



UNIVERSIDAD NACIONAL DE COLOMBIA

# **Pronóstico de la volatilidad usando ensambles de redes neuronales basados en familias de funciones de error**

**Johanna Marcela Orozco Castañeda, MSc**

Universidad Nacional de Colombia  
Facultad de Minas, Departamento de Ciencias de la Computación y de la Decisión  
Programa de doctorado en Ingeniería-Sistemas e Informática  
Medellín, Colombia  
2016



# **Pronóstico de la volatilidad usando ensambles de redes neuronales basados en familias de funciones de error**

**Johanna Marcela Orozco Castañeda, MSc**

Tesis presentada como requisito parcial para optar al título de:  
**Doctor en Ingeniería - Sistemas e Informática**

Director:

Prof. Juan David Velásquez Henao, PhD

Línea de Investigación:

Inteligencia Artificial

Grupo de Investigación:

Sistemas e Informática

Universidad Nacional de Colombia  
Facultad de Minas, Departamento de Ciencias de la Computación y de la Decisión  
Medellín, Colombia  
2016



A mi mamá, a mi papá

A mis hermanos

A mis abuelos



# Agradecimientos

Agradezco a la Universidad Nacional por permitirme iniciar y culminar mi doctorado. De manera especial quiero agradecer al profesor Juan David Velásquez, mi director, por orientarme en este proceso de formación, por su tiempo, su paciencia, por sus ideas y su capacidad para ver luz donde hay oscuridad.

Agradezco a Colciencias por el apoyo económico que tuve durante gran parte del tiempo que duró este proyecto.

Quiero agradecer de manera particular a la profesora Lilian De Menezes y el Dr Eduardo Alonso de la “City University London” que me recibieron de la mejor manera durante mi pasantía en Londres, y por sus aportes que fueron fundamentales en algunas partes de esta tesis.

Gracias a todos mis amigos y compañeros de almuerzos por las discusiones académicas, los ratos de risas y las tertulias.

A mi familia toda mi gratitud, por el apoyo incondicional que recibí todo el tiempo, por creer en mí antes que nadie.

A César gracias por acompañarme y alentarme en los momentos difíciles, gracias por estar siempre dispuesto a hacer mis días felices.



## Resumen

En esta tesis se estudia el pronóstico de la volatilidad condicional de series de tiempo financieras mediante ensambles de pronóstico de redes neuronales artificiales. La primera contribución de esta tesis es el planteamiento de un proceso de especificación de modelos de volatilidad condicional por medio de una técnica de remuestreo por bootstrapping, donde cada modelo se estima usando diferentes funciones de error como criterios de optimización. La segunda contribución de este trabajo es la combinación no lineal de los conjuntos de pronósticos de volatilidad por medio de una red neuronal artificial, así como la aplicación del remuestreo por bootstrapping para hallar la distribución muestral de los parámetros de la red. La tercera contribución de esta tesis es la definición y aplicación empírica de un contraste estadístico de dominancia de pronósticos en combinaciones no lineales de pronósticos de volatilidad condicional para determinar la significancia de las variables de entrada de la red.

**Palabras clave:** redes neuronales, combinación de pronósticos, series de tiempo no lineales, volatilidad.

## Abstract

In this thesis the aim is to study the conditional volatility forecasting of financial time series with forecasts ensemble of artificial neural networks. The first contribution of this thesis is the proposal of a specification process for conditional volatility models with a bootstrapping technique, where each model is estimated using different error functions as optimization criteria. The second contribution of this work is the nonlinear combination of volatility forecast sets by a neural network; likewise, the application of bootstrapping technique to find the sample distribution of the neural network weights. The third contribution of this thesis is the definition and empirical application of a statistical test for forecasts encompassing in nonlinear forecasts combination of conditional volatility forecasts to evaluate the significance of the input variables in the neural network.

**Keywords:** Neural networks, forecasts combination, nonlinear time series, conditional heteroskedasticity, forecasts encompassing.

# Contenido

<b>Agradecimientos</b>	<b>vii</b>
<b>Resumen</b>	<b>ix</b>
<b>1 Introducción</b>	<b>2</b>
1.1 Pronóstico de la volatilidad usando redes neuronales	4
1.1.1 Redes neuronales en pronóstico de series financieras	4
1.1.2 Metodología de investigación	5
1.1.3 Resultados obtenidos	6
1.1.4 Discusión	7
1.1.5 Conclusiones	11
1.2 Pronóstico de la volatilidad usando combinación de pronósticos	12
1.2.1 Combinación de pronósticos	12
1.2.2 Metodología de investigación	14
1.2.3 Resultados obtenidos	15
1.2.4 Discusión	15
1.2.5 Conclusiones	19
1.3 Preguntas emergentes de investigación	19
1.4 Hipótesis de investigación	20
1.5 Objetivos de la tesis	20
1.5.1 Objetivo general	21
1.5.2 Objetivos específicos	21
1.6 Contribuciones de la tesis	21
1.7 Estructura de la tesis	22
<b>2 Sistema de combinación de pronósticos para predecir la volatilidad</b>	<b>24</b>
2.1 Metodología propuesta	26
2.1.1 Modelo GARCH	26
2.1.2 Estructura general	27
2.1.3 Procedimiento de estimación alternativo de los parámetros para el modelo GARCH	27
2.1.4 Criterios de optimización para estimar los parámetros de un modelo GARCH	28
2.1.5 Combinador de pronósticos	29

2.2	Caso de aplicación	30
2.2.1	Datos y análisis preliminar	30
2.2.2	Montaje experimental	30
2.2.3	Precisión de pronóstico de los modelos individuales	31
2.2.4	Precisión de pronóstico de los pronósticos combinados usando promedio aritmético simple	32
2.2.5	Precisión de la combinación de pronósticos usando regresión lineal múltiple	33
2.2.6	Precisión de la combinación de pronósticos usando redes neuronales artificiales	33
2.2.7	Desempeño del modelo elegido	33
2.3	Conclusiones	40
<b>3</b>	<b>Proceso de especificación de modelos de pronóstico de volatilidad implementando diferentes criterios de optimización</b>	<b>41</b>
3.1	Justificación	41
3.1.1	Funciones de error como criterio de optimización	42
3.1.2	Caracterización de las funciones de error	43
3.2	Metodología	45
3.2.1	Valor proxy de la volatilidad	45
3.2.2	Especificaciones GARCH	45
3.2.3	Proceso de especificación del modelo	46
3.2.4	Estimación de los parámetros del modelo TGARCH(1,1)	47
3.3	Aplicación a series de tiempo reales	49
3.3.1	Datos	49
3.3.2	Ajuste de las series de tiempo	49
3.3.3	Estimación de los parámetros por bootstrapping	52
3.3.4	Evaluación de los modelos	55
3.4	Conclusiones	57
<b>4</b>	<b>Combinación y dominancia de pronósticos</b>	<b>59</b>
4.1	Métodos tradicionales de combinación de pronósticos	59
4.2	Dominancia de pronósticos	61
4.2.1	Prueba de dominancia de pronósticos en una regresión lineal	62
4.2.2	Dominancia de pronósticos con redes neuronales artificiales	63
4.3	Proceso de entrenamiento de una red neuronal artificial	64
4.4	Prueba de Wald	66
4.5	Aplicación de la prueba de Wald en una combinación no lineal de pronósticos	68
4.5.1	Métodos de combinación	69
4.5.2	Evaluación de la combinación	69
4.5.3	Cálculo de la prueba de Wald	70

---

4.6	Discusión . . . . .	72
<b>5</b>	<b>Conclusiones y trabajo futuro</b>	<b>73</b>
5.1	Respuesta a las preguntas de investigación . . . . .	73
5.1.1	Pregunta 1 . . . . .	73
5.1.2	Pregunta 2 . . . . .	74
5.1.3	Pregunta 3 . . . . .	74
5.1.4	Pregunta 4 . . . . .	74
5.2	Cumplimiento de los objetivos específicos de esta tesis . . . . .	75
5.2.1	Objetivo específico 1 . . . . .	75
5.2.2	Objetivo específico 2 . . . . .	76
5.2.3	Objetivo específico 3 . . . . .	76
5.2.4	Objetivo específico 4 . . . . .	76
5.3	Cumplimiento del objetivo general de esta tesis . . . . .	77
5.4	Trabajo futuro . . . . .	78
5.5	Difusión de resultados . . . . .	78
5.5.1	Artículo de investigación . . . . .	78
5.5.2	Ponencias . . . . .	78
	<b>Bibliografía</b>	<b>80</b>

# 1 Introducción

En términos generales, no es una tarea fácil dar pronósticos precisos de los valores de una serie de tiempo, y más aún, teniendo en cuenta que los pronósticos de una serie de tiempo, particularmente en el sector financiero sirven de base en la toma de decisiones. Sin embargo, para el caso de los pronósticos obtenidos mediante modelos de predicción hay quienes afirman que estos solo deberían dar cuenta de una mínima parte en la toma de las decisiones y no deberían usarse como un recurso cien por ciento confiable ([Makridakis et al., 1998](#)). Este tipo de afirmaciones toman fuerza debido a que cuando se intenta modelar un conjunto de datos, no es posible saber a priori cuál método resulta más preciso para determinar el proceso generador de datos y por ende tener pronósticos cuantitativos de la serie de tiempo de interés.

Teniendo en cuenta que una serie de tiempo  $\{x_t\}$  es una realización de un proceso estocástico,  $\{X_t\}_{t \in \mathbb{Z}}$ , intrínsecamente lo que se desea con una única observación en cada instante de tiempo, es determinar la distribución de la variable aleatoria del proceso estocástico. En algunos casos, solo es posible determinar la media y varianza de dicha variable aleatoria. Es así como construir modelos de pronóstico de series de tiempo que den cuenta del proceso generador de los datos (PGD), usualmente desconocido, es un problema que ha captado la atención de académicos y profesionales de diversas áreas del conocimiento.

Específicamente, el pronóstico de volatilidad de los rendimientos de un activo financiero es inherentemente un problema de predicción difícil debido a la naturaleza de la variable que se quiere predecir y por la importancia de la administración del riesgo en mercados financieros. La no observabilidad directa de la serie de las volatilidades ha llevado al desarrollo de diversos modelos de pronóstico, así como al estudio de su desempeño. El modelado y pronóstico de la volatilidad juegan un papel preponderante para la toma de decisiones en el ámbito financiero ya que los modelos de volatilidad son una herramienta para la administración del riesgo; asimismo, el conocimiento acerca del impacto de los factores económicos sobre la volatilidad de tasas de cambio es importante para la formulación de políticas ([Tsangari, 2007](#)).

Se define la volatilidad como una medida de dispersión del rendimiento de un activo financiero. En este trabajo se considera la varianza como tal medida de dispersión.

En la literatura se reportan diferentes características de la volatilidad de las series de tiempo financieras que hacen particularmente difícil encontrar pronósticos precisos. Se sabe, por ejemplo:

---

que las series de rendimientos financieros exhiben persistencia en su varianza (Bollerslev y Engle, 1993; Franses y Dijk, 2000); que se caracterizan por tener varianza cambiante en el tiempo; que hay agrupamientos de volatilidad, lo que indica que si en un período la volatilidad es alta o baja tiende a seguir siendo alta o baja para el período siguiente. Que tiene una reacción diferente ante un incremento del precio y ante un descenso del precio, esto es el efecto de apalancamiento o respuesta asimétrica al signo de los rendimientos pasados, las colas pesadas y exceso de curtosis y el comportamiento no lineal y no estacionario (Bollerslev y Engle, 1993; Donaldson y Kamstra, 1997; Franses y Dijk, 2000; Nelson, 1991; Poon y Granger, 2003; Rabemananjara y Zakoian, 1993). Los agrupamientos de volatilidad o la persistencia sugieren un modelo de series de tiempo en el cual las perturbaciones sucesivas son dependientes, aunque no estén correlacionadas.

Estas, entre otras, son las razones que motivan el uso de un esquema de combinación de pronósticos para hallar predicciones más precisas. Naturalmente, se desea capturar la mayor parte de las características de una serie con miras a tomar la mejor decisión en el mercado, por ejemplo comprar o vender. Tener una gama de predicciones obtenidas de diferentes modelos, que dan cuenta del comportamiento del mercado es, sin duda, una ventaja, ya que representaría en el mejor escenario, información complementaria que serviría para llegar a un pronóstico combinado más próximo al valor real.

La combinación de pronósticos se ha convertido en un método plausible cuando se tienen disponibles una variedad de pronósticos de una misma variable. En la literatura este enfoque de combinación ha sido implementado obteniéndose resultados empíricos en algunos casos satisfactorios en cuanto a precisión se refiere, ver por ejemplo (Becker y Clements, 2008; Clemen y Winkler, 1986; Clemen, R., 1989; Hashem, 1997; Lux y Morales-Arias, 2010; Zhang et al., 2011). Así, el objetivo de esta tesis está dirigido a proponer un esquema de pronóstico de la volatilidad de los rendimientos de activos financieros basado en el uso de combinación de pronósticos; en este esquema, también conocido como ensamble de pronósticos basado en redes neuronales artificiales, el elemento clave es la diversidad de los expertos, la cual se logra usando diferentes criterios de optimización para obtener los parámetros de los modelos individuales.

En este capítulo se presenta una revisión sistemática de la literatura en pronóstico de la volatilidad usando redes neuronales y en pronóstico de la volatilidad usando combinación de pronósticos. Lo anterior permitirá tener una visión general acerca de los trabajos realizados en esta área del conocimiento, además de la pertinencia y actualidad del tema. Como resultado de dicha revisión se presentan la hipótesis de investigación, los objetivos de esta tesis, los logros y estructura de la misma.

## 1.1 Pronóstico de la volatilidad usando redes neuronales

En esta sección se presenta una revisión del estado actual de las metodologías que se han usado hasta este momento para el pronóstico de la volatilidad de series de tiempo financieras que incluyen el uso de redes neuronales artificiales. Si bien desde principios de la década de 1980 aparecieron los primeros trabajos en modelado de volatilidad fue a finales cuando comenzaron a aparecer los primeros trabajos en pronóstico así como la idea de redes neuronales artificiales. Las cuales ganarían importancia en este y otros campos dada su flexibilidad para modelar características complejas inherentes a las series de tiempo financieras. Una de las principales contribuciones de esta revisión es destacar la vigencia del tema de combinación de pronósticos tanto a nivel práctico como teórico. Además, reunir los aportes existentes en la literatura en pronóstico de la volatilidad mediante el uso de redes neuronales artificiales.

### 1.1.1 Redes neuronales en pronóstico de series financieras

Las redes neuronales artificiales se han convertido en una herramienta ampliamente usada para el pronóstico de series de tiempo. Su uso se ha extendido a varias disciplinas del conocimiento como son la ingeniería, las ciencias sociales, la economía y las finanzas con un creciente nivel de éxito en pronóstico ([Zhang et al., 1998](#)). En particular, su uso en pronóstico de volatilidad de series de tiempo financieras evidencia mayor precisión en el pronóstico en comparación con los pronósticos obtenidos de los modelos de volatilidad econométricos convencionales ([Donaldson y Kamstra, 1996](#); [Hyup Roh, 2007](#)).

Los modelos econométricos fueron los primeros en intentar modelar las características complejas inherentes a las series de tiempo financieras, así mismo se destacan algunos modelos de espacio de estados y de cambio de régimen que se han usado para el pronóstico de la volatilidad condicional [Frühwirth-Schnatter \(2006\)](#); una revisión completa se puede encontrar en [Poon y Granger \(2003\)](#). Seguidamente las redes neuronales comenzaron a mostrar buenos resultados en pronóstico debido a que satisfacen la propiedad de aproximación universal ([Donaldson y Kamstra, 1996](#); [Dunis y Huang, 2002](#); [Gonzalez y Burgess, 1997](#); [Ormoneit y Neuneier, 1996](#)). Sin embargo, la mayoría de los modelos fallan al intentar dar cuenta, simultáneamente, de los hechos estilizados observados en series de tiempo financieras, como por ejemplo: la alta curtosis o distribución de colas pesadas, la alta persistencia de las innovaciones cuadradas o memoria larga, Los agrupamientos de volatilidad, la respuesta asimétrica de los rendimientos ante innovaciones positivas y negativas y las relaciones no lineales presentes en los datos ([Engle y Bollerslev, 1986](#); [Franses y Dijk, 2000](#); [Rabemananjara y Zakoian, 1993](#); [Rapach y Strauss, 2008](#)).

Los modelos no lineales se han usado para intentar describir mejor las características de las series financieras y por ello las redes neuronales han recibido gran atención por parte de los investigadores, ya que estas pueden describir mejor estas no linealidades, gracias a su propiedad de

aproximación universal y a los resultados exitosos obtenidos en pronóstico (Zhang et al., 1998). En este sentido, se considera que el trabajo en esta dirección es alentador y motiva a investigar nuevos modelos basados en redes neuronales que mejoren la precisión del pronóstico de la volatilidad. No obstante, existen problemas asociados al proceso de especificación correcta del modelo (Zhang et al., 1998), puesto que en general los investigadores han recurrido a su experiencia y a juicios más subjetivos al momento de decidir sobre parámetros libres como por ejemplo la arquitectura del modelo o el método de entrenamiento.

### 1.1.2 Metodología de investigación

Se aplicó la metodología de Revisión Sistemática de la Literatura (RSL) desarrollada por Kitchenham et al. (2009); adicionalmente las preguntas de investigación y los criterios de calidad aplicados están basados en la revisión sistemática de la literatura realizada por Cogollo y Velásquez (2014). Esta estrategia de búsqueda permite efectuar un análisis de la literatura disponible más relevante. Para ello se plantearon unas preguntas de investigación que se esperan resolver con la revisión.

- Preguntas de investigación

Las preguntas de investigación planteadas en este estudio son:

- RQ1 ¿Cuáles son los modelos que se han usado para el pronóstico de la volatilidad que involucran el uso de redes neuronales artificiales?
- RQ2 ¿Qué modelos capturan los hechos estilizados de la volatilidad presentes en los datos?
- RQ3 ¿Cómo se obtienen los pronósticos de volatilidad condicional de un modelo propuesto?
- RQ4 ¿Cómo se realiza la evaluación del pronóstico de volatilidad en las diferentes metodologías?
- RQ5 ¿Cuáles son las principales dificultades pendientes por resolver en las diferentes metodologías de pronóstico de volatilidad condicional?

Para dar respuesta a las preguntas de investigación planteadas y cumplir con el objetivo principal de esta sección se implementó una metodología de RSL en la cual se hace un análisis de la literatura disponible más relevante.

Entre los aspectos que se destacan en esta investigación es que no hay en la literatura una revisión hecha en torno a la temática de pronóstico de volatilidad de series financieras que incluya redes neuronales. Adicionalmente, se busca identificar nuevos problemas de investigación a partir de la literatura publicada.

- Proceso de búsqueda

El proceso de búsqueda se hizo usando la herramienta bibliográfica SCOPUS, la cual es la base de datos de referencias más amplia disponible actualmente. Se buscan los documentos

que presentan la metodología de pronóstico de volatilidad y que incluyen una metodología de redes neuronales publicados en revistas seriadadas entre 1990 y 2013.

La ecuación de búsqueda aplicada para responder las preguntas de investigación fue: TITLE-ABS-KEY(volatility AND (forecasting OR forecast) AND neural networks) AND PUBYEAR > 1989.

Adicionalmente, se hace un seguimiento a las referencias bibliográficas y se identifican algunos documentos que fueron incluidos en la revisión.

- **Criterios de Inclusión y Exclusión**

Se define como criterio de inclusión que los artículos revisados tuvieran las temáticas:

1. Pronóstico de volatilidad de series de activos financieros (rendimientos o tasas de cambio).
2. Uso de redes neuronales.

Adicionalmente, se aplican los criterios de exclusión:

1. Los artículos que consideran pronóstico de volatilidad de otros activos como petróleo, electricidad.
2. Los artículos que hacen pronóstico de precios de activos financieros.

- **Evaluación de la Calidad**

Cada artículo seleccionado se evalúa utilizando las siguientes preguntas:

- QA1 ¿Es explícita la formulación matemática del modelo?
- QA2 ¿Es definido el proceso de estimación de parámetros para el nuevo modelo?
- QA3 ¿El estudio presenta un método para determinar la complejidad apropiada (número de capas o neuronas) del modelo?
- QA4 ¿Hay una evaluación o diagnóstico para el modelo?
- QA5 ¿Se examina la factibilidad de la metodología propuesta por medio de una aplicación a un caso real?
- QA6 ¿Se especifica un procedimiento para el entrenamiento del modelo (transformación de datos, valores iniciales de los parámetros, criterio de parada, etc.)?

### **1.1.3 Resultados obtenidos**

A continuación se describen los resultados obtenidos en esta investigación.

- **Resultados de la búsqueda**

Mediante la búsqueda aplicando la ecuación en la base de referencias bibliográficas de SCOPUS se obtuvieron un total de 94 documentos a los cuales se les aplicó manualmente los

criterios de inclusión y exclusión en una revisión del título, resumen, palabras clave y conclusiones. Finalmente, se seleccionaron 27 artículos cuyas referencias aparecen en la Tabla **1-1**.

- Evaluación de la Calidad

Se evaluaron los 27 documentos seleccionados usando los criterios definidos arriba. Los resultados obtenidos están en la Tabla **1-2**. Las columnas de la Tabla **1-2** muestran el grado de cumplimiento de cada artículo de los criterios de calidad definidos por las preguntas Q1 a Q6. Los resultados muestran que todos los artículos tienen puntuación mayor a 2, y solo dos artículos cumplieron todos los criterios y alcanzaron la puntuación máxima de 6. Los resultados del análisis de calidad muestran que el puntaje promedio de los artículos es de 4.3, aproximadamente el 62% de los artículos seleccionados contiene una estrategia adecuada para la selección de un modelo apropiado de pronóstico de volatilidad vía redes neuronales artificiales.

- Factores de la Calidad

En los artículos seleccionados se encontró que para el pronóstico de volatilidad se usan las RNA (Redes Neuronales Artificiales) de manera distinta: modelo de RNA como combinador no lineal de pronósticos, modelos basados en RNA y modelos híbridos de RNA con GARCH y EWMA (medias móviles exponencialmente ponderadas).

Se analizó la relación entre el puntaje de cada artículo y ambos la fecha de publicación y el número de citas recibidas por año. El coeficiente de correlación entre el puntaje obtenido y la razón citas por año de los estudios es 0.089 (valor- $p = 0.6603$ ), lo que indica que la calidad de la especificación del modelo no tiene influencia sobre el número de citas recibidas por año.

En la Tabla **1-3** se muestra que el número de artículos publicados por año ha sido estable, aunque con un incremento en los últimos años. El promedio del puntaje total de calidad parece haber ido disminuyendo. El coeficiente de correlación de Spearman entre año y puntaje fue de -0.2349 ( $p < 0.238$ ), lo cual nos indica que la calidad de los artículos es estadísticamente estable (no varía significativamente).

### 1.1.4 Discusión

En esta sección se responden las preguntas de investigación planteadas.

RQ1 ¿Qué modelos se usan para el pronóstico de la volatilidad y cómo involucran el uso de redes neuronales artificiales?

En general se destaca el uso de los modelos de la familia GARCH como los modelos base para hacer el pronóstico de volatilidad condicional; adicionalmente, modelos como caminata aleatoria, MAV, EWMA y máquinas de vectores de soporte también son considerados para hacer pronóstico de volatilidad por los autores de los diferentes artículos. Seis artículos

Tabla 1-1 Descripción de los documentos seleccionados

ID	Referencia	Modelo propuesto	# Citaciones
D1	<a href="#">Donaldson y Kamstra (1996)</a>	RNA como combinador no lineal de pronósticos	85
D2	<a href="#">Malliaris y Salchenberger (1996)</a>	RNA como modelo de pronóstico	21
D3	<a href="#">Donaldson y Kamstra (1997)</a>	Modelo basado en RNA	61
D4	<a href="#">Harrald y Kamstra (1997)</a>	RNA como combinador no lineal de pronósticos	32
D5	<a href="#">Brooks (1998)</a>	RNA como modelo de pronóstico	64
D6	<a href="#">Donaldson y Kamstra (1999)</a>	RNA como combinador no lineal de pronósticos	13
D7	<a href="#">Hu y Tsoukalas (1999)</a>	RNA como combinador no lineal de pronósticos	23
D8	<a href="#">Schittenkopf et al. (2000)</a>	Modelo basado en RNA	29
D9	<a href="#">Refenes y Holt (2001)</a>	Modelo basado en RNA	10
D10	<a href="#">Dunis y Huang (2002)</a>	Modelo basado en RNA	32
D11	<a href="#">Gavrishchaka y Ganguli (2003)</a>	Modelo basado en RNA	28
D12	<a href="#">Hamid y Iqbal (2004)</a>	RNA como modelo de pronóstico	46
D13	<a href="#">Slim (2004)</a>	RNA como modelo de pronóstico	2
D14	<a href="#">Hyup Roh (2007)</a>	Modelo basado en RNA	50
D15	<a href="#">Li et al. (2008)</a>	Modelo basado en RNA y RNA como combinador no lineal de pronósticos	7
D16	<a href="#">Tsangari (2007)</a>	Modelo basado en RNA	3
D17	<a href="#">Wang (2008)</a>	Modelo basado en RNA	2
D18	<a href="#">Bildirici y Ersin (2009)</a>	RNA como modelo de pronóstico	7
D19	<a href="#">Li et al. (2009)</a>	Modelo basado en RNA	5
D20	<a href="#">Marcek et al. (2009)</a>	RNA como modelo de pronóstico	3
D21	<a href="#">Sun y Xie (2009)</a>	RNA como modelo de pronóstico	0
D22	<a href="#">Chen et al. (2010)</a>	Modelo basado en RNA	9
D23	<a href="#">Dhamija y Bhalla (2010)</a>	RNA como modelo de pronóstico	5
D24	<a href="#">McAleer y Medeiros (2011)</a>	RNA como modelo de pronóstico	1
D25	<a href="#">Hajizadeh et al. (2012)</a>	RNA como modelo de pronóstico	4
D26	<a href="#">Sermpinis et al. (2013)</a>	RNA como modelo de pronóstico	1
D27	<a href="#">Orozco y Velásquez (2013)</a>	RNA como combinador no lineal de pronósticos	0

**Tabla 1-2** Evaluación de la calidad de los documentos.

ID	Referencia	Q1	Q2	Q3	Q4	Q5	Q6	Puntaje
D1	Donaldson y Kamstra (1996)	S	P	S	S	S	S	5.5
D2	Malliaris y Salchenberger (1996)	N	S	P	S	S	P	4.0
D3	Donaldson y Kamstra (1997)	S	S	N	S	S	S	5.0
D4	Harrald y Kamstra (1997)	S	S	N	S	S	S	5.0
D5	Brooks (1998)	S	P	N	S	S	N	3.5
D6	Donaldson y Kamstra (1996)	S	P	S	S	S	P	5.0
D7	Hu y Tsoukalas (1999)	S	S	N	S	S	S	5.0
D8	Schittenkopf et al. (2000)	S	P	P	S	S	P	4.5
D9	Refenes y Holt (2001)	N	S	P	S	S	S	4.5
D10	Dunis y Huang (2002)	S	S	P	S	S	S	5.5
D11	Gavrishchaka y Ganguli (2003)	S	P	P	P	S	N	3.5
D12	Hamid y Iqbal (2004)	N	P	N	S	S	P	3.0
D13	Slim (2004)	S	S	N	S	S	N	4.0
D14	Hyup Roh (2007)	S	N	N	S	S	N	3.0
D15	Li et al. (2008)	S	P	P	S	S	N	4.0
D16	Tsangari (2007)	S	S	S	S	S	S	6.0
D17	Wang (2008)	N	S	N	S	S	N	3.0
D18	Bildirici y Ersin (2009)	S	S	N	S	S	S	5.0
D19	Li et al. (2009)	N	P	P	S	S	P	3.5
D20	Marcek et al. (2009)	P	S	N	S	S	N	3.5
D21	Sun y Xie (2009)	S	N	P	S	S	N	3.5
D22	Chen et al. (2010)	S	S	N	S	S	S	5.0
D23	Dhamija y Bhalla (2010)	S	S	P	S	S	S	5.5
D24	McAleer y Medeiros (2011)	S	S	S	S	S	S	6.0
D25	Hajizadeh et al. (2012)	N	P	N	S	S	N	2.5
D26	Sermpinis et al. (2013)	P	P	N	S	S	N	3.0
D27	Orozco y Velásquez (2013)	S	S	N	S	S	P	4.5

**Tabla 1-3** Puntajes de calidad de los estudios por fecha de publicación.

	Años			
	1996-1999	2000-2003	2004-2007	2008-2013
<b>Número de estudios</b>	7	4	5	11
<b>Promedio puntaje total de calidad</b>	4.7	4.5	4.0	4.1
<b>Desviación estándar puntaje total de calidad</b>	0.699	0.816	1.225	1.158

presentan una red neuronal como metodología de combinación de pronósticos obtenidos de diferentes modelos (Donaldson y Kamstra, 1999, 1996; Harrald y Kamstra, 1997; Hu y Tsoukalas, 1999; Li et al., 2008; Orozco y Velásquez, 2013). Once artículos usan las RNA como modelos de pronósticos (Bildirici y Ersin, 2009; Brooks, 1998; Dhamija y Bhalla, 2010; Hajizadeh et al., 2012; Hamid y Iqbal, 2004; Malliaris y Salchenberger, 1996; Marcek et al., 2009; McAleer y Medeiros, 2011; Sermpinis et al., 2013; Slim, 2004; Sun y Xie, 2009). Once artículos usan modelos híbridos basados en RNA con modelos de la familia GARCH y modelos EWMA, (Chen et al., 2010; Donaldson y Kamstra, 1997; Dunis y Huang, 2002; Gavrishchaka y Ganguli, 2003; Hyup Roh, 2007; Li et al., 2009, 2008; Refenes y Holt, 2001; Schittenkopf et al., 2000; Tsangari, 2007; Wang, 2008). En el artículo de Tsangari (2007), se implementa un método no lineal y no paramétrico basado en kernel para combinar pronósticos provenientes de modelos GARCH y RNA.

RQ2 ¿Qué hechos estilizados son capturados por los modelos propuestos?

Los diferentes modelos propuestos en la literatura han sido parametrizados para dar cuenta de algunos de los hechos estilizados presentes en los datos. Por ejemplo, los modelos GARCH tienen la capacidad para reproducir series con alta curtosis y decaimiento lento en la función de autocorrelación de las observaciones cuadradas (Baillie et al., 1996; Bollerslev, 1986; Teräsvirta, 1996). Los modelos IGARCH y FIGARCH permiten reproducir series con varianza persistente Baillie et al. (1996). Los modelos EGARCH permiten responder asimétricamente a valores positivos y negativos de los rezagos de las innovaciones, así como los modelos TGARCH o GJR-GARCH tienen una respuesta asimétrica dependiente del tamaño y el signo de las innovaciones (Nelson, 1991; Rabemananjara y Zakoian, 1993). No obstante, estos modelos no capturan toda la dinámica de proceso generador de los datos. Alternativamente, los modelos basados en RNA o modelos híbridos presentan mayor capacidad de generalización y pueden detectar las relaciones no lineales que en algunos casos de aplicaciones empíricas se refleja en ajustes superiores a los de los modelos convencionales (Donaldson y Kamstra, 1997; Harrald y Kamstra, 1997; Kuan y White, 1994).

RQ3 ¿Cómo se obtienen los pronósticos de volatilidad condicional de un modelo propuesto?

Más del 75% de los artículos realizan pronóstico un paso adelante para los datos fuera de la muestra con diferentes ventanas de pronóstico. Algunos de estos efectúan un proceso recursivo donde reestiman los parámetros del modelo cada vez que se realiza un pronóstico, actualizando el conjunto de datos de entrenamiento con una observación cada vez. En algunos casos se realiza pronóstico un paso adelante mediante validación cruzada.

Aproximadamente un 20% de los artículos revisados hacen pronóstico para más de dos pasos adelante.

RQ4 ¿Cómo se realiza la evaluación del pronóstico de volatilidad en las diferentes metodologías?

Los estadísticos de evaluación usados son: Error Absoluto Medio y su raíz cuadrada (MAE y RMAE), Error cuadrático Medio y su Raíz cuadrada (MSE, RMSE), Error Porcentual Ab-

soluta Medio (MAPE), el estadístico Theil-U, CDC (Dirección Correcta de Cambio), Error Absoluto Medio Normalizado (NMAE) y Error Cuadrático Medio Normalizado (NMSE), DA (Indicador de Dirección), HR (Tasa de Aciertos que mide Dirección Correcta de Cambio). Los primeros criterios de evaluación los más frecuentes.

Adicionalmente, en tres artículos ([McAleer y Medeiros, 2011](#); [Orozco y Velásquez, 2013](#); [Sun y Xie, 2009](#)) se aplicó la prueba de precisión de pronósticos de [Diebold y Mariano \(1995\)](#), la cual sirve para hacer una evaluación de la calidad de los pronósticos obtenidos desde diferentes modelos. En tres artículos se aplicaron pruebas de dominancia de pronósticos ([Donaldson y Kamstra, 1997, 1996](#); [Harrald y Kamstra, 1997](#)), estas pruebas se realizan para determinar si una combinación de pronósticos reduce significativamente el error de pronóstico producido por los modelos con respecto a los errores de pronóstico individuales.

RQ5 ¿Cuáles son las principales dificultades pendientes por resolver en las diferentes metodologías de pronóstico de volatilidad condicional?

Entre las características más repetidas en los diferentes artículos se destaca la falta de criterios para la selección de la arquitectura de red neuronal adecuada para predecir la volatilidad de las series de tiempo financieras. Se requiere estudiar más el cómo seleccionar los parámetros óptimos de los modelos híbridos, cómo ganar información adicional de las series de tiempo para mejorar el desempeño del pronóstico. Se requiere más desarrollo teórico y práctico sobre los modelos híbridos ([Li et al., 2008](#)). Probar otras variables y otras arquitecturas de las redes para mejorar la calidad de los pronósticos ([Malliaris y Salchenberger, 1996](#)).

### 1.1.5 Conclusiones

De la revisión hecha para el pronóstico de la volatilidad de series de tiempo financieras cuando se usan redes neuronales artificiales se puede concluir que:

- En los 27 artículos seleccionados se encontró que en el proceso de construcción de un modelo de pronóstico basado en redes neuronales artificiales no se sigue un procedimiento estadístico formal de especificación del modelo. En parte, debido a la ausencia de criterios para la selección de una arquitectura de red neuronal, lo cual ha sido un problema común a muchas investigaciones.
- Los modelos de la familia GARCH presentan mayor frecuencia de uso y en general se consideran como el modelo base de comparación cuando se implementa otro tipo de modelos para predecir la volatilidad.
- Las redes neuronales se comportan bien en relación a los modelos econométricos cuando son usadas como modelo de pronóstico de volatilidad. Si se usan como modelo de combi-

nación no lineal de pronósticos en algunos casos se reporta mayor precisión con respecto a combinaciones lineales de pronósticos.

- En el pronóstico de la volatilidad, el uso combinado de modelos de la familia GARCH y redes neuronales artificiales se complementan debido a que permitirían explotar individualmente sus ventajas. Los modelos GARCH por su parte representando los hechos estilizados de los rendimientos y las redes neuronales las no linealidades.

## 1.2 Pronóstico de la volatilidad usando combinación de pronósticos

Debido a la gran importancia del pronóstico de la volatilidad y a los múltiples esfuerzos hechos por académicos, profesionales, inversionistas, entre otros, para mejorar la precisión de su pronóstico, en esta sesión se realiza una revisión sistemática de la literatura para clasificar y valorar los estudios existentes en pronóstico de volatilidad mediante el esquema de ensamble de pronósticos o combinación de pronósticos. Más específicamente, aquellas investigaciones donde el modelo, experto o el método de combinación requieren técnicas de la inteligencia artificial.

### 1.2.1 Combinación de pronósticos

Cuando se tiene varios pronósticos provenientes de diferentes modelos, surge la alternativa de considerarlos simultáneamente para obtener un pronóstico único. Esta idea de combinar varios pronósticos individuales se justifica en diversos resultados empíricos que indican que la precisión del pronóstico combinado es con frecuencia mayor que la de cada pronóstico individual (Becker y Clements, 2008; Clemen y Winkler, 1986; Clemen, R., 1989; Hashem, 1997; Lux y Morales-Arias, 2010). Una definición apropiada para ensamble es dada por Zhang et al. (2011) quienes lo consideran como un grupo de modelos de aprendizaje cuyos pronósticos se juntan para dar un pronóstico final. La idea conceptual detrás de estos enfoques es la diversificación de los modelos de pronóstico. Se considera que modelos diferentes capturan información diferente del conjunto de datos y por ende la composición mezclará toda la información para así obtener un pronóstico compuesto de mejor calidad. Esta diversificación puede lograrse en diferentes formas, las principales son:

- Usando modelos diferentes.
- Usando el mismo tipo de modelo con configuración interna idéntica pero con valores de los parámetros diferentes.
- Usando diferentes subconjuntos de datos para entrenar cada modelo.

Un estudio a fondo del estado del arte de ensamble de pronósticos o metodología de combinación de pronósticos fue realizada por [Rodríguez et al. \(2012\)](#).

Hay muchas razones aceptadas para explicar por qué la combinación de pronósticos es efectiva:

- Primero, cada predicción puede ser obtenida usando diferentes fuentes de información o cada modelo se basa en diferentes supuestos sobre el tipo de relación en los datos.
- Segundo, cada modelo no está en la capacidad de identificar el proceso real de generación de los datos, pero captura información complementaria en relación con otros modelos alternativos ([Terui y van Dijk, 2002](#)).
- Tercero, Cuando hay nueva información disponible pueden ocurrir cambios fuertes en la estructura o en los parámetros del modelo; puede deberse a una especificación pobre del modelo o a la presencia de cambios estructurales en los modelos individuales ([Hendry y Clements, 2004](#); [Timmermann, 2006](#)).
- En el caso de un ensamble con una regresión, [Brown et al. \(2005\)](#) probaron que la precisión del ensamble sobre un conjunto de datos esta relacionado con la precisión y diversidad de los expertos individuales y que es posible controlar la precisión del ensamble a través de la adaptación de expertos individuales por precisión y diversidad.

Por su parte [Makridakis \(1989\)](#), explica los factores que contribuyen a realizar una combinación de pronósticos. El primer factor, es que los pronósticos individuales no entregan buenos resultados, esto se ve complementado con los valores de los errores. El segundo son los patrones inestables, o los cambios constantes que por lo general se presenta en una serie de tiempo, las cuales no pueden ser tenidas en cuenta en un pronóstico individual. Tercero, en la combinación de pronóstico se utilizan modelos que minimizan el error.

Cabe resaltar que combinar los pronósticos provenientes de un conjunto de modelos no necesariamente conlleva a una predicción con desempeño mejor que el del mejor modelo en la combinación. Sin embargo, sí disminuye la probabilidad escoger un pronóstico con desempeño pobre.

Si bien en la literatura de combinación de pronósticos se han usado múltiples técnicas de combinación de pronósticos, que van desde el uso del promedio aritmético simple (PAS) de pronósticos, las combinaciones lineales –estimando los pesos por el método de mínimos cuadrados ordinarios u otros- hasta las combinaciones no lineales de los pronósticos –como por ejemplo usando las RNA- aún carecemos de fundamentos teóricos que permitan decidir qué metodología de combinación funciona mejor. Adicionalmente, a pesar de que el Promedio Aritmético Simple (PAS) podría funcionar bien no es posible recomendar su uso en general, porque existen casos en donde el PAS falla ([Wei y Yang, 2012](#)). Algunos autores como por ejemplo ([Clemen, R., 1989](#)) argumentan que la combinación de pronósticos usando el PAS funciona mejor que otras técnicas de combinación, pero en la práctica no se mantiene esta afirmación para todos los casos de aplicación. Sin embargo, los resultados empíricos de combinación de pronósticos que implementan combinaciones no lineales muestran tener mejor desempeño que el PAS ver por ejemplo [Tsangari \(2007\)](#), [Donaldson y Kamstra \(1996\)](#), [Coshall \(2009\)](#), [Preminger et al. \(2006\)](#).

## 1.2.2 Metodología de investigación

En esta sesión se realiza una revisión sistemática de la literatura siguiendo metodologías aplicadas en otras disciplinas como la medicina, economía y ciencias sociales (Kitchenham et al., 2009, 2010) para clasificar y valorar los estudios existentes en pronóstico de volatilidad mediante el esquema de ensamble de pronósticos o combinación de pronósticos.

- Preguntas de investigación

Las preguntas de investigación planteadas en este estudio son:

RQ1 ¿Qué técnicas de inteligencia artificial se usan para obtener pronósticos individuales de volatilidad, en un ensamble de pronósticos?

RQ2 ¿Cuáles son las metodologías de combinación más usadas para obtener pronósticos compuestos de la volatilidad?

RQ3 ¿Cómo se ha obtenido la diversificación de los expertos para el pronóstico de volatilidad en el ensamble?

RQ4 ¿Cuáles son los casos de aplicación reportados sobre la combinación de pronósticos para volatilidad?

RQ5 ¿Cómo se evalúa el pronóstico de la volatilidad en los ensambles de pronósticos?

- Bases de datos utilizadas

Para extraer la información y responder las preguntas de investigación se usaron las herramientas bibliográficas Scopus, Jstor y IEEE con el fin de encontrar la mayor cantidad de artículos que están relacionados con el pronóstico de volatilidad bajo la metodología de ensambles. Se limitó la búsqueda a las áreas de Ciencias computacionales, Matemáticas, Administración de negocios y Contabilidad, Ingeniería y Economía, econometría y finanzas. Y publicaciones en revistas y conferencias a octubre de 2011. Adicionalmente, se hizo un seguimiento manual a las referencias de los artículos para complementar la búsqueda y validar los resultados de la revisión.

- Proceso de búsqueda

Los criterios de búsqueda que se implementaron considerando el título, el resumen y las palabras clave de los artículos y las conferencias publicados a octubre de 2011:

1. (“Composite forecast” OR “forecast combination” OR “ensemble”) AND volatility.
2. (ensemble OR “neural networks”) AND volatility AND forecast.

- Criterios de inclusión y exclusión

Los criterios de inclusión tenidos en cuenta son:

1. Las investigaciones en las cuales se usó el método de combinación de pronósticos para predecir la volatilidad de series económicas y/o financieras.

2. Los artículos donde el modelo, experto o el método de combinación requieren técnicas de la inteligencia artificial.

Los criterios de exclusión que se consideran son:

1. Artículos que consideran series de tiempo que no son financieras, como por ejemplo demanda de turismo, temperatura, número de pasajeros, entre otras.
2. Artículos que no presentan la técnica de ensamble o combinación de pronósticos para el pronóstico de la volatilidad.

### 1.2.3 Resultados obtenidos

Se obtuvieron después del refinamiento automático de las búsquedas un total de 93 resultados. Realizando una revisión manual y aplicando los criterios de inclusión y exclusión se obtuvieron 20 artículos. La Tabla 1-4 muestra los autores de los documentos obtenidos en la revisión con su respectivo título, tipo y año de publicación y número de citas. Se observa que el tema es actual y no ha sido abordado desde el enfoque que se propone en este trabajo.

### 1.2.4 Discusión

En esta sección se presentan las respuestas obtenidas para las preguntas de investigación planteadas en la revisión sistemática de literatura desarrollada en este aparte.

RQ1 ¿Qué técnicas de inteligencia artificial se usan para obtener pronósticos individuales de volatilidad, en un ensamble de pronósticos?

En muchos de los trabajos revisados se encontró el uso de redes neuronales como modelo de pronóstico de volatilidad. Por ejemplo, [Tsangari \(2007\)](#) compara el desempeño de modelos de redes neuronales (perceptrón de una capa oculta) con el modelo GARCH para el pronóstico de volatilidad de tasas de cambio y muestra que las RN superan el modelo GARCH. Adicionalmente, propone una metodología de combinación de estos dos modelos que resulta exitosa para el pronóstico de la volatilidad de tasas de cambio de distintos países. [Donaldson y Kamstra \(1999\)](#) usan los modelos de RNA para capturar los efectos de interacción que inactivan un pronóstico a favor de otro. [Liu y Liu \(2006\)](#) comparan cuatro modelos de pronóstico de la volatilidad, uno de los cuales es un Perceptrón de una capa oculta.

RQ2 ¿Cuáles son las metodologías de combinación más usadas para obtener pronósticos compuestos de la volatilidad?

En general, se emplean el promedio aritmético simple, combinaciones lineales de pronósticos y otras combinaciones no lineales de pronósticos de volatilidad mencionados a continuación:

**Tabla 1-4** Tipo de publicación y número de citas en combinación de pronósticos con RNA para volatilidad.

<b>Autor</b>	<b>Título del documento</b>	<b>Tipo de publicación</b>	<b># citas</b>
Donaldson y Kamstra (1996)	Forecast combining with neural networks	Artículo	69
Hansen y Lunde (2005)	A realized variance for the whole day based on intermittent high-frequency data	Artículo	31
Harrald y Kamstra (1997)	Evolving artificial neural networks to combine financial forecasts	Artículo	25
Hu y Tsoukalas (1999)	Combining conditional volatility forecasts using neural networks: An application to the EMS exchange rates	Artículo	19
Chang y Tsai (2008)	Forecast approach using neural network adaptation to support vector regression grey model and generalized auto-regressive conditional heteroscedasticity	Artículo	18
Rapach y Strauss (2008)	Structural breaks and GARCH models of exchange rate volatility	Artículo	11
Donaldson y Kamstra (1999)	Neural network forecast combining with interaction effects	Artículo	9
Wang (2008)	Using neural network to forecast stock index option price: A new hybrid GARCH approach	Artículo	4
Patton y Sheppard (2009)	Optimal combinations of realised volatility estimators	Artículo	4
Fuertes et al. (2009)	On forecasting daily stock volatility: The role of intraday information and market conditions	Artículo	4
Becker y Clements (2008)	Are combination forecasts of S&P 500 volatility statistically superior?	Artículo	4
Amendola y Storti (2008)	A GMM procedure for combining volatility forecasts	Artículo	2

Continuación Tabla 1-4.

Autor	Título del documento	Tipo de publicación	# citaciones
Liu y Liu (2006)	Currency options volatility forecasting with shift-invariant wavelet transform and neural networks	Conferencia	1
Kisinbay (2010)	Predictive ability of asymmetric volatility models at medium-term horizons	Artículo	1
Lux y Morales-Arias (2010)	Forecasting volatility under fractality, regime-switching, long memory and student-t innovations	Artículo	1
Preminger et al. (2007)	The extended switching regression model: Allowing for multiple latent state variables	Artículo	1
Wenbing et al. (2008)	Forecasting Chinese stock markets volatility based on neural network combining	Conferencia	0
Tsangari (2007)	An alternative methodology for combining different forecasting models	Artículo	0
Dong (2008)	Application of an EWMA combining technique to the prediction of stock market volatility	Conferencia	0
Preminger et al. (2006)	Extended switching regression models with time-varying probabilities for combining forecasts	Artículo	0

- [Amendola y Storti \(2008\)](#) proponen el uso del método de los momentos generalizados para estimar los pesos de una combinación lineal de pronósticos de la volatilidad.
- [Becker y Clements \(2008\)](#) usan dos estrategias de combinación: promedio de pronósticos (ponderaciones iguales a cada pronóstico) y el método propuesto que consiste en una combinación como una regresión donde las ponderaciones deben ser estimadas.
- [Hu y Tsoukalas \(1999\)](#) emplean tres estrategias de combinación: promedio aritmético simple, combinación usando el método de mínimos cuadrados ordinarios para estimar los pesos y la tercera técnica usada es RNA.
- [Harrald y Kamstra \(1997\)](#) implementan los métodos de combinación de pronósticos tradicionales, el método de Kernel el cual es no paramétrico y la combinación usando redes neuronales artificiales y emplea técnicas de optimización evolutiva tales como AG y EP para el entrenamiento de RNA.
- [Wenbing et al. \(2008\)](#) usan un perceptrón de una capa oculta como método no lineal de combinación de pronósticos.
- [Tsangari \(2007\)](#) presenta un nuevo método no lineal, no paramétrico, basado en el método del kernel, donde la forma funcional de la combinación no requiere ser conocida, el método propuesto es más preciso en el pronóstico de la volatilidad de las tasas de cambio consideradas en comparación con el pronóstico obtenido por el modelo GARCH y un modelo de redes neuronales implementado.

RQ3 ¿Cómo se logra la diversificación de los expertos para el pronóstico de volatilidad en el ensemble?

En general se observa que la diversificación en el ensemble se logra al considerar diferentes modelos de volatilidad condicional como por ejemplo en [Hu y Tsoukalas \(1999\)](#), usan los modelos de la familia GARCH (GARCH, IGARCH, EGARCH y MAV) y los combinan por medio de métodos paramétricos (mínimos cuadrados ordinarios) y no paramétricos (RNA) para obtener un pronóstico de la volatilidad condicional, en este caso la diversificación se hizo usando expertos diferentes. [Wenbing et al. \(2008\)](#) Combina cuatro pronósticos de volatilidad de los modelos GARCH, EGARCH, TGARCH y MAV por tres métodos específicos de combinación: promedio aritmético simple, combinación lineal usando el método de mínimos cuadrados para la estimación de los parámetros y combinación por medio de una red neuronal artificial. [Harrald y Kamstra \(1997\)](#) combinan pronósticos producidos por dos modelos de volatilidad el MAV y el GARCH. También, considerando un modelo de volatilidad condicional de la familia GARCH y una red neuronal ([Tsangari, 2007](#)).

RQ4 ¿Cuáles son las aplicaciones de combinación de pronósticos para volatilidad?

Entre las aplicaciones más recurrentes se encontró el pronóstico de la volatilidad de los

rendimientos diarios y mensuales, la volatilidad de los precios de opciones financieras, la volatilidad de opciones sobre divisas, la volatilidad de las tasas de cambio y volatilidad de las tasas de cambio del Sistema Monetario Europeo (ver bibliográfica comentada).

RQ5 ¿Cómo se evalúa el pronóstico de la volatilidad?

Los criterios de evaluación que más se usan son: MAE, MAPE, MSE, RMSE, U-Theil, CDC y MSE normalizado. En general, en todas las investigaciones revisadas es común encontrar más de un criterio de evaluación para mostrar robustez de los resultados.

### 1.2.5 Conclusiones

Las conclusiones que se derivan de la revisión sistemática de la literatura en pronóstico de volatilidad usando combinaciones de pronósticos son:

- En los estudios que implementan una red neuronal artificial como experto para el pronóstico de la volatilidad, en general éste tiene mejor desempeño que los métodos de combinación convencionales. Igualmente, cuando se usa un modelo de RNA para la combinación de los pronósticos se observa, con respecto a la mayoría de criterios de evaluación, mayor precisión en el pronóstico de la volatilidad comparada con los modelos individuales y con otros métodos de combinación.
- En la mayoría de los artículos revisados la diversificación de los expertos en los ensambles se da al usar modelos de volatilidad diferentes.
- Se usan los mismos criterios de evaluación sin tener en cuenta la estructura de los expertos usados en el ensamble ni la forma del combinador. Esto es, la selección de la función de error se hace por las propiedades estadísticas y no por preferencia al modelo.
- En los artículos explorados no hay evidencia que justifique que los modelos de combinación de pronósticos o los ensambles implementados capturen alguna o todas las características de la volatilidad.

## 1.3 Preguntas emergentes de investigación

De las revisiones de literatura realizadas y las cuales han sido reportadas al inicio de este capítulo, se puede afirmar que no se ha explorado el uso de un mismo experto o modelo con diferentes valores de los parámetros para una combinación de pronósticos. En esta tesis se estudia el ensamble de pronósticos, propuesto desde la óptica de diversificación obtenida por el uso de diferentes funciones de pérdida como criterios de optimización de los modelos individuales. Algunos casos de aplicaciones empíricas sugieren que la combinación de pronósticos de modelos que han sido contruidos basados en un elemento de diversidad supera los pronósticos de modelos individuales, pero no se tiene una explicación clara de por qué en algunos casos funciona bien. Así que la primera

pregunta de esta tesis es “¿Por qué y como los criterios de optimización generan diversidad en los conjuntos de pronóstico?”. La diversidad de pronósticos puede ser examinada con varias medidas, una que ha sido ampliamente usada es la correlación entre los errores de pronóstico. No obstante, muchas veces ésta no necesariamente da información de qué tan diversos son los conjuntos de pronósticos. La segunda pregunta de esta tesis es entonces “¿Por qué la correlación de los errores de pronósticos de volatilidad no necesariamente es una medida de diversificación y que otra medida puede ser propuesta?”. En consecuencia, una tercera pregunta de investigación es “¿Por qué la combinación de pronósticos de volatilidad tiende a ser más precisa que los pronósticos de modelos de volatilidad individuales?”. Para combinaciones lineales de pronósticos, se cuenta con una prueba de hipótesis de dominancia de pronósticos, en inglés *forecasts encompassing*; Sin embargo, para combinaciones no lineales como por ejemplo con RNA apenas se define un contraste de hipótesis por extensión del anterior. En este caso es deseable determinar si esta prueba es confiable en orden de tomar la decisión de si se combinan varios conjuntos de pronósticos disponibles; es así como la cuarta pregunta de esta tesis es “¿Qué tan confiable es la prueba de Wald para dominancia de pronósticos en aras de decidir si un conjunto de pronósticos debe incluirse en la combinación no lineal?”

## 1.4 Hipótesis de investigación

La hipótesis de investigación planteada en esta tesis es:

Una esquema de ensamble de modelos para el pronóstico de volatilidad de rendimientos financieros logrará predecir la volatilidad con más precisión que los pronósticos obtenidos con los modelos individuales. La diversificación de los modelos en el ensamble se logra mediante el uso de diferentes funciones de error para la estimación de los parámetros de los expertos. De esta forma, el uso de diferentes funciones de error permite que cada experto capture información o relaciones en los datos no incluidas en los demás expertos. La combinación de los pronósticos obtenidos de cada experto se lleva a cabo usando un modelo de redes neuronales artificiales, el cual identifica y combina la información relevante entregada por cada modelo individual; consecuentemente, es posible proponer una metodología de ensamble de pronósticos de volatilidad de rendimientos financieros. Esta metodología de ensamble de pronósticos de volatilidad permite mayor precisión en el pronóstico de la volatilidad que las técnicas convencionales de combinación de pronósticos.

## 1.5 Objetivos de la tesis

En esta sección se presentan el objetivo general de esta tesis y los objetivos específicos de la misma.

### 1.5.1 Objetivo general

Proponer un esquema para la predicción de la volatilidad de rendimientos financieros, basada en el uso de ensamble de expertos, en el cual, la combinación de los pronósticos individuales es realizada usando una red neuronal artificial y la obtención de los parámetros de cada experto se logra al aplicar diferentes funciones de error; el ensamble tiene como fin obtener pronósticos más precisos de la volatilidad que los obtenidos por cada experto considerado y que los calculados usando técnicas tradicionales de combinación de pronósticos.

### 1.5.2 Objetivos específicos

1. Seleccionar las familias de funciones de error que se usarán para el entrenamiento de los expertos en el ensamble, con el fin de lograr su diversificación.
2. Diseñar una metodología para especificar cada experto en el ensamble, que incorpora las familias de funciones de error seleccionadas en el objetivo anterior.
3. Diseñar una metodología para la combinación de pronósticos individuales generados por los expertos obtenidos en el objetivo anterior.
4. Evaluar, al menos, tres casos de aplicación con el fin de determinar si la metodología propuesta permite obtener resultados más precisos que los principales métodos reportados en la literatura.

## 1.6 Contribuciones de la tesis

La principal contribución de la tesis es un mejor entendimiento de como las funciones de error vistas como criterio de optimización puede producir modelos de pronóstico que tienen información complementaria de los datos para ser combinada. Un resumen de las principales contribuciones es:

- Una revisión de la literatura muestra que el uso de los modelos de combinación de pronósticos puede mejorar las predicciones y disminuir el riesgo en la toma de decisiones al considerar diferentes fuentes.
- Una revisión de la literatura que identifica los principales modelos que se han implementado para el pronóstico de la volatilidad de series de tiempo financieras.
- Una aplicación empírica del uso de una estructura de combinación de pronósticos que respalda la hipótesis de que un modelo de combinación de pronósticos con una red neuronal supera los modelos de pronóstico individuales.

- Una metodología de remuestreo en el proceso de entrenamiento de los modelos individuales para identificar la distribución de los parámetros de los modelos considerados, así como la distribución de las medidas de exactitud usadas para evaluar su desempeño. Esto se realiza debido a que para un conjunto de datos, estimar un modelo una sola vez produce un único conjunto de parámetros y con una sola muestra en cada modelo, no es suficiente para sacar conclusiones acerca de si los modelos son diferentes.
- El análisis de un contraste de hipótesis de dominancia de pronósticos (en inglés “*forecast encompassing*”) para determinar si un conjunto de pronósticos debe incluirse en un modelo de combinación. Se extiende la definición de dominancia a combinaciones no lineales y se propone un contraste de hipótesis para este caso.
- Una evidencia empírica de la posible dependencia del desempeño de los modelos de volatilidad con los datos. Cuando se presentan agrupamientos de altas volatilidades y qué tan altos son los valores extremos con respecto al grueso de los datos.
- Una primera aplicación empírica del contraste de hipótesis para cuatro conjuntos de datos y combinación de pronósticos con una red neuronal artificial.
- Una análisis sobre la correlación entre los errores de pronóstico de dos modelos sobre un mismo conjunto de datos sugiere que esta medida no debe usarse para decidir si se combinan los pronósticos de diferentes modelos, esto es, no es un buen indicador de diversidad entre los conjuntos de pronósticos. El hecho de que dos conjuntos de datos estén altamente correlacionados no indica que necesariamente estos sean iguales. Lo que se puede concluir es que los datos tienen la misma tendencia. En el caso de datos de volatilidad, un modelo puede capturar muy bien las bajas volatilidades y tener la misma tendencia de otro que captura muy bien las altas volatilidades.

## 1.7 Estructura de la tesis

Este documento está dividido en 5 capítulos. En el segundo capítulo, se encuentra una aplicación empírica de combinación de pronósticos con diferentes métodos de combinación: promedio aritmético simple, regresión lineal y redes neuronales artificiales. En el capítulo 3, una metodología de especificación de los modelos de pronóstico individuales encaminada a generar la diversidad en los conjuntos de pronóstico. Debido a que ese es uno de los puntos claves de esta tesis, se tiene como foco principal la optimización con diferentes funciones de error, con la cual se afirma y se muestra evidencia empírica de que ello conduce a la diversidad de pronósticos. En el capítulo 4 se presentan algunos resultados teóricos de combinación de pronósticos así como la prueba de Wald que es una extensión de un contraste ya conocido para combinaciones lineales en dominancia de pronósticos. Adicionalmente, una aplicación empírica del contraste de hipótesis para la dominancia de pronósticos en combinaciones no lineales de pronósticos. Finalmente, en el capítulo 5 se

concluyen los resultados de esta tesis mostrando el cumplimiento de los objetivos propuestos, las respuestas a las preguntas de investigación planteadas y las oportunidades de trabajo futuro.

## 2 Sistema de combinación de pronósticos para predecir la volatilidad

La volatilidad de los rendimientos es una medida estadística de dispersión del precio de un instrumento financiero y puede medirse como la varianza de los rendimientos del activo. En una serie de tiempo, se entiende como una medida de la magnitud de los cambios de una observación a la siguiente. En econometría financiera se usa para describir la variabilidad de la componente aleatoria de una serie de tiempo (Andersen et al., 2005). En este capítulo se estudian las características de la volatilidad y el objetivo principal es predecir mediante el uso de modelos de inteligencia computacional la volatilidad condicional de los rendimientos de un activo financiero, esto es, modelar la varianza condicional del error  $\varepsilon_t$ .

Predecir la volatilidad de los rendimientos de una serie de tiempo es un problema que merece ser investigado debido a la importancia del manejo del riesgo en el mercado financiero, donde se sabe que la volatilidad de los rendimientos es una variable estocástica que evoluciona en el tiempo.

En esta investigación el objetivo es presentar una metodología de combinación de pronósticos de volatilidad condicional basada en los siguientes puntos:

1. Reescribir el modelo GARCH en la forma de un ARMA, tales que, los parámetros del modelo sean estimados minimizando una función de la diferencia entre las innovaciones al cuadrado y la volatilidad, ver Tsay (2010).
2. El uso de diferentes criterios de optimización para la obtención de los parámetros de los modelos GARCH.
3. Cuando varios pronósticos de una misma variable están disponibles simultáneamente, es posible obtener un pronóstico compuesto usando técnicas de combinación de pronósticos. En este capítulo se consideran: el promedio aritmético simple, regresión lineal múltiple y redes neuronales, la cual es una técnica de inteligencia computacional.

La metodología definida por los puntos anteriores se plantea de cara a estudiar la combinación de pronósticos explorando diferentes criterios de optimización para estimar los modelos individuales. Este enfoque no ha sido estudiado antes como un elemento de diversificación y en la aplicación empírica presentada más adelante se nota que los modelos para el pronóstico de la volatilidad

---

condicional, ajustados con esta metodología capturan diferentes características de la serie de tiempo, dando cuenta de las no linealidades presentes en este tipo de series.

Para predecir la volatilidad, entre todas las técnicas usadas sobresalen dos enfoques cuando se considera un único modelo: el primer enfoque usa los modelos econométricos rigurosos cuyo objetivo es entender las propiedades dentro de la muestra y explicar el comportamiento histórico de la volatilidad. La mayoría de estos modelos están basados en el trabajo seminal de Engle (1982), quien introdujo el modelo autorregresivo de heterocedasticidad condicional, ARCH por su sigla en inglés. Por ejemplo, Bollerslev (1986) postuló el modelo generalizado ARCH, denominado GARCH, el cual resulta ser más parsimonioso que el modelo ARCH de Engle; otros modelos, como el GJR-GARCH de Glosten et al. (1993), el GARCH exponencial (EGARCH) de Nelson (1991), los A-GARCH, NA-GARCH y V-GARCH de Engle y Ng (1993), los IGARCH y FIGARCH de Engle y Bollerslev (1986) y Baillie et al. (1996), se desarrollaron para tener en cuenta otras características complejas que presentan las series de volatilidad como lo son la respuesta asimétrica ante shocks positivos y negativos, la distribución de colas pesadas de los rendimientos y la persistencia del proceso de volatilidad. El modelado de volatilidad es ampliamente abordado por Andersen et al. (2005).

En contraste, el segundo enfoque para el pronóstico de la volatilidad con un único modelo, está relacionado con pronósticos de otras metodologías, ver (Poon y Granger, 2003; Taylor, 2004; Verhoeven et al., 2002). Por ejemplo, Taylor (2004) desarrolla un modelo exponencial de transición suave en el cual los parámetros varían en el tiempo con el objeto de adaptarse a las características de la serie de tiempo, en este caso el modelo propuesto se aleja de los supuestos estadísticos que tienen los modelos econométricos. En el modelo de volatilidad estocástica, la volatilidad se modela como una variable latente cuya evolución se rige por un proceso estocástico autorregresivo, esto es, permite un elemento estocástico en el proceso de varianza condicional de la serie de tiempo. Las redes neuronales artificiales, en adelante RNA, también hacen parte de este enfoque ya que se han usado como modelos de pronóstico. Diferentes estudios empíricos con RNA han mostrado resultados exitosos en el modelado y pronóstico de la volatilidad. Kuan y White (1994) las proponen e introducen como herramientas para los economistas y econometristas en aplicaciones de modelado y pronóstico de series de tiempo. Sin embargo, aunque estas detectan las relaciones no lineales y proveen ajustes superiores a los de los modelos convencionales de series de tiempo lineales, debido a su flexibilidad pueden presentar problemas de sobreajuste a los datos; esto significa que un ajuste superior dentro de la muestra puede conllevar a un sobreajuste lo cual no es garantía de que estos modelos tengan buen desempeño en el pronóstico fuera de la muestra. Adicionalmente, si bien los parámetros en una RNA pueden describir la relación entre las variables y determinar la influencia relativa de la información que es procesada por los pesos no existe una interpretación clara de los mismos, la cual es una de las críticas más sobresalientes en este tipo de modelos.

En consecuencia, debido a que muchas veces los pronóstico individuales no logran capturar la dinámica completa de las series de tiempo, es posible considerar pronósticos de varios modelos independientes para obtener un único pronóstico de la variable objetivo. La idea conceptual de combinar simultáneamente varios pronósticos tiene sus ventajas y esta sustentada en la diversificación de los mismos. Algunas de las razones que justifican la combinación fueron presentadas en el Capítulo 1 de este documento. En lo que se refiere a la diversificación esta puede lograrse por varias vías: usando diferentes modelos o el mismo tipo de modelo con la misma configuración interna y valores diferentes de los parámetros. Esta diversificación en la combinación de pronósticos ayuda a reducir los sesgos de mala especificación de los modelos individuales.

En este capítulo se presentan el modelo GARCH, su parametrización y los criterios de optimización que pueden implementarse en el proceso de especificación del modelo. Igualmente, el modelo de combinación de pronósticos propuesto y un experimento numérico usando una serie de tiempo de referencia para ejemplificar y probar el desempeño de los modelos de volatilidad.

## 2.1 Metodología propuesta

### 2.1.1 Modelo GARCH

El modelo generalizado autorregresivo de heterocedasticidad condicional, GARCH, propuesto por [Bollerslev \(1986\)](#) es uno de los más usados para modelado y pronóstico de la volatilidad. La varianza actual  $\sigma_t^2$ , depende de las innovaciones pasadas al cuadrado y de las varianzas pasadas de la siguiente forma:

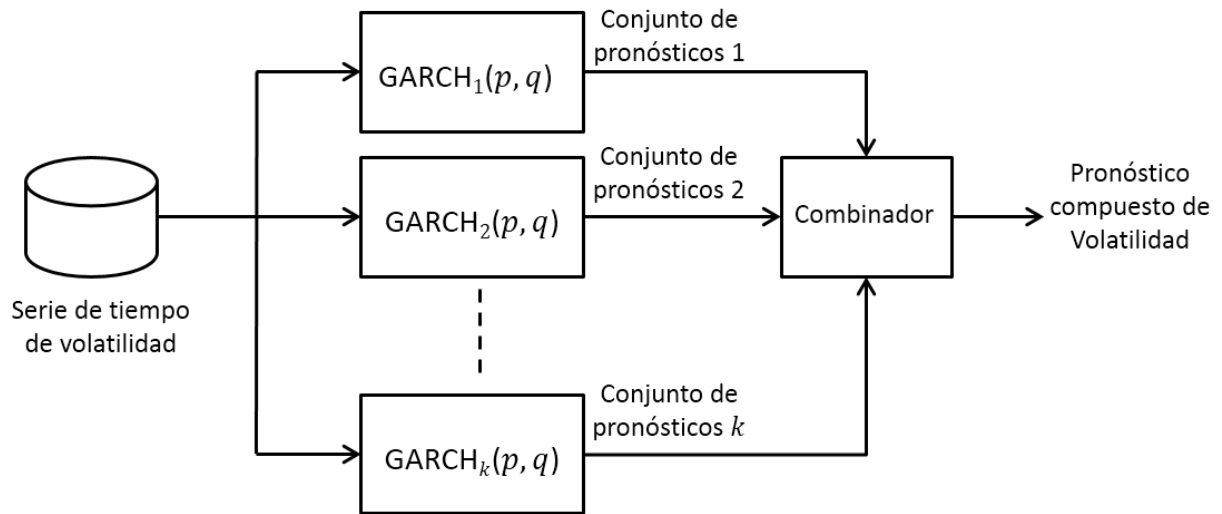
$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \cdot a_{t-i}^2 + \sum_{j=1}^q \beta_j \cdot \sigma_{t-j}^2 \quad (2-1)$$

Con  $r_t = \mu + a_t$ ,  $a_t = \sigma \cdot \varepsilon_t$  y  $\varepsilon_t \sim N(0, 1)$ , donde los parámetros  $\omega$ ,  $\alpha_i$  y  $\beta_j$ , están sujetos a las restricciones  $\omega > 0$ ,  $\alpha_i \geq 0$ ,  $\beta_j \geq 0$  y  $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_j) < 1$ .

El modelo GARCH(1,1) ha sido, en general, el más apropiado de los modelos estándar de la familia ARCH para datos de rendimientos y puede escribirse como:

$$r_t = \mu + a_t, a_t \sim N(0, \sigma_t^2) \text{ y } \sigma_t^2 = \omega + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (2-2)$$

Como se ha mencionado antes, el modelo GARCH es uno de los más usados para modelado y pronóstico de la volatilidad condicional. El modelo general cuya varianza actual,  $\sigma_t^2$ , es una función de las innovaciones cuadradas anteriores,  $\{a_{t-i}^2; i = 1, \dots, p\}$ , y las varianzas pasadas,  $\sigma_{t-j}^2; j = 1, \dots, q$  tiene un procedimiento clásico de estimación de parámetros bajo el supuesto de que los residuales normalizados  $\varepsilon_t$  siguen una distribución normal estándar, ver [Tsay \(2010\)](#). Los valores óptimos de los parámetros se calculan por lo tanto maximizando el logaritmo natural de la



**Figura 2-1** Estructura del sistema de pronóstico de volatilidad.

función de verosimilitud de los residuales:

$$\ln L = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^N \left[ \ln \sigma_t^2 + \frac{a_t^2}{\sigma_t^2} \right] \quad (2-3)$$

donde  $N$  es la longitud de la serie de tiempo.

Sin embargo, en la sección siguiente se presenta un método alternativo de estimación de los parámetros.

### 2.1.2 Estructura general

La estructura general del sistema propuesto se presenta en la Figura 2-1. Este sistema está compuesto de  $k$  modelos con la misma estructura,  $\text{GARCH}(p, q)$ , pero con diferentes valores de los parámetros. Los modelos  $\text{GARCH}(p, q)$  se estiman usando la misma muestra y diferentes criterios de optimización con el objetivo de capturar diferentes características de la dinámica de la volatilidad. Cada modelo produce un pronóstico individual de la volatilidad condicional para el tiempo  $t + 1$  con base en la información disponible hasta el tiempo  $t$ . Finalmente, todos los pronósticos individuales se combinan para obtener un pronóstico compuesto de la volatilidad.

### 2.1.3 Procedimiento de estimación alternativo de los parámetros para el modelo GARCH

Un procedimiento de estimación alternativo está basado en la minimización de una función de las diferencias entre la volatilidad y su pronóstico, ver Tsay (2010). Definiendo  $u_t = a_t^2 - \sigma_t^2$  el modelo GARCH presentado en la ec. (2-1) puede re-escribirse como:

$$a_t^2 = \omega + \left[ \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) a_{t-i}^2 \right] - \left[ \sum_{j=1}^q \beta_j u_{t-j} \right] + u_t \quad (2-4)$$

De este modo, la ecuación previa tiene la misma estructura de un modelo ARMA donde la primera expresión corresponde a la componente autorregresiva mientras que la segunda corresponde a la componente de medias móviles. Como consecuencia, los parámetros del modelo de la ec. (2-4) pueden estimarse minimizando la expresión:

$$\min \sum_t u_t^2 \quad (2-5)$$

donde  $u_t = a_t^2 - \sigma_t^2$ .

Note que maximizar la ec. (2-3) es diferente a minimizar la ec. (2-5) y los parámetros óptimos en cada caso son diferentes. Las implicaciones teóricas y las ventajas prácticas de este enfoque son desconocidas y es necesario más investigación.

#### 2.1.4 Criterios de optimización para estimar los parámetros de un modelo GARCH

El uso de (2-3) para estimar los valores de los parámetros óptimos en una modelo GARCH( $p, q$ ) esta basado en el supuesto de normalidad de los residuales estandarizados  $\varepsilon_t$ ; cuando el logaritmo de la función de verosimilitud es grande, entonces la distribución de probabilidad de  $\varepsilon_t$  es más próxima a la distribución normal estándar. Sin embargo, esto no significa que los parámetros sean óptimos en términos de la precisión del pronóstico de la volatilidad.

El proceso de estimación propuesto por Tsay (2010) esta enfocado en minimizar el error de pronóstico; esto es, los parametros en el modelo definido en la ec. (2-1) se estiman minimizando la expresión en la ec. (2-5). No obstante, los parámetros de un modelo GARCH( $p, q$ ) pueden ser estimados minimizando cualquier función de perdida.

En esta sección se usan cuatro criterios de optimización, algunos de los cuales hacen parte de la familia de funciones de pérdida  $L_p$ , que con frecuencia se encuentran en la literatura para la evaluación de pronósticos de volatilidad (De Gooijer y Hyndman, 2006; Dunis y Huang, 2002; Patton, 2011). Algunas de estas funciones no son robustas ante ruido Patton (2011), por lo que, el pronóstico óptimo no es la verdadera varianza condicional. Sin embargo, estas funciones se usan como criterio de optimización para estimar el modelo GARCH( $p, q$ ) del sistema acá propuesto debido a que son capaces de capturar otras características estadísticas de la serie de tiempo que son ignoradas cuando el modelo GARCH es estimado maximizando la función de verosimilitud. Estas funciones son una medida que da una base para comparar el pronóstico de la volatilidad con la volatilidad. No obstante, bajo los supuestos de la distribución condicional de los rendimientos

diarios dada en Patton (2011), la función de pérdida MSE genera un pronóstico óptimo igual a la varianza condicional, y así satisface la condición necesaria para la robustez. La función Linex, propuesta por Varian (1974) y usada por Zellner (1986) y Christoffersen y Diebold (1994), es incluida para modelar la posible existencia de la respuesta asimétrica de la volatilidad ante el signo de las innovaciones; en términos prácticos, esto significa que el modelo es capaz de modelar las diferencias grandes positivas y negativas entre las innovaciones al cuadrado y las varianzas. En la ec. (2-9) el parámetro  $c$  controla el nivel de asimetría. La función MAD es más robusta ante observaciones extremas, permite modelar mejor los periodos de tranquilidad o dinámicas bajas generando pronósticos bajos. Por otro lado, la función de pérdida MCE es muy sensible a los valores extremos y captura mejor las dinámicas de alta volatilidad.

- Desviación absoluta media:

$$MAE = \frac{1}{N} \sum_{i=1}^N |a_t^2 - \sigma_t^2| = \frac{1}{N} \sum_{i=1}^N |u_t| \quad (2-6)$$

- Error cuadrático medio:

$$MSE = \frac{1}{N} \sum_{i=1}^N |a_t^2 - \sigma_t^2|^2 = \frac{1}{N} \sum_{i=1}^N |u_t|^2 \quad (2-7)$$

- Error cúbico medio:

$$MCE = \frac{1}{N} \sum_{i=1}^N |a_t^2 - \sigma_t^2|^3 = \frac{1}{N} \sum_{i=1}^N |u_t|^3 \quad (2-8)$$

- Función de pérdida Linex

$$LLF = \frac{1}{N} \sum_{i=1}^N \exp c(a_t^2 - \sigma_t^2) - c(a_t^2 - \sigma_t^2) - 1; c \neq 0 \quad (2-9)$$

### 2.1.5 Combinador de pronósticos

Se analiza la literatura más relevante acerca de los métodos de ensamble y técnicas de combinación de pronósticos, se seleccionan las siguientes técnicas para combinar pronósticos de volatilidad,  $f_i$  con  $i = 1, \dots, N$ , de los modelos  $GARCH_i(p, q)$ :

- Promedio aritmético simple:

$$f_* = \frac{\sum_{i=1}^N f_i}{N}$$

- Combinación por regresión lineal múltiple:

$$f_* = \beta_0 + \beta_1 f_1 + \dots + \beta_k f_k$$

donde  $\beta_i$  son los parámetros.

- Combinación no lineal por redes neuronales artificiales:

$$f_* = \phi_0 + \left( \sum_{i=1}^k \phi_i f_i \right) + \sum_{h=1}^H \beta_h \cdot g \left( \gamma_{0h} + \sum_{i=1}^k \gamma_{ih} \cdot f_i \right) \quad (2-10)$$

Esta red neuronal propuesta también en [Hu y Tsoukalas \(1999\)](#) será usada como método de combinación de pronósticos en este trabajo.  $\phi_0, \phi_i, \beta_h, \gamma_{ih}$  son los parámetros y  $g(\cdot)$  es la función de transferencia logística:

$$g(u) = \frac{1}{1 + \exp(-u)}$$

Note que el modelo de RNA descrito en (2-10) es una mezcla entre un modelo de regresión lineal y un perceptrón multicapa; este modelo tiene  $kH + 2H + k + 1$  parámetros para ser estimados. Este modelo permitirá combinar no linealmente los pronósticos obtenidos de los modelos hallados mediante la metodología propuesta.

## 2.2 Caso de aplicación

### 2.2.1 Datos y análisis preliminar

Para este experimento se usan los excesos de los rendimientos mensuales del índice S&P500 a partir de enero de 1926 hasta diciembre de 1991; el conjunto de datos contiene 792 observaciones y se usa comúnmente como punto de referencia para ejemplificar y probar modelos de volatilidad; ver [Tsay \(2010\)](#) y [Verhoeven et al. \(2002\)](#). Los cálculos de este caso de aplicación se implementaron en Microsoft Excel y en el programa R.

[Tsay \(2010\)](#) analiza esta serie de tiempo usando toda la información disponible. Primero, se postula un modelo AR(3)-GARCH(1,1) para representar la dinámica de los datos; los parámetros de este modelo se estiman maximizando la Ec. (2-3). Después, todos los parámetros no significativos se eliminan y el modelo se reduce a un GARCH(1,1). El modelo final se escribe como:

$$\sigma_t^2 = \omega + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (2-11)$$

### 2.2.2 Montaje experimental

En el experimento numérico, se usaron las primeras 708 observaciones para estimar los parámetros de todos los modelos GARCH(1,1) y las 84 observaciones restantes para pronóstico. La precisión del pronóstico, para las muestras de ajuste y pronóstico, se mide usando la desviación absoluta media (MAD) y el error cuadrático medio (MSE) previamente definido en las Ecs. (2-6) y (2-7) respectivamente.

La configuración del experimento se describe en los siguientes pasos:

1. El modelo GARCH se ajusta usando las primeras 708 observaciones y maximizando el logaritmo natural de la función de .
2. Se reescribe el modelo GARCH(1, 1) como:

$$a_t^2 = \omega + (\alpha_1 + \beta_1)a_{t-1}^2 - \beta_1 u_{t-1} + u_t \quad (2-12)$$

Y luego, se estima el conjunto de parámetros óptimo  $\{\omega, (\alpha_1 + \beta_1), \beta_1\}$  minimizando cada una de las funciones de pérdida de las Ecs. (2-6), (2-7), (2-8) y (2-9). Así, se obtienen cuatro conjuntos de parámetros para el modelo presentado en la Ec. (2-12).

3. Para cada modelo GARCH considerado en este experimento, se pronostica la volatilidad un mes adelante para ambas muestras, la de ajuste y la de pronóstico.
4. Los pronósticos combinados se calculan, usando los pronósticos individuales obtenidos de los modelos estimados, en dos pasos de la siguiente forma:
  - El promedio aritmético simple de los pronósticos de cada modelo GARCH.
  - Usando una combinación lineal de los pronósticos de cada modelo GARCH y minimizando la función de error MSE para la estimación de los pesos.
  - Usando modelos de redes neuronales con  $H = 1, \dots, 4$  y  $k = 3$ . Ver la ec. (2-10).  $f_1, f_2, f_3$  y  $f_4$  son los conjuntos de pronóstico de los modelos estimados en el paso 2 usando los criterios de optimización MAD, MSE, MCE y Linex respectivamente. Cada modelo de RNA es entrenado 50 veces con pesos iniciales aleatorios. Se usa el modelo de ANN con el conjunto de parámetros que minimizan el MSE para la muestra de ajuste.
5. Se calcula el MSE y MAD en la muestra de ajuste y de pronóstico. Los resultados obtenidos, multiplicando el MSE por  $10^4$  y el MAD por  $10^2$ , se presentan en las Tablas 2-2, 2-3, 2-4 y 2-5.

En todos los casos, se usa el MSE como criterio de evaluación para probar la precisión del pronóstico de volatilidad. Como muestra Patton (2011), la función de error MSE es insesgada y satisface la condición necesaria para que una función de pérdida sea robusta ante el ruido en la volatilidad proxy.

El conjunto de parámetros para los modelos maximizando la ec. (2-3) y minimizando las ecs. (2-6), (2-7), (2-8) y (2-9) se presenta en la Tabla 2-1. Se considera como modelo base el modelo  $M_1$  cuyos parámetros se estiman con el método tradicional.

### 2.2.3 Precisión de pronóstico de los modelos individuales

En la Tabla 2-2 se muestran los valores de las medidas de precisión de los cinco modelos individuales para las muestras de ajuste y pronósticos usando dos criterios: el error cuadrático medio

**Tabla 2-1** Parámetros de los modelos GARCH(1, 1) individuales

Criterio de optimización	Modelo obtenido
Maximizar el log L	$\sigma_t^2 = 1.3023 + 0.1400a_{t-1}^2 + 0.8180\sigma_{t-1}^2$
Minimizar el MSE	$a_t^2 = 1.3800 + (0.1119 + 0.8494)a_{t-1}^2 - 0.8494u_{t-1}$
Minimizar el MAD	$a_t^2 = 0.4279 + (0.0472 + 0.8464)a_{t-1}^2 - 0.8464u_{t-1}$
Minimizar el MCE	$a_t^2 = 3.9451 + (0.1413 + 0.8953)a_{t-1}^2 - 0.8953u_{t-1}$
Minimizar LINEX	$a_t^2 = 1.0572 + (0.1279 + 0.8534)a_{t-1}^2 - 0.8534u_{t-1}$

**Tabla 2-2** Evaluación del pronóstico de la volatilidad de los modelos individuales

Modelo GARCH(1,1)	Ajuste		Pronóstico	
	MSE	MAD	MSE	MAD
$M_1$ : Ajustado maximizando el log L	12294.0	38.1	3718.6	28.7
$M_2$ : Ajustado minimizando el MSE	12262.3	38.6	* 3678.7	29.2
$M_3$ : Ajustado minimizando el MAD	13426.5	31.3	3770.2	23.0
$M_4$ : Ajustado minimizando el MCE	15768.1	73.2	5905.2	60.6
$M_5$ : Ajustado minimizando LINEX	12326.1	40.1	3740.0	30.1

\*Menor valor del MSE en la columna.

MSE y el error absoluto medio MAD. En este caso, el modelo  $M_2$ , ajustado minimizando el MSE, presenta un valor bajo para el MSE de pronóstico entre los cinco modelos. Los valores del MAD para las muestras de ajuste y pronóstico no son tan buenos. Sin embargo, el modelo estándar  $M_1$  es superado por el modelo  $M_2$  en términos de los valores del MSE y por el modelo  $M_3$  en los valores del MAD; esta conclusión es directa debido a los criterios de optimización. Adicionalmente, el modelo  $M_5$  tiene menor valor del MSE que el modelo  $M_3$  en ambas muestras. En este sentido, se busca un modelo que alcance un balance entre los valores del MSE y MAD tanto en la muestra de ajuste como en la de pronóstico.

#### 2.2.4 Precisión de pronóstico de los pronósticos combinados usando promedio aritmético simple

En la Tabla 2-3, se presentan los valores del MSE y MAD para las muestras de ajuste y pronóstico. En este caso, se considera el promedio de las posibles combinaciones considerando 2, 3, 4 y 5 modelos. Se encuentra que el modelo  $M_{26}$  tiene el menor valor del MSE en la muestra de pronóstico en comparación con los modelos competentes y los modelos individuales. Sin embargo, en la muestra de ajuste esto no se mantiene; este mismo modelo tiene los menores valores de MAD en ambas muestras. El siguiente modelo con un desempeño similar en términos del MSE es el  $M_{19}$ .

Una práctica común en la literatura de combinación de pronósticos es considerar el promedio del pronóstico de todos los modelos disponibles. Para este caso, correspondería al modelo  $M_6$  cuyos resultados indican que este modelo solo supera en términos del MSE a los modelos  $M_3, M_4$  y  $M_5$ .

### 2.2.5 Precisión de la combinación de pronósticos usando regresión lineal múltiple

En esta sección se consideran la mayoría de las posibles combinaciones lineales de los pronósticos individuales. La estimación de los parámetros de la regresión se realiza minimizando el MSE. En la Tabla 2-4 se encuentran los valores de las medidas de precisión usadas para la evaluación del desempeño de la combinación de pronósticos. En este conjunto de combinaciones el mejor modelo en términos del valor MSE en la muestra de pronóstico es el modelo  $M_{49}$ , que combina los pronósticos provenientes de los modelos  $M_1$  y  $M_3$ . Adicionalmente, este modelo supera los modelos de pronóstico individuales en términos del MSE pero no del MAD.

### 2.2.6 Precisión de la combinación de pronósticos usando redes neuronales artificiales

En este caso, se consideran la mayoría de las posibles combinaciones no lineales de los pronósticos individuales usando el modelo de red neuronal descrito en la ec. (2-10). Para cada conjunto de entradas posibles, se consideraron los modelos de RNA con  $H = 1, \dots, 4$  neuronas en la capa oculta. En la Tabla 2-5, se presentan los valores de los estadísticos MSE y MAD para las muestras de ajuste y pronóstico. Las conclusiones que se derivan de la Tabla 2-5 son:

1. Todos los modelos considerados tienen valores más pequeños del MSE en la muestra de ajuste que cada modelo individual de la Tabla 2-2. Esto sugiere que cada pronosticador no lineal es capaz de capturar información valiosa sobre la dinámica no lineal de la volatilidad.
2. Para la mayoría de las posibles combinaciones de las entradas, la configuración óptima de la red neuronal artificial en términos del valor MSE en la muestra de pronóstico tiene dos neuronas en la capa oculta ( $H=2$ ).
3. El modelo  $M_{87}$  es el preferido entre todas las posibles combinaciones de pronóstico de volatilidad. En este caso, el pronóstico compuesto se obtiene por la combinación no lineal de los pronósticos calculados con los modelos  $M_2, M_3$  y  $M_4$ .

### 2.2.7 Desempeño del modelo elegido

En esta sección se implementa un test estadístico para mostrar la diferencia en la precisión entre el modelo elegido y el modelo base  $M_1$ . Adicionalmente, se presenta una evaluación del modelo

**Tabla 2-3** Evaluación de la combinación de pronósticos de la volatilidad por medio del promedio aritmético simple

Promedio de pronósticos	Ajuste		Pronóstico	
	MSE	MAD	MSE	MAD
$M_6$ : Promedio de $M_1, M_2, M_3, M_4, M_5$	12314.0	41.8	3738.6	32.1
$M_7$ : Promedio de $M_1, M_2, M_3, M_4$	12321.5	42.3	3744.1	32.6
$M_8$ : Promedio de $M_1, M_2, M_3, M_5$	12304.5	36.1	3657.9	27.2
$M_9$ : Promedio de $M_1, M_2, M_4, M_5$	12540.6	46.0	3902.4	35.8
$M_{10}$ : Promedio de $M_1, M_3, M_4, M_5$	12341.2	42.7	3762.2	32.8
$M_{11}$ : Promedio de $M_2, M_3, M_4, M_5$	12343.7	42.9	3759.0	33.0
$M_{12}$ : Promedio de $M_1, M_2, M_3$	12381.9	35.0	3647.3	26.2
$M_{13}$ : Promedio de $M_1, M_2, M_4$	12662.3	48.2	3997.3	37.9
$M_{14}$ : Promedio de $M_1, M_2, M_5$	12278.2	38.9	3710.0	29.3
$M_{15}$ : Promedio de $M_1, M_3, M_4$	12364.3	43.7	3783.2	33.9
$M_{16}$ : Promedio de $M_1, M_3, M_5$	12336.8	35.4	3659.5	26.5
$M_{17}$ : Promedio de $M_1, M_4, M_5$	12753.7	48.7	4030.8	38.3
$M_{18}$ : Promedio de $M_2, M_3, M_4$	12373.7	43.9	3781.4	34.1
$M_{19}$ : Promedio de $M_2, M_3, M_5$	12342.1	35.5	* 3645.4	26.7
$M_{20}$ : Promedio de $M_2, M_4, M_5$	12744.6	49.0	4030.8	38.5
$M_{21}$ : Promedio de $M_3, M_4, M_5$	12373.7	43.9	3781.4	34.1
$M_{22}$ : Promedio de $M_1, M_2$	12271.3	38.3	3696.7	28.9
$M_{23}$ : Promedio de $M_1, M_3$	12530.3	33.6	3662.1	24.9
$M_{24}$ : Promedio de $M_1, M_4$	13154.9	53.7	4307.1	42.9
$M_{25}$ : Promedio de $M_1, M_5$	12297.1	39.1	3727.0	29.4
$M_{26}$ : Promedio de $M_2, M_3$	12553.1	33.6	* 3639.3	25.2
$M_{27}$ : Promedio de $M_2, M_4$	13145.3	54.2	4315.9	43.3
$M_{28}$ : Promedio de $M_2, M_5$	12278.3	39.3	3708.1	29.6
$M_{29}$ : Promedio de $M_3, M_4$	12506.5	47.0	3891.4	36.9
$M_{30}$ : Promedio de $M_3, M_5$	12441.9	34.2	3651.8	25.6
$M_{31}$ : Promedio de $M_4, M_5$	13341.1	55.0	4376.3	43.8

\*Menor valor del MSE en la columna.

**Tabla 2-4** Evaluación de la combinación de pronósticos de la volatilidad usando regresión lineal múltiple

Combinación lineal de pronósticos	Ajuste		Pronóstico	
	MSE	MAD	MSE	MAD
$M_{32}$ : Regresión lineal de $M_1, M_2, M_3, M_4, M_5$	12233.9	38.7	3691.2	29.7
$M_{33}$ : Regresión lineal de $M_1, M_2, M_3, M_4$	12233.3	38.7	3690.1	29.7
$M_{34}$ : Regresión lineal de $M_1, M_2, M_3, M_5$	12238.1	38.6	3694.0	29.7
$M_{35}$ : Regresión lineal de $M_1, M_2, M_4, M_5$	12233.9	38.7	3691.0	29.7
$M_{36}$ : Regresión lineal de $M_1, M_3, M_4, M_5$	12235.8	38.6	3691.7	29.7
$M_{37}$ : Regresión lineal de $M_2, M_3, M_4, M_5$	12232.4	38.7	3688.1	29.8
$M_{38}$ : Regresión lineal de $M_1, M_2, M_3$	12258.8	38.5	3677.9	29.2
$M_{39}$ : Regresión lineal de $M_1, M_2, M_4$	12235.9	38.7	3688.4	29.6
$M_{40}$ : Regresión lineal de $M_1, M_2, M_5$	12261.2	38.7	3678.0	29.3
$M_{41}$ : Regresión lineal de $M_1, M_3, M_4$	12239.1	38.6	3686.8	29.5
$M_{42}$ : Regresión lineal de $M_1, M_3, M_5$	12241.1	38.7	3692.7	29.7
$M_{43}$ : Regresión lineal de $M_1, M_4, M_5$	12241.2	38.6	3684.9	29.4
$M_{44}$ : Regresión lineal de $M_2, M_3, M_4$	12248.5	38.9	3686.2	29.7
$M_{45}$ : Regresión lineal de $M_2, M_3, M_5$	12262.3	38.6	3678.6	29.2
$M_{46}$ : Regresión lineal de $M_2, M_4, M_5$	12256.5	38.6	3681.5	29.2
$M_{47}$ : Regresión lineal de $M_3, M_4, M_5$	12247.1	38.6	3681.5	29.3
$M_{48}$ : Regresión lineal de $M_1, M_2$	12262.1	38.6	3677.0	29.2
$M_{49}$ : Regresión lineal de $M_1, M_3$	12261.5	38.6	* 3676.8	29.2
$M_{50}$ : Regresión lineal de $M_1, M_4$	12270.4	38.6	3679.9	29.2
$M_{51}$ : Regresión lineal de $M_2, M_3$	12262.3	38.6	3679.0	29.2
$M_{52}$ : Regresión lineal de $M_2, M_4$	12261.1	38.6	3683.1	29.2
$M_{53}$ : Regresión lineal de $M_3, M_4$	12262.0	38.5	3682.8	29.2

\*Menor valor del MSE en la columna.

**Tabla 2-5** Evaluación de la combinación de pronósticos de la volatilidad usando ANN con diferente número de neuronas en la capa oculta

Combinación de pronósticos	Ajuste		Pronóstico	
	MSE	MAD	MSE	MAD
Combinación de los modelos $M_1, M_2, M_3, M_4, M_5$ usando ANN				
$M_{54}$ : ANN(H=1)	10978.0	38.2	3724.2	26.0
$M_{55}$ : ANN(H=2)	4521.2	30.3	* 3667.4	29.2
$M_{56}$ : ANN(H=3)	4383.2	29.9	3985.2	30.1
$M_{57}$ : ANN(H=4)	4210.2	29.1	4010.7	32.0
Combinación de los modelos $M_1, M_2, M_3, M_4$ usando ANN				
$M_{58}$ : ANN(H=1)	11434.8	38.8	4159.3	32.3
$M_{59}$ : ANN(H=2)	4521.2	30.3	* 3665.3	29.1
$M_{60}$ : ANN(H=3)	4351.0	29.9	3857.2	30.2
$M_{61}$ : ANN(H=4)	4291.6	29.2	3988.4	29.5
Combinación de los modelos $M_1, M_2, M_3, M_5$ usando ANN				
$M_{62}$ : ANN(H=1)	10372.0	39.5	4378.5	34.4
$M_{63}$ : ANN(H=2)	9507.4	38.4	3629.5	26.0
$M_{64}$ : ANN(H=3)	8057.3	33.9	3784.0	31.3
$M_{65}$ : ANN(H=4)	5321.1	34.5	* 3620.8	28.7
Combinación de los modelos $M_2, M_3, M_4, M_5$ usando ANN				
$M_{66}$ : ANN(H=1)	11408.8	38.8	3941.8	29.9
$M_{67}$ : ANN(H=2)	6529.8	32.5	3723.5	28.1
$M_{68}$ : ANN(H=3)	6526.7	32.4	* 3713.1	27.9
$M_{69}$ : ANN(H=4)	6238.4	32.5	3795.0	27.2
Combinación de los modelos $M_1, M_2, M_3$ usando ANN				
$M_{70}$ : ANN(H=1)	11694.5	39.0	4191.2	33.1
$M_{71}$ : ANN(H=2)	6629.7	32.6	* 3731.6	29.3
$M_{72}$ : ANN(H=3)	6541.8	33.5	3860.3	30.4
$M_{73}$ : ANN(H=4)	6368.2	33.1	3911.6	31.4
Combinación de los modelos $M_1, M_2, M_4$ usando ANN				
$M_{74}$ : ANN(H=1)	10970.7	38.7	3707.2	25.5
$M_{75}$ : ANN(H=2)	10436.2	40.7	* 3661.6	33.4
$M_{76}$ : ANN(H=3)	10232.8	39.4	3723.3	33.4
$M_{77}$ : ANN(H=4)	8009.7	33.4	3724.7	30.7

\*Menor valor del MSE en la columna.

Continuación Tabla 2-5

Combinación de pronósticos	Ajuste		Pronóstico	
	MSE	MAD	MSE	MAD
Combinación de los modelos $M_1, M_3, M_4$ usando ANN				
$M_{78}$ : ANN(H=1)	11562.2	39.5	4180.0	32.3
$M_{79}$ : ANN(H=2)	10438.4	39.1	* 3603.2	25.3
$M_{80}$ : ANN(H=3)	10329.3	38.8	3619.0	24.6
$M_{81}$ : ANN(H=4)	9704.5	36.6	3880.7	27.4
Combinación de los modelos $M_1, M_3, M_4$ usando ANN				
$M_{78}$ : ANN(H=1)	11562.2	39.5	4180.0	32.3
$M_{79}$ : ANN(H=2)	10438.4	39.1	* 3603.2	25.3
$M_{80}$ : ANN(H=3)	10329.3	38.8	3619.0	24.6
$M_{81}$ : ANN(H=4)	9704.5	36.6	3880.7	27.4
Combinación de los modelos $M_1, M_3, M_5$ usando ANN				
$M_{82}$ : ANN(H=1)	11660.3	39.2	4261.1	34.1
$M_{83}$ : ANN(H=2)	7024.6	35.5	3952.9	32.1
$M_{84}$ : ANN(H=3)	6414.3	32.7	* 3873.4	30.7
$M_{85}$ : ANN(H=4)	6260.8	33.0	3915.3	31.8
Combinación de los modelos $M_2, M_3, M_4$ usando ANN				
$M_{86}$ : ANN(H=1)	11939.1	38.3	3786.2	29.8
$M_{87}$ : ANN(H=2)	4954.1	30.6	* 3580.6	27.2
$M_{88}$ : ANN(H=3)	4489.7	30.2	3878.6	30.8
$M_{89}$ : ANN(H=4)	2619.8	37.7	4147.0	30.7
Combinación de los modelos $M_2, M_3, M_5$ usando ANN				
$M_{90}$ : ANN(H=1)	11244.5	38.3	3895.7	26.4
$M_{91}$ : ANN(H=2)	10391.0	37.9	3796.4	29.1
$M_{92}$ : ANN(H=3)	9249.5	37.7	* 3742.9	28.9
$M_{93}$ : ANN(H=4)	9221.9	37.8	3834.3	30.9
Combinación de los modelos $M_1, M_2$ usando ANN				
$M_{94}$ : ANN(H=1)	11195.7	38.7	3755.5	25.6
$M_{95}$ : ANN(H=2)	6688.4	32.9	* 3734.4	29.4
$M_{96}$ : ANN(H=3)	6486.6	32.4	3803.4	30.0
$M_{97}$ : ANN(H=4)	6483.9	32.9	3800.8	30.6

\*Menor valor del MSE en la columna.

Continuación Tabla 2-5

Combinación de pronósticos	Ajuste		Pronóstico	
	MSE	MAD	MSE	MAD
Combinación de los modelos $M_1, M_3$ usando ANN				
$M_{98}$ : ANN(H=1)	11278.6	38.0	3936.8	28.7
$M_{99}$ : ANN(H=2)	6573.8	32.8	3768.5	29.6
$M_{100}$ : ANN(H=3)	6569.0	32.7	* 3767.7	29.6
$M_{101}$ : ANN(H=4)	6249.9	32.5	3733.4	30.3
Combinación de los modelos $M_1, M_4$ usando ANN				
$M_{102}$ : ANN(H=1)	11287.6	38.5	3888.6	27.4
$M_{103}$ : ANN(H=2)	11203.6	37.6	* 3684.1	27.2
$M_{104}$ : ANN(H=3)	10283.1	37.0	3906.4	28.3
$M_{105}$ : ANN(H=4)	10281.2	36.7	3875.0	28.7
Combinación de los modelos $M_2, M_3$ usando ANN				
$M_{106}$ : ANN(H=1)	11776.8	38.2	3963.8	31.2
$M_{107}$ : ANN(H=2)	6731.2	34.2	* 3746.0	30.4
$M_{108}$ : ANN(H=3)	6502.6	32.4	3763.0	29.5
$M_{109}$ : ANN(H=4)	6481.4	31.6	3808.3	29.2
Combinación de los modelos $M_2, M_4$ usando ANN				
$M_{110}$ : ANN(H=1)	11646.0	38.7	4209.4	33.2
$M_{111}$ : ANN(H=2)	6677.1	32.7	* 3740.8	29.4
$M_{112}$ : ANN(H=3)	6379.9	33.3	3885.4	31.6
$M_{113}$ : ANN(H=4)	6348.8	32.7	3836.4	30.2
Combinación de los modelos $M_3, M_4$ usando ANN				
$M_{114}$ : ANN(H=1)	11331.5	38.7	3871.8	27.2
$M_{115}$ : ANN(H=2)	6462.1	32.5	3787.4	29.8
$M_{116}$ : ANN(H=3)	6459.0	32.4	3784.5	29.8
$M_{117}$ : ANN(H=4)	6455.6	32.3	* 3777.6	29.3

\*Menor valor del MSE en la columna.

preferido y el modelo base GARCH usando una evaluación del valor en riesgo (VaR por sus siglas en inglés).

- Prueba de precisión estadística

Una demostración de la efectividad del modelo  $M_{87}$  sobre el modelo  $M_1$  se muestra en una prueba estadística. El contraste de hipótesis cuya hipótesis nula es la no diferencia entre la precisión del modelo base y el modelo elegido es aplicado mediante el test de [Diebold y Mariano \(1995\)](#) el cual usa el estadístico MGN de Morgan-Granger-Newbold. La prueba consiste en considerar los errores de pronóstico puntual  $e_{1t}$  y  $e_{2t}$  de ambos modelos respectivamente. El estadístico implementado tiene una distribución  $t$ -student con  $T - 1$  grados de libertad. El estadístico  $t$  está definido por:

$$t = \frac{\rho_{xz}}{\sqrt{\frac{1-\rho_{xz}^2}{T-1}}} \quad (2-13)$$

donde,  $\rho_{xz} = \frac{\sum_t x_t z_t}{\sqrt{\sum_t x_t^2 \sum_t z_t^2}}$ ;  $x_t = e_{1t} + e_{2t}$ ,  $t = 1 : T$ ; y  $T$  es el número de observaciones.

Los resultados de la prueba muestran que la hipótesis nula de igual precisión predictiva entre los modelos  $M_1$  y  $M_{87}$  es rechazada a un nivel de significancia del 7% en ambas muestras, con un valor  $p$  de  $7.06694 \times 10^{-54}$  en la muestra de ajuste y 0.0663 en la muestra de pronóstico.

- Evaluación del valor en riesgo

Se ha comparado el modelo base GARCH(1,1) ajustado maximizando el logaritmo de la función de verosimilitud y el modelo elegido utilizando el método del valor en riesgo VaR. El VaR se define como la pérdida potencial máxima de un portafolio bajo condiciones normales del mercado con una probabilidad predefinida. Asumiendo que los rendimientos son normalmente distribuidos, el VaR de un 95% se calcula como el cuantil 95 de la distribución de los rendimientos:

$$\widehat{VaR}_{t+1} = -1.00 \times 1.65 \times \hat{\sigma}_{t+1} \quad (2-14)$$

donde  $\hat{\sigma}_{t+1}$  es el pronóstico de la desviación estándar dada toda la información disponible hasta el tiempo  $t$ .

Para evaluar el desempeño de los estimadores del VaR al 95% se usa el criterio de tasa de fallos. La tasa de fallo ( $F$ ) es el número de veces en que las pérdidas exceden el VaR estimado. Este criterio se define como  $F = \sum_t D_t$ , donde la variable indicadora  $D_t$  es uno si la pérdida excede el VaR estimado y cero en otro caso.

La evaluación del desempeño del VaR es realizada; la tasa de fallo del modelo  $M_{87}$  es 4.8% con 38 casos y es menor que la tasa de fallo del modelo  $M_1$  que es 6.2% con 49 casos. Se concluye que el modelo obtenido de la combinación con ANN puede mejorar la precisión predictiva.

## 2.3 Conclusiones

En este capítulo se ha considerado la combinación de pronósticos como una alternativa para reducir el error de la volatilidad de los rendimientos mensuales del índice S&P500. Varios modelos GARCH(1, 1) fueron estimados, considerando diferentes criterios de optimización, y combinados usando promedio aritmético simple, regresión lineal múltiple y ANN para producir un nuevo modelo que integra toda la información de los pronósticos individuales. Adicionalmente, se implementó una prueba de precisión estadística y una evaluación del valor en riesgo VaR entre el modelo de combinación con mejor desempeño y el modelo base considerado. Los resultados obtenidos son alentadores ya que muestran que un enfoque de combinación puede predecir con mayor precisión la volatilidad en comparación con los modelos de pronóstico tradicionales. En el siguiente capítulo se ahondará en temas tanto teóricos como prácticos que implican el uso de un sistema de combinación de pronósticos por medio de redes neuronales artificiales.

# 3 Proceso de especificación de modelos de pronóstico de volatilidad implementando diferentes criterios de optimización

En este capítulo se presenta un proceso de especificación para el ajuste de los modelos de volatilidad condicional, usando conjuntamente varios criterios de optimización y una técnica de remuestreo por bootstrapping. Algunas funciones de error que se usan como criterios de optimización son analizadas y se presentan algunas propiedades. Por su parte, el remuestreo surge como una necesidad ante la falta de propiedades teóricas de los estimados de los parámetros del modelo ajustado con criterios de optimización no convencionales.

Cuando se optimiza el modelo con el método de mínimos cuadrados, los estimados de los parámetros tienen propiedades estadísticas que permiten hacer inferencias acerca de los parámetros poblacionales. Cuando no se tienen dichas propiedades, éstas deben construirse; el método de remuestreo tienen como objetivo generar réplicas de las estimaciones de los parámetros, permitiendo encontrar su distribución de probabilidad y dar cuenta de los parámetros poblacionales.

El objetivo en este capítulo es mostrar que implementando el enfoque de diferentes criterios de optimización para estimar un modelo se generan conjuntos de pronósticos diferentes y se puede mejorar el desempeño de los modelos estimados con el método convencional. A lo largo de este capítulo se describe la propuesta para el proceso de especificación de modelos de volatilidad condicional y la técnica de remuestreo que se aplica para hacer inferencia de los parámetros poblacionales de dichos modelos. Además, se presenta una aplicación de esta propuesta a varios conjuntos de datos y se compara con los resultados reportados en el artículo de [Taylor \(2004\)](#).

## 3.1 Justificación

Si bien algunos autores han sugerido prestar mayor atención a las funciones de error y su importancia en el proceso de especificación de un modelo ([Christoffersen y Jacobs, 2004](#); [Elliott](#)

y Timmermann, 2008), no existe una metodología que explore las ventajas que conlleva el uso de estas como criterio de optimización para la estimación de los parámetros. Por otro lado, se debe tener en cuenta que debido a que el proceso generador de los datos (PGD) es desconocido y no es posible conocerlo plenamente, cualquier modelo que se proponga como estimador producirá un error. Es así como la forma en que se mida dicho error puede dar información relevante para la construcción del modelo.

Las funciones de error se encargan de penalizar la diferencia entre el valor realizado y el valor pronosticado; en series de tiempo financieras donde los datos extremos son comunes, se observan diferentes respuestas dependiendo del error que se mida; por ejemplo, es bien sabido que el error cuadrático medio es bastante sensible a valores extremos en comparación al error absoluto medio, así que intuitivamente al usar una función de error como método de optimización que cada una tendrá un efecto diferente en el modelo y por ende en los parámetros que se estiman. Es así como es de interés explorar bajo la perspectiva de las funciones de error como métodos de estimación el impacto en los pronósticos obtenidos. En consecuencia, el uso de una función de error para estimar un modelo permitirá que el modelo identifique ciertas características de la serie que pueden ser diferentes a las capturadas en un modelo optimizado con otra función de error.

### 3.1.1 Funciones de error como criterio de optimización

Es necesario prestar atención a la metodología de modelado y pronóstico de la volatilidad en cuanto a la selección de la forma funcional de las funciones de error ya que debe depender en gran medida del tipo de datos y del problema en consideración. Adicionalmente, no existe una metodología formal que permita seleccionar una función de error para la estimación y evaluación de pronósticos en un modelo específico. Esta afirmación tiene como fundamento los siguientes hechos:

- Por lo general la selección de la función de error se hace por conveniencia estadística y no por preferencia al modelo o a los objetivos del pronóstico (Granger y Machina, 2006).
- Las funciones de error tienen un alto impacto sobre la especificación de un modelo estadístico, y comúnmente se sugiere por aplicaciones empíricas que los modelos de pronóstico deben ser entrenados y evaluados con la misma función de error (Christoffersen y Jacobs, 2004; Patton y Timmermann, 2007).
- La escogencia implícita de la función de pérdida MSE es difícil de justificar en el campo de la economía ya que representa dificultades en el costo de sub y sobre-predicción (Elliott y Timmermann, 2008).

Lo anterior sugiere que el proceso de estimación convencional que usa mínimos cuadrados es meramente por conveniencia estadística ya que presenta características deseables en los estimadores de los parámetros. La justificación de esta práctica convencional requiere supuestos que difícilmente

se cumplen, como por ejemplo normalidad de los errores. Particularmente, en series de tiempo financieras esta escogencia resulta difícil de sustentar pero se usa como punto de referencia. Así, cualquier otra forma funcional puede ser usada con la desventaja de la falta de teoría estadística. Sin embargo, en algunos casos es posible identificar ciertas propiedades como se presenta en la siguiente sesión.

Considere que se tiene una sucesión de  $n$  observaciones de una variable aleatoria,  $\{y_t\}$ , y una función  $\{f_t = f_t(\theta, y_{t-1})\}$  que genera predicciones de los valores reales de la variable para cada instante de tiempo cuando se ha estimado  $\theta$ .

En términos generales, una función de pérdida  $\mathcal{L}(f(Y, \theta), Y)$  es una función de valor real definida en  $\Theta \times \mathbf{A}$ , donde  $\Theta$  representa el conjunto de los posibles estados de la naturaleza también conocido como el espacio paramétrico y  $\mathbf{A}$  es el conjunto de las posibles salidas de la variable  $Y$ . Una función de error cumple las propiedades:

- $\mathcal{L}(f(Y, \theta), Y) \geq 0$  para todo  $\theta \in \Theta$  y  $y \in \mathbf{A}$
- $\mathcal{L}(f(Y, \theta), Y) = 0$  cuando  $f(y, \theta) = y$ .

La familia de funciones de pérdida  $\mathcal{L}_p$  está definida como:

$$L_p(f(Y, \theta), Y) = 1/N \sum_{t=1}^N |Y_t - f_t(Y_t, \theta)|^p$$

Es decir, la suma de los errores absolutos elevados a la potencia  $p$ . Para algunos valores de  $p$ ,  $p = 1, 2, 3, 1/2$ , se encuentran anidadas dentro de esta familia las siguientes funciones: La pérdida del Error Absoluto Medio (MAE, por sus siglas en inglés), el Error Cuadrático Medio (MSE), el Error Cúbico Medio, el Error Cuarto Medio (MFE). Otras funciones de pérdida que permiten cuantificar el error son por ejemplo las funciones basadas en errores porcentuales y la función propuesta en [Taylor \(2004\)](#) usada para estimar los parámetros de un modelo de volatilidad cuando se cuenta con datos de alta frecuencia.

### 3.1.2 Caracterización de las funciones de error

El método de mínimos cuadrados, el cual es el método clásico de estimación, consiste en minimizar la suma del error cuadrático:

$$\min \sum_{t=1}^n (y_t - f_t)^2 \quad (3-1)$$

Para ello, se deriva la ecuación (3-1) con respecto a  $f_t$  y se obtiene que el estimador de  $f_t$  es la media condicional de  $Y_t$ , esto es,  $\hat{f}_t = E(Y_t | I_t - 1, \theta) = \bar{y}_t$ .

El estimador anterior no es robusto ante valores atípicos en las observaciones debido a que penaliza fuertemente los errores grandes, haciéndolo más sensible a los valores extremos; tampoco es robusto ante desviaciones de los supuestos de los datos. No obstante, tiene propiedades estadísticas que han sido reportadas en la literatura, [Lehmann y Casella \(2003\)](#).

En términos generales es posible proponer otros métodos para estimar un modelo, minimizando funciones de error diferentes al error cuadrático medio y en algunos casos se encuentra que estos pueden llevar a estimaciones que no producen estadísticos conocidos.

En el caso del error absoluto medio se minimiza la expresión:

$$\min \sum_{t=1}^n |y_t - f_t| \quad (3-2)$$

diferenciando con respecto a  $f_t$  e igualando a cero se obtiene:

$$-\sum_{t=1}^n \frac{y_t - \hat{f}_t}{|y_t - \hat{f}_t|} = 0 \quad (3-3)$$

el cociente en la sumatoria (3-3) es 1 o -1, por lo tanto la suma total será cero cuando haya igual cantidad de 1 y -1, indicando que el estimador es la mediana condicional de  $Y$   $\hat{f}_t = \text{mediana}(y_t)$ . Su característica principal es reducir la influencia de los valores extremos y se usa en estadística robusta por considerar que no es excesivamente afectada por los valores atípicos y la violación de supuestos del proceso generador de los datos.

Para cuando se usan las funciones de pérdida basadas en errores porcentuales se tiene la ventaja de independencia de escala y se usan con frecuencia para compara el desempeño de pronóstico a través de diferentes conjuntos de datos. Una de sus desventajas es que son indefinidas si  $Y_t = 0$  para cualquier período de interés, y tienen una distribución demasiado asimétrica cuando cualquier valor de  $Y_t$  es cercano a cero.

Otro método consiste en minimizar la suma del cuadrado del logaritmo de la razón de precisión, siguiendo el procedimiento anterior, es posible mostrar que el estimador es la media geométrica [Tofallis \(2015\)](#); este autor sugiere además, que el uso de la función de error MAPE en un modelo de regresión como método de optimización arrojará predicciones muy bajas. En este trabajo se considera que esta afirmación puede extenderse para modelos diferentes a los modelos de regresión.

Para la demás funciones de error consideradas acá no es posible obtener estimadores que conlleven a un estadístico simple, aunque se observa, en algunos casos, mas o menos susceptibilidad a los valores extremos. Así que estimar los parámetros de un modelo minimizando distintas funciones de error puede conllevar a modelos que capturen diferentes información contenida en los datos,

esto debido a que se valora en forma diferente el error, o dicho en otros términos, no se miden los mismos aspectos de precisión del pronóstico.

En este capítulo se considerarán diferentes funciones de error como criterios de optimización. La selección de la mismas se hace con base a las funciones disponibles en la literatura y que se han usado como mecanismo de evaluación.

## 3.2 Metodología

Se consideran datos de varias series de tiempo de precios de activos financieros usadas en [Taylor \(2004\)](#). Se realiza el proceso de especificación de los modelos de volatilidad con el método convencional para cada una de las series y se determina cual es el modelo seleccionado de acuerdo a los criterios de Akaike y Bayes, en este paso se seleccionaron modelos  $TGARCH(p, q)$  con diferentes valores de  $p$  y  $q$ , esto se hizo al comparar el desempeño de los modelos GARCH, TAGRCH y EGARCH. Una vez escogido el orden del modelo, se aplica el proceso de especificación y el remuestreo, conocido como bootstrapping, descritos más adelante. Todo el proceso en estos casos de aplicación se hace con el programa R.

El remuestreo se implementa como un mecanismo de construcción de la distribución de probabilidad de los parámetros poblacionales del modelo para cada serie de tiempo.

### 3.2.1 Valor proxy de la volatilidad

Considere la serie  $P_t$  del precio en el tiempo  $t$  de un activo financiero; existen varias definiciones de los rendimientos de un activo, no obstante en este trabajo es de interés el logaritmo natural del rendimiento bruto simple también conocido como rendimiento compuesto continuo o logaritmo del rendimiento, el cual esta dado por

$$r_t = \ln \left( \frac{P_t}{P_{t-1}} \right). \quad (3-4)$$

Puesto que la volatilidad ha sido definida en este trabajo como la varianza de los rendimientos de un activo financiero, una medida para la varianza condicional o volatilidad condicional será el cuadrado de los rendimientos,  $\sigma_t^2 = r_t^2$ , ver ([Hansen y Lunde, 2005](#); [McAleer y Medeiros, 2011](#); [Patton, 2011](#)).

### 3.2.2 Especificaciones GARCH

Inicialmente se consideraron los modelos GARCH, EGARCH y TGARCH y se realizaron pruebas de diagnóstico, las cuales llevaron a escoger un modelo TGARCH como la mejor especificación

para modelar y predecir los datos.

EL modelo de umbral TGARCH, fue propuesto por Rabemananjara y Zakoian (1993) y es también conocido como el modelo GJR-GARCH debido a que Glosten et al. (1993) propusieron el mismo modelo. El modelo TGARCH( $p, q$ ) esta dado por la expresión:

$$\sigma_t^2 = \omega + \sum_{i=1}^p (\alpha_i + \gamma_i I_{t-i}) \cdot a_{t-i}^2 + \sum_{j=1}^q \beta_j \cdot \sigma_{t-j}^2 \quad (3-5)$$

Donde  $I_{t-i} = 1$  si  $a_{t-i} < 0$  y cero en otro caso. Para garantizar la positividad de  $\sigma_t^2$  y la estacionariedad del modelo se imponen las restricciones sobre los parámetros:  $\omega > 0$ ,  $\alpha_i, \beta_j \geq 0$ ,  $\alpha_i + \gamma_i \geq 0$  y  $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j + \frac{1}{2} \sum_{i=1}^p \gamma_i < 1$ .

Este modelo, a diferencia del modelo GARCH, tiene en cuenta el efecto de asimetría o apalancamiento (*leverage*), presente en las series de tiempo financieras, por medio del parámetro  $\gamma_i$ , ya que tiene en cuenta el signo de la innovación  $a_{t-i}$ .

El modelo TGARCH(1,1) está representado por la ecuación:

$$\sigma_t^2 = \omega + \alpha_1 \cdot a_{t-1}^2 + \beta_1 \cdot \sigma_{t-1}^2 + \gamma_1 a_{t-1}^2 I_{t-1}. \quad (3-6)$$

El cual hereda las restricciones en los parámetros descritas para el modelo TGARCH( $p, q$ ).

### 3.2.3 Proceso de especificación del modelo

El procedimiento de construcción del modelo seguido en este trabajo esta propuesto por Tsay (2010) y consiste en: primero, determinar si existe correlación serial en los rendimientos con la prueba de Ljung-Box, si existe alguna dependencia lineal esta debe ser modelada. La función de autocorrelación, FAC, de los rendimientos es también usada para probar la correlación serial de los rendimientos y para seleccionar el orden del modelo autorregresivo AR de la media.

Segundo, determinar si existe efecto ARCH mediante la prueba de Ljung-Box aplicada a los residuales cuadrados,  $a_t^2$ , de la ecuación de la media.

Tercero, especificar un modelo de volatilidad si el efecto ARCH resulta significativo. Los modelos TGARCH( $p, q$ ) para valores  $p, q = 1, 2$ , se estiman conjuntamente con el modelo de media. La FAC y la FAC Parcial de los rendimientos al cuadrado sugieren el posible orden del modelo ARMA( $p, q$ ) para la media. Sin embargo, no existe un procedimiento formal para seleccionar el orden  $p$  y  $q$  de la ecuación TGARCH para modelar la volatilidad de los rendimientos. El procedimiento empírico consiste en iniciar con valores pequeños de  $p$  y  $q$ ,  $p, q = 0, 1, 2$ , realizar una estimación conjunta del modelo ARMA( $p, q$ )-TGARCH( $p, q$ ) y hacer la evaluación con los

criterios de información Akaike y Bayes, AIC y BIC respectivamente. Se asume que los errores siguen una distribución t-Student estandarizada, la cual es una distribución de colas pesadas. Los parámetros de las ecuaciones de la media y de la volatilidad se prueban a un nivel de significancia de 5%. Se usa la misma muestra de entrenamiento para todos los modelos estimados.

Después de seleccionar el modelo conjunto que resulta más adecuado, el procedimiento de especificación termina probando la adecuación de las ecuaciones de media y de volatilidad. Esto es, probar que los residuales estandarizados sean ruido blanco y que el estadístico de Ljung-Box sobre los residuales estandarizados al cuadrado muestre la validez de la ecuación de volatilidad.

### 3.2.4 Estimación de los parámetros del modelo TGARCH(1,1)

El procedimiento estándar de estimación de parámetros es vía el método de máxima verosimilitud, el cual bajo supuestos de normalidad de los errores es equivalente a método de mínimos cuadrados. Sin embargo, como se afirmó antes, la estimación de parámetros puede llevarse a cabo minimizando otras funciones de error.

En este capítulo se aplica el mismo procedimiento de estimación de parámetros considerado en el capítulo anterior. Para ello se realiza la transformación  $u_t = a_t^2 - \sigma_t^2$  para el modelo TGARCH( $p, q$ ) de la ecuación (3-5) para obtener la estructura de un modelo ARMA. En el caso de un modelo TGARCH(1,1), este puede reescribirse como:

$$a_t^2 = \omega + (\alpha_1 + \beta_1 + \gamma I_{t-1})a_{t-1}^2 - \beta_1 u_{t-1} + u_t \quad (3-7)$$

En general,  $u_t$  no es una sucesión independiente idénticamente distribuida pero se puede mostrar que es una serie martingala en diferencias, esto es,  $E[u_t] = 0$  y  $\text{cov}(u_t, u_{t-j}) = E(u_t u_{t-j}) = 0$  para  $j \geq 1$ .

El proceso de estimación consiste en minimizar diferentes funciones de pérdida de  $u_t$ . Para ello se proponen seis funciones de pérdida que se usan como criterios de optimización para obtener los parámetros del mismo modelo TGARCH. El error de predicción está dado por  $u_t = a_t^2 - \sigma_t^2$  y el error porcentual por  $p_t = 100 * (a_t^2 - \sigma_t^2) / a_t^2$ .

Los criterios de optimización usados en este trabajo han sido aplicados en la literatura para evaluar el desempeño de modelos de pronóstico, ver [De Gooijer y Hyndman \(2006\)](#). Sin embargo, la función denominada MSER fue inicialmente propuesta por [Taylor \(2004\)](#) para estimar parámetros usando datos de alta frecuencia. Entre las funciones de error a minimizar se encuentran:

1. Error Absoluto Medio MAE =  $1/N \sum_1^N |u_t|$
2. Raíz del Error Cuadrático Medio RMSE =  $\sqrt{(1/N \sum_1^N u_t^2)}$

3. Raíz cúbica del Error Cúbico Medio  $RMCE = \left( \frac{1}{N} \sum_1^N |u_t|^3 \right)^{1/3}$
4. Raíz cuarta del Error Cuarto Medio  $RMFE = \left( \frac{1}{N} \sum_1^N |u_t|^4 \right)^{1/4}$
5. Promedio del Error de la Raíz Cuadrática al Cuadrado  $MSSER = 1/N \sum_1^N (\sqrt{a_t^2} - \sqrt{\sigma_t^2})^2$
6. Error Porcentual de la media Absoluta Simétrica  $SMAPE = \frac{200}{N} \sum_1^N \left( \frac{a_t^2 - \sigma_t^2}{a_t^2 + \sigma_t^2} \right)$

Por otro lado la precisión de los modelos se mide con las métricas: MAE y RMSE en ambas muestras, ajuste y pronóstico. Una vez estimados los parámetros del modelo con cada uno de los criterios de optimización, se fijan los parámetros estimados y el pronóstico obtenido en la muestra de prueba se realiza un paso adelante. Entre las ventajas prácticas que se obtienen al usar diferentes criterios de optimización en un modelo con una estructura fija se encuentran: en primer lugar, se obtienen pronósticos que identifican características diferentes de los datos en aras de explotar la mayor cantidad de información de los datos y estimar mejor el proceso generador de datos. En segundo lugar, sirve como elemento de diversificación de los pronósticos, elemento clave para la combinación de pronósticos.

### Método de Bootstrapping

Resulta de interés comparar estadísticamente la distribución de los parámetros del modelo TGARCH a través de los diferentes criterios de optimización, para ello es necesario aplicar un método bootstrapping para generar muestras de los parámetros estimados en cada método de optimización. Así mismo, se puede obtener la distribución de los pronósticos en cada instante de tiempo y de las medidas de desempeño de cada modelo en ambas muestras. Con este procedimiento se pretende diversificar los modelos en dos formas, a saber: primero, con el uso de un criterio de optimización diferente para cada uno y segundo, generando para cada modelo una submuestra al adicionar un ruido de estructura fija a la serie de tiempo original.

Los pasos que se siguen para generar las distintas muestras de los parámetros son:

1. Para cada criterio de optimización, se construyen  $n_b$  submuestras de la serie de tiempo agregando un ruido gaussiano a cada una de las serie de los rendimientos en la muestra de ajuste. Agregar un ruido con una distribución conocida sirve para simular diferentes realizaciones de la serie de tiempo, es decir, el proceso estocástico; además de conservar la estructura de la serie de tiempo original.
2. Se estiman los parámetros mediante optimización numérica, minimizando las funciones de error descritas en la sección anterior, denominadas criterios de optimización, para las  $n_b$  series de tiempo.
3. Para cada criterio de optimización se obtiene:  $n_b$  vectores con los parámetros estimados del modelo TGARCH, esto es, la distribución de cada parámetro en los diferentes criterios de

optimización,  $n_b$  vectores con los valores pronosticados un paso adelante de la volatilidad y  $n_b$  vectores con las medidas de desempeño.

Una vez llevado a cabo este proceso, se comparan estadísticamente los parámetros y las medidas de desempeño de los pronósticos individuales obtenidos de cada modelo TGARCH en la muestra de prueba. En adelante, se usará la notación TGARCH-A para hacer explícito que el modelo TGARCH se ha optimizado minimizando la función de error A.

### 3.3 Aplicación a series de tiempo reales

En esta aplicación consideramos cuatro series de tiempo que han sido ajustadas en [Taylor \(2004\)](#) con el fin de comparar el desempeño en el pronóstico de la metodología propuesta en este trabajo.

#### 3.3.1 Datos

Los datos corresponden a los precios de cuatro índices bursátiles de los principales mercados de valores: New York (S&P500), París (CAC40), Singapur (Singapore all Shares) y Tokio (Nikkei). El período de la muestra utilizada en este estudio consistió en 2000 días de cotización, del 30 de diciembre de 1987 al 30 de agosto de 1995. El objetivo es predecir la volatilidad semanal del logaritmo de los rendimientos, lo que equivale a una muestra de tamaño  $N = 400$  observaciones. Siguiendo [Taylor \(2004\)](#) y en aras de comparar el desempeño de los modelos en la muestra de prueba, se divide en dos conjuntos la muestra, el 50% de los datos de la muestra se usa para entrenamiento de los modelos y el 50% restante para prueba.

#### 3.3.2 Ajuste de las series de tiempo

Se procede siguiendo el proceso de especificación declarado en la sección anterior. Se denota por  $r_t$  la serie del logaritmo de los rendimientos que en adelante denominaremos rendimientos. Estas se muestran en la [Figura 3-1](#). La [Figura 3-2](#) muestra la FAC muestral de  $r_t$  indicando que no existe correlación serial de los rendimientos en ninguno de los cuatro casos.

En la [Tabla 3-1](#) se observa que para los índice bursátiles de New York, París y Singapur por medio de la prueba de Ljung-Box la correlación serial de los rendimientos es no significativa. Esto sugiere que un modelo constante para la media de los rendimientos en los tres casos es adecuado. No obstante, para el índice de Japón la correlación serial debe ser modelada. El efecto ARCH se revisa con la prueba de Ljung-Box en los rendimientos al cuadrado, la cual, es significativa e indica la presencia de efecto ARCH para los índices de París, Singapur y Japón con un nivel de significancia de 5% y para New York después del rezago 15; por lo tanto la volatilidad condicional en todos los casos debe ser modelada.

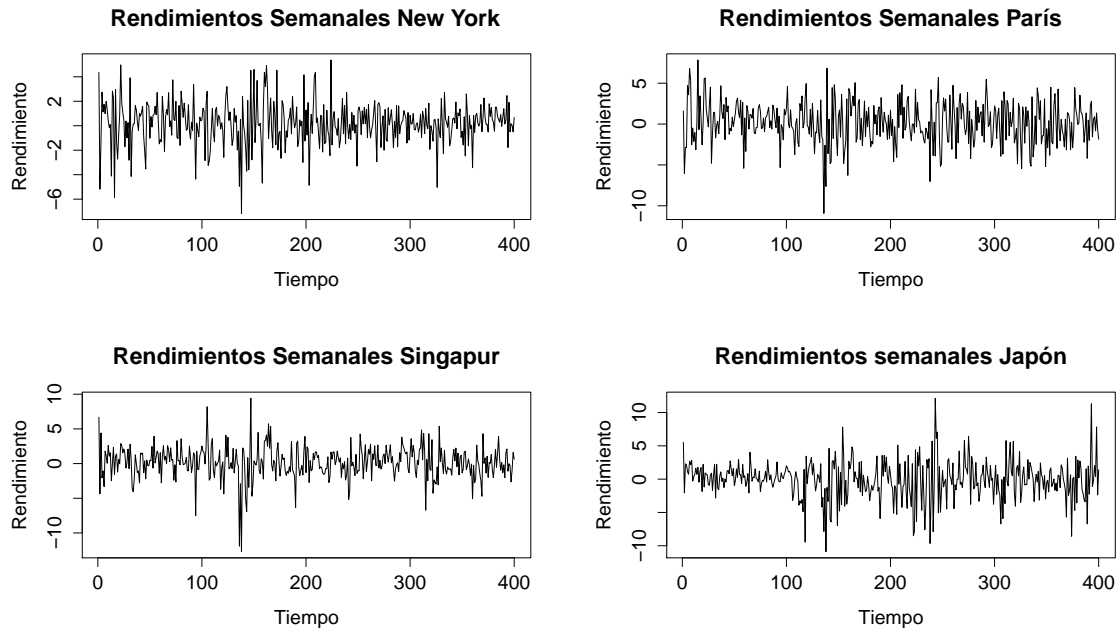


Figura 3-1 Rendimientos semanales de los índices bursátiles.

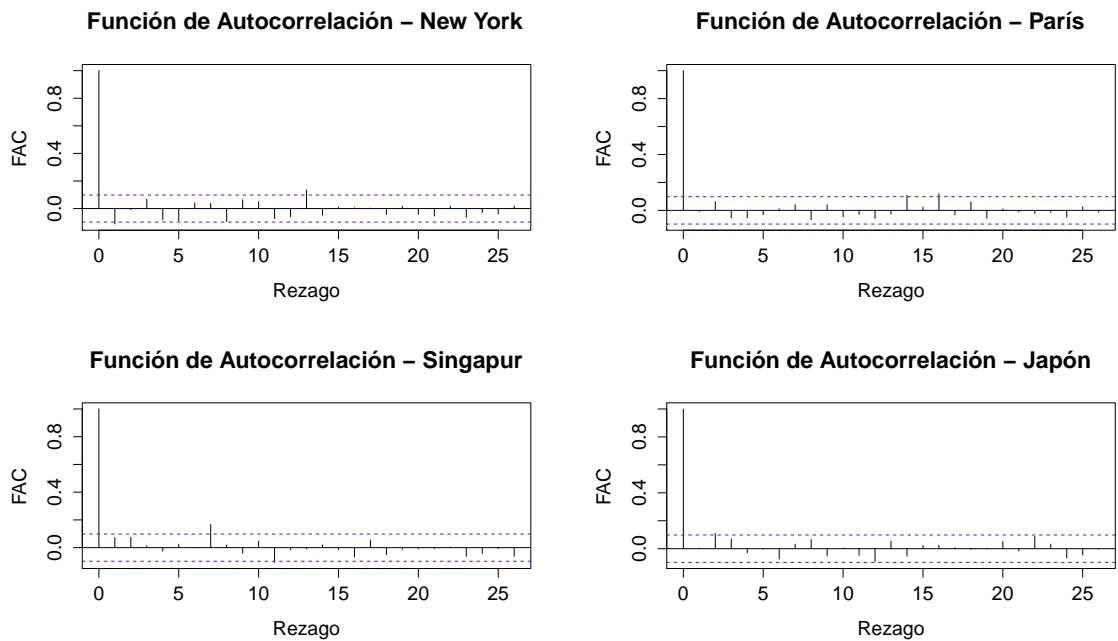


Figura 3-2 Función de Autocorrelación de  $r_t$ .

**Tabla 3-1** Prueba de Ljung-Box para los rendimientos y rendimientos al cuadrado

Prueba de Ljung-Box en los rendimientos				Prueba de Ljung-Box en los rendimientos cuadrados			
Rezagos	Estadístico	df	valor-p	Rezagos	Estadístico	df	valor-p
<b>New York</b>							
5	6.979678	5	0.2221561	5	7.34546	5	0.196189
10	9.504420	10	0.4849921	10	10.24846	10	0.418972
15	16.839902	15	0.3285248	15	36.03772	15	0.001746
20	18.520466	20	0.5531634	20	37.30252	20	0.010762
25	21.717977	25	0.6519725	25	43.47372	25	0.012411
<b>París</b>							
5	5.920460	5	0.3140356	5	21.95409	5	5.342246e-04
10	8.169383	10	0.6122959	10	36.01560	10	8.365808e-05
15	17.791279	15	0.2737976	15	39.01823	15	6.368473e-04
20	24.001631	20	0.2423209	20	47.51254	20	4.977417e-04
25	26.689720	25	0.3715611	25	52.60586	25	1.004058e-03
<b>Singapur</b>							
5	2.505597	5	0.77565184	5	26.86332	5	6.064676e-05
10	16.510063	10	0.08593271	10	34.89203	10	1.302746e-04
15	20.087632	15	0.16859672	15	45.66305	15	6.014187e-05
20	24.491518	20	0.22158062	20	46.35867	20	7.194933e-04
25	27.031235	25	0.35432587	25	47.95233	25	3.778941e-03
<b>Japón</b>							
5	18.12532	5	0.00279338	5	48.67256	5	2.588907e-09
10	21.72748	10	0.01655471	10	57.67076	10	9.970590e-09
15	33.98152	15	0.00342577	15	69.66958	15	5.115093e-09
20	39.42301	20	0.00590432	20	116.61996	20	1.221245e-15
25	54.11442	25	0.00064214	25	123.44921	25	5.440093e-15

**Tabla 3-2** Criterios de selección para TGARCH( $p, q$ )

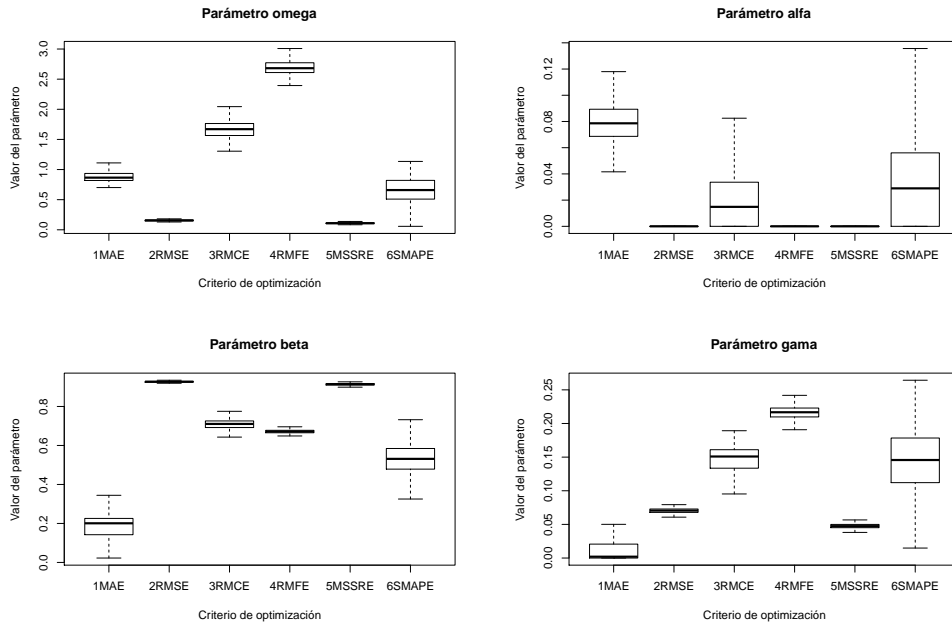
Modelo TGARCH	New York		París		Singapur		Japón	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
TGARCH(0,1)	4.2863	4.3523	4.7554	4.8544	<b>4.6132</b>	<b>4.6792</b>	4.3372	4.4196
TGARCH(1,0)	4.2941	<b>4.3435</b>	4.8024	4.8518	4.6361	4.6856	4.4737	4.5232
<b>TGARCH(1,1)</b>	<b>4.2863</b>	4.3523	<b>4.7433</b>	<b>4.8092</b>	<b>4.6132</b>	<b>4.6792</b>	<b>4.3283</b>	<b>4.3942</b>
TGARCH(1,2)	4.2969	4.3794	4.7519	4.8344	4.6228	4.7053	4.3372	4.4196
TGARCH(2,1)	4.3069	4.4059	4.7554	4.8544	4.6329	4.7318	4.3472	4.4462
TGARCH(2,2)	4.3006	4.4161	4.7635	4.8789	4.6281	4.7436	4.3526	4.4680

Siguiendo el método estándar para el ajuste del modelo, en este trabajo se usa el paquete garchFit de R y se ajustan varios modelos TGARCH( $p, q$ ) con diferentes valores de  $p$  y  $q$ . Se obtienen los valores para los criterios de información de Akaike y Bayes, AIC y BIC respectivamente, para los diferentes modelos y en cada una de las series de tiempo. Se obtienen los valores de AIC y BIC de la Tabla 3-2 para el modelo TGARCH( $p, q$ ), se selecciona el modelo TGARCH(1,1) para las cuatro series ya que con ambos criterios se obtienen los valores, en general, menores; exceptuando el valor del BIC para la serie de New York, cuyo menor valor del BIC se obtiene para el modelo TGARCH(0,1) y una igualdad de los valores en ambos criterios para la serie de Singapur de los modelos de orden (0, 1) y (1, 1). Los modelos TGARCH de orden uno han sido usados con mayor frecuencia en la literatura debido a que son fácilmente interpretables y representan de manera adecuada la heteroscedasticidad condicional.

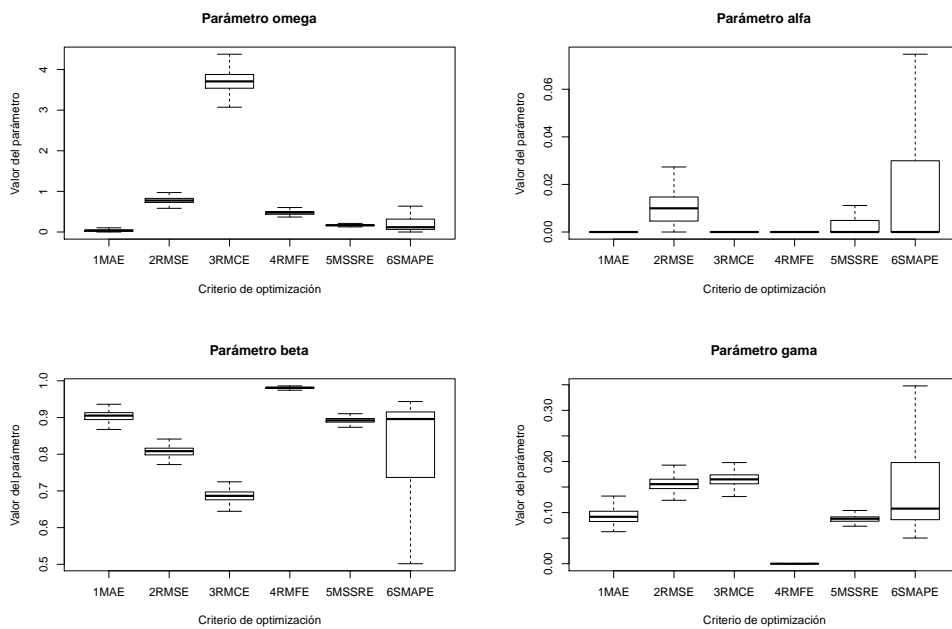
### 3.3.3 Estimación de los parámetros por bootstrapping

Siguiendo el procedimiento de estimación por bootstrapping descrito en la sección anterior, y considerando  $n = 500$  sub-muestras de la serie de tiempo se obtienen los diagramas de cajas y bigotes para cada uno de los parámetros a través de los seis criterios de optimización definidos. Las Figuras (3-3), (3-4), (3-5) y (3-6) para los parámetros estimados de los modelos muestran que existen diferencias significativas entre algunos de los parámetros dependiendo del criterio de optimización usado. Por ejemplo, para el caso de New York, los modelos 2 y 5 optimizados con los criterios de optimización RMSE y MSSRE, respectivamente, tienen valores de los parámetros estadísticamente iguales. Los demás modelos son diferentes debido a que al menos tienen un parámetro estadísticamente diferente de los demás. Para los datos de Singapur, París y Japón, se observa diferencia significativa entre algunos de los parámetros de cada modelo. Lo anterior indica que los modelos estimados con distintas funciones objetivo son diferentes entre sí, lo que significaría que los modelos dan cuenta de información complementaria contenida en los datos.

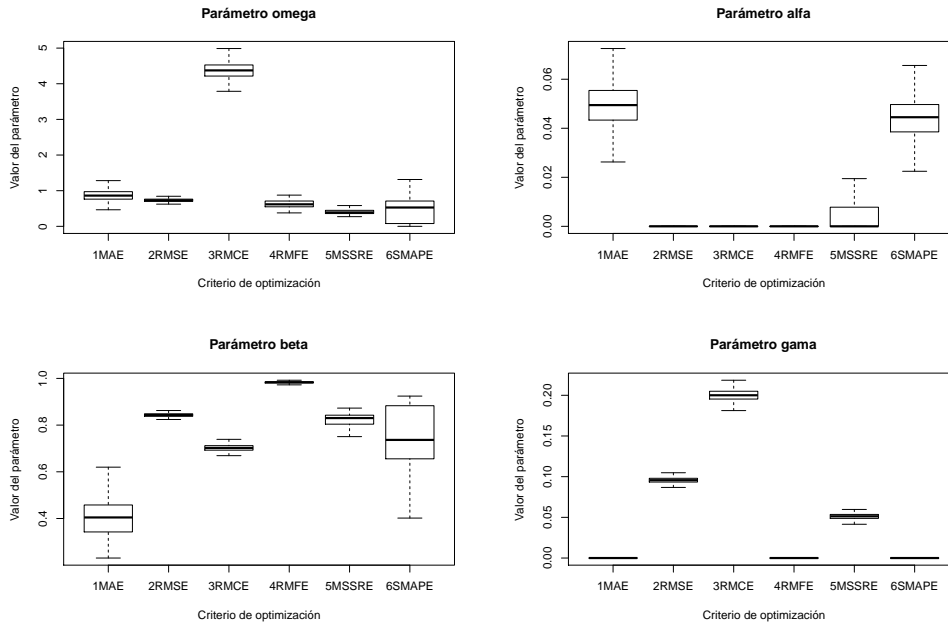
La prueba de Ljung-Box sobre los residuales estandarizados y los residuales estandarizados cuadrados para cada uno de los seis modelos considerados en cada serie de tiempo muestran que los modelos de media y de varianza son adecuados.



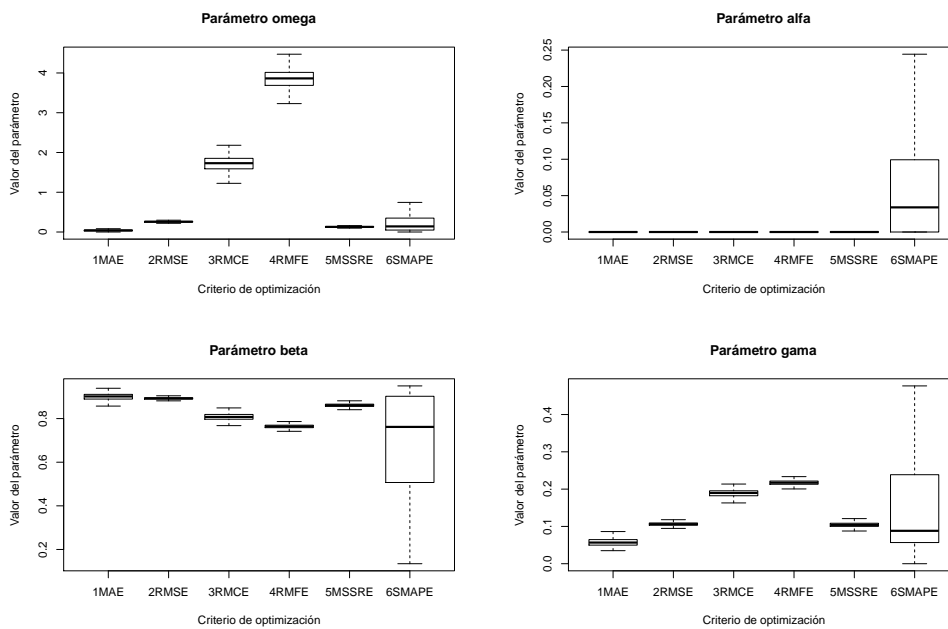
**Figura 3-3** Diagramas de cajas y bigotes de los parámetros del modelo TGARH con cada criterio de optimización para los datos de New York.



**Figura 3-4** Diagramas de cajas y bigotes de los parámetros del modelo TGARH con cada criterio de optimización para los datos de París.



**Figura 3-5** Diagramas de cajas y bigotes de los parámetros del modelo TGARH con cada criterio de optimización para los datos de Singapur.



**Figura 3-6** Diagramas de cajas y bigotes de los parámetros del modelo TGARH con cada criterio de optimización para los datos de Japón.

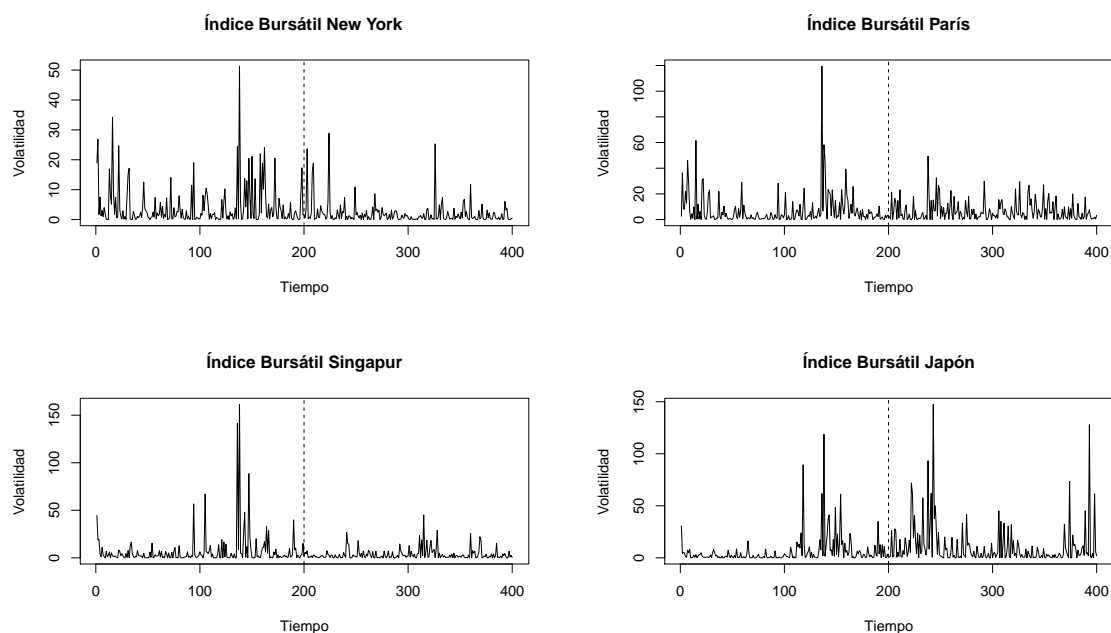
### 3.3.4 Evaluación de los modelos

La evaluación de los modelos se hace con dos medidas de exactitud, MAE, RMSE en ambas muestras: entrenamiento y pronóstico para cada serie de tiempo considerada y para cada uno de los modelos propuestos. En la Tabla 3-3 se reportan los valores promedio del MAE y RMSE así como las respectivas desviaciones estándar obtenidas con la metodología de remuestreo considerada. Se comparan los valores del RMSE con los de los mejores modelos del artículo de Taylor (2004) en la muestra de prueba de cada una de las cuatro series de tiempo consideradas. En la mayoría de los modelos propuestos en este trabajo, se obtiene una mayor exactitud con respecto a lo reportado en el artículo. Por ejemplo, para New York en la muestra de prueba comparando el valor del RMSE obtenido en los seis modelos entrenados, cuatro de ellos tienen valores promedio del RMSE de: 393.1, 400.6, 386.3 y 392.2, que están alrededor del valor del mejor modelo reportado en el artículo base, que es de 385, pero ninguno superó ese nivel de exactitud. No obstante, para los datos de París, se superó el mejor modelo del artículo de referencia con un RMSE de 783, con el modelo 2 el cual es el TGARCH(1, 1) optimizado con el RMSE. Así mismo, para los datos de Singapur el valor 620 del RMSE en el artículo de referencia es superado por los modelos 5 y 6, con valores promedio del RMSE de 600.8 y 610.3 respectivamente. Para la serie de tiempo de Japón, cuatro de los modelos propuestos superan al mejor modelo reportado en el trabajo de referencia que tiene un RMSE de 2023, aunque el modelo 5 con un RMSE de 2018 y una desviación estándar de 47.2 indica que estadísticamente este modelo tiene el mismo desempeño que el del artículo con respecto al RMSE.

La comparación en la muestra de prueba de los valores del RMSE en la Tabla 3-3 se realiza ante la ausencia de esta medida de exactitud en la muestra de entrenamiento en el trabajo de referencia. No obstante, para los resultados presentados en esta Tabla se indican los valores más pequeños de cada una de las medidas de exactitud en cada columna.

Se observa que para las series de tiempo de New York, París y Singapur, los modelos 3 y 4 estimados usando como función objetivo RMCE y RMFE respectivamente, producen valores del RMSE muy altos y además alejados de los valores del RMSE de los modelos restantes. Esto puede explicarse así: en primer lugar, debido a que estas funciones objetivo son bastante sensibles a los valores extremos, entonces hacen que los parámetros se re-acomoden para que el modelo de cuenta de estos valores extremos; si además hay muchos valores de bajas volatilidades, el error que se mide entre el valor observado y lo que el modelo predice tiende a ser grande. También dependerá del tamaño y el número de valores extremos presentes en la serie de las volatilidades.

A partir de la Figura 3-7 y de la Tabla 3-3 se observan varios hechos: Primero, para la serie de tiempo de New York una mayor volatilidad en la muestra de entrenamiento con respecto a la muestra de prueba, por lo cual se espera mayor sensibilidad en la primera muestra, de los modelos que se estiman con los criterios de optimización que castigan más fuertemente los valores



**Figura 3-7** Volatilidad semanal de los índices bursátiles.

extremos, como por ejemplo los modelos estimados con los criterios de optimización RMCE y RMFE, los cuales tuvieron mayores valores en la medida de exactitud RMSE.

Segundo, en los datos del índice bursátil de Singapur se observan bajas volatilidades sobretodo en la muestra de pronóstico, por lo que los valores pequeños de MAE y RMSE para varios de los modelos están acorde a los datos. Mientras que, mayor volatilidad en la muestra de entrenamiento produce mayores valores de estas medidas de exactitud en los modelos.

Tercero, para el índice Japonés, se observan muy bajas volatilidades en la primera mitad de la muestra de entrenamiento en comparación con el resto de los datos. Esta puede ser la razón de valor nulo en el parámetro  $\alpha$  en 5 de los 6 modelos considerados en este trabajo. Otra característica de interés en estos datos, se debe a la poca variación de los valores del RMSE de los seis modelos en ambas muestras; ya que en los demás conjuntos de datos, se tiene una variación mayor con respecto a esta medida de exactitud.

Con base en los resultados obtenidos se deben considerar la influencia de la variabilidad de los datos sobre los modelos. De acuerdo a la Figura 3-7 los valores extremos en los datos del índice de New York alcanzan el valor de 50 con baja frecuencia ya que la mayoría de los agrupamientos de volatilidad que se presentan alcanzan solo el valor 20. En estos datos, ninguno de los modelos propuestos supera estadísticamente el mejor modelo del artículo base (Con respecto a la medida

de RMSE en la muestra de prueba). Sin embargo, cuatro de los seis modelos sí tuvieron valores estadísticamente iguales de esta medida.

Para el índice Parisino, los valores extremos están alrededor 50 y 60, con excepción de un par de datos alrededor de 100. En este caso, uno de los seis modelos supera el mejor modelo del artículo.

Mientras que en los índices de Singapur y Japón se presentan valores extremos alrededor de 150 y el primero de estos con un agrupamiento de alta volatilidad y aproximadamente tres agrupamientos de baja volatilidad, dos de los modelos propuestos resultan mejores que el mejor del artículo con respecto al RMSE. Por otro lado, para el índice Japonés se observan alrededor de seis agrupamientos de volatilidad de los cuales cuatro son de alta volatilidad. Para este caso, hay cuatro modelos que superan el modelo del artículo.

### 3.4 Conclusiones

Se ha considerado el proceso de especificación convencional para el modelo de volatilidad condicional con el fin de obtener la estructura general del modelo para cada conjunto de datos. Una vez definida la estructura, se ha implementado la propuesta de estimación de los modelos con diferentes criterios de optimización y con el método de remuestreo. Lo anterior ha permitido hacer comparaciones estadísticas múltiples de los parámetros poblacionales basadas en lo obtenido del mecanismo de remuestreo. Además, de concluir que existen diferencias significativas entre estos modelos. Se ha superado el desempeño, en la muestra de prueba, de los mejores modelos para cada serie de tiempo del artículo de referencia. Con base en la evidencia empírica se concluiría además que los criterios de optimización permiten obtener conjuntos de pronóstico diferentes. Se concluye que, aunque algunos modelos tienen una medida de exactitud estadísticamente mayor al resto, estos capturan los picos más altos de volatilidad. Otros de los modelos estimados con esta metodología son menos sensibles a las altas volatilidades y capturan el comportamiento medio y bajo.

**Tabla 3-3** Medidas de Exactitud Artículo base vs modelos TGARCH(1,1) con diferentes criterios de optimización.

Modelo	Activo Financiero							
	New York		París		Singapur		Japón	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
<b>Muestra de Entrenamiento</b>								
CO-MAE	377.6*	747.4	585.5*	1287.9	649.4*	1884.1	526.0*	1340.5
	(4.4)	(12.0)	(8.4)	(28.6)	(7.8)	(32.7)	(9.4)	(34.7)
CO-RMSE	439.6	690.7*	671.6	1237.0*	801.5	1789.3*	612.5	1276.8*
	(6.0)	(11.4)	(11.1)	(28.2)	(10.8)	(31.8)	(11.4)	(32.9)
CO-RMCE	603.8	750.3	1086.4	1395.9	1531.4	2069.0	1041.4	1444.9
	(10.4)	(13.7)	(28.5)	(38.0)	(28.6)	(39.2)	(31.1)	(41.3)
CO-RMFE	775.1	877.0	1690.0	1912.0	2513.6	2842.0	1642.7	1875.5
	(19.9)	(21.8)	(57.7)	(63.8)	(68.9)	(65.1)	(66.5)	(69.8)
CO-MSSRE	387.1	711.9	594.1	1269.8	674.6	1831.6	539.9	1302.4
	(4.8)	(11.6)	(9.1)	(28.3)	(7.8)	(30.4)	(9.0)	(30.0)
CO-SMAPE	386.6	721.5	597.7	1273.1	664.2	1846.2	557.5	1342.8
	(6.1)	(13.6)	(13.2)	(31.0)	(11.1)	(32.1)	(25.6)	(53.7)
<b>Muestra de pronóstico</b>								
<b>Mejor modelo (Taylor)</b>		<b>385</b>		<b>783</b>		<b>620</b>		<b>2023</b>
CO-MAE	183.4	393.1	504.4	816.2	322.8	628.7	953.5	2057.8
	(2.1)	(1.5)	(1.8)	(7.8)	(0.6)	(3.5)	(3.1)	(16.7)
CO-RMSE	265.9	400.6	574.1	<b>770.4*</b>	469.7	626.7	1010.0	<b>1915.5*</b>
	(9.7)	(5.0)	(3.5)	(2.0)	(4.3)	(2.1)	(4.3)	(3.8)
CO-RMCE	528.6	583.1	948.2	1050.3	1294.2	1347.2	1334.7	<b>1939.1*</b>
	(10.7)	(8.9)	(20.6)	(19.6)	(24.4)	(24.3)	(16.9)	(7.3)
CO-RMFE	754.5	785.5	1866.2	1975.3	3317.7	3363.6	1795.2	2197.6
	(22.0)	(20.7)	(102.3)	(101.8)	(163.6)	(162.5)	(56.9)	(39.8)
CO-MSSRE	199.3	386.3	507.5	790.5	334.7	<b>600.8*</b>	947.5	<b>1988.1*</b>
	(3.0)	(0.6)	(1.2)	(2.6)	(1.4)	(1.1)	(1.9)	(5.2)
CO-SMAPE	203.3	392.2	514.2	796.7	328.6	<b>610.3.7*</b>	983.9	<b>2018.4*</b>
	(9.6)	(2.6)	(9.1)	(12.8)	(6.4)	(6.9)	(31.3)	(47.2)

Las medidas de exactitud fueron multiplicadas por  $10^6$

## 4 Combinación y dominancia de pronósticos

Del capítulo anterior se tienen disponibles varios conjuntos de pronósticos de una misma variable para cada serie de tiempo considerada. Se tienen dos opciones: la primera es seleccionar un único pronóstico, por ejemplo, el del modelo con mejor desempeño; sin embargo, esto puede ser bastante riesgoso ya que no estaría teniendo en cuenta la información proporcionada por los demás modelos. La segunda opción es combinar los pronósticos, lo que equivale a combinar la información contenida en los diferentes modelos; En esta, la combinación de pronósticos se ve como un mecanismo para controlar el riesgo financiero.

No obstante, como ha sido reportado en la literatura, no siempre se obtiene un pronóstico combinado con mejor desempeño. Así que es de interés saber cuando es conveniente hacer la combinación de los pronósticos; que para el caso de combinaciones lineales de pronósticos se cuenta con una prueba de dominancia de pronósticos en inglés “*forecasts encompassing*”. Sin embargo, cuando se plantea realizar una combinación no lineal de los conjuntos de pronóstico individuales no existe una prueba estadística específica que permita determinar la contribución de cada uno a la combinación.

El objetivo de este capítulo es presentar brevemente los principales métodos de combinación de pronósticos y la teoría existente en la literatura en pruebas de dominancia de pronósticos para determinar la significancia de un conjunto de pronósticos en una combinación. Así, el aporte principal de este capítulo es presentar y aplicar a nivel conceptual y experimental un contraste de hipótesis estadística de “variables relevante” para combinaciones no lineales de pronósticos con redes neuronales artificiales.

### 4.1 Métodos tradicionales de combinación de pronósticos

Uno de los métodos más comunes para obtener el pronóstico combinado  $f^*(t)$  es el promedio aritmético simple, el cual fue utilizado originalmente en los trabajos de [Reid \(1968\)](#) y [Barnard \(1963\)](#):

$$f_c(t) = \frac{1}{N} \sum_{i=1}^N f_i(t) \quad (4-1)$$

Una variación a este método es el promedio aritmético con recorte [Jose y Winkler \(2008\)](#); en este caso, se desecha un porcentaje de los modelos con peor desempeño antes de calcular el promedio. El modelo de combinación de pronósticos descrito en la ec. (4-1) puede ser generalizado al considerar que la combinación puede ser realizada como un promedio ponderado:

$$f_c(t) = \sum_{i=1}^N w_i(t) f_i(t) \quad (4-2)$$

Surge, entonces, el problema de cómo calcular los pesos  $w_i$  de forma óptima. Existen diferentes maneras de abordar este problema, por ejemplo, [Bates y Granger \(1969\)](#) consideran este problema cuando se tienen dos conjuntos de pronósticos y las varianzas de sus errores constantes,  $\sigma_1^2$  y  $\sigma_2^2$ . Se asignan pesos  $\lambda$  y  $1 - \lambda$  a los pronósticos del primer y segundo modelo respectivamente. Bajo estos supuestos, la varianza del error de la combinación de los pronósticos queda determinada por:

$$\sigma_c^2 = \lambda^2 \sigma_1^2 + (1 - \lambda)^2 \sigma_2^2 + 2\rho \lambda \sigma_1 (1 - \lambda) \sigma_2 \quad (4-3)$$

donde  $\rho$  es la correlación entre los errores de los pronósticos de los modelos. El valor óptimo de  $\lambda$  se obtiene minimizando  $\sigma_c^2$  y está dado por:

$$\lambda = \frac{\sigma_2^2 - \rho \sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho \sigma_1 \sigma_2} \quad (4-4)$$

En el caso en el que  $\rho = 0$ , la expresión anterior se reduce a:

$$\lambda = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad (4-5)$$

[Bates y Granger \(1969\)](#) demuestran que el valor óptimo de  $\sigma_c^2$  no podrá ser mayor que la menor de las varianzas individuales. Es decir, que el pronóstico combinado produce un mejoramiento con respecto al error cuadrático medio excepto cuando los pronósticos individuales son igual de precisos,  $\sigma_1^2 = \sigma_2^2$ , y están altamente correlacionados  $\rho = 1$ .

Otras aproximaciones parten del supuesto de que el pronóstico combinado puede calcularse como una combinación lineal óptima de los pronósticos individuales que minimiza una medida de error entre el pronóstico combinado y el valor real observado  $y(t)$ . Otros autores proponen el método de ensamble generalizado, en el cual, el pronóstico combinado es calculado usando la ec. (4-2), y los pesos son estimados como:

$$w_i = \frac{\sum_j C_{ij}^{-1}}{\sum_k \sum_j C_{kj}^{-1}} \quad (4-6)$$

donde  $C_{ij}$  son los elementos de la matriz de covarianzas de los errores de los pronósticos individuales  $f_i(t)$  y  $f_j(t)$ ; donde se cumple que  $\sum_{i=1}^N w_i = 1$ .

[Hashem \(1997\)](#) considera la determinación de los pesos en la ec. (4-2) como un problema de regresión en el que se desea minimizar el error cuadrático medio entre el pronóstico combinado y el valor real  $y(t)$ ; Además, considera los siguientes problemas de regresión para la determinación de los pesos:

- Combinación lineal sin restricciones con un término constante.
- Combinación lineal restringida a que los pesos deben sumar la unidad y con un término constante.
- Combinación lineal sin restricciones sin término constante.
- Combinación lineal restringida a que los pesos deben sumar la unidad y sin término constante.

Adicional, a las restricciones anteriores, se suele incorporar la restricción de que los pesos sean mayores o iguales a cero.

## 4.2 Dominancia de pronósticos

En combinación de pronósticos, es deseable determinar la contribución de un conjunto de pronósticos individual a la combinación ([Anders y Korn, 1999](#); [Clements, M. y Harvey, D., 2008](#)). Debido a que existen evidencias empíricas de que el pronóstico combinado es, con frecuencia, mejor que los individuales, es necesario determinar si debe incluirse un conjunto de pronósticos en la combinación.

Entre los primeros trabajos publicados se encuentran el de [Bates y Granger \(1969\)](#) en el cual combina linealmente dos conjuntos de pronósticos minimizando la varianza del error del pronóstico combinado. [Nelson \(1972\)](#) y [J. Phillip Cooper \(1975\)](#) aportan los primeros resultados en miras a la construcción de una prueba de especificación en la combinación de pronósticos por medio de una regresión lineal. Posteriormente, [Chong y Hendry \(1986\)](#) introducen el concepto de dominancia de pronósticos en el contexto de combinación de pronósticos y lo usan para la evaluación del pronóstico. Algunos trabajos posteriores socializan la relación entre los conceptos combinación de pronósticos y dominancia de pronósticos, ver por ejemplo [Diebold \(1989\)](#); Una prueba adicional de precisión predictiva es presentada por [Diebold y Mariano \(1995\)](#). Finalmente, [Harvey et al. \(1998\)](#) y [Harvey y Newbold \(2000\)](#) introducen la prueba de dominancia de pronósticos en combinaciones lineales de dos o mas pronósticos.

Esta prueba es una herramienta útil para saber si varios conjuntos de pronóstico pueden combinarse para producir un pronóstico compuesto superior [Fang \(2003\)](#). Hay dos enfoques cuando se tiene una gama de pronósticos y se desea combinarlos. El primero consiste en realizar una prueba de dominancia de pronósticos antes de combinarlos; está basada en una prueba de igualdad de precisión de los conjuntos de pronósticos presentada por [Harvey et al. \(1998\)](#), la cual es una modificación de la prueba propuesta por [Diebold y Mariano \(1995\)](#). Este enfoque puede usarse como un criterio para decidir si se agregan o no pronósticos a la combinación. El segundo enfoque, es una prueba de dominancia aplicada después de efectuar la combinación; puede ser vista

como una estrategia de selección del modelo ya que prueba si el pronóstico combinado resulta en una reducción estadísticamente significativa en relación a cada uno de los pronósticos individuales con respecto a una medida de error, por ejemplo MSE (Clements y Hendry, 2008; Timmermann, 2006). En este trabajo es de interés el segundo enfoque y para ello se presentan las definiciones subsiguientes.

#### 4.2.1 Prueba de dominancia de pronósticos en una regresión lineal

Se encuentran varias definiciones de dominancia de pronósticos de acuerdo al tipo de combinación lineal. Una descripción breve de cada una de ellas se presenta a continuación.

En una combinación lineal de pronósticos y usando una función de error cuadrática se presentan dos definiciones de *dominancia de pronósticos*, en donde,  $\varepsilon_t$  es el error del pronóstico compuesto y  $e_{1t}$ ,  $e_{2t}$  representan los errores de los pronósticos individuales  $f_{1t}$  y  $f_{2t}$ .

**Definición 1.**  $f_{1t}$  domina a  $f_{2t}$  si  $f_{2t}$  recibe un peso óptimo de cero,  $\beta_2 = 0$ , en la regresión

$$y_t = \alpha + \beta_1 f_{1t} + \beta_2 f_{2t} + \varepsilon_t.$$

**Definición 2.** Imponiendo las restricciones  $\alpha = 0$  y  $\beta_1 + \beta_2 = 1$ ;  $f_{1t}$  domina a  $f_{2t}$  si  $\lambda = 0$  en la regresión

$$e_{it} = \lambda(e_{1t} - e_{2t}) + \varepsilon_t.$$

Se sigue de las definiciones que el contraste de hipótesis de dominancia de pronósticos está dado por

$$\begin{aligned} H_0 : \beta_2 = 0 & \quad f_{1t} \text{ domina a } f_{2t} \\ H_1 : \beta_2 > 0 & \quad f_{1t} \text{ no domina a } f_{2t} \end{aligned}$$

y cambiando  $\beta$  por  $\lambda$  para la segunda definición.

Bajo el supuesto de errores de pronóstico con distribución normal bivariada, la prueba se llevaría a cabo como en el caso de una regresión estándar. No obstante, este supuesto es difícil de sostener por lo que debe construirse un estadístico robusto ante desviaciones del supuesto. El estadístico que se usa para contrastar estas hipótesis resulta entonces de una modificación del estadístico de prueba de una regresión estándar y está dado por

$$R = \frac{\hat{\lambda}}{\sqrt{\frac{\sum_{\tau=-(h-1)}^{h-1} \sum_{t=|\tau|+1}^n (e_{1t} - e_{2t}) \hat{\varepsilon}_t (e_{1,t-|\tau|} - e_{2,t-|\tau|}) \hat{\varepsilon}_{t-|\tau|}}{\sum_{t=1}^n (e_{1t} - e_{2t})^2}}}}$$

donde  $\hat{\lambda}$  es el estimador de mínimos cuadrados ordinarios de  $\lambda$  en la segunda regresión,  $\hat{\varepsilon}_t$  es el residual de la regresión ajustada; y bajo el supuesto de dependencia de los  $h - 1$  errores de

pronóstico cuando se hace pronóstico  $h$  pasos adelante, el estadístico tiene una distribución normal asintótica bajo la hipótesis nula de dominancia de pronósticos. Para profundizar en el proceso de construcción y propiedades asociadas se puede consultar [Harvey et al. \(1998\)](#).

El estadístico anterior ha sido construido con base a una regresión lineal y sirve para contrastar hipótesis de dominancia de pronósticos en combinaciones lineales, como se muestra en la definición. Para el caso de combinaciones no lineales de pronósticos no existe una prueba, o incluso una definición explícita, para dominancia de pronósticos. Por tal razón, en la sección siguiente se introduce una extensión de tal definición y un contraste de hipótesis para dominancia de pronósticos en combinaciones con redes neuronales artificiales.

### 4.2.2 Dominancia de pronósticos con redes neuronales artificiales

En primer lugar, se define la forma no lineal que se usa como combinador y posteriormente se propone una definición de dominancia de pronósticos para este tipo de combinación. Una prueba estadística de dominancia es también presentada.

El combinador seleccionado es una red neuronal tipo perceptrón de una capa oculta definido por

$$Y_t = f(\mathbf{X}_t; \theta) + \varepsilon_t \quad (4-7)$$

donde

$$f(\mathbf{X}_t; \theta) = \phi_0 + \mathbf{X}_t' \phi + \sum_{j=1}^q \beta_j (1 + \exp(-(\gamma_{0j} + \mathbf{X}_t' \gamma_j)))^{-1} \quad (4-8)$$

donde  $\theta$  es el vector de parámetros que contiene todos los parámetros  $\phi$ ,  $\beta$  y  $\gamma$ , con  $\gamma_j' = (\gamma_{1j}, \dots, \gamma_{pj})$  y función de transferencia logística,  $Y_t$  es la variable objetivo.  $f(\mathbf{X}_t; \theta)$  es una función no lineal de valor real,  $f: \mathbb{R}^p \times \Theta \rightarrow \mathbb{R}$ ,  $\Theta$  denota el espacio de parámetros.  $\mathbf{X}_t$  es un vector con  $p$  variables explicativas  $\mathbf{X}_t = (Y_{t1}, Y_{t2}, \dots, Y_{tp})$  y  $\varepsilon_t$  es el error entre el valor observado  $Y_t$  y lo que el modelo predice.

En el contexto de redes neuronales artificiales, esta especificación contiene  $p$  unidades en la capa de entrada,  $q$  unidades en la capa oculta, con función de activación logística  $G_j = G(\gamma_{0j} + \mathbf{X}_t' \gamma_j)$ , y una unidad en la capa de salida. Un modelo de ANN de la forma (4-8) se le conoce como red neuronal artificial aumentada de una capa oculta, ver [Kuan y White \(1994\)](#). El vector de parámetros contiene  $l = p + 1 + q(p + 2)$  parámetros, donde:

$$\begin{aligned} \phi' &= (\phi_1, \dots, \phi_p) \\ \beta' &= (\beta_1, \dots, \beta_q) \\ \gamma' &= (\gamma_1', \dots, \gamma_p'); \text{ cada } \gamma_i' = (\gamma_{i1}, \dots, \gamma_{iq}) \\ \gamma_0' &= (\gamma_{01}, \dots, \gamma_{0q}) \end{aligned}$$

Esta estructura ha sido considerada gracias a que los autores [Hornik et al. \(1989\)](#) han probado que un modelo de la forma (4-8) puede aproximar cualquier función medible a cualquier grado de precisión (bajo una métrica apropiada) y con suficientes unidades en la capa oculta. Esta arquitectura de red neuronal ha sido considerada en el análisis de series de tiempo econométricas ([Franses y Dijk, 2000](#)).

Los parámetros relacionados con una entrada particular  $Y_{ti}$  son seleccionados de un vector completo de  $l$  parámetros,  $\theta$ , usando una matriz de selección  $S_i$ , de tal forma que  $S_i\theta = (\phi_i, \gamma_{i1}, \dots, \gamma_{iq})$ . La matriz de selección de tamaño  $(q+1) \times l$  contiene ceros y unos en sus entradas.

La definición propuesta en este capítulo para dominancia de pronósticos es una extensión natural de la definición existente para dominancia de pronósticos en combinaciones lineales.

**Definición 3.** Un conjunto de pronósticos  $Y_{ti}$  es dominado por sus competidores, basados en combinación no lineal de  $p$  pronósticos definida en la ecuación (4-8), si todos los parámetros relacionados con la entrada  $Y_{ti}$  son cero.

El contraste de hipótesis que se asocia a esta definición tiene la forma:

$$H_0 : S_i\theta = \mathbf{0} \quad Y_{ti} \text{ es dominado por los demás conjuntos de pronósticos}$$

vs

$$H_1 : S_i\theta \neq \mathbf{0} \quad Y_{ti} \text{ no es dominado por los demás conjuntos de pronósticos}$$

El estadístico de prueba idóneo que permite contrastar esta hipótesis se define mas adelante.

Este contraste será la base para construir la prueba de Wald que permitirá realizar una prueba de dominancia de pronósticos en combinaciones no lineales.

## 4.3 Proceso de entrenamiento de una red neuronal artificial

El proceso de estimación de los parámetros de la red está basado en el método de mínimos cuadrados no lineales y consiste en encontrar una superficie que ajuste el conjunto de datos  $(y_t, \mathbf{x}_t)$ ,  $t = 1, \dots, N$ , donde  $y_t$  es el valor observado de la serie en el tiempo  $t$  y  $\mathbf{x}_t$  es el vector que contiene los  $p$  pronósticos en el tiempo  $t$ . En términos generales es deseable tener propiedades de los estimadores de los parámetros de un modelo, en este caso para los estimadores de la red se da la definición de distribución asintótica de la teoría de inferencia asintótica dada por [White \(1994\)](#).

**Definición 4.** Sean  $\{X_n\}$  y  $\{Z_n\}$  sucesiones de variables aleatorias con funciones de distribución  $F_n(x) = P(X_n \leq x)$ ,  $G_n(x) = P(Z_n \leq x)$ . Sea  $\mathcal{D}_n$  el conjunto de discontinuidades de  $G_n$ , y defina  $\mathcal{D} \equiv \cup_{n=1}^{\infty} \mathcal{D}_n$ . Entonces,  $X_n$  se distribuye asintóticamente como  $Z_n$ ,  $X_n \overset{A}{\sim} Z_n$ , si para cada  $x \in \mathcal{D}^c$  se tiene que  $F_n(x) - G_n(x) \rightarrow 0$  cuando  $n \rightarrow \infty$ .

El estimador  $\hat{\theta}$  de mínimos cuadrados no lineales que se considera para este problema de optimización es la solución de

$$\min_{\theta \in \Theta} Q_N(\theta) = \min_{\theta \in \Theta} N^{-1} \sum_{t=1}^N [Y_t - f(\mathbf{X}_t; \theta)]^2 \quad (4-9)$$

donde  $f$  se considera que tiene la forma de la expresión (4-8). Considerando la definición anterior para el vector de parámetros  $\hat{\theta}$  de mínimos cuadrados y el vector de parámetros  $\theta^*$  de mínimos cuadrados óptimos, entonces se tiene que  $\hat{\theta}$  converge asintóticamente al vector de parámetros  $\theta^*$  que se obtiene al minimizar la expresión  $E[(E[Y_t|X_t] - f(\mathbf{X}_t; \theta))^2]$  White (1994).

Bajo la teoría asintótica de White, permiten concluir que el estimador de mínimos cuadrados no lineales

$$\hat{\theta}_N = \arg \min Q_N(\theta)$$

es generalmente consistente para  $\theta^*$  y tiene distribución normal asintótica. En otros términos,  $\hat{\theta}_N \rightarrow \theta^*$  cuando  $N \rightarrow \infty$ . En particular,  $N^{1/2}(\hat{\theta}_N - \theta^*) \overset{A}{\sim} N(0, C^*)$ , donde la matriz de varianzas y covarianzas está dada por  $C^* = A^{*-1} B^* A^{*-1}$ , con  $A^* = E(\nabla^2 Q_N(\theta^*))$  y  $B^* = \text{Var}(N^{1/2} \nabla Q_N(\theta^*))$ , donde  $\nabla$  denota el operador gradiente con respecto al vector de parámetros  $\theta$ .

En este caso es necesario contar con estimadores para estas expresiones y por lo tanto para la matriz de varianzas y covarianzas. Los estimadores consistentes de  $A^*$  y  $B^*$  están dados por  $\hat{A}_N = \nabla^2 Q_N(\hat{\theta}_N)$  y  $\hat{B}_N = N^{-1} \sum_{t=1}^N \hat{\varepsilon}_t^2 \nabla \hat{f}_t' \nabla \hat{f}_t$  donde  $\hat{\varepsilon}_t = Y_t - f_t(\hat{\theta}_N)$  es la diferencia entre el valor observado y lo que el modelo predice y  $\nabla \hat{f}_t = \nabla f(X_t, \hat{\theta}_N)$ .

Por otro lado, para mantener la consistencia de los estimadores es necesario imponer algunas condiciones formales, que para el caso funciones de tipo redes neuronales con una capa oculta son:

1. La sucesión de entrenamiento es una sucesión de vectores aleatorios independientes idénticamente distribuidos cuya norma euclidiana es finita.
2. La función de transferencia en la ecuación (4-8) es continuamente diferenciable de orden dos en  $\mathbb{R}$ .

Además, una solución de este problema de minimización debe satisfacer dos condiciones de orden como muestra Kuan (2008); la primera es  $\nabla_{\theta} Q_N(\theta) = 0$ , la cual es un sistema de  $l$  ecuaciones no lineales con  $l$  variables desconocidas correspondientes a las conexiones de la red neuronal. La segunda condición de orden consiste en garantizar que  $\nabla^2 Q_N(\theta)$  sea definida positiva.

Estas condiciones son necesarias para asegurar que se puede hallar un mínimo. Sin embargo, no garantizan que la solución hallada sea única. Es más, lo que se tiene es que existe una solución que minimiza la función no lineal (4-9). No obstante, el problema de minimización puede no tener una solución de forma cerrada o analítica debido a que la primera condición de orden es un

sistema de ecuaciones no lineales. En este caso, los parámetros deben estimarse usando métodos numéricos.

Los algoritmos usados para optimizar una función no lineal son, en general, iterativos, inician en un valor semilla de la variable a estimar y lo evalúan en la función objetivo hasta que un nuevo valor sea obtenido de acuerdo a una fórmula particular. Los algoritmos más aplicados son Newton-Rapson, Gauss-Newton, Steepest descent, the Quasi-Newton, el método BFGS, Nelder-Mead, entre otros. Sin embargo, estos algoritmos usan solo el término de primer orden de la expansión de Taylor para aproximar el hessiano de la función objetivo.

## 4.4 Prueba de Wald

Para probar una hipótesis acerca de un conjunto de parámetros que han sido estimados a partir de un conjunto de datos existen tres pruebas diferentes que asintóticamente son equivalentes. Estas son: la prueba de Wald, de Razón de Verosimilitud y la prueba de los Multiplicadores de Lagrange. Los estadísticos asociados a cada una siguen una distribución Chi-cuadrado con grados de libertad igual al número de parámetros especificados en la hipótesis nula. Las diferencias entre estas pruebas son principalmente de costo computacional en el cálculo del estadístico y de procedimiento con respecto a las hipótesis que se prueban.

La prueba de Wald asume el modelo de la hipótesis alternativa y mide que tanto se aleja de la hipótesis nula y con base en la muestra rechaza o no la hipótesis nula. Contrariamente, la prueba de los Multiplicadores de Lagrange, comienza asumiendo el modelo de la hipótesis nula y considera moverse al modelo de la hipótesis alternativa, es también llamada una prueba de hipótesis hacia adelante. Por otro lado la prueba de la Razón de Verosimilitud compara las dos hipótesis directamente. Más acerca de estas pruebas puede encontrarse en [Engle \(1984\)](#). En este trabajo se implementará el enfoque de la prueba de Wald

La prueba de Wald tiene la ventaja de ser aplicable en muchos ámbitos, además sirve para contrastar la coherencia de afirmar un valor concreto de un conjunto de parámetros de un modelo que ya ha sido previamente ajustado. Así que, para determinar si un nodo de entrada en el modelo (4-8) es importante para lograr la mejor aproximación local se establece una hipótesis de “variables irrelevantes” mediante la restricción  $H_0 : S_i\theta = \mathbf{0}$  contra  $H_1 : S_i\theta \neq \mathbf{0}$ , donde  $S_i$  es la matriz de selección de tamaño  $(q + 1) \times (p + 1 + q(p + 2))$ , que selecciona los elementos  $\phi_1$  y  $\gamma_{i,j}$ ,  $i = 1, \dots, p$ ,  $j = 1, \dots, q$  relacionados con la variable de entrada  $y_{ti}$ . Entre otros autores, [White \(1989\)](#), [Kuan y White \(1994\)](#) y [Franses y Dijk \(2000\)](#) proponen el estadístico de Wald para contrastar esta hipótesis.

Estas hipótesis para probar “variables irrelevantes” justamente coinciden con las hipótesis de la

definición de dominancia de pronósticos dada anteriormente. Por lo tanto el contraste propuesto a continuación es válido para probar la dominancia de pronósticos en una combinación no lineal de los mismos.

El estadístico de Wald tiene la forma:

$$W_i = N(S_i\theta)'(S_i'\Sigma S_i)^{-1}(S_i\theta) \quad (4-10)$$

donde  $\Sigma = A_n^{-1}B_nA_n^{-1}$  representa la matriz de varianzas y covarianzas asintótica de  $N^{-1/2}(\theta_n - \theta^*)$ ; tal y como se considera en la sección anterior los estimadores de  $A$  y  $B$  son consistentes y están dados por  $\hat{A} = \nabla^2 Q_N(\hat{\theta}_n)$  y  $\hat{B} = N^{-1} \sum_{t=1}^N \hat{\epsilon}_t^2 \nabla f(\mathbf{x}_t, \hat{\theta}_n) \nabla' f(\mathbf{x}_t, \hat{\theta}_n)$  respectivamente.

Notese que el estadístico de la prueba de Wald es una forma cuadrática en el vector  $S_i\theta$  y la inversa de una matriz que estima la matriz de varianzas y covarianzas de este vector de parámetros. Bajo la hipótesis nula,  $W_i \stackrel{A}{\sim} \chi_{q+1}^2$ , tiene una distribución asintótica chi-cuadrado con  $q + 1$  grados de libertad, donde  $q + 1$  representa el número de parámetros relacionados con una entrada, o lo que es equivalente, el número de parámetros de la hipótesis nula. Se rechazará la hipótesis nula cuando el valor del estadístico de prueba,  $W_i > \chi_{1-\alpha/2}^2(q + 1)$  o  $W_i < \chi_{\alpha/2}^2(q + 1)$ .

Bajo el supuesto que  $\hat{\theta}_N = \arg \min_{\theta \in \Theta} Q_N(\theta)$  es el estimador de mínimos cuadrados no lineales, se tiene que  $\hat{\theta}_N$  es generalmente consistente para  $\theta^*$ .

En este sentido, y con el objetivo de aplicar la prueba de Wald como un método de selección del modelo de combinación definido previamente, se aplica esta prueba para determinar la significancia de los parámetros asociados a una variable de entrada de la red. Se advierte que el hecho de que un contraste de hipótesis de dominancia arroje unos parámetros cero, no significa que el pronóstico asociado a los pronósticos cuyos parámetros son diferentes de cero sea el óptimo; lo que indicaría es que la combinación de estos pronósticos no puede ser mejorada en términos del error cuadrático medio al incluir el primer conjunto de pronósticos.

Para calcular completamente el estadístico, se requiere el cálculo de las primeras y segundas derivadas de la función objetivo con respecto a cada uno de los parámetros. La aplicación de la prueba de Wald para probar si las variables de entrada de una red neuronal son significativas, consiste en calcular el estadístico para cada una de las entradas de la red y eliminar las entradas que no resulten estadísticamente significativas.

Las expresiones necesarias para calcular el estadístico de prueba de Wald dado que la combinación no lineal se da mediante la función 4-8, tienen la forma:

$$\frac{\partial Q_N(\theta)}{\partial \phi_0} = \frac{-2}{N} \sum_{t=1}^N (Y_t - f(\mathbf{x}_t; \theta))$$

$$\frac{\partial Q_N(\theta)}{\partial \phi_i} = \frac{-2}{N} \sum_{t=1}^N [Y_t - f(\mathbf{x}_t; \theta)] \cdot y_{ti}$$

$$\frac{\partial Q_N(\theta)}{\partial \beta_j} = \frac{-2}{N} \sum_{t=1}^N [Y_t - f(\mathbf{x}_t; \theta)] \cdot G(\gamma_{0j} + \mathbf{x}_t \gamma_j)$$

$$\frac{\partial Q_N(\theta)}{\partial \gamma_{ij}} = \frac{-2}{N} \beta_j \sum_{t=1}^N [Y_t - f(\mathbf{x}_t; \theta)] \cdot y_{ti} \cdot G(\gamma_{0j} + \mathbf{x}_t \gamma_j) \cdot (1 - G(\gamma_{0j} + \mathbf{x}_t \gamma_j))$$

$$\frac{\partial Q_N(\theta)}{\partial \gamma_{0j}} = \frac{-2}{N} \beta_j \sum_{t=1}^N [Y_t - f(\mathbf{x}_t; \theta)] \cdot G(\gamma_{0j} + \mathbf{x}_t \gamma_j) \cdot (1 - G(\gamma_{0j} + \mathbf{x}_t \gamma_j))$$

$$\nabla^2 Q_N(\theta) = \frac{-2}{N} \sum_{t=1}^N [\varepsilon_t \nabla_{\theta}^2 f(\mathbf{x}_t; \theta)] + \frac{2}{N} \nabla_{\theta} f(\mathbf{x}_t; \theta) \nabla'_{\theta} f(\mathbf{x}_t; \theta)$$

No obstante, calcular la matriz de varianzas y covarianzas teóricamente depende de la función que sea usada como combinador y del vector de parámetros estimado. En muchos casos este cálculo resulta costoso computacionalmente, es por esto que en general se recurre a calcular una aproximación de dicha matriz.

## 4.5 Aplicación de la prueba de Wald en una combinación no lineal de pronósticos

En este capítulo se retoman los conjuntos de pronóstico individuales generados a partir de los diferentes modelos en el Capítulo 3 y se combinan mediante el uso del promedio aritmético simple y con una red neuronal artificial. Adicionalmente, se aplica la prueba de Wald para medir la importancia de las variables de entrada en la combinación no lineal.

Es así como el objetivo de este capítulo es evaluar la contribución potencial de cada conjunto de pronósticos como variables de entrada en una red neuronal artificial por medio de la aplicación de pruebas de dominancia de pronósticos para determinar la contribución parcial de cada variable de entrada sobre el pronóstico combinado. Para ello se implementa una prueba de dominancia como un criterio de selección de entradas en una combinación no lineal de pronósticos.

### 4.5.1 Métodos de combinación

Se implementan dos métodos de combinación, uno lineal y el otro no lineal.

1. Combinación con promedio aritmético simple:

Se verifica en qué casos la combinación con promedio aritmético simple supera el desempeño de los pronósticos individuales.

2. Combinación con redes neuronales artificiales:

Se combinan los pronósticos obtenidos de los modelos individuales considerando la especificación no lineal presentada en la ec. (4-8).

Se retoman los conjuntos de pronósticos de la volatilidad de los rendimientos de los activos financieros de New York, París, Singapur y Japón que fueron obtenidos en el capítulo anterior, provenientes de modelos ajustados con diferentes criterios de optimización y se combinan mediante los dos esquemas de combinación descritos. En el caso del combinador de red neuronal, se usan las mismas dos muestras. Con la primera muestra se entrena el modelo de red neuronal y con la segunda se evalúa el desempeño del modelo.

En ambos métodos de combinación se consideran tres casos, a saber, considerando 6, 5 y 4 conjuntos de pronósticos; en la primera combinación se incluyen los seis conjuntos de pronósticos, en el segundo caso se excluye el conjunto de pronósticos provenientes del modelo con menor desempeño en la muestra de entrenamiento (correspondiente en cada caso al modelo optimizado con el RMFE), y en el tercer caso se excluyen los dos conjuntos de pronósticos con menor desempeño en la muestra de entrenamiento (corresponden a los obtenidos con los modelos optimizados con RMCE y RMFE).

El proceso de estimación de la red neuronal, como se indicó anteriormente, no garantiza que se pueda encontrar el vector que minimiza la función objetivo. Por lo tanto, en este trabajo se estima el modelo varias veces poniendo diferentes vectores semilla hasta que un valor menor de la función objetivo es encontrado, para ello se usa el algoritmo de Nelder y Mead. Un elemento clave en este proceso de entrenamiento del modelo de red neuronal esta en seleccionar de manera adecuada estos puntos semilla, puesto que en un modelo no lineal la solución óptima depende fuertemente del punto inicial proporcionado para su entrenamiento. En este trabajo, se usa un método Heurístico, llamado evolución diferencial, para encontrar el vector semilla de parámetros.

### 4.5.2 Evaluación de la combinación

Se presentan los resultados obtenidos para ambas combinaciones y en ambas muestras: ajuste y pronóstico. En las Tablas 4-1 y 4-2 se muestran las medidas de desempeño de los modelos individuales y de las combinaciones para cada una de las series de tiempo. Se indica en cada

**Tabla 4-1** Distribución de las Medidas de Exactitud en las combinaciones de pronósticos con Promedio Aritmético Simple PAS.

Índice	Entrenamiento			Pronóstico		
	$H = 6$	$H = 5$	$H = 4$	$H = 6$	$H = 5$	$H = 4$
New York	459.5 (3.0)	414.8 (2.1)	<b>386.9</b> (1.4)	322.8 (5.2)	253.9 (4.2)	<b>207.8</b> (3.3)
	<b>690.3*</b> (0.6)	692.3 (0.8)	706.0 (1.6)	430.6 (3.0)	397.1 (1.6)	<b>386.6*</b> (0.5)
París	786.2 (7.8)	656.9 (3.9)	<b>596.6</b> (2.7)	699.8 (11.2)	558.4 (3.7)	<b>513.9</b> (2.0)
	1259.7 (2.7)	<b>1233.1*</b> (1.4)	1250.0 (2.7)	828.9 (8.0)	<b>765.5*</b> (1.8)	780.7 (2.3)
Singapur	1005.3 (10.4)	785.7 (4.8)	<b>680.0</b> (2.4)	873.5 (23.8)	470.7 (5.9)	<b>345.1</b> (3.7)
	1815.9 (2.8)	<b>1791.9</b> (1.1)	1820.6 (2.8)	932.0 (22.0)	625.7 (0.3)	<b>596.6*</b> (0.8)
Japón	731.5 (10.4)	605.3 (5.5)	<b>543.2</b> (4.9)	1092.6 (10.0)	996.1 (7.0)	<b>956.2</b> (5.1)
	1296.2 (4.7)	<b>1278.8</b> (3.9)	1294.7 (5.8)	<b>1910.3*</b> (5.3)	1938.6 (7.7)	1982.9 (11.2)

Cada celda contiene el valor promedio de MAE y RMSE y sus respectivas desviaciones estándar entre paréntesis. Las medidas de exactitud fueron multiplicadas por  $10^6$ .

serie de tiempo, en negrita, el menor valor de cada medida de exactitud y con asterisco cuando el modelo indicado supera los modelos individuales. Notemos, por ejemplo, que en la muestra de entrenamiento para la serie de tiempo de New York, el modelo que combina con el PAS para  $H = 6$  supera todos los modelos individuales con respecto al RMSE; mientras que en la muestra de prueba, se observa que el modelo que supera todos los modelos individuales es el PAS con  $H = 4$ .

El modelo de Red neuronal que combina los conjuntos de pronóstico individuales ANN4 supera todos los modelos individuales en todas las series de tiempo consideradas en la muestra de entrenamiento, en este caso la configuración de la red neuronal es  $p = 4$  entradas y  $q = 4$  nodos en la capa oculta. Sin embargo, el mejor desempeño en el pronóstico no se obtiene necesariamente con esta red neuronal. Por ejemplo, en la muestra de pronóstico, el mejor modelo del artículo de referencia es superado: en la serie de París por el modelo ANN6; y en la serie de Japón por los modelos ANN6, ANN5 y ANN4.

### 4.5.3 Cálculo de la prueba de Wald

Una vez obtenida la combinación de los pronósticos, se procede a aplicar la prueba de dominancia de pronósticos en la combinación no lineal como se definió en anteriormente, yendo de un modelo complejo a uno más simple.

Para calcular el estadístico que permite aplicar la prueba de Wald, se requiere de la matriz de varianzas y covarianzas de los parámetros, que en general es difícil de calcular. Una aproximación

**Tabla 4-2** Distribución de las Medidas de Exactitud en las combinaciones de pronósticos con una Red Neuronal Artificial.

Índice	Entrenamiento			Pronóstico		
	ANN6	ANN5	ANN4	ANN6	ANN5	ANN4
New York	438.4 (9.6)	451.3 (13.2)	<b>433.7</b> (13.2)	<b>271.1</b> (35.8)	324.6 (26.8)	301.9 (30.4)
	675.0 (7.9)	692.0 (12.0)	<b>668.7*</b> (11.7)	<b>413.3</b> (18.8)	437.9 (18.5)	425.2 (17.9)
París	699.9 (24.6)	673.3 (11.5)	<b>665.6</b> (11.1)	688.0 (49.7)	<b>585.6</b> (22.1)	587.3 (36.3)
	1284.2 (24.5)	1232.9 (14.0)	<b>1224.3*</b> (11.8)	<b>759.8*</b> (38.9)	786.3 (15.6)	797.4 (28.5)
Singapur	<b>794.9</b> (18.9)	800.6 (21.6)	804.1 (16.1)	561.1 (122.7)	<b>453.1</b> (42.1)	484.2 (28.0)
	1782.6 (15.6)	1789.2 (19.2)	<b>1778.4*</b> (13.9)	698.5 (86.4)	<b>625.7</b> (25.9)	640.3 (18.5)
Japón	644.2 (34.3)	634.8 (31.2)	<b>606.6</b> (21.9)	<b>1042.6</b> (65.0)	1067.3 (89.1)	1052.6 (43.3)
	1318.2 (36.9)	1305.0 (41.4)	<b>1250.2*</b> (29.4)	1954.5 (38.3)	1976.8(69.5)	<b>1946.7</b> (31.1)

Cada celda contiene el valor promedio de MAE y RMSE y sus respectivas desviaciones estándar entre paréntesis. Las medidas de exactitud fueron multiplicadas por  $10^6$ .

para esta matriz se obtiene hallando la distribución de los parámetros de la red neuronal. Por lo tanto, teniendo en cuenta el proceso de remuestreo del capítulo anterior en el cual se generaron  $n_b$  muestras de pronósticos de la volatilidad, se entrena la red neuronal  $n_b$  veces y se obtienen  $n_b$  muestras de parámetros estimados de la red haciendo posible calcular su matriz de varianzas y covarianzas.

En la Tabla 4-3 se muestra el valor del estadístico de prueba para cada una de las entradas de la red neuronal artificial y para cada una de las series de tiempo que se consideran en esta aplicación. Se indica con asterisco cuando la entrada resulta significativa, a un nivel de significancia del 5% y del 10%, esto es, cuando la entrada debe ser incluida en la combinación debido a que aporta información que no esta presente en los demás conjuntos de pronóstico. Note por ejemplo, que para la serie de New York, las entradas uno y dos resultan no significativas cuando se consideran los seis conjuntos de entrada. Cuando se excluyen el conjunto de pronósticos del modelo cuatro (MSFE), la entrada 6 es excluida.. Pero cuando se consideran solo cuatro entradas, donde se excluyen los conjuntos de pronóstico de los modelos tres y cuatro (MSCE y MSFE respectivamente), todos los cuatro conjuntos deben incluirse en la combinación no lineal.

Para la serie de Japón, cuando se consideran los seis conjuntos de entrada, la prueba muestra que no hay evidencia para considerar los pronósticos del modelo tres en la combinación no lineal. Sin embargo, al eliminar esa variable de entrada junto con el conjunto de pronósticos del modelo cuatro (considerando que estos son los dos modelos con menor desempeño), solo resultan significativas las entradas dos y seis.

**Tabla 4-3** Prueba de Wald para encompasamiento de pronósticos.

Índice	Estadístico de Wald					
	$W_1$	$W_2$	$W_3$	$W_4$	$W_5$	$W_6$
Red Neuronal con 6 entradas						
<b>New York</b>	5.01	3.64	11.30*	1.01*	9.49*	20.94**
<b>París</b>	10.35*	1.86	11.21*	1.08*	5.31	2.38
<b>Singapur</b>	10.56*	9.69*	15.96*	8.48	8.01	3.54
<b>Japón</b>	122.92**	331.55**	19.98	25.98**	17.85**	292.52**
	$W_1$	$W_2$	$W_3$	$W_5$	$W_6$	
Red Neuronal con 5 entradas						
<b>New York</b>	31.19**	46.47**	162.34**	110.96**	5.13	
<b>París</b>	1.75	1.04*	1.15*	2.02	0.91*	
<b>Singapur</b>	3.07	0.63*	6.64	1.79	3.12	
<b>Japón</b>	51.07**	8.18	34.63**	77.54**	13.33*	
	$W_1$	$W_2$	$W_5$	$W_6$		
Red Neuronal con 4 entradas						
<b>New York</b>	115.73**	24.37**	62.99**	29.77**		
<b>París</b>	8.67	4.20	6.24	3.47		
<b>Singapur</b>	59.36**	95.67**	7.42	0.72*		
<b>Japón</b>	5.38	0.46*	2.27	1.33*		

Cada celda contiene el valor del estadístico de Wald para cada entrada. Nivel de significancia \*(0.05) \*\*(0.10)

## 4.6 Discusión

En el proceso de ajuste de la red neuronal se identifica que los parámetros estimados tienen una alta dependencia de los puntos semilla usados para inicializar la red. Este hecho, conllevaría a una alta variabilidad de los parámetros estimados usando los datos del proceso de remuestreo. En segundo lugar, debido a que la superficie a optimizar probablemente tiene múltiples puntos de mínimo afecta directamente el cálculo del estadístico de la prueba de Wald ya que la matriz de varianzas y covarianzas, la cual es una matriz aproximada, tendría valores muy grandes y la conclusión con la prueba estadística estaría sesgada. No obstante, cabe resaltar que el método de remuestreo aplicado surge como la mejor alternativa para calcular una aproximación del estadístico de prueba del contraste de dominancia de pronósticos en la combinación no lineal, ante la imposibilidad de hacerlo analíticamente.

# 5 Conclusiones y trabajo futuro

En este capítulo se da respuesta a las preguntas de investigación de esta tesis, se identifica el cumplimiento de los objetivos propuestos y se plantean las oportunidades de trabajo futuro.

## 5.1 Respuesta a las preguntas de investigación

Se retoman las preguntas de investigación planteadas en el capítulo 1 y se presentan las respectivas respuestas.

### 5.1.1 Pregunta 1

**“¿Por qué y como los criterios de optimización generan diversidad en los conjuntos de pronóstico?”**

El método estándar para la estimación de parámetros de un modelo es el de mínimos cuadrados. Sin embargo, se sabe que el uso de este se debe en gran medida a las propiedades estadístico-matemáticas establecidas; más aún se sabe que se deben satisfacer ciertos supuestos para obtener el mejor estimado del vector de parámetros del modelo y para mantener las características deseables de los estimadores. Debido a que, cuando se trabaja con datos reales, difícilmente los supuestos se cumplen, sería válido considerar otros métodos de estimación de los parámetros que tengan valoraciones distintas del error. Naturalmente, este tipo de propuestas no siempre son bienvenidas porque se alejan del terreno de la teoría estadística. Pero, por otro lado pueden ser vistas como oportunidades de investigación para el desarrollo de nuevas teorías y como un método para explorar la información contenida en los datos.

En el Capítulo 3, se exponen las distintas características que posee cada función de error y se observa que la valoración del error influye de manera directa en los parámetros estimados del modelo. Se observa cómo con algunas funciones de error como criterio de optimización se modelan las bajas o altas volatilidades indicando que el modelo, de acuerdo a la penalización que se haga, será más o menos sensible a las altas volatilidades.

Una de las principales contribuciones en este capítulo es la implementación de la metodología definida por: el criterio de optimización y el mecanismo de remuestreo. Se logra hacer inferencias

acerca de la diferencia significativa de los modelos. Además, se muestra que se logra superar el desempeño de los mejores modelos del artículo de referencia con los modelos de esta propuesta.

### 5.1.2 Pregunta 2

**“¿Por qué la correlación de los errores de pronósticos de volatilidad no necesariamente es una medida de diversificación y que otra medida puede ser propuesta?”**

En esta tesis, se ha encontrado que dos modelos de pronóstico de volatilidad pueden ser distintos porque valoran el error de forma diferente y tener una correlación alta entre los errores de pronóstico. En la literatura es común encontrar que uno de los principales criterios de diversificación para combinar conjuntos de pronósticos es la correlación entre los errores de pronóstico. No obstante, por la particularidad de los datos de volatilidad, que tienen como valor mínimo el cero y se presentan picos de alta volatilidad, puede pasar que se tengan dos modelos diferentes con la misma tendencia que generan errores de pronóstico altamente correlacionados. Ambos pronósticos están ofreciendo información diferente de la serie de tiempo y no debería descartarse uno de ellos. Una medida que podría servir para identificar que se tienen conjuntos de pronósticos es la distancia euclidiana normalizada calculada punto a punto entre los conjuntos de pronóstico disponibles. Debe normalizarse esta distancia para determinar entre todos los conjuntos de pronóstico cuáles son mas distantes y cuales más similares, obteniendo valores entre cero y uno; distancias cercanas a cero mas similares son los conjuntos de pronóstico y valores próximos a uno, mayor diferencia entre los conjuntos de pronóstico.

### 5.1.3 Pregunta 3

**“¿Por qué la combinación de pronósticos de volatilidad tiende a ser más precisa que los pronósticos de modelos de volatilidad individuales?”**

En la revisión de la literatura presentada en el capítulo uno se encuentra que en términos generales, la combinación puede mejorar los pronósticos individuales y se exponen varias razones que permiten ver en la combinación de pronósticos una estrategia prometedora. Por otro lado, Cuando se habla de series de tiempo financieras y en particular de pronosticar la volatilidad condicional, lo que se desea implícitamente es minimizar el riesgo financiero. Es así como tener información de diferentes fuentes o que da cuenta de la volatilidad condicional de los rendimientos calculada desde distintos puntos de vista, equivalente a tener modelos de pronóstico de volatilidad diversos, es una forma de evaluar la rentabilidad del activo financiero y de minimizar el riesgo.

### 5.1.4 Pregunta 4

**“¿Qué tan confiable es la prueba de Wald para dominancia de pronósticos en aras de decidir si un conjunto de pronósticos debe incluirse en la combinación no lineal?”**

La prueba de dominancia de pronósticos es un contraste de hipótesis que se aplica a los parámetros estimados de un modelo de combinación de pronósticos. La construcción de la prueba de Wald requiere del cumplimiento de unos supuestos acerca de los parámetros estimados, así que mientras se tenga el cumplimiento de todos ellos, esta prueba será confiable y permitirá dar conclusiones acorde a los datos. Uno de los supuestos más fuertes es la convergencia asintótica de estos estimadores de los parámetros a los parámetros poblacionales, puesto que en el caso de una red neuronal, tal y como se explicó anteriormente, es muy difícil llegar al mínimo global y por lo general se obtienen mínimos locales comienza a debilitarse la prueba.

Cuando se tienen datos reales y se hace todo el proceso de construcción de modelos de pronóstico de volatilidad condicional se está intentando resolver un problema inverso, esto es, obtener los valores de los parámetros a partir de los datos observados. Este problema, debido a las características del modelo de combinación, es un problema mal condicionado o como se conoce en matemáticas un problema mal planteado. En el proceso de estimación de los parámetros de un modelo no lineal, se encuentran fallas principalmente en las condiciones de unicidad y estabilidad de las soluciones.

En vista de lo expuesto se tiene que, partiendo de datos reales y de un problema mal planteado es posible aplicar en la práctica todo el desarrollo que se propone desde el ámbito teórico, siempre teniendo presente que los resultados que se obtengan en la aplicación práctica pueden no estar acorde a las conclusiones que se obtendrían desde la teoría.

## 5.2 Cumplimiento de los objetivos específicos de esta tesis

En esta sección se describen las conclusiones asociadas a cada uno de los objetivos específicos que se trazaron para el cumplimiento del objetivo general de esta tesis.

### 5.2.1 Objetivo específico 1

**“Seleccionar las familias de funciones de error que se usarán para el entrenamiento de los expertos en el ensamble, con el fin de lograr su diversificación”.**

En el Capítulo 3 se presentan algunas de las funciones de pérdida que con frecuencia se usan para evaluar el desempeño de diferentes modelos de series de tiempo. En esta tesis, es deseable que cada modelo valore de forma diferente los errores en la muestra de entrenamiento, se exhiben las principales características de cada una de las funciones de error; ya que estas son las que definen las características del pronóstico. Un elemento clave que debe tenerse en cuenta es el hecho de que se desea modelar la volatilidad y con ella ciertos hechos estilizados, como por ejemplo los agrupamientos de altas y bajas volatilidades, es decir, se desea un modelo capaz de dar cuenta por los valores extremos, ya que en este caso no deben considerarse datos “outliers”. Las funciones

más sensibles a valores extremos deben ser incluidas en el conjunto de funciones de error. Es por ello que un conjunto de funciones de la familia  $\mathcal{L}_p$  es considerado. Con respecto a las funciones que miden errores porcentuales, se ha seleccionado una de ellas por considerar que puede modelar los valores bajos de la volatilidad.

### 5.2.2 Objetivo específico 2

**“Diseñar una metodología para especificar cada experto en el ensamble, que incorpora las familias de funciones de error seleccionadas en el objetivo anterior”.**

El cumplimiento de este objetivo específico se logra en el capítulo 3, en donde se realiza un proceso de especificación estadística de los modelos de volatilidad individuales. Las funciones de error seleccionadas en el objetivo 1 son fundamentales en este proceso de especificación debido a que se usan como criterio de optimización para la estimación de los parámetros del modelo. En este proceso de especificación se usa una técnica de re-muestro que permite construir la distribución de los parámetros de cada modelo así como la distribución de las medidas de exactitud consideradas en la evaluación del desempeño.

### 5.2.3 Objetivo específico 3

**“Diseñar una metodología para la combinación de pronósticos individuales generados por los expertos obtenidos en el objetivo anterior”.**

En el capítulo 4 se presenta la función no lineal que se implementa como combinador de los pronósticos individuales. Se consideran los conjuntos de pronóstico de los distintos expertos como las variables de entrada de la función de red neuronal artificial. El proceso de entrenamiento de la red neuronal artificial se hace con mínimos cuadrados no lineales. Debido a la influencia de los puntos semilla sobre la solución óptima que se encuentra, el experto se estima varias veces con el objetivo de que no se atasque en mínimos locales. Adicionalmente, los puntos semilla se determinan con un algoritmo de evolución diferencial, ya que en general, funciona mejor que la inicialización con puntos aleatorios.

### 5.2.4 Objetivo específico 4

**“Evaluar, al menos, tres casos de aplicación con el fin de determinar si la metodología propuesta permite obtener resultados más precisos que los principales métodos reportados en la literatura”.**

En el Capítulo 2, se presenta un caso de una aplicación empírica exitosa, en donde la combinación con RNA supera los pronósticos individuales. En el capítulo 4, se presentan cuatro casos de

aplicación para cuatro series de tiempo de los rendimientos de índices bursátiles conocidos, en estos casos, se tiene como referencia un artículo con estos mismos datos con el cual se puede comparar al menos en la muestra de prueba el desempeño tanto de los modelos individuales estimados en esta tesis como de los modelos de combinación de pronósticos. En dos de las cuatro series de tiempo, se encontró que cuatro modelos no lineales de los presentados en el capítulo 4 superan el mejor modelo del artículo de Taylor (2004). Adicionalmente, todos los modelos no lineales para las series de París, Singapur y Japón superan el mejor modelo del artículo en la muestra de pronóstico.

Mientras que, en la muestra de ajuste solo se puede comparar el desempeño entre los modelos individuales estimados con la metodología propuesta en el Capítulo tres de esta tesis con los modelos de combinación del Capítulo cuatro. En este caso, el modelo, ANN4, de RNA con 4 conjuntos de pronóstico de entrada supera todos los modelos individuales en cada una de las cuatro series de tiempo consideradas, con respecto a la medida RMSE. El modelo ANN5, es mejor que los modelos individuales en al menos 2 de las series de tiempo. Cuando se incrementa el número de entradas a la red neuronal, de cinco a seis se encontró un deterioro en el pronóstico; para el modelo ANN6 no hubo una ganancia significativa en el desempeño del pronóstico con respecto a la muestra de entrenamiento.

### 5.3 Cumplimiento del objetivo general de esta tesis

**“Proponer un esquema para la predicción de la volatilidad de rendimientos financieros, basada en el uso de ensamble de expertos, en el cual, la combinación de los pronósticos individuales es realizada usando una red neuronal artificial y la obtención de los parámetros de cada experto se logra al aplicar diferentes funciones de error; el ensamble tiene como fin obtener pronósticos más precisos de la volatilidad que los obtenidos por cada experto considerado y que los calculados usando técnicas tradicionales de combinación de pronósticos”.**

En el Capítulo dos, se propone un esquema de combinación no lineal de pronósticos de volatilidad en el cual se muestra que la combinación mejora el pronóstico de los modelos individuales, así como las combinaciones con el promedio aritmético simple y con regresión lineal. Así mismo, en el Capítulo 4 se implementa el mismo esquema de combinación no lineal, considerando una red neuronal con la misma configuración. No obstante, se agregan otras funciones de error como criterios de optimización para los expertos individuales. Se procede con la aplicación en cuatro series de tiempo y se obtiene en algunas de las configuraciones de la red neuronal mayor exactitud en el pronóstico que en los modelos individuales.

## 5.4 Trabajo futuro

Dentro de las oportunidades de investigación que deja esta tesis se tienen:

- Estudiar métodos que permitan encontrar los mejores valores de inicialización para el entrenamiento de un modelo no lineal, puesto que se ha encontrado que existe una relación de dependencia de las semillas con los estimados de los parámetros de un modelo de pronóstico.
- Hay una falta de caracterización de las series de tiempo que presentan efecto leverage, puesto que no siempre esto es garantía de que un modelo de pronóstico de volatilidad capture toda su información.
- Analizar otros métodos que miden la influencia de las variables de entrada en una red neuronal y compararlos con el desempeño de la prueba de Wald.
- Se puede extender la metodología de pronóstico de volatilidad condicional planteada en esta tesis modelar para el caso de predecir la media condicional de una serie de tiempo. Así mismo, integrar dos metodologías para pronosticar la volatilidad condicional sujeto a un modelo de media condicional no constante.

## 5.5 Difusión de resultados

Los resultados de esta tesis se han divulgado en diferentes entidades a nivel nacional e internacional, a saber:

### 5.5.1 Artículo de investigación

Orozco, J. M., & Velásquez, J. D. (2013). A new forecasting combination system for predicting volatility. *Innovar*, 23(50), 5–16.

### 5.5.2 Ponencias

1. Orozco-Castañeda, J.M., De Menezes, L., Alonso, E., Velásquez, J.D. (2015) *Nonlinear Forecast Encompassing with Artificial Neural Networks*. XXVII European Conference on Operational Research, University of Strathclyde, Glasgow, Reino Unido.
2. Orozco-Castañeda, J.M., Velásquez, J.D. (2014) *A Wald Test for the Significance of Inputs in a Neural Network Based on Combining Forecast*. XIII CLAPEM Latin American Congress of Probability and Mathematical Statistics, Cartagena, Colombia.
3. Orozco-Castañeda, J.M., Velásquez, J.D. (2013) *Volatility forecast combinations using multidimensional fuzzy inference systems*. VIII Congreso Colombiano de la Computación, Universidad del Quindío, Armenia, Colombia.

4. Orozco-Castañeda, J.M., Velásquez, J.D. (2013) *Pruebas de Dominancia para Modelos de Ensamble de Pronósticos de Volatilidad de Series de Tiempo Financieras*. XXIII Simposio Internacional de Estadística, Bogotá, Colombia (2013).
5. Orozco-Castañeda, J.M., Velásquez, J.D. (2013) *Nonlinear combination of volatility forecasts based on ANFIS*. XXVI European Conference on Operational Research, Università Sapienza, Roma, Italia.

# Bibliography

- Amendola, A. y Storti, G. (2008). A GMM procedure for combining volatility forecasts. *Computational Statistics & Data Analysis*, 52(6):3047–3060.
- Anders, U. y Korn, O. (1999). Model selection in neural networks. *Neural Networks*, 12(2):309 – 323.
- Andersen, T. G., Bollerslev, T., Christoffersen, P., y Diebold, F. X. (2005). Volatility Forecasting. *PIER Working Paper No. 05-011*.
- Baillie, R. T., Bollerslev, T., y Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30.
- Barnard, G. A. (1963). New methods of quality control. *Journal of the Royal Statistical Society. Series A (General)*, 126(2):255–258.
- Bates, J. M. y Granger, C. (1969). The Combination of Forecasts. *Operations Research Quarterly*, 20:451–468.
- Becker, R. y Clements, A. E. (2008). Are combination forecasts of S&P 500 volatility statistically superior? *International Journal of Forecasting*, 24(1):122–133.
- Bildirici, M. y Ersin, z. m. (2009). Improving forecasts of GARCH family models with the artificial neural networks: An application to the daily returns in Istanbul Stock Exchange. *Expert Systems with Applications*, 36(4):7355–7362.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.
- Bollerslev, T. y Engle, R. F. (1993). Common Persistence in Conditional Variances. *Econometrica*, 61(1):167–186. ArticleType: research-article / Full publication date: Jan., 1993 / Copyright © 1993 The Econometric Society.
- Brooks, C. (1998). Predicting stock index volatility: can market volume help? *Journal of Forecasting*, 17(1):59–80.
- Brown, G., Wyatt, J. L., y Tiño, P. (2005). Managing Diversity in Regression Ensembles. *J. Mach. Learn. Res.*, 6:1621–1650.

- Chang, B. y Tsai, H. (2008). Forecast approach using neural network adaptation to support vector regression grey model and generalized auto-regressive conditional heteroscedasticity. *Expert Systems with Applications*, 34(2):925–934.
- Chen, S., Härdle, W. K., y Jeong, K. (2010). Forecasting volatility with support vector machine-based GARCH model. *Journal of Forecasting*, 29:406–433.
- Chong, Y. Y. y Hendry, D. F. (1986). Econometric Evaluation of Linear Macro-Economic Models. *The Review of Economic Studies*, 53(4):671–690.
- Christoffersen, P. y Jacobs, K. (2004). The importance of the loss function in option valuation. *Journal of Financial Economics*, 72(2):291–318.
- Christoffersen, P. F. y Diebold, F. X. (1994). Optimal Prediction Under Asymmetric Loss.
- Clemen, R. T. y Winkler, R. L. (1986). Combining Economic Forecasts. *Journal of Business & Economic Statistics*, 4(1):39–46. ArticleType: research-article / Full publication date: Jan., 1986 / Copyright © 1986 American Statistical Association.
- Clemen, R. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4):559–583.
- Clements, M. y Hendry, D. F. (2008). Economic Forecasting in a Changing World. *Capitalism and Society*, 3(2):1–17.
- Clements, M. y Harvey, D. (2008). Forecast combination and encompassing.
- Cogollo, M. R. y Velásquez, J. D. (2014). Recent Artificial Neural Networks for Time Series Forecasting: A Systematic Literature Review.
- Coshall, J. T. (2009). Combining volatility and smoothing forecasts of UK demand for international tourism. *Tourism Management*, 30(4):495–511.
- De Gooijer, J. G. y Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, 22(3):443–473.
- Dhamija, A. y Bhalla, V. (2010). Financial time series forecasting: Comparison of neural networks and ARCH models. *International Research Journal of Finance and Economics*, 49:194–212.
- Diebold, F. X. (1989). Forecast combination and encompassing: Reconciling two divergent literatures. *International Journal of Forecasting*, 5(4):589–592.
- Diebold, F. X. y Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263. ArticleType: research-article / Full publication date: Jul., 1995 / Copyright © 1995 American Statistical Association.

- Donaldson, R. y Kamstra, M. (1997). An artificial neural network-GARCH model for international stock return volatility. *Journal of Empirical Finance*, 4(1):17–46.
- Donaldson, R. y Kamstra, M. (1999). Neural network forecast combining with interaction effects. *Journal of the Franklin Institute*, 336(2):227–236.
- Donaldson, R. G. y Kamstra, M. (1996). Forecast Combining with Neural Networks. *Journal of Forecasting*, 15(1):49–61.
- Dong, J.-R. (2008). Application of an EWMA combining technique to the prediction of stock market volatility. In *Proceedings of 2007 International Conference on Management Science and Engineering, ICMSE'07 (14th)*, pages 1844–1848.
- Dunis, C. L. y Huang, X. (2002). Forecasting and trading currency volatility: an application of recurrent neural regression and model combination. *Journal of Forecasting*, 21(5):317–354.
- Elliott, G. y Timmermann, A. (2008). Economic Forecasting. *Journal of Economic Literature*, 46(1):3–56. ArticleType: research-article / Full publication date: Mar., 2008 / Copyright © 2008 American Economic Association.
- Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987–1007.
- Engle, R. F. (1984). Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. *Handbook of econometrics*, 2:775–826.
- Engle, R. F. y Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, 5(1):1–50.
- Engle, R. F. y Ng, V. K. (1993). Measuring and Testing the Impact of News on Volatility. *The Journal of Finance*, 48(5):1749–1778. ArticleType: research-article / Full publication date: Dec., 1993 / Copyright © 1993 American Finance Association.
- Fang, Y. (2003). Forecasting combination and encompassing tests. *International Journal of Forecasting*, 19(1):87–94.
- Franses, P. H. y Dijk, D. V. (2000). *Nonlinear Time Series Models in Empirical Finance*. Cambridge University Press.
- Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer series in statistics. Springer.
- Fuertes, A.-M., Izzeldin, M., y Kalotychou, E. (2009). On forecasting daily stock volatility: The role of intraday information and market conditions. *International Journal of Forecasting*, 25(2):259–281.

- Gavrishchaka, V. V. y Ganguli, S. B. (2003). Volatility forecasting from multiscale and high-dimensional market data. *Neurocomputing*, 55(1-2):285–305.
- 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*, 48(5):1779–1801.
- Gonzalez, M. F. y Burgess, N. (1997). Modelling market volatilities: the neural network perspective. *The European Journal of Finance*, 3(2):137–157.
- Granger, C. W. y Machina, M. J. (2006). *Chapter 2 Forecasting and Decision Theory*, volume Volume 1, pages 81–98. Elsevier.
- Hajizadeh, E., Seifi, A., Fazel Zarandi, M., y Turksen, I. (2012). A hybrid modeling approach for forecasting the volatility of S&P 500 index return. *Expert Systems with Applications*, 39(1):431–436.
- Hamid, S. A. y Iqbal, Z. (2004). Using neural networks for forecasting volatility of S&P 500 Index futures prices. *Journal of Business Research*, 57(10):1116–1125.
- Hansen, P. R. y Lunde, A. (2005). A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data. *Journal of Financial Econometrics*, 3(4):525–554.
- Harrald, P. y Kamstra, M. (1997). Evolving artificial neural networks to combine financial forecasts. *Evolutionary Computation, IEEE Transactions on*, 1(1):40–52.
- Harvey, D. y Newbold, P. (2000). Tests for Multiple Forecast Encompassing. *Journal of Applied Econometrics*, 15(5):pp. 471–482.
- Harvey, D. I., Leybourne, S. J., y Newbold, P. (1998). Tests for Forecast Encompassing. *Journal of Business & Economic Statistics*, 16(2):pp. 254–259.
- Hashem, S. (1997). Optimal Linear Combinations of Neural Networks. *Neural Networks*, 10(4):599–614.
- Hendry, D. F. y Clements, M. P. (2004). Pooling of forecasts. *The Econometrics Journal*, 7(1):1–31.
- Hornik, K., Stinchcombe, M., y White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366.
- Hu, M. Y. y Tsoukalas, C. (1999). Combining conditional volatility forecasts using neural networks: an application to the EMS exchange rates. *Journal of International Financial Markets, Institutions and Money*, 9(4):407–422.

- Hyup Roh, T. (2007). Forecasting the volatility of stock price index. *Expert Systems with Applications*, 33(4):916–922.
- J. Phillip Cooper, C. R. N. (1975). The Ex Ante Prediction Performance of the St. Louis and FRB-MIT-PENN Econometric Models and Some Results on Composite Predictors. *Journal of Money, Credit and Banking*, 7(1):1–32.
- Jose, V. R. R. y Winkler, R. L. (2008). Simple robust averages of forecasts: Some empirical results. 24(1):163 – 169.
- Kisinbay, T. (2010). Predictive ability of asymmetric volatility models at medium-term horizons. *Applied Economics*, 42(30):3813–3829.
- Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., y Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, 51(1):7–15.
- Kitchenham, B., Pretorius, R., Budgen, D., Pearl Brereton, O., Turner, M., Niazi, M., y Linkman, S. (2010). Systematic literature reviews in software engineering – A tertiary study. *Information and Software Technology*, 52(8):792–805.
- Kuan, C.-M. (2008). artificial neural networks. In Durlauf, S. N. y Blume, L. E., editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke.
- Kuan, C.-M. y White, H. (1994). Artificial neural networks: an econometric perspective \*. *Econometric Reviews*, 13(1):1–91.
- Lehmann, E. y Casella, G. (2003). *Theory of Point Estimation*. Springer Texts in Statistics. Springer New York.
- Li, N., Liang, X., Li, X., Wang, C., y Wu, D. D. (2009). Network Environment and Financial Risk Using Machine Learning and Sentiment Analysis. *Human and Ecological Risk Assessment: An International Journal*, 15(2):227–252.
- Li, W., Luo, Y., Zhu, Q., Liu, J., y Le, J. (2008). Applications of AR\*-GRNN model for financial time series forecasting. *Neural Computing and Applications*, 17(5-6):441–448.
- Liu, F.-Y. y Liu, F.-X. (2006). Currency options volatility forecasting with shift-invariant wavelet transform and neural networks. In *Proceedings of the 13th international conference on Neural information processing - Volume Part III*, pages 461–468. Springer-Verlag.
- Lux, T. y Morales-Arias, L. (2010). Forecasting volatility under fractality, regime-switching, long memory and student- innovations. *Computational Statistics & Data Analysis*, 54(11):2676–2692.

- Makridakis, S. (1989). Why combining works? *International Journal of Forecasting*, 5(4):601–603.
- Makridakis, S., Wheelwright, S., y Hyndman, R. (1998). *Forecasting: Methods and Applications*. Wiley series in management. Wiley. LCCN: 97044416.
- Malliaris, M. y Salchenberger, L. (1996). Using neural networks to forecast the S&P 100 implied volatility. *Neurocomputing*, 10(2):183–195.
- Marcek, D. b., Marcek, M. d., y Babel, J. (2009). Granular RBF NN approach and statistical methods applied to modelling and forecasting high frequency data. *International Journal of Computational Intelligence Systems*, 2(4):353–364.
- McAleer, M. y Medeiros, M. C. (2011). Forecasting realized volatility with linear and nonlinear univariate models: forecasting realized volatility with linear/nonlinear models. *Journal of Economic Surveys*, 25(1):6–18.
- Nelson, C. R. (1972). The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy. *The American Economic Review*, 62(5):902–917.
- Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59(2):347–370.
- Ormoneit, D. y Neuneier, R. (1996). Experiments in predicting the German stock index DAX with density estimating neural networks. In *Computational Intelligence for Financial Engineering, 1996., Proceedings of the IEEE/IAFE 1996 Conference on*, pages 66–71.
- Orozco, J. y Velásquez, J. (2013). A new forecasting combination system for predicting volatility [Un nuevo sistema de combinación de pronósticos para la predicción de la volatilidad]. *Innovar*, 23(50):5–16.
- Patton, A. b. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256. cited By (since 1996)37.
- Patton, A. J. y Sheppard, K. (2009). Optimal combinations of realised volatility estimators. *International Journal of Forecasting*, 25(2):218–238.
- Patton, A. J. y Timmermann, A. (2007). Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics*, 140(2):884–918.
- Poon, S.-H. y Granger, C. W. J. (2003). Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature*, 41(2):478–539.
- Preminger, A., Ben-Zion, U., y Wettstein, D. (2006). Extended switching regression models with time-varying probabilities for combining forecasts. *The European Journal of Finance*, 12(6-7):455–472.

- Preminger, A., Ben-zion, U., y Wettstein, D. (2007). The extended switching regression model: allowing for multiple latent state variables. *Journal of Forecasting*, 26(7):457–473.
- Rabemananjara, R. y Zakoian, J. M. (1993). Threshold arch models and asymmetries in volatility, Threshold arch models and asymmetries in volatility. *Journal of Applied Econometrics*, *Journal of Applied Econometrics*, 8, 8(1, 1):31, 31–49, 49.
- Rapach, D. E. y Strauss, J. K. (2008). Structural breaks and GARCH models of exchange rate volatility. *Journal of Applied Econometrics*, 23(1):65–90.
- Refenes, A.-P. y Holt, W. (2001). Forecasting volatility with neural regression: A contribution to model adequacy. *Neural Networks, IEEE Transactions on*, 12(4):850–864.
- Reid, D. J. (1968). Combining three estimates of gross domestic product. *Economica*, 35(140):431–444.
- Rodríguez, L. F., Velásquez, y Franco, C. J. (2012). *Una investigación científica acerca del progreso de métodos de ensamble basados en inteligencia computacional para predicción de series de tiempo económicas y financieras*, pages 17–45. Universitaria.
- Schittenkopf, C., Dorffner, G., y Dockner, E. J. (2000). Forecasting Time-dependent Conditional Densities: A Semi-non-parametric Neural Network Approach. *Journal of Forecasting*, 19:355–374.
- Sermpinis, G., Laws, J., y Dunis, C. L. (2013). Modelling and trading the realised volatility of the FTSE100 futures with higher order neural networks. *The European Journal of Finance*, 19(3):165–179.
- Slim, C. (2004). *Forecasting the volatility of stock index returns: A stochastic neural network approach*, pages 935–944. Springer.
- Sun, B. y Xie, C. (2009). RMB Exchange Rate Forecasting in the Context of the Financial Crisis. *Systems Engineering-Theory & Practice*, 29(12):53–64.
- Taylor, J. W. (2004). Volatility forecasting with smooth transition exponential smoothing. *International Journal of Forecasting*, 20(2):273–286.
- Teräsvirta, T. (1996). Two stylized facts and the garch (1,1) model. SSE/EFI Working Paper Series in Economics and Finance 96, Stockholm School of Economics.
- Terui, N. y van Dijk, H. K. (2002). Combined forecasts from linear and nonlinear time series models. *International Journal of Forecasting*, 18(3):421–438.
- Timmermann, A. (2006). *Chapter 4 Forecast Combinations*, volume Volume 1, pages 135–196. Elsevier.

- Tofallis, C. (2015). A better measure of relative prediction accuracy. 66(8):1352–1362. bibtex: Tofallis2015.
- Tsangari, H. (2007). An Alternative Methodology for Combining Different Forecasting Models. *Journal of Applied Statistics*, 34(3-4):403–421.
- Tsay, R. S. (2010). *Analysis of Financial Time Series*. John Wiley & Sons, 3 edition.
- Varian, H. R. (1974). *A Bayesian approach to real estate assessment*, pages 195–208.
- Verhoeven, P., Pilgram, B., McAleer, M., y Mees, A. (2002). Non-linear modelling and forecasting of S&P 500 volatility. *Mathematics and Computers in Simulation*, 59(1–3):233–241.
- Wang, Y.-H. (2008). Using neural network to forecast stock index option price: a new hybrid GARCH approach. *Quality & Quantity*, 43(5):833–843.
- Wei, X. y Yang, Y. (2012). Robust forecast combinations. *Journal of Econometrics*, 166(2):224–236.
- Wenbing, X., Qi, F., y Weimin, C. (2008). Forecasting Chinese stock markets volatility based on neural network combining. In *Proceedings - 4th International Conference on Natural Computation, ICNC 2008*, volume 7, pages 23–27.
- White, H. (1989). Some Asymptotic Results for Learning in Single Hidden-Layer Feedforward Network Models. *Journal of the American Statistical Association*, 84(408):pp. 1003–1013.
- White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge University Press.
- Zellner, A. (1986). Bayesian Estimation and Prediction Using Asymmetric Loss Functions. *Journal of the American Statistical Association*, 81(394):446–451. ArticleType: research-article / Full publication date: Jun., 1986 / Copyright © 1986 American Statistical Association.
- Zhang, G., Patuwo, B. E., y Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1):35–62.
- Zhang, G., Zhang, S., Yin, J., y Cheng, L. (2011). Regularization based ordering for ensemble pruning. In *Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference on*, volume 2, pages 1325 –1329.