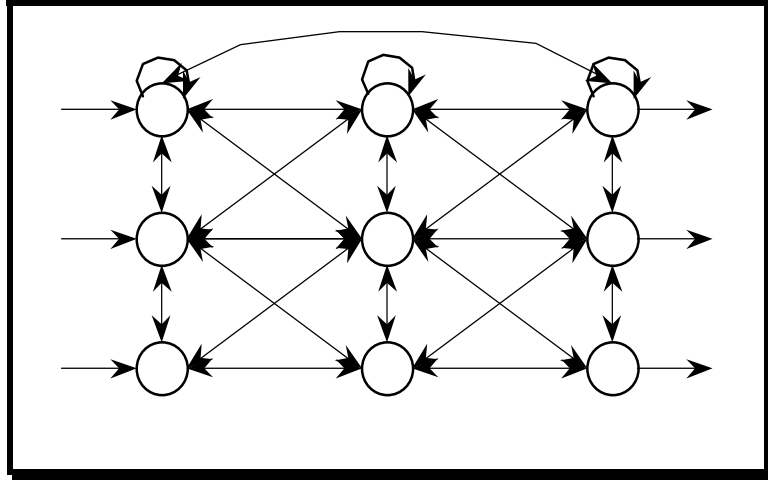


*Capa de entrada*

*Capas ocultas*

*Capa de salida*

**Figura 2: Red con interconexión total**



Una red neuronal artificial no es más que la conexión de neuronas, que se organizan en varias capas. Las conexiones entre elementos de proceso llevan asociadas un *peso* o *fuerza de conexión*,  $W$ , que determina cuantitativamente el efecto que producen unos elementos sobre otros. Es decir, en los pesos se almacena la información de la red, al igual que sucede en las redes de neuronas biológicas.

El que una entrada tenga un efecto excitatorio o inhibitorio, depende de que el signo del peso correspondiente sea, respectivamente, positivo o negativo. La efectividad de las entradas está determinada por la fuerza de la conexión, representada por el valor absoluto de los pesos. Así, cada uno de los elementos,  $W_{ij}$ , de la matriz de pesos,  $W$ , conocida como patrón de conexiones, representa la intensidad y sentido de la relación del elemento de proceso  $j$ , con respecto al elemento de proceso  $i$ .

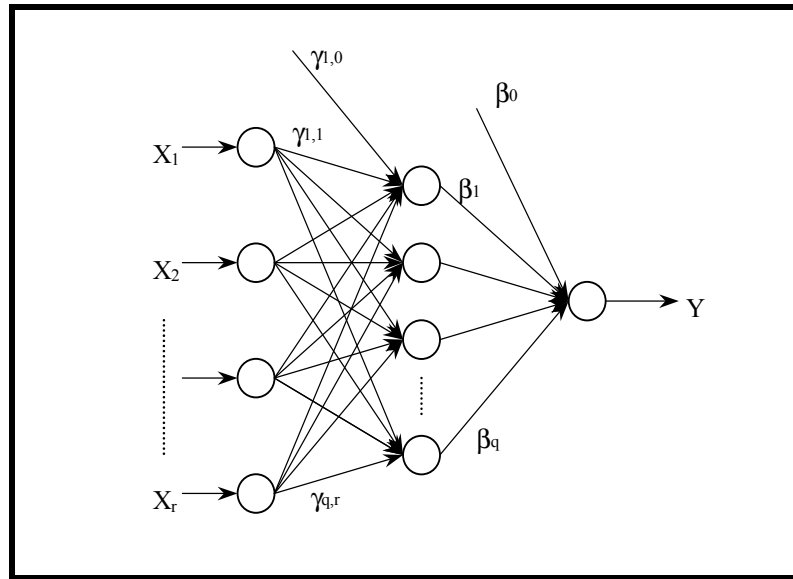
Siguiendo el planteamiento de Kuan y White (1994), el proceso de transformación de las entradas en salidas, en una red neuronal artificial alimentada hacia delante, con  $r$  entradas, una única capa oculta, compuesta de  $q$  elementos de proceso, y una unidad de salida (figura 3), puede resumirse en la siguiente formulación:

$$\hat{f}(x, W) = F \left( \beta_0 + \sum_{j=1}^q \beta_j G(x' \gamma_j) \right) \quad [1]$$

donde,  $\hat{f}(x, W)$  es la salida de la red, el vector  $x = (1, x_1, x_2, \dots, x_r)'$  representa las entradas de la red (el 1 se corresponde con el sesgo de un modelo tradicional),

$\gamma_j = (\gamma_{j0}, \gamma_{j1}, \dots, \gamma_{ji}, \dots, \gamma_{jr})' \in \mathfrak{R}^{r+1}$  son los pesos de las neuronas de la capa de entrada a las de la intermedia u oculta,  $\beta_j, j = 0, \dots, q$ , representa la fuerza de conexión de las unidades ocultas a las de salida ( $j=0$  indexa la unidad sesgo),  $q$  es el número de unidades intermedias, es decir, el número de nodos de la capa oculta,  $F: \mathfrak{R} \rightarrow \mathfrak{R}$  es la función de activación de la unidad de salida y  $G: \mathfrak{R} \rightarrow \mathfrak{R}$  se corresponde con la función de activación de las neuronas intermedias.  $W$  es un vector que incluye todos los pesos de la red, es decir,  $\gamma_j$  y  $\beta_j$ .

**Figura 3: Red neuronal artificial alimentada hacia delante, con  $r$  entradas, una única capa oculta, compuesta de  $q$  elementos de proceso, y una unidad de salida**



La función de activación permite incorporar en los modelos de redes neuronales artificiales una de las propiedades de los sistemas biológicos que, a su vez, introduce la importante característica de respuesta no lineal en las redes de neuronas artificiales. Este fenómeno puede resumirse en la tendencia de ciertas clases de neuronas a ponerse en actividad sólo cuando el nivel de actividad de su entrada supera un umbral determinado, en caso contrario permanecen estables.

Históricamente, en principio se emplearon funciones de umbral, cuyo efecto es que las unidades se activan bruscamente, esto es, o no se activan, o se activan de golpe. La respuesta sólo puede ser blanco o negro, por ello, éstas funciones son adecuadas para



tareas de clasificación y reconocimiento. Con el tiempo, se introdujeron funciones de activación que permiten que las neuronas se activen gradualmente a medida que el nivel de actividad de sus entradas aumenta, en lugar de que su estado pueda ser, únicamente, activación - desactivación. En concreto, la función que se propone es la sigmoideal o logística [Cowan, 1967],  $G(a) = 1/(1+\exp(-a))$ , que produce una respuesta sigmoideal alisada. Como podemos observar, si volvemos a la notación de la expresión 1 y consideramos que  $a = x'\gamma_j$ , nos encontramos con que  $G(x'\gamma_j)$  se corresponde con el conocido modelo logit de probabilidad binaria<sup>3</sup>.

En general, las funciones  $F$  y  $G$  pueden adoptar cualquier forma. Ahora bien, es práctica habitual considerar, bien que la función de activación de las neuronas de salida y de las intermedias es idéntica,  $F(a) = G(a)$ , y que se corresponde con la función sigmoideal, o bien, que  $F(a) = a$ , es decir, que es la función identidad y que  $G(a)$  se corresponde con la función logística o sigmoideal, lo que es equivalente a considerar que sólo existe función de activación (la sigmoideal) en las unidades ocultas. Esta última hipótesis es la que suponemos, a partir de ahora, en nuestro planteamiento, porque, además de simplificar enormemente la notación, es la que con mayor frecuencia se adopta en la construcción de redes neuronales artificiales. También debemos apuntar que empleamos la función sigmoideal, que es la más habitualmente utilizada, ya que sus propiedades le permiten adaptarse a algoritmos de aprendizaje como el de retropropagación de errores, que utilizamos en nuestro trabajo. La justificación formal de este hecho reside en que una RNA con una capa oculta y función de activación sigmoideal es capaz de aproximar cualquier función medible [Hornik et al., 1989].

Suponiendo, como hemos indicado, que sólo existe función de activación en las neuronas intermedias y que ésta se corresponde con la sigmoideal, la expresión 1 se transforma en:

$$\hat{f}(x, W) = \beta_0 + \sum_{j=1}^q \beta_j G(x'\gamma_j) \quad [2]$$

Otra posibilidad, de gran utilidad en aplicaciones econométricas, es considerar que en la red que representamos, una red neuronal artificial alimentada hacia delante, con  $r$  entradas, una única capa oculta, compuesta de  $q$  elementos de proceso, y una

---

<sup>3</sup> Para un análisis detallado, véase, por ejemplo, Amemiya (1981, p.268).

unidad de salida, también existen conexiones directas entre la capa de entrada y la de salida. En este caso, la salida de la red se obtiene mediante la siguiente expresión:

$$\hat{f}(x, W) = x'\alpha + \beta_0 + \sum_{j=1}^q \beta_j G(x'\gamma_j) \quad [3]$$

donde  $\alpha$  es un vector de dimensión  $r \times 1$  que representa los pesos de las conexiones directas entre las capas de entrada y salida. Como es lógico, ahora  $W$ , que recoge la totalidad de pesos de la red, se compone de  $\alpha$ ,  $\beta_j$  y  $\gamma_j$ .

Para interpretar la expresión 3 podemos descomponerla en dos partes. La primera de ellas, que se corresponde con los dos primeros términos, representa un modelo lineal, de manera que, si tomamos como variables de entrada  $r$  retardos de la variable  $x$ , se convierte en una regresión lineal sobre las variables de entrada retardadas, que actúan como variables explicativas, y una constante ( $\beta_0$ ). Es decir, en un contexto de series temporales en que  $x = x_{t-1}, x_{t-2}, \dots, x_{t-r}$  puede considerarse como un modelo autorregresivo de orden  $r$ ,  $AR(r)$ . Esta primera parte, como es lógico, capta las dependencias lineales entre los patrones de entrada y las salidas de la red. La segunda parte, que es el tercer término de la formulación anterior, recoge, en caso de que existan, las dependencias no lineales entre las variables de entrada y la salida de la red, dado que la función empleada es no lineal. Concretando, este tercer término es una composición, ponderada con los pesos sinápticos de las neuronas intermedias a las de salida ( $\beta_j$ ), de funciones sigmoideas de las entradas de la red, ponderadas, éstas últimas, por la fuerza de conexión de las unidades de entrada a las intermedias. Este modelo puede considerarse una extensión de los conocidos y, tan frecuentemente utilizados, modelos lineales, ya que se compone de un modelo lineal, aumentado con términos no lineales.

Como podemos apreciar, la red que describimos mediante la expresión 2, goza de tal grado de flexibilidad que permite ajustar todo tipo de funciones, por ello se caracteriza a las redes neuronales artificiales como “aproximadores universales”. Es decir, una red neuronal artificial es capaz de aprender cualquier función. En este sentido y, según demuestran Carroll y Dickinson (1989), Cybenko (1989), Hecht-Nielsen (1989), Hornik et al. (1989, 1990) y Stinchcombe y White (1989), entre otros, una red neuronal artificial alimentada hacia delante, con una única capa oculta compuesta de un número suficiente de neuronas intermedias, puede aproximar en cualquier grado a cualquier función medible. Este resultado, en principio, se demuestra para el caso en que la función de transferencia es sigmoidea, no obstante, también se puede demostrar

que se verifica para cualquier función no lineal. Sin embargo, debemos considerar que para los diferentes tipos de funciones, las propiedades de aproximación difieren.

Tras diseñar una red neuronal artificial, lo que pretendemos conseguir con la misma es que, para ciertas entradas, o patrones ejemplo que suministramos a la red, ésta sea capaz de generar una salida deseada. Para ello, además de que la topología de la red (entendida como la estructura de la red) sea adecuada, se requiere que la misma aprenda a proporcionar soluciones correctas, es decir, es necesario someter a la red a un proceso de aprendizaje o entrenamiento. El aprendizaje puede entenderse como un procedimiento de prueba y error que permite la estimación estadística de los parámetros del modelo de red neuronal empleado.

Podemos distinguir tres tipos básicos de aprendizaje; cuando el entrenador proporciona a la red la salida deseada, se dice que el *aprendizaje es supervisado*. En caso contrario, nos encontramos ante un *aprendizaje no supervisado*. Por último, un tipo intermedio de *aprendizaje* es el *reforzado*, en el cuál el entrenador sólo proporciona a la red una indicación de si la respuesta a una entrada dada es buena o mala.

Puesto que en nuestro trabajo únicamente empleamos RNA alimentadas hacia adelante y con aprendizaje supervisado, solamente nos detendremos en exponer el aprendizaje en este caso.

El proceso de *aprendizaje supervisado* de una red neuronal artificial pretende que la misma llegue a determinar la relación existente entre dos fenómenos de interés, las entradas y salidas deseadas, que denotamos, respectivamente, por  $X$  e  $Y$ . Considerando el caso más general (suponemos que son variables aleatorias) podemos agrupar los pares entrada-salida deseada en una variable aleatoria  $Z = (X, Y)$ . Dado que los pesos sinápticos ( $W$ ) recogen la relación entre las entradas y salidas de cada una de las unidades de una red neuronal, podemos definir de manera sencilla el aprendizaje como el proceso por el cual la red modifica dichos pesos hasta encontrar la configuración adecuada de los mismos, es decir, la que mejor capta la mencionada relación. Esto se consigue mediante reglas de aprendizaje que, a partir de una serie de  $n$  patrones-ejemplo,  $z^n = (z_1, z_2, \dots, z_n)$ , siendo  $z_i = (x_i, y_i)$  con  $x_i \in \mathfrak{R}^r$ ,  $y_i \in \mathfrak{R}^p$ , que proporcionan a la red una muestra de  $n$  ejemplos de la relación existente entre la entrada ( $x_i$ ) y la salida deseada ( $y_i$ ), van cambiando los pesos ( $W$ ), hasta que sus valores son los adecuados.

Generalmente, el método que emplea el algoritmo de aprendizaje para conocer el grado de adecuación de los pesos consiste en comparar la salida de la red ( $\hat{f}(x_t, W)$ ) con la deseada ( $y_t$ ), es decir, calcular el error cometido ( $e_t = e(\hat{f}(x_t, W), y_t)$ ). De este modo, cuando se producen errores se reajustan los pesos y reinicializa el proceso, hasta lograr la estructura de pesos idónea, esto es, la que minimiza el error de salida de la red ( $e_t \in \mathfrak{R}^p$ ) a lo largo del conjunto de ejemplos ( $z^n$ ), al que normalmente nos referimos como conjunto de entrenamiento o aprendizaje.

En este punto de la exposición, podemos definir el concepto de generalización. Una red neuronal generaliza en un sentido óptimo, si para cualquier observación seleccionada aleatoriamente que se le presente, aunque no forme parte del conjunto de entrenamiento, la salida que genera minimiza el error de salida de la red.

Formalmente, el proceso del aprendizaje consiste en resolver un problema de mínimos cuadrados no lineales. Para ello, hay que emplear métodos numéricos de optimización como el de *retropropagación de errores* (“Backpropagation”), que surge de la mano de Werbos (1974), atribuible también a Parker (1982, 1985) y Le Cun (1985). Este método se fundamenta en el algoritmo de aproximación estocástica de Robbins y Monro (1951). Actualmente, viene siendo el algoritmo más utilizado.

Como hemos comentado, el proceso de aprendizaje consiste en modificar los pesos de la red (donde se acumula la información), hasta que el patrón de conexiones,  $W$ , resulta óptimo. En este sentido, si la red aprende, el conocimiento acumulado en el momento  $t+1$ ,  $W_{t+1}$ , debe ser mayor que el acumulado en el momento  $t$ ,  $W_t$ , de manera que,  $W_{t+1} = W_t + \Delta_t \cdot \Delta_t$  representa la cantidad de conocimiento aprendida por la red, o modificación de los pesos, entre los momentos  $t$  y  $t+1$ . Esta variación de los pesos depende, lógicamente, de los pares entrada - salida deseada que se proporcionan a la red y de los pesos de la misma, es decir,  $\Delta_t = \psi_t(Z_t, W_t)$ . Pues bien, un algoritmo de aprendizaje, como es lógico, debe proporcionar la forma de ir cambiando los pesos al objeto de optimizar el comportamiento de la red. El algoritmo de retropropagación de errores propone la siguiente regla, basada en el descenso del gradiente, para modificar adecuadamente los pesos:

$$W_{t+1} = W_t + a \nabla \hat{f}(x_t, W_t) (Y_t - \hat{f}(x_t, W_t)) \quad [4]$$

siendo  $a$  una constante denominada tasa de aprendizaje y  $\nabla \hat{f}(x_t, W_t)$  el gradiente de la salida que genera la red en el momento  $t$ ,  $\hat{f}(x_t, W_t)$ , en relación con los pesos en dicho instante,  $W_t$ . Esta expresión muestra que el valor de un peso en un determinado momento depende del valor adoptado en el instante anterior, de la tasa de aprendizaje, del error cometido por la red en el momento anterior y de la sensibilidad del output del modelo ante las variaciones de dicho peso.

Antes de iniciarse el proceso de aprendizaje, los pesos toman valores aleatorios relativamente pequeños, para que la red no se sature con grandes pesos, y de este modo comenzar en un punto aleatorio de la superficie de error.

Una vez finalizado el aprendizaje se debe proceder a testear la red. La fase de test consiste en introducir nuevos patrones de entrada y comprobar la eficacia del sistema generado. Si no resulta aceptable se repite la fase de entrenamiento utilizando nuevos patrones-ejemplo, e incluso puede ser necesario modificar la estructura de la red.

Para finalizar, y analizada la estructura y funcionamiento de las redes neuronales, debemos puntualizar un problema que las mismas pueden presentar y que es común a todos los métodos no paramétricos, el del sobreaprendizaje o sobreajuste, que está relacionado con el tamaño o número de parámetros de la red. No existe una regla específica para determinar el tamaño que debe tener una red de neuronas artificiales y, el encontrar la estructura óptima consiste en un intercambio sesgo-varianza [Geman et al., 1992], lo cuál entraña un alto grado de dificultad. Como señala Olmeda (1993), se trata de una cuestión de vital importancia, pues sólo cuando existe equilibrio entre la estructura del modelo y el problema a resolver se logra una generalización adecuada.

Una red infradimensionada, es decir, cuyo número de parámetros es menor al que requiere el problema, nunca puede llegar a aprender la relación entre las variables entrada-salida deseada del conjunto de entrenamiento. En consecuencia, es incapaz de generalizar correctamente. Por el contrario, una red sobredimensionada memoriza el conjunto de aprendizaje y lo reproduce. Sin embargo, al no aprender la relación entre las variables de interés, es incapaz de generalizar.

En general, análogamente al conocido dicho “lo bueno, si breve, dos veces bueno”, la mejor red, con mayor poder de generalización, es la más sencilla, siempre que su comportamiento a lo largo del conjunto de aprendizaje sea eficiente.

El número de unidades que deben utilizarse en las capas de entrada y de salida suele ser bastante evidente; tantas neuronas de entrada como variables explicativas y de salida como variables a predecir. El problema reside en determinar el número de nodos de las capas ocultas. Una de las técnicas utilizadas para determinar el tamaño óptimo de la red consiste en comenzar con una red de un nodo y añadir otros cuando es necesario. La técnica opuesta empieza con una red muy grande, cuyos nodos innecesarios se eliminan paulatinamente.

De entre los métodos más conocidos para evitar el sobreaprendizaje podemos destacar: la detención prematura del aprendizaje, el establecimiento de un control estadístico del tamaño de la red, la reducción el tamaño de la red por cualquiera de los métodos posibles y el empleo de arquitecturas autogenerativas [Olmeda, 1993].

### **3. MODELOS DE VOLATILIDAD**

La variedad de modelos que permiten predecir y modelizar la volatilidad es muy amplia, por ello, ante la dificultad de abarcar el gran abanico de técnicas existente, en la presente sección nos centramos en cinco de las alternativas paramétricas más populares para la modelización de la volatilidad (modelos ARCH, GARCH, EGARCH, GJR y homocedástico) y en dos potentes herramientas no paramétricas, Redes Neuronales Artificiales y modelos de agregación mediante RNA. Además, nuestro objetivo no es proporcionar una comparación completa de los modelos existentes, sino ofrecer algunos resultados de aquéllos que se emplean con mayor frecuencia.

En el presente trabajo ofrecemos dos aplicaciones de las RNA a la predicción de volatilidades. En nuestra primera aportación empleamos las RNA como una herramienta no paramétrica para encontrar la forma funcional de la volatilidad. En segundo lugar proponemos un modelo “híbrido” de volatilidad consistente en la integración, mediante una RNA, de modelos paramétricos de la clase ARCH. Como veremos, para ambos enfoques, es posible obtener, en algunos casos, modelos predictivos más exactos que los paramétricos anteriormente mencionados.

En lo que sigue, asumimos que los rendimientos logarítmicos pueden representarse como:

$$R_t = \mu + \lambda R_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2) \quad [5]$$

de manera que la media condicional de los rendimientos sigue un proceso AR(1).

Un análisis preliminar de las series con las que trabajamos, efectuado en [Marco, 1998] muestra que el primer coeficiente de autocorrelación es significativo, sin embargo, los de orden superior no lo son. Por esta razón, como mencionamos con anterioridad, modelizamos la media condicional como un proceso autorregresivo de orden 1. Obviamente, aunque es muy comúnmente utilizado en la literatura financiera, este modelo puede resultar inapropiado para algunas de las series, por lo que sería necesario incluir un mayor número de retardos, o bien, especificaciones alternativas mediante otros modelos. Sin embargo, modelos lineales más complicados no cubren todas las posibilidades plausibles de modelización capaces de generar dependencia débil en la media condicional, pero claramente dependiente en la varianza condicional (procesos estocásticos no lineales en media o memoria a largo plazo cíclica, por citar dos alternativas propuestas recientemente), característicos de las series temporales financieras y, en particular, series temporales de tipos de cambio, por tanto, una identificación más intensiva del modelo lineal para las series particulares, también obviaría alternativas plausibles.

Además, una identificación exhaustiva de la media condicional haría computacionalmente inviable nuestro problema. Por estas dos razones, nuestra elección representa un compromiso entre una adecuada modelización de la media condicional, como un paso previo a la modelización de la varianza condicional, y la obtención de resultados significativos empleando un número de predicciones considerable.

### **3.1. Modelos paramétricos de predicción**

Como los modelos que utilizamos son bien conocidos en la literatura financiera<sup>4</sup>, en las líneas siguientes sólo realizamos una breve descripción de la parametrización específica de los mismos.

El modelo *homocedástico* es el más sencillo de los propuestos en la literatura financiera. Según dicho modelo, la varianza de un activo permanece constante. Si llamamos  $\sigma_t^2$  a la varianza condicional del activo, entonces el modelo homocedástico supone que:

---

<sup>4</sup> Para profundizar sobre los modelos empleados pueden referirse a Bollerslev et al. (1992).

$$\sigma_t^2 = \omega \quad [6]$$

Los modelos *ARCH* [Engle, 1982] sugieren una forma de heterocedasticidad, es decir, no estacionariedad en varianzas, en la cual la varianza condicional del error de predicción depende del tamaño de las perturbaciones precedentes. Un proceso *ARCH(p)*, es decir, de Heterocedasticidad Condicional Autorregresiva de orden  $p$ , está caracterizado por la siguiente expresión:

$$x_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \sum_{j=1}^p \alpha_j \varepsilon_{t-j}^2 \quad [7]$$

donde los parámetros  $\omega$ ,  $\alpha_j$  y  $p$  (orden del proceso *ARCH* que indica el número de retardos) se estiman mediante procedimientos de máxima verosimilitud.

Como podemos observar, la varianza condicional es un proceso autorregresivo en los residuos al cuadrado, que además es heterocedástica y, por ello, recibe la denominación de proceso de Heterocedasticidad Condicional Autorregresiva.

Para construir la ecuación de la varianza condicional, como es lógico, necesitamos disponer de una serie de residuos ( $\varepsilon_t$ ) o término perturbación, los cuales están definidos por una ecuación de la media condicional como [5].

El error de predicción o perturbación aleatoria representa la nueva información disponible para los agentes.

Un problema que plantea la formulación original del modelo *ARCH* es la restricción de no negatividad a que están sujetos los parámetros  $\omega$  y  $\alpha_j$  para asegurar que la varianza condicional sea positiva. Sin embargo, las restricciones de no negatividad pueden ser violadas en los casos en que es necesario incluir numerosos retardos para modelizar correctamente el proceso. A pesar de ello, en muchas de las aplicaciones con el modelo lineal *ARCH(p)* es necesario introducir un elevado número de retardos ( $p$ ), para que la modelización resulte adecuada, lo cual puede plantear problemas, debido a la posible violación de la mencionada restricción de no negatividad.



Una estructura de retardos alternativa y más flexible, que exige un número de variables retardadas mucho menor, es a menudo proporcionada por el modelo *ARCH generalizado* o *GARCH(p,q)*, sugerido en Bollerslev (1986). Los modelos GARCH generalizan la dependencia de la varianza condicional de los residuos incluyendo los propios valores pasados de la varianza condicional. Así, ésta depende sólo de los valores pasados de la variable dependiente. Un proceso *GARCH(p,q)* tiene por expresión:

$$\sigma_t^2 = \omega + \sum_{j=1}^p \alpha_j \varepsilon_{t-j}^2 + \sum_{k=1}^q \beta_k \sigma_{t-k}^2 ; \quad \varepsilon_t \sim N(0, \sigma^2) \quad [8]$$

donde se exige la condición de que  $\omega, \alpha_j$  y  $\beta_k > 0$ , con el fin de eliminar la posibilidad de que la varianza condicional sea negativa.

El modelo GARCH representa que la varianza condicional en el momento  $t$  depende de tres factores:

- una constante,  $\omega$ ,
- la información disponible en relación con la volatilidad pasada, representada por el cuadrado de los residuos de períodos anteriores, conocido como término ARCH,  $\varepsilon_{t-j}^2$ , y
- las predicciones de la varianza para los períodos anteriores a  $t$ , conocidas como término GARCH,  $\sigma_{t-k}^2$ .

Esta expresión de la varianza incorpora, al igual que los modelos ARCH, la tendencia a la agrupación de la volatilidad, que es tan evidente en las series de rendimientos financieros.

Siguiendo a Mandelbrot (1963) cambios grandes de precios en mercados financieros suelen ir seguidos de cambios grandes de cualquier signo y lo mismo ocurre con cambios pequeños; sin embargo, el signo de estos cambios es impredecible<sup>5</sup>. Las variables financieras, en particular, a menudo exhiben períodos de tranquilidad seguidos por períodos de alta volatilidad, es decir, la volatilidad es variable en el tiempo.

---

<sup>5</sup> Ver Granger y Ding (1994).

Un fenómeno a menudo observado en los mercados financieros es que los rendimientos de los activos están correlacionados negativamente con las variaciones en la volatilidad de los mismos<sup>6</sup>. Así, la respuesta de la volatilidad es diferente ante malas y buenas noticias, es decir, ésta tiende a aumentar cuando los rendimientos caen por debajo de lo esperado y a disminuir si los rendimientos son mayores de lo esperado.

No obstante, la respuesta asimétrica de la volatilidad, observada en las series de rendimientos financieros, muestra la existencia de un efecto apalancamiento que ocasiona variaciones diferentes en el riesgo previsto dependiendo del signo de la innovación. Esto sugiere que para las aplicaciones en la valoración de activos financieros sería preferible un modelo en el que  $\sigma_t^2$  responda asimétricamente ante residuos positivos y negativos. Sin embargo, tanto en los procesos ARCH como en los GARCH la varianza condicional depende sólo del tamaño de las  $\varepsilon_t$ , pero no de su signo. Además, estos modelos imponen restricciones a los parámetros para asegurar la positividad de  $\sigma_t^2$ . El modelo *EGARCH* o *GARCH exponencial*, propuesto en Nelson (1991) trata de superar estas limitaciones.

Si llamamos  $\sigma_t^2$  a la varianza de  $\varepsilon_t$ , condicionada al conjunto de información disponible en el momento  $t$ , lógicamente tiene que ser no negativa con probabilidad igual a uno. Los modelos GARCH aseguran esta condición al determinar  $\sigma_t^2$  como una combinación lineal de variables aleatorias positivas, con ponderaciones también positivas. Sin embargo, Nelson (1991) asegura la no negatividad de  $\sigma_t^2$  de un modo diferente. Trabaja con el logaritmo de  $\sigma_t^2$  haciendo que éste sea una función lineal de las innovaciones.

La especificación para la varianza del EGARCH(1,1) es:

$$\log(\sigma_t^2) = \omega + \beta \log(\sigma_{t-1}^2) + \alpha \left( \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right) + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \quad [9]$$

la cual es una función asimétrica de las innovaciones si  $\gamma \neq 0$ , lo que permite que el impacto sobre la volatilidad sea diferente según los valores retardados sean positivos o

---

<sup>6</sup> Multitud de investigadores, comenzando con Black (1976), han encontrado evidencia empírica de este hecho.

negativos. El término  $\sqrt{\frac{2}{\pi}}$  se introduce para centrar la media de  $\left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right|$  con respecto de cero.

Esta formulación presenta las siguientes ventajas:

- i) debido a la transformación logarítmica, garantiza que la varianza sea siempre positiva,
- ii) como consecuencia de lo anterior, no impone restricciones a los parámetros  $\alpha$  y  $\beta$  para asegurar la no negatividad de la varianza condicional,
- iii) introduce efectos asimétricos y no lineales de las innovaciones sobre la varianza de la serie (lineales sobre el logaritmo). El proceso EGARCH, como  $\gamma$  puede adoptar valores distintos de cero, permite que la respuesta de la varianza condicional sea diferente ante variaciones de aumento o disminución de los rendimientos superiores a lo esperado, mientras que, los modelos ARCH y GARCH imponen una respuesta simétrica.

El modelo EGARCH recoge el impacto de los residuos más recientes de manera exponencial en lugar de cuadrática y, aunque la función de verosimilitud<sup>7</sup> resulta mucho más difícil de maximizar, se considera que es un modelo muy ventajoso.

Otro modelo que describe el impacto asimétrico de las innovaciones sobre la varianza es el *GJR*, también conocido como *TARCH* o proceso *GARCH asimétrico de umbral*, introducido por Glosten et al. (1993) y Zakoïan (1994). Este modelo permite diferentes reacciones de la volatilidad según cual sea el signo y la magnitud de los shocks.

El modelo GJR(1,1) representa la varianza de la siguiente forma :

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \gamma\varepsilon_{t-1}^2 d_{t-1} + \beta\sigma_{t-1}^2 \quad [10]$$

donde  $d_t = 1$ , si  $\varepsilon_t < 0$  y  $d_t = 0$ , en cualquier otro caso.

---

<sup>7</sup> Nelson (1991), pág. 355.

Así, cuando llegan buenas noticias, el signo de la innovación es positivo, es decir,  $\varepsilon_t > 0$  y, en consecuencia,  $d_t = 0$ , por lo que el impacto de dicha innovación es de  $\alpha$ . Por el contrario, ante la incorporación de innovaciones negativas,  $d_t = 1$  y su efecto sobre la varianza es de  $\alpha + \gamma$ .

Al realizar estimaciones en base a este modelo, cuando el valor del parámetro  $\gamma$  es significativamente distinto de cero, quiere decir que el efecto apalancamiento o asimetría está presente. Por otra parte, al predecir, el modelo GJR supone que la distribución de los residuos es simétrica, lo que implica que  $d_t = 1$  durante la mitad del horizonte temporal de la predicción, sin embargo, es imposible prever cuándo lo es.

### 3.2. *RNA como modelos no paramétricos de predicción*

Puesto que, las herramientas tradicionales, como la función de autocorrelación, no permiten esclarecer ciertos tipos de comportamientos no lineales, en un trabajo previo al que ahora presentamos aplicamos a las series de volatilidades del tipo de cambio que utilizamos una nueva técnica capaz de determinar en mayor medida la existencia de no linealidad, el *estadístico BDS* [Brock et al., 1987], concluyendo que existen comportamientos no lineales en las series de volatilidades del tipo de cambio de la peseta con que trabajamos. Es por ello por lo que utilizamos RNA, que constituyen una potente técnica de modelización no paramétrica, capaz de captar dinámicas como la no linealidad.

En nuestro estudio emplearemos los rendimientos al cuadrado de las series de tipos de cambio como una aproximación a la volatilidad. En nuestro primer análisis, el problema consiste en encontrar una función  $f$  tal que

$$\sigma_t^2 = f(\sigma_{t-1}^2, \sigma_{t-2}^2, \dots, \sigma_{t-5}^2) \quad [11]$$

donde  $\sigma_{t-p}^2$  son los rendimientos al cuadrado para el instante  $t-p$ . Es decir, en nuestra formulación particular intentamos, mediante los modelos de RNA, encontrar una función que permita predecir la volatilidad a partir de las cinco observaciones inmediatamente anteriores de la misma.

### 3.3. RNA como modelos de agregación

En nuestro segundo estudio proponemos el empleo de las RNA como modelo de agregación de otros modelos individuales de volatilidad. La agregación de modelos de predicción fue propuesta originalmente por Bates y Granger (1969) con el fin de explotar las posibles ventajas comparativas de cada uno de los modelos. Granger y Ramanathan (1984) extienden el enfoque anterior proponiendo la construcción de modelos combinados mediante una regresión lineal de la variable a predecir sobre las predicciones de los diferentes modelos. Sin embargo, el enfoque lineal no es necesariamente óptimo y es posible que otras agregaciones no lineales permitan combinar los diferentes modelos más eficientemente.

El problema de la agregación no lineal de modelos de predicción ha sido abordado únicamente en fechas muy recientes: por ejemplo, Kamstra y Kennedy (1998) proponen un modelo de agregación paramétrico (basado en el logit), mientras que Olmeda y Fernández (1997) y Kumar y Olmeda (1997) proponen la agregación con métodos no paramétricos. En nuestro trabajo proponemos el empleo de RNA como modelo de agregación. A conocimiento nuestro, no existe hasta la fecha una propuesta de combinar las predicciones de los modelos de volatilidad a través de RNA.

Hemos de hacer notar que nuestra propuesta incluye otras formulaciones (por ejemplo la agregación mediante logit [Kamstra y Kennedy, 1998] y lineal [Granger y Ramanathan, 1984] como casos particulares). Como resulta obvio de nuestra exposición anterior sobre las RNA, los modelos lineales y logit pueden ser vistos como modelos anidados de una RNA; en el primer caso no se trata más que de una RNA sin unidades intermedias y con función de activación lineal, mientras que en el segundo caso es posible entender el modelo logit como una RNA con una sola unidad intermedia con función de activación logística y función de salida lineal. Pensamos, por tanto, que nuestra propuesta puede resultar atractiva, ampliando, aún más, las herramientas disponibles de predicción.

Formalmente, el problema consiste en encontrar  $f$  tal que:

$$\sigma_t^2 = f(\sigma_{(ARCH)_t}^2, \sigma_{(GARCH)_t}^2, \sigma_{(EGARCH)_t}^2, \sigma_{(GJR)_t}^2, \sigma_{(HOM)_t}^2) \quad (12)$$

mediante el empleo de una RNA capaz de aproximar  $f$ , donde  $\sigma_{(EGARCH)_t}^2$  es, por ejemplo, la predicción del modelo EGARCH para la observación  $t$ . Obsérvese que, en este caso, el procedimiento de construcción del modelo es notablemente más complejo,

puesto que cada una de las estimaciones de los modelos exige, a su vez, un conjunto muestral sobre el que obtener estimaciones y predicciones de cada uno de los modelos que integran el modelo híbrido propuesto.

#### **4. BASE DE DATOS, METODOLOGÍA Y RESULTADOS**

En el trabajo que presentamos, y dividimos en dos partes, estudiamos la modelización y predicción de volatilidades. Proponemos analizar en este aspecto la capacidad de las RNA como modelo no paramétrico (1er. estudio) y de las RNA como modelo de agregación (2º estudio). En un primer análisis, comparamos la capacidad predictiva de las RNA con la de otros modelos: ARCH, GARCH, EGARCH y GJR, además de comparar la eficiencia de estos modelos entre sí y en relación con el homocedástico. En una segunda aplicación, proponemos un modelo híbrido de RNA, que basado en una sencilla combinación de diferentes modelos, permite construir modelizaciones más complejas que aprovechan las ventajas de cada uno de los modelos individuales. Concretamente, para cada moneda, construimos dos modelos híbridos: una RNA híbrida y un simple modelo lineal, que combinan las predicciones de los modelos homocedástico, ARCH, GARCH, EGARCH y GJR.

##### **4.1. Base de datos**

La base de datos de que disponemos, y que utilizamos en trabajos anteriores, se compone de 8.427 observaciones, desde el 1-1-1965 hasta el 21-4-1997, que se corresponde con el tipo de cambio medio diario, al cierre de la jornada de contratación, de la peseta frente al marco alemán (DMK), franco francés (FFR), lira italiana (LIT), libra esterlina (STG), dólar estadounidense (USD) y yen japonés (YEN). Sin embargo, debido a que su enorme amplitud hubiese hecho casi imposible la obtención de predicciones mediante el procedimiento dinámico que empleamos, para determinar el período predictivo se partió de la observación correspondiente al 1-9-1995 (observación 8000 de nuestra base de datos).

Los inputs o variables explicativas que emplean los modelos que aplicamos son los siguientes. En el caso de los procesos de la clase ARCH, los residuos obtenidos a partir de la ecuación de la media condicional de los rendimientos logarítmicos (expresión 5). Las RNA utilizan, en el 1er. estudio, las volatilidades históricas observadas, calculadas a partir de la serie de los cuadrados de los rendimientos

logarítmicos del tipo de cambio; en el 2º análisis, las entradas de la red son las predicciones obtenidas mediante los 5 modelos paramétricos comentados.

## **4.2. Metodología**

En nuestro primer estudio, RNA como modelos no paramétricos de predicción, los modelos son estimados de manera dinámica de la siguiente forma: para cada una de las monedas e instantes de tiempo consideramos las últimas 100 observaciones pasadas y estimamos una RNA que empleamos para predecir la observación siguiente. Acto seguido desplazamos nuestra ventana observacional un día (eliminando la observación más antigua e incluyendo otra nueva disponible) y repetimos el procedimiento. De esta manera, cada una de las predicciones se basa en un modelo independiente estimado sobre las 100 observaciones anteriores y el modelo va incorporando la información que aparece día a día.

Tras experimentar exhaustivamente con diferentes parámetros (número de neuronas de la capa oculta, funciones de activación, tasa de aprendizaje, etc.) determinamos que una RNA con una única capa oculta con 2 neuronas, una tasa de aprendizaje igual a  $10^{-3}$  y funciones de activación sigmoidales para la capa oculta y lineales para la de salida, ofrecía los mejores resultados (en particular, también se experimentó con 3, 4 y 5 neuronas en la capa oculta). Dicha especificación se mantuvo constante para todas las monedas, con el fin de evitar problemas asociados al “data snooping” (el empleo particularizado de especificaciones de modelos con el fin de obtener mejoras predictivas).

Como mencionamos con anterioridad, nuestro período predictivo comienza en la observación 8000 de nuestra base de datos original (la correspondiente al 1-9-1995). Por tanto, para cada una de las monedas, la primera red emplea las observaciones 8000 a 8.099 para estimar el modelo y predice la observación 8.100. Después empleamos las observaciones 8.001 a 8.100 para estimar y predecimos la observación 8.101, y así sucesivamente. Cuando hemos efectuado 100 predicciones (de la 8.100 a la 8.199) computamos el error absoluto medio y lo etiquetamos con EAMRNA(1). El proceso descrito anteriormente continúa, desplazando una unidad las sucesivas ventanas involucradas. Así, el EAMRNA (2) corresponde al error absoluto medio cometido por 100 RNA para el horizonte diario de predicción 8.101-8.200.

Para comparar la capacidad predictiva del modelo de RNA estimamos, de manera similar a las redes, cada uno de los siguientes modelos ARCH(1), GARCH(1,1),

EGARCH(1,1), GJR(1,1) y homocedástico y procedemos del mismo modo. De esta forma, EAMARCH(1) corresponde al error absoluto medio cometido por 100 modelos ARCH(1), etc. Obsérvese que, por ejemplo, el ratio EAMRNA(t)/ EAMARCH(t) proporciona una medida de la ventaja comparativa del empleo de una RNA frente a un modelo ARCH para las 100 sesiones anteriores al instante  $t$ . Si a lo largo del horizonte predictivo dicho ratio es consistentemente menor que uno, existe evidencia de una superior capacidad predictiva de las RNA frente al modelo ARCH. Si por el contrario este ratio es muy variante, existe evidencia de que la optimalidad de uno u otro modelo es inestable. Finalmente, un ratio superior a uno indica peor comportamiento predictivo de las RNA.

Para la construcción de las predicciones en nuestra segunda aplicación, RNA como modelos de agregación, actuamos del siguiente modo; en cada instante de tiempo empleamos las 250 observaciones anteriores para estimar cada uno de los modelos paramétricos. Dichos modelos son entonces utilizados para predecir la siguiente observación. Finalmente construimos un 5-vector compuesto por las predicciones de los diferentes modelos. Seguidamente, empleamos 120 5-vectores de predicciones para estimar una RNA con cuatro unidades ocultas (en una única capa oculta) e intentamos predecir la volatilidad contemporánea de las predicciones. Obsérvese que en el caso anterior nos enfrentamos a un problema de regresión dinámica (en concreto, a un problema de estimar un modelo autorregresivo no lineal de orden 5) mientras que en el presente contexto el problema se plantea como uno de regresión no lineal estática. El resto del procedimiento es análogo al caso anterior, esto es, EAM(t) tiene la misma interpretación que antes: se trata del error absoluto medio de las 100 predicciones anteriores a  $t$ , por lo que las figuras tienen la misma significación.

Al igual que en el caso anterior, se probaron diferentes parámetros para las redes, así como distintos conjuntos de información; en concreto, se estimaron los modelos de RNA sobre 60 5-predicciones en vez de sobre 120 5-predicciones, en todos los casos los resultados resultaban inferiores a los aquí presentados.

Por último, con el fin de comparar las posibles ventajas de agregar linealmente las predicciones o, por el contrario, emplear una RNA para conseguir una agregación no lineal, también aplicamos, en cada uno de los estudios, la metodología descrita, empleando funciones lineales.



### 4.3. *Resultados*

En el análisis efectuado predecimos la volatilidad del tipo de cambio de la peseta frente a las seis monedas contempladas, utilizando, para ello, como comentamos anteriormente, un método recursivo, mediante los siguientes modelos: homocedástico, ARCH, GARCH, EGARCH, GJR, RNA y regresión lineal, empleando como conjuntos de información las volatilidades pasadas, en el primer estudio, y las predicciones de los modelos paramétricos, en el segundo.

La presentación de los resultados obtenidos la efectuamos separadamente para cada uno de los dos estudios efectuados<sup>8</sup>, cuya metodología acabamos de exponer. En el primero de ellos (RNA como modelos no paramétricos de predicción) utilizamos, únicamente, observaciones históricas de las series con que trabajamos, a partir de las cuáles predecimos la volatilidad del tipo de cambio de la peseta frente a cada una de las seis monedas. Los modelos que empleamos son el homocedástico, ARCH, GARCH, EGARCH y GJR. Obtenidas las predicciones que genera cada uno de los anteriores modelos, calculamos el cociente entre el error correspondiente a cada modelo y el error del modelo homocedástico, al que denominamos ratio (por ejemplo, RATARCH representa dicho ratio para el modelo ARCH). En consecuencia, un ratio menor que uno significa que el modelo en cuestión predice mejor que el homocedástico. Los resultados de este análisis, que nos permiten comparar la capacidad predictiva de cada uno de los modelos paramétricos de la clase ARCH empleados en relación con el homocedástico y, también, entre sí, los ilustramos en las figuras A.1 a A.6.

Para la volatilidad del tipo de cambio peseta/marco alemán (figura A.1), durante los sesenta días que comprende nuestro período de evaluación de predicciones, todos los modelos superan al homocedástico en cuanto a capacidad predictiva, a excepción del EGARCH hasta el tercer día en que evaluamos nuestras predicciones. Debemos destacar que la mejora es significativa, entre el 5 y el 15%, aproximadamente, dependiendo de los modelos. El modelo GARCH es el que mejor se comporta hasta el día 32 y, a partir de este momento, el modelo GJR supera a los demás. De entre los modelos tipo ARCH, el EGARCH y el ARCH son los que presentan menor calidad predictiva.

---

<sup>8</sup> Los resultados los presentamos en las figuras que aparecen en el Anexo del trabajo.

La figura A.2 muestra los resultados para el caso del franco francés. Como podemos observar, en todo el horizonte de predicción, las cuatro modelizaciones ARCH superan al modelo homocedástico en términos predictivos. Generalmente, las mejoras son significativas. El modelo que peor se comporta es el ARCH, no obstante, llega a conseguir mejoras de alrededor del 6%. El modelo que predice con mayor eficiencia es el GARCH hasta el día 48, en que es relevado por el GJR. Estos modelos llegan a superar al homocedástico, en términos aproximados, en un 13%.

La figura A.3 muestra que, para la volatilidad del tipo de cambio peseta/lira italiana, el comportamiento del modelo homocedástico es peor que el de los demás modelos analizados, a lo largo de todo el período considerado. A excepción del modelo ARCH, todos consiguen mejoras significativas, que para el caso del proceso GARCH llegan a superar el 30%. Los modelos con mayor capacidad predictiva son el GARCH y, tras cuarenta días de evaluación, el GARCH y el GJR, cuyo comportamiento es idéntico.

Los resultados son completamente distintos en el caso de la libra esterlina (figura A.4). Por lo general todos los modelos se muestran incapaces de superar al homocedástico. Los modelos que mejor se comportan son el GARCH, hasta el vigésimo quinto día de evaluación de las predicciones y el EGARCH desde entonces hasta el día cuarenta y siete, en que se sigue mostrando como el mejor, junto con la modelización GJR. El empeoramiento, en ningún caso supera el 3%.

Para la volatilidad del tipo de cambio de la peseta frente al dólar estadounidense (figura A.5), el único modelo que durante los sesenta días no supera al homocedástico es el EGARCH. Los procesos ARCH, GARCH y GJR, normalmente se comportan mejor, siendo los dos últimos, en algunos momentos, significativamente mejores. Los modelos más eficientes son, durante los primeros 21 días, el GARCH junto con el GJR, mostrándose, a partir de este momento, superior el GJR.

La figura A.6 muestra que, en relación con la moneda japonesa, todos los modelos mejoran al homocedástico, salvo el EGARCH en ciertos tramos. Por lo general, el modelo que mejor se comporta es el GARCH, seguido muy de cerca del GJR. Estos modelos llegan a superar casi en un 10% al homocedástico. Todo lo que acabamos de exponer permite concluir que ningún modelo se muestra absolutamente superior para todas las monedas, sin embargo, todos ellos parecen tener capacidad predictiva al superar, en la mayor parte de las ocasiones, al homocedástico.

A continuación, siguiendo con nuestro primer análisis, construimos, para cada una de las monedas, una red neuronal artificial. Recordemos que esta red es empleada para predecir, también de manera dinámica, la volatilidad del tipo de cambio de nuestra moneda, a partir de las observaciones históricas. Calculamos además la media de los errores de los cuatro modelos tipo ARCH considerados. Al objeto de comparar el modelo de red neuronal con la media de los ARCH, determinamos el ratio (RATMED) entre el error de la red y el error medio de los modelos paramétricos, de modo que si éste es menor que uno, significa que, en media, las redes neuronales artificiales poseen mayor capacidad predictiva que los modelos paramétricos de heterocedasticidad condicional que hemos analizado. En las figuras A.7 a A.12 mostramos el ratio entre el EAM de una RNA y el EAM medio de los modelos paramétricos (excluyendo al modelo homocedástico), estas figuras proporcionan una indicación de la mejora o empeoramiento del empleo de una red frente a los modelos paramétricos en su conjunto.

Como podemos observar, la red supera clara y significativamente a los modelos paramétricos considerados en los casos del marco, lira y libra. Las mejoras llegan a situarse en el 10%, 13% y 4%, respectivamente, para las tres monedas mencionadas. En cuanto a la predicción de la volatilidad del tipo de cambio peseta/franco francés, la red neuronal artificial, en media, mejora a los modelos paramétricos sólo hasta el décimo día de evaluación de las predicciones. Desde ese momento, la media de los modelos de la clase ARCH supera a la red, aunque las mejoras, en ningún momento, alcanzan el 3%. Podemos observar, también, que en los casos del dólar estadounidense y yen japonés los modelos paramétricos, en media, presentan una mayor eficiencia predictiva que la RNA que construimos. Para la volatilidad del tipo de cambio peseta/dólar, la mejora sobre la red se sitúa alrededor del 6%, sin embargo, en el caso del yen, la cuantía de esta mejora es, por lo general, muy reducida.

En el segundo estudio (RNA como modelos de agregación) construimos, para cada moneda, dos modelos: una red neuronal artificial y un simple modelo lineal. Recordemos que, estos modelos se construyen ahora combinando las predicciones de los cinco modelos paramétricos anteriores, es decir, el homocedástico, ARCH, GARCH, EGARCH y GJR. Esto significa, que dichas predicciones son las variables explicativas (entradas de las RNA) de ambos modelos. Las redes construidas, según el procedimiento expuesto, constituyen una forma de modelización híbrida que proponemos en el presente trabajo, al objeto de mejorar la calidad de las predicciones que generan los demás modelos.

En este segundo análisis, en primer lugar, comparamos la capacidad predictiva de las RNA y del modelo lineal (agregados) con la de la media de los modelos paramétricos tipo ARCH, siguiendo un procedimiento idéntico al de la parte anterior de este estudio. Las figuras A.13 a A.18 ilustran los resultados de los ratios de errores obtenidos, a los que denominamos RATMEDRNA y RATMEDLIN, para la RNA y modelo lineal, respectivamente. Podemos observar que, durante todo el horizonte temporal considerado, la capacidad predictiva del modelo híbrido de red neuronal que hemos diseñado es claramente superior a la de la media de modelos paramétricos para todas las monedas, a excepción de la libra esterlina y el dólar estadounidense. En los casos del marco alemán y la lira italiana, las mejoras son espectaculares, pues llegan a situarse, respectivamente, en el 32% y 18%. Para la moneda estadounidense, la red neuronal domina claramente a los modelos paramétricos, en media, hasta el punto 55, a partir del cuál los empeora ligeramente. En el caso de la volatilidad del tipo de cambio peseta/libra esterlina, la red que construimos se muestra claramente inferior (hasta un 16%) a la media de los modelos paramétricos.

El modelo lineal se muestra claramente superior a la media de los ARCH para el marco, lira y yen, durante todo el horizonte de predicción. En los casos del dólar y franco, transcurridos respectivamente 40 y 48 días de evaluación de predicciones, el modelo lineal aparece como inferior, no siendo dicha inferioridad significativa. Para la volatilidad del tipo de cambio peseta/libra esterlina, sin embargo, los modelos paramétricos, en media, superan significativamente al modelo lineal.

Estas figuras también nos permiten comparar la capacidad predictiva de los modelos lineal y de RNA que hemos diseñado. Así, podemos observar que las redes neuronales, por lo general, superan claramente al modelo lineal para el tipo de cambio de la peseta frente a la lira italiana y al dólar. En el caso del dólar, esta superioridad es, en algunos momentos, altamente significativa. Para las monedas francesa y japonesa no podemos afirmar la superioridad clara de alguno de los modelos. Finalmente, el modelo de red queda superado por el lineal, generalmente, en los casos del marco y la libra. La mejora del modelo lineal es muy reducida para la moneda alemana, sin embargo, es bastante elevada para la moneda británica.

En segundo lugar, comparamos las redes neuronales artificiales y los modelos lineales (híbridos) con el mejor de los modelos paramétricos (obtenido en el primer estudio). Debemos destacar que este es el juicio más duro que se puede hacer a las redes neuronales, pues seleccionamos como patrón comparativo el modelo que, a posteriori, ha resultado ser el más eficiente de entre los paramétricos. Para cada uno de los modelos

(red, lineal y mejor), calculamos el cociente (ratio) entre el error correspondiente a cada modelo y el error del modelo homocedástico, a los que llamamos RATRNA, RATLIN y MEJOR, respectivamente. En consecuencia, un ratio menor que uno significa que el modelo en cuestión predice mejor que el homocedástico. Los resultados están reflejados en las figuras A.19 a A.24. Debemos considerar que como las comparaciones de las redes neuronales frente al modelo lineal y modelos de la clase ARCH versus homocedástico las efectuamos con anterioridad, a continuación nos centramos en comparar la eficiencia predictiva de las redes y el modelo lineal, frente al homocedástico y mejor modelo de entre los paramétricos.

Como ilustran las figuras anteriores, para todas las monedas, a excepción del dólar estadounidense y libra esterlina, tanto las redes neuronales, como el modelo lineal y el mejor de entre los paramétricos, muestran una eficiencia predictiva, por lo general, significativamente mayor que la del modelo homocedástico. En algunos casos la mejora se cifra en el 40%. En el caso del dólar el modelo homocedástico domina a lo largo de un tramo muy reducido, desde el día 37 al modelo lineal, desde el 51 a la red neuronal y desde el 54 al mejor paramétrico. Para el tipo de cambio peseta/libra esterlina, el modelo homocedástico supera claramente a todos los demás.

En relación con el análisis comparativo efectuado entre las redes neuronales, la combinación lineal y el mejor de los modelos paramétricos podemos observar que, para el tipo de cambio peseta/marco los modelos con mayor capacidad predictiva son el lineal y la RNA. La red neuronal domina entre los días 37 y 50, sin embargo, durante el resto del período la combinación lineal se muestra superior. Las mejoras de estos modelos frente al mejor de los paramétricos son significativas, pues llegan a situarse en el nivel del 30%, aproximadamente. En el caso del franco francés, el modelo lineal se comporta mejor que los demás hasta el punto 37 y, a partir de dicho momento, domina la red neuronal. En los últimos diez días del período la mejora de la red es significativa, llegando a alcanzar, aproximadamente, el 6%. Para la moneda italiana, por lo general, el mejor modelo paramétrico se muestra superior a los demás, mientras que, el modelo lineal es el que peor se comporta. Las mejoras, sin embargo, no son muy significativas. Generalmente, el mejor modelo paramétrico supera a la RNA en un 2%, aproximadamente, y la mejora de la red frente a la combinación lineal es algo menor. En el caso de la libra esterlina el mejor modelo de entre los paramétricos domina clara y significativamente (mejoras de hasta el 20%) durante todo el período de predicción. En este caso, el modelo que peor se comporta es el de red neuronal. Para el tipo de cambio peseta/dólar estadounidense, entre los días cuatro y cuarenta y ocho del período que

consideramos, la RNA supera a los demás modelos, ofreciendo mejoras de hasta el 6%. Durante el resto del período el modelo dominante es el mejor paramétrico. Para finalizar, en el caso del yen, la red neuronal supera al resto de modelos hasta el punto 26, a partir del cuál domina el modelo lineal. Las mejoras que consiguen estos modelos en relación con el mejor de los paramétricos son del 4%, en términos aproximados.

A la luz de los resultados anteriores, debemos destacar que no podemos afirmar que ninguno de los modelos considerados (red neuronal, combinación lineal y mejor modelo paramétrico) domine sistemáticamente a los demás. El dominio de algún modelo frente a otros depende de la moneda y momento del tiempo de que se trate. Sin embargo estos resultados muestran que, generalmente, los modelos que utilizamos tienen una capacidad predictiva superior a la de los modelos paramétricos habituales.

## **5. CONCLUSIONES**

En este trabajo predecimos las volatilidades del tipo de cambio de la peseta frente a seis monedas mediante una potente herramienta no paramétrica, con gran capacidad para modelizar y predecir dinámicas no lineales, las RNA, comparando su calidad predictiva con la de otros modelos: homocedástico, ARCH, GARCH, EGARCH, GJR y modelo lineal. Además, proponemos un modelo híbrido de redes neuronales artificiales, que basado en una sencilla combinación de diferentes modelos, permite construir modelizaciones más complejas que aprovechan las ventajas de cada uno de los modelos individuales.

El primer estudio, en que aplicamos RNA a la predicción de la volatilidad del tipo de cambio de la peseta frente a las seis monedas contempladas, permite extraer numerosas e interesantes conclusiones. Debemos destacar que efectuamos un exhaustivo análisis comparativo, en cuanto a capacidad predictiva, de las RNA frente a otros tipos de modelizaciones y que las predicciones se obtienen de forma dinámica. En la primera parte de este estudio, que descomponemos en dos etapas, comparamos la capacidad predictiva de cada uno de los modelos paramétricos de la clase ARCH empleados en relación con el homocedástico y, también, entre sí. En este análisis, las predicciones de la volatilidad del tipo de cambio de la peseta se obtienen a partir de observaciones históricas de las series con que trabajamos. Los resultados que obtenemos permiten concluir que ningún modelo se muestra absolutamente superior

para todas las monedas, sin embargo, todos ellos parecen tener capacidad predictiva al superar, en la mayor parte de las ocasiones, al homocedástico.

A continuación, comparamos el modelo de red neuronal que construimos, para cada una de las monedas, con la media de los procesos ARCH. Las predicciones de la volatilidad del tipo de cambio de la peseta, las obtenemos a partir de las observaciones históricas. La conclusión que podemos obtener es que la capacidad predictiva de las redes que hemos diseñado, en media, parece ser similar a la de los modelos tipo ARCH, ya que la red se muestra superior en la mitad de las ocasiones. Ahora bien, los resultados apuntan a una superioridad de las RNA, ya que la cuantía de las mejoras que consiguen es bastante mayor que el importe de los empeoramientos frente a los modelos ARCH.

En el segundo análisis de RNA construimos, para cada moneda, dos modelos de agregación, es decir, una RNA y un simple modelo lineal, que combinan las predicciones de los modelos homocedástico, ARCH, GARCH, EGARCH y GJR. Con ello, pretendemos mejorar la calidad de las predicciones, por tanto, realizamos una extensa comparación de estos modelos frente a los paramétricos en media y, también, frente al modelo paramétrico que mejor se comporta, con el fin de verificar si efectivamente logramos nuestro objetivo. Podemos concluir que, en general, las redes híbridas superan claramente a la media de los modelos paramétricos, presentando en algunos casos mejoras espectaculares. El modelo lineal de agregación también mejora, generalmente, a la media de los paramétricos, ahora bien, la mejora es más clara y significativa en el caso de las RNA. En cuanto a la capacidad predictiva de las redes neuronales frente a la combinación lineal, debemos apuntar que las redes neuronales parecen comportarse algo mejor. También, debemos destacar que el comportamiento de ambos modelos muestra claras diferencias, ya que en algunos casos domina la RNA, mientras que en otros el modelo lineal se muestra superior, según el horizonte de predicción y moneda tomada en consideración. Por tanto, la selección de un modelo u otro con fines predictivos, depende de los factores mencionados.

Además, con generalidad, podemos afirmar que los modelos de RNA analizados tienen capacidad predictiva, dado que superan al homocedástico y a los modelos paramétricos en general. Por otro lado, queremos resaltar que el modelo híbrido que proponemos, sí parece comportarse mejor que la media de los paramétricos y que los modelos de RNA que predicen a partir de la serie histórica de observaciones.

Como los antecedentes de la aplicación de RNA en el ámbito predictivo de volatilidad de series cambiarias son, hasta nuestros días, prácticamente inexistentes, debemos destacar que éste es un campo cuasi virgen para la investigación y donde, actualmente, todavía queda mucho por hacer.



## ANEXO

### Análisis comparativo de los modelos de Heterocedasticidad Condicional frente al Homocedástico

Figura A.1: DMK

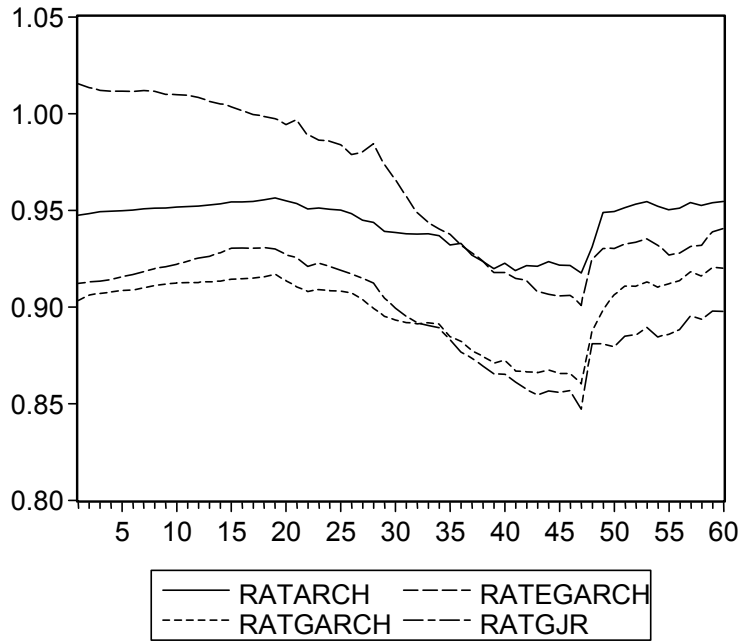
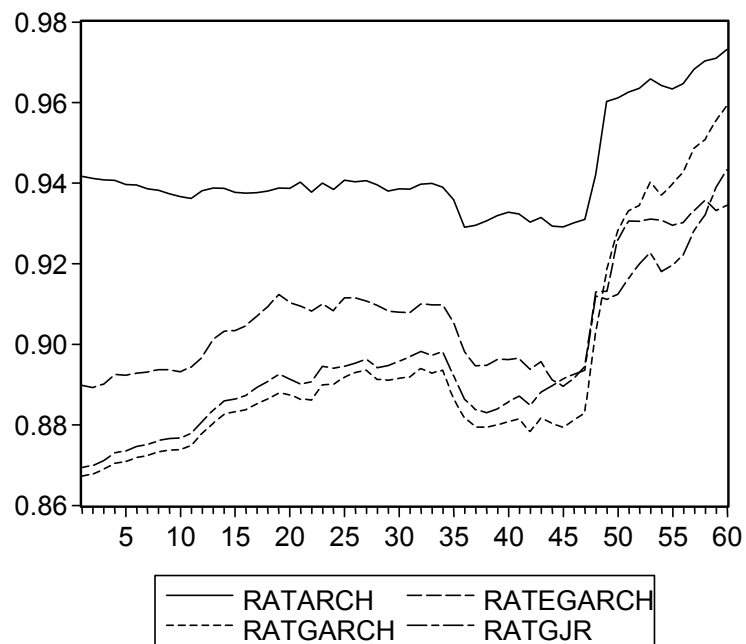


Figura A.2: FFR



## Análisis comparativo de los modelos de Heterocedasticidad Condicional frente al Homocedástico

Figura A.3: LIT

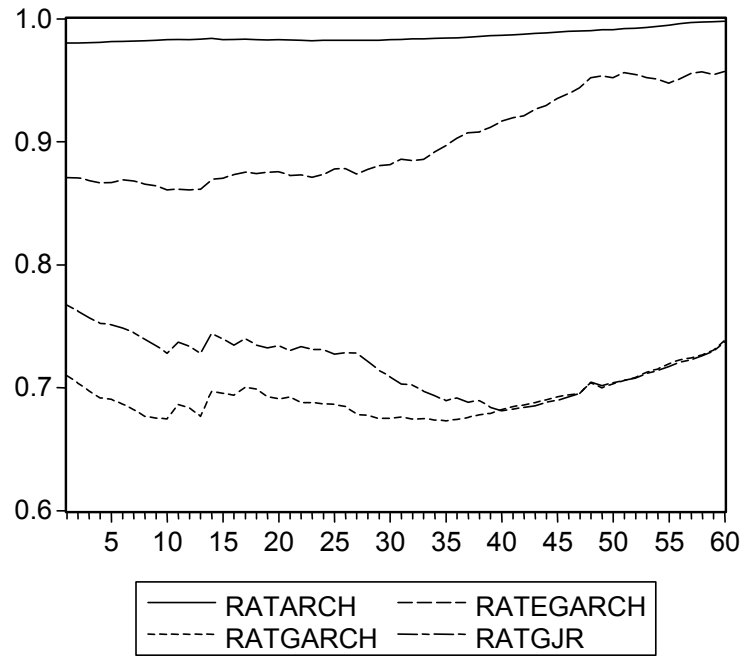
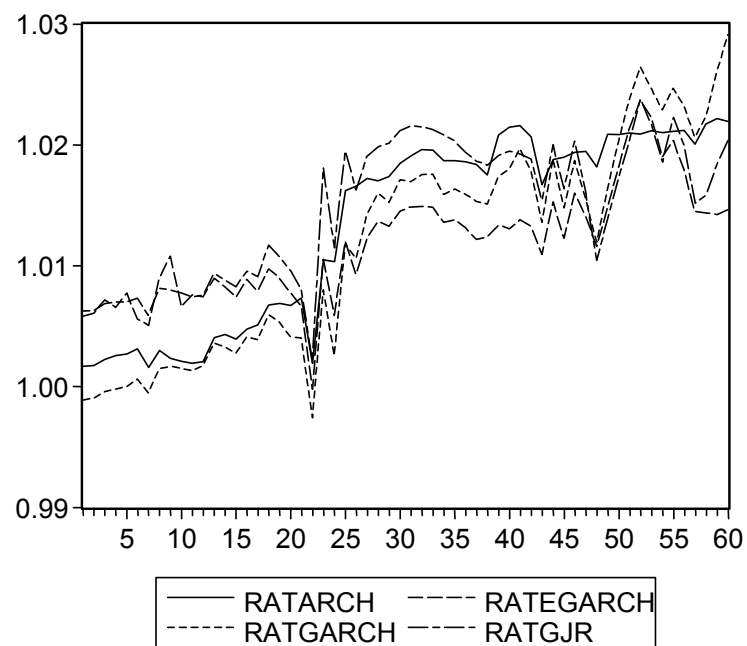


Figura A.4: STG



## Análisis comparativo de los modelos de Heterocedasticidad Condicional frente al Homocedástico

Figura A.5: USD

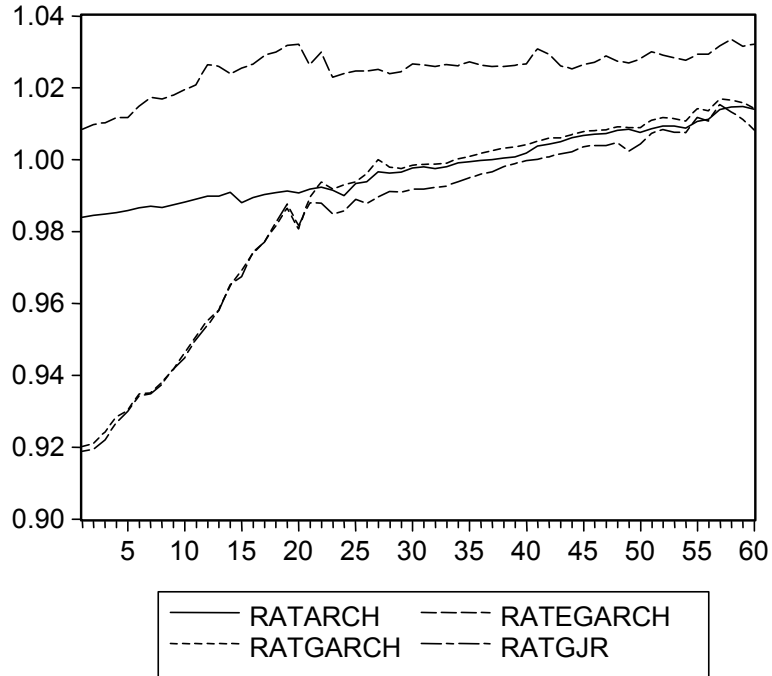
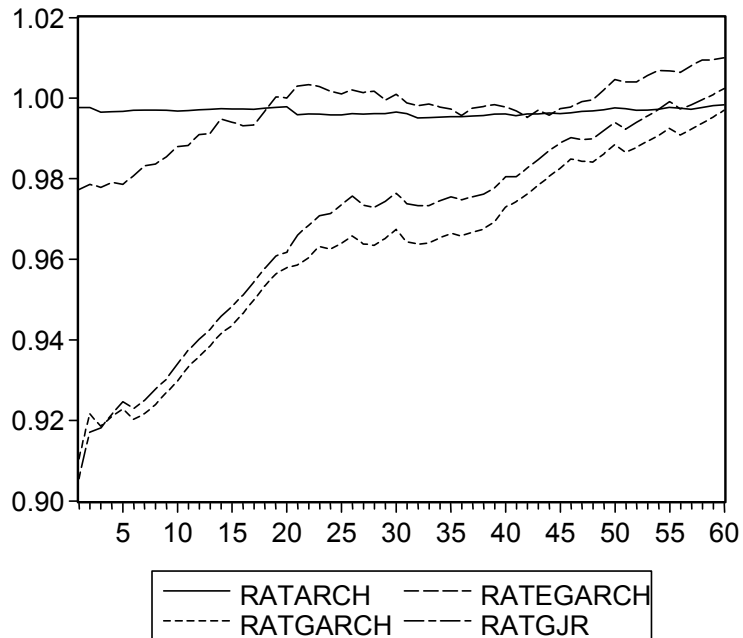
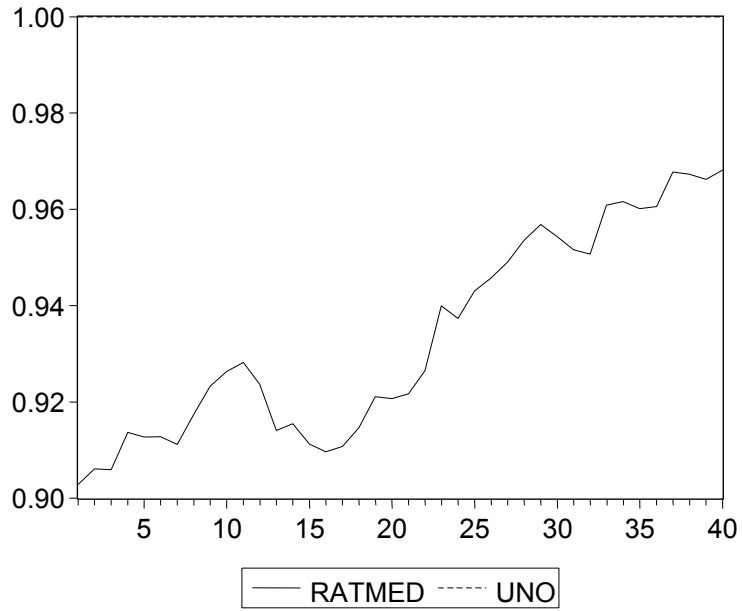


Figura A.6: YEN

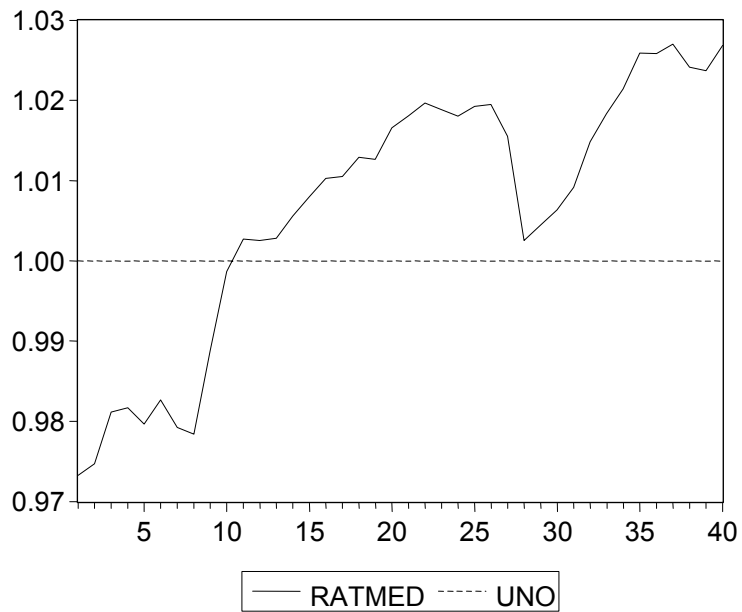


**Análisis comparativo de los RNA en relación con la media de los modelos de la clase ARCH**

**Figura A.7: DMK**

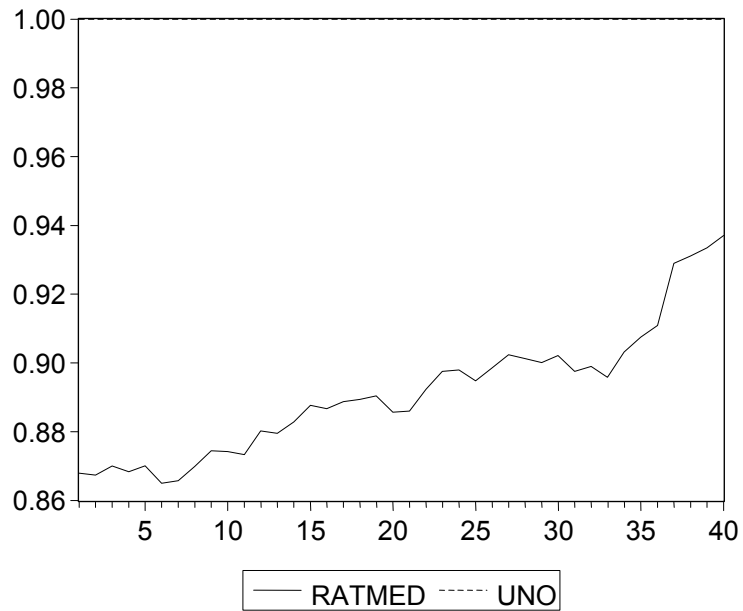


**Figura A.8: FFR**

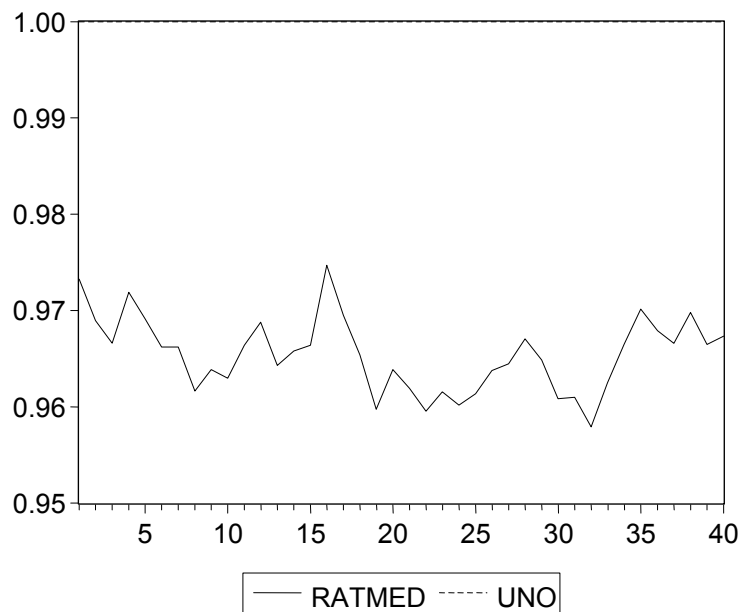


**Análisis comparativo de los RNA en relación con la media de los modelos de la clase ARCH**

**Figura A.9: LIT**

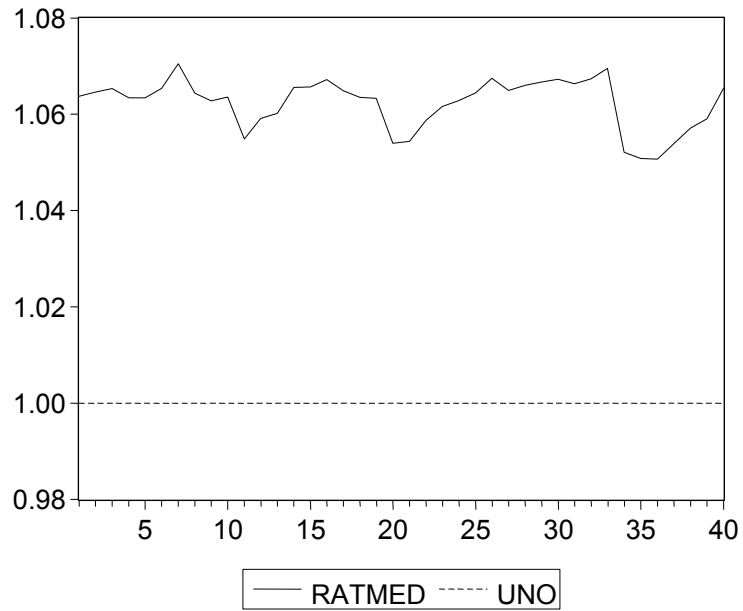


**Figura A.10: STG**

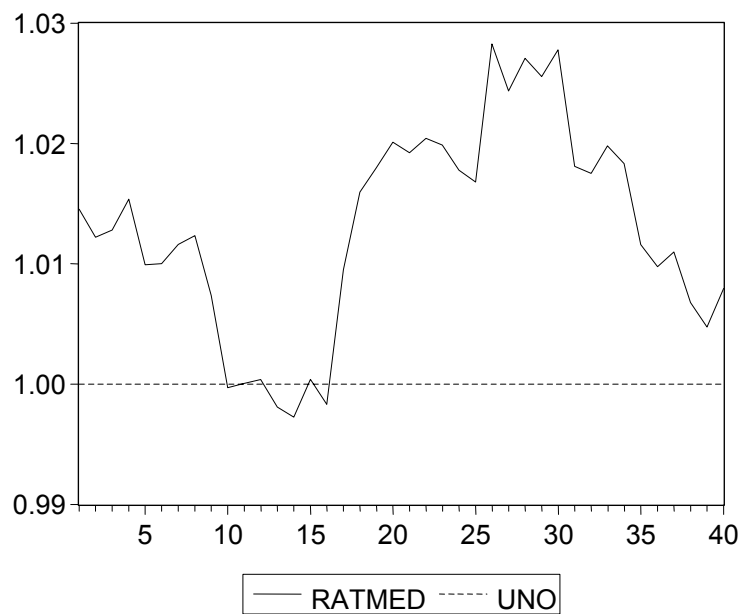


**Análisis comparativo de los RNA en relación con la media de los modelos de la clase ARCH**

**Figura A.11: USD**

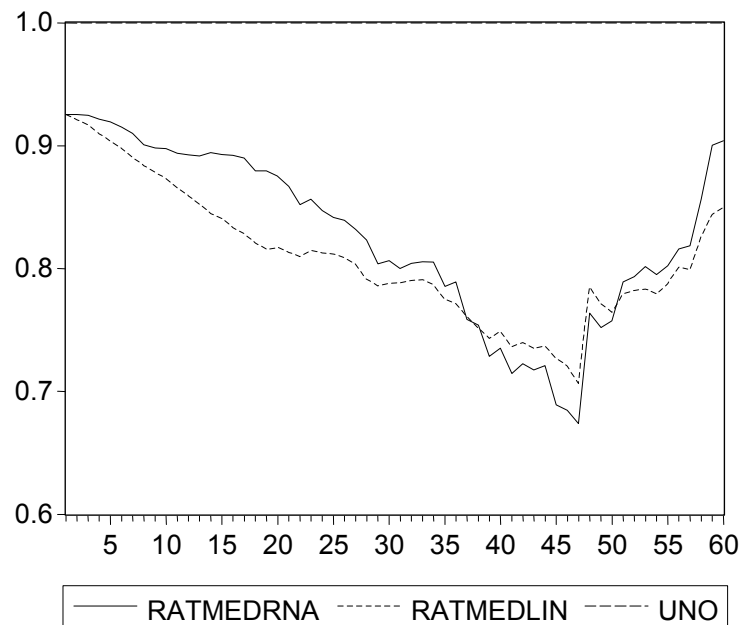


**Figura A.12: YEN**

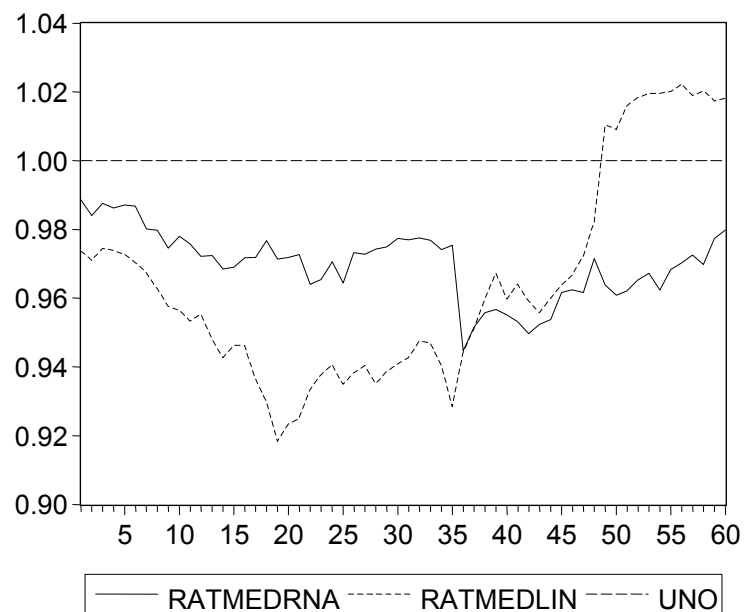


**Análisis comparativo de los modelos híbridos (RNA y Lineal) frente a la media de los modelos de Volatilidad Condicional**

**Figura A.13: DMK**



**Figura A.14: FFR**



## Análisis comparativo de los modelos híbridos (RNA y Lineal) frente a la media de los modelos de Volatilidad Condicional

Figura A.15: LIT

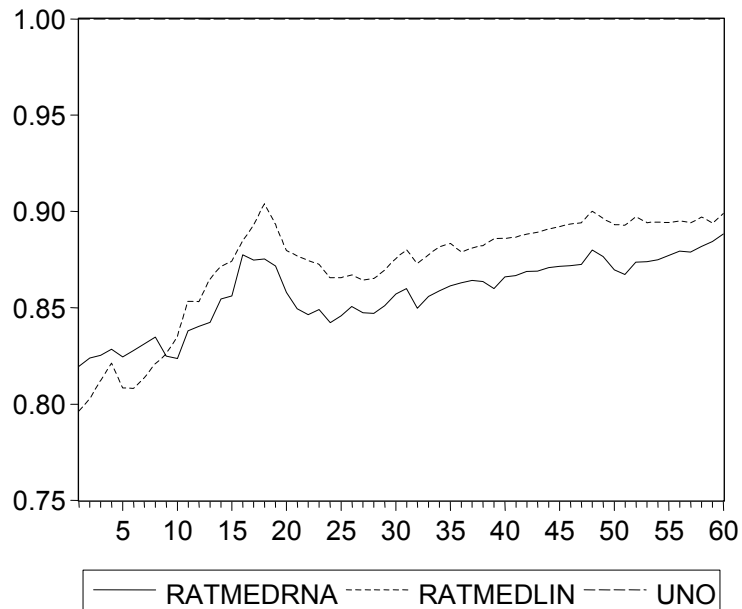
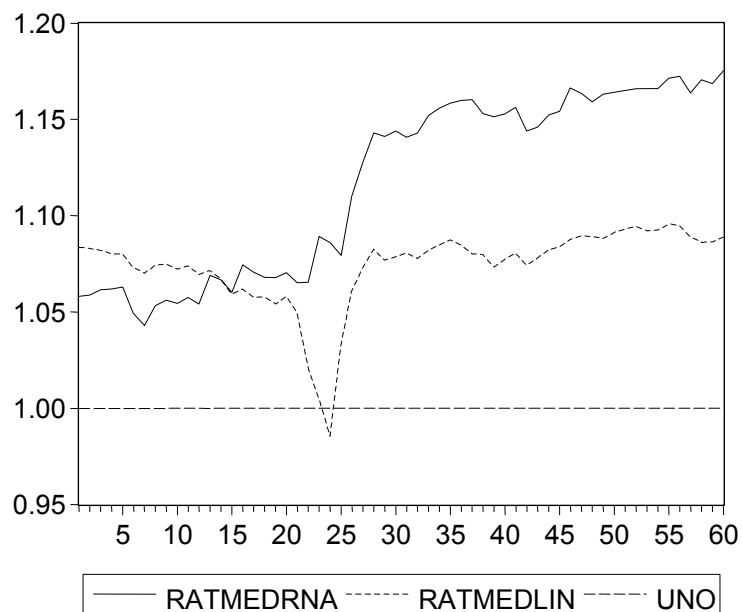


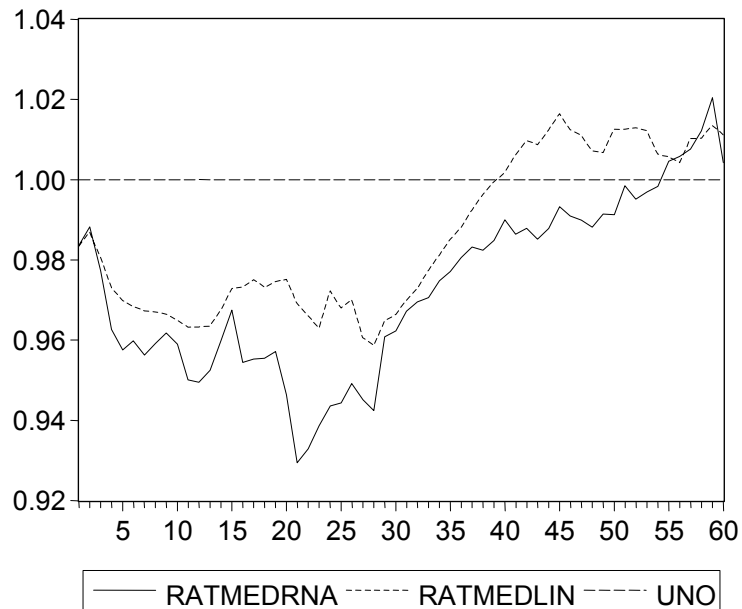
Figura A.16: STG



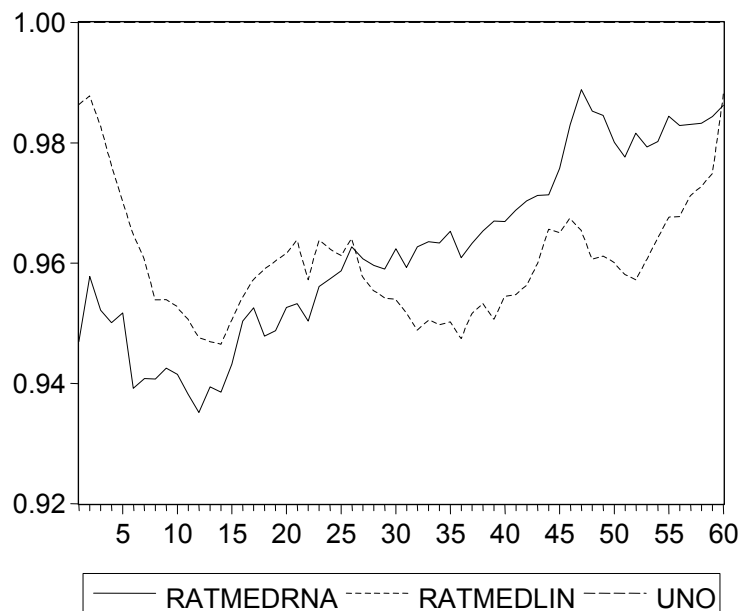


**Análisis comparativo de los modelos híbridos (RNA y Lineal) frente a la media de los modelos de Volatilidad Condicional**

**Figura A.17: USD**

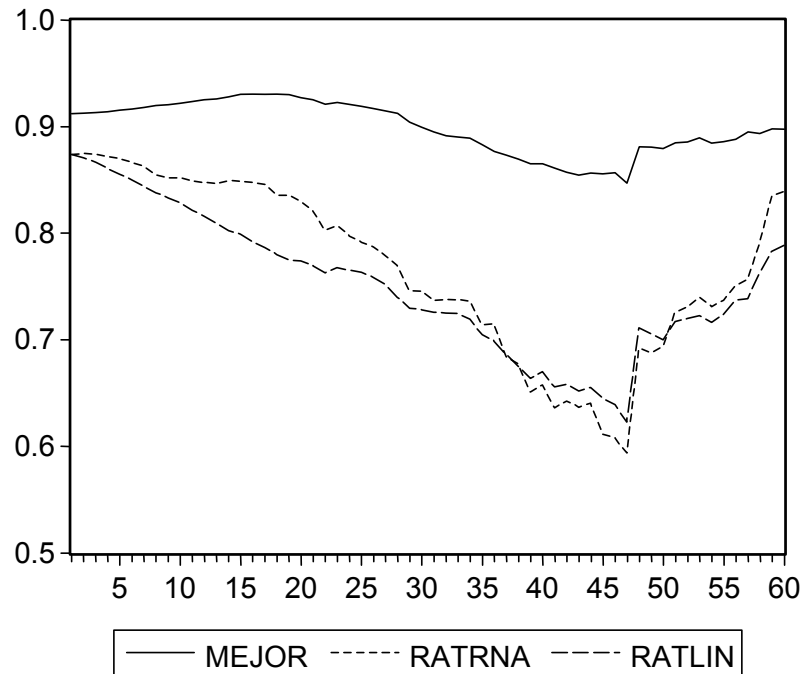


**Figura A.18: YEN**

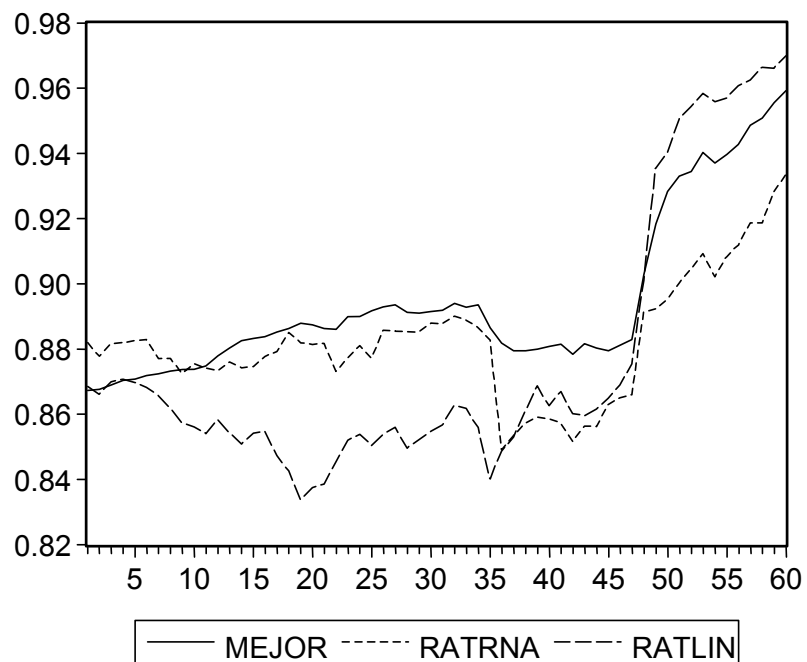


**Análisis comparativo de los modelos híbridos (RNA y Lineal) frente al modelo paramétrico que mejor se comporta**

**Figura A.19: DMK**

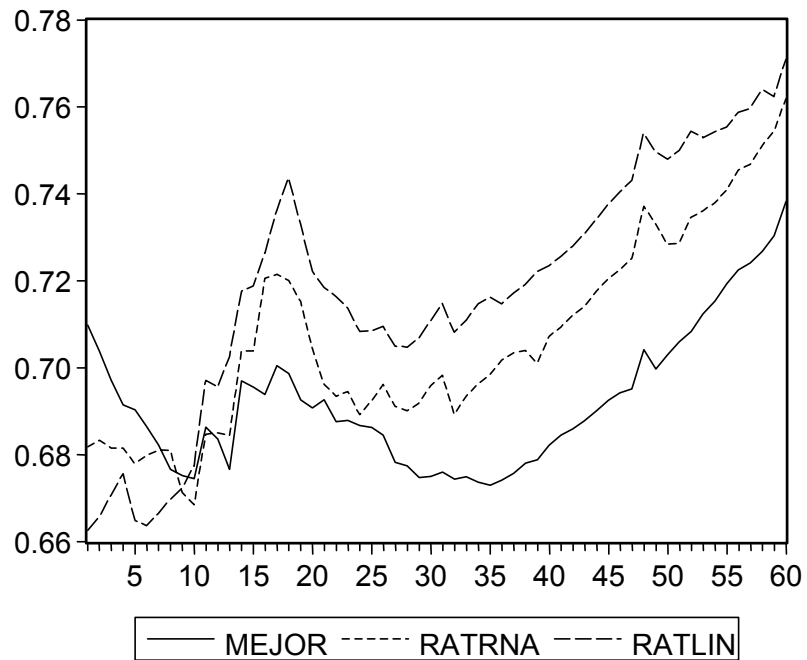


**Figura A.20: FFR**

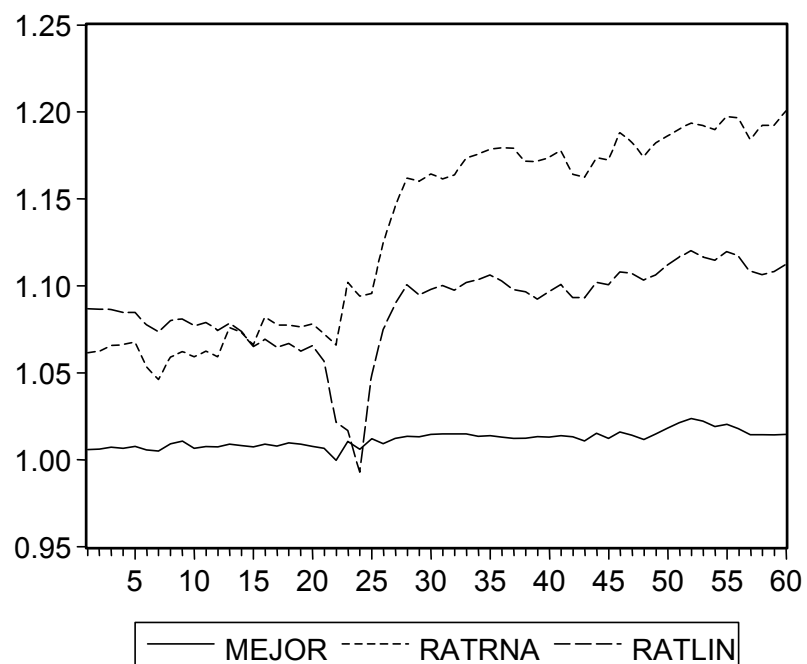


**Análisis comparativo de los modelos híbridos (RNA y Lineal) frente al modelo paramétrico que mejor se comporta**

**Figura A.21: LIT**

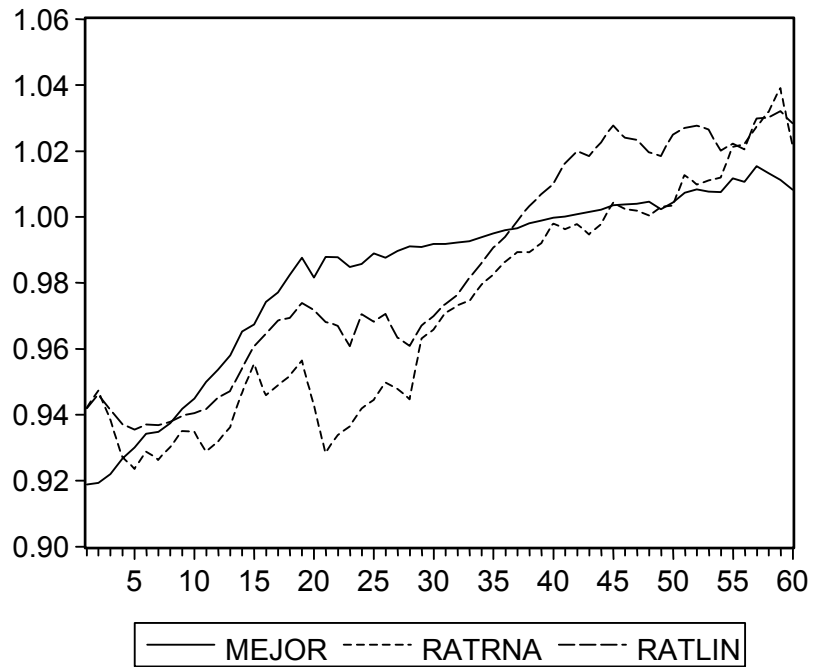


**Figura A.22: STG**

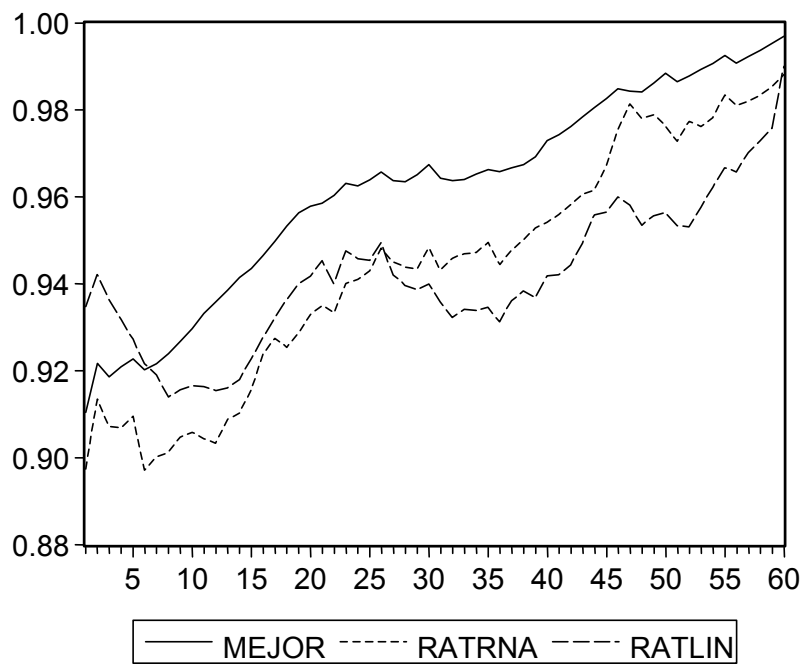


**Análisis comparativo de los modelos híbridos (RNA y Lineal) frente al modelo paramétrico que mejor se comporta**

**Figura A.23: USD**



**Figura A.24: YEN**



## REFERENCIAS BIBLIOGRÁFICAS

- AMEMIYA, T. [1981]: "Qualitative response models: a survey", *Journal of Economic Literature*, núm. 19, pp. 1483-1536.
- BATES, J.M., y GRANGER, C.W.J. [1969]: "The combination of forecasts", *Operational Research Quarterly*, núm. 20, pp. 451-468.
- BLACK, F. [1976]: "Studies of Stock Market Volatility Changes", *1976 Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pp. 177-181.
- BOLLERSLEV, T. [1986]: "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, núm. 31, pp. 307-327.
- BOLLERSLEV, T.; CHOU, R.Y., y KRONER, K.F. [1992]: "ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence", *Journal of Econometrics*, núm. 52, pp. 5-59.
- BROCK, W.; DECHERT, W., y SCHEINKMAN, J. [1987]: "A Test for Independence Based on the Correlation Dimension", *Working Paper*, University of Chicago.
- CARROLL, S.M., y DICKINSON, B.W. [1989]: "Construction of Neural Nets Using the Random Transform", en *Proceedings of the International Joint Conference on Neural Networks*, Washington D.C., New York: IEEE Press, pp. I: 607-611.
- COWAN, J. [1967]: "A Mathematical Theory of Central Nervous Activity", *Unpublished Ph.D. Dissertation*, University of London.
- CYBENKO, G. [1989]: "Approximation by Superpositions of a Sigmoid Function", *Mathematics of Control, Signals and Systems*, núm. 2, pp. 303-314.
- ENGLER, R.F. [1982]: "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, núm. 50, pp. 987-1008.
- GEMAN, S.; BIENENSTOCK, E., y DOURSAT, R. [1992]: "Neural Networks and the Bias/Variance Dilemma", *Neural Computation*, núm. 4, pp. 1-58.
- GLOSTEN, L.R.; JAGANNATHAN, R., y RUNKLE, D.E. [1993]: "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", *Journal of Finance*, núm. 48, 5, pp.1779-1801.
- GRANGER, C.W.J., y DING, Z. [1994]: "Stylized Facts on the Temporal and Distributional Properties of Daily Data from Speculative Markets", *Working Paper*, 94. Department of Economics, University of California, San Diego, USA.
- GRANGER, C.W.J., y RAMANATHAN, R. [1984]: "Improved Methods of Forecasting", *Journal of Forecasting*, núm. 3, pp. 197-204.

- HECHT-NIELSEN, R. [1989]: "Theory of the Back-Propagation Neural Network", en *Proceedings of the International Joint Conference on Neural Networks*, Washington D.C., New York: IEEE Press, pp. I: 593-606.
- HORNIK, K.; STINCHCOMBE, M., y WHITE, H. [1989]: "Multilayer feedforward networks are universal approximators", *Neural Networks*, núm. 2, pp. 359-366.
- HORNIK, K.; STINCHCOMBE, M., y WHITE, H. [1990]: "Universal Approximation of an Unknown Mapping and its Derivatives Using Multilayer Feedforward Networks", *Neural Networks*, núm. 3, pp. 551-560.
- HSIEH, D.A. [1988]: "The Statistical Properties of Daily Foreign Exchange Rates: 1974-1983", *Journal of International Economics*, núm. 24, pp. 129-145.
- KAMSTRA, M., y KENNEDY, P. [1998]: "Combining qualitative forecasts using logit", *International Journal of Forecasting*, núm. 14, pp. 83-93.
- KUAN, C.M., y WHITE, H. [1994]: "Artificial Neural Networks: an Econometric Perspective", *Econometric Reviews*, núm. 13, pp. 1-91.
- KUMAR, A., y OLMEDA, I. [1997]: "A Study of Hybrid Models for Knowledge Discovery", *INFORMS Conference on Information Systems and Technology*, San Diego, CA, 1-5.
- LE CUN, Y. [1985]: "Une Procédure D'Apprentissage pour Réseau a Seuil Assymétrique", *Proceedings of Cognitiva*, núm. 85, pp. 599-604.
- MANDELBROT, B. [1963]: "The Variation of Certain Speculative Prices", *Journal of Business*, núm. 36, pp. 394-419.
- MAÑAS, L.A. [1986]: "Empirical Behavior of Flexible Exchange Rates: Statistical Analysis and Consistent Models", *Ph.D. Dissertation*, University of Chicago.
- MARCO, P. [1998]: "Modelos de Volatilidad en el Mercado Español de Divisas", *Tesis Doctoral no Publicada*, Universitat de València.
- MARCO, P., y PUERTAS, R. [1999]: "Aplicaciones Financieras de las Redes Neuronales Artificiales", en E. Fernández, e I. Olmeda (Eds.) [1999]: *Aplicaciones de la Inteligencia Artificial* (Madrid: Díaz de Santos), próxima publicación.
- NELSON, D.B. [1991]: "Conditional Heteroskedasticity in Asset Returns: A New Approach", *Econometrica*, vol.59, núm. 2, pp. 347-370.
- OLMEDA, I., y BARBA-ROMERO, S. [1993]: *Redes Neuronales Artificiales: Fundamentos y Aplicaciones* (Madrid: Servicio de Publicaciones de la Universidad de Alcalá de Henares).
- OLMEDA, I., y FERNÁNDEZ, E. [1997]: "Hybrid classifiers for financial multicriteria decision making: the case of Bankruptcy prediction", *Computational Economics*, núm. 10, pp. 317-335.

- OLMEDA, I. [1993]: “Aprendizaje y Generalización”, en I. Olmeda, y S. Barba-Romero (Eds.) [1993]: *Redes Neuronales Artificiales: Fundamentos y Aplicaciones* (Madrid: Servicio de Publicaciones de la Universidad de Alcalá de Henares), pp. 43-63.
- PARKER, D.B. [1982]: “Learning Logic Invention”, *Report S81-64*, file 1, Office of Technology Licensing, Stanford University.
- PARKER, D.B. [1985]: “Learning Logic”, MIT Center for Computational Research in Economics and Management Science, *Technical Report TR-47*.
- ROBBINS, H., y MONRO, S. [1951]: “A stochastic approximation method”, *The annals of Mathematical Statistics*, núm. 22, pp. 400-407.
- STINCHCOMBE, M., y WHITE, H. [1989]: “Universal Approximation Using Feedforward Networks with Non-Sigmoid Hidden Layer Activation Functions”, en *Proceedings of the International Joint Conference on Neural Networks*, Washington D.C., New York: IEEE Press, pp. I: 612-617.
- WERBOS, P. [1974]: “Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences”, *Ph. D. Dissertation*, Harvard University, Department of Applied Mathematics.
- ZAKOÏAN, J.M. [1994]: “Threshold Heteroskedastic Models”, *Journal of Economic Dynamics and Control*, núm. 18, pp. 931-955.