UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INDUSTRIAS
VALPARAÍSO - CHILE

PRONÓSTICO DE VOLATILIDAD DE TIPO DE CAMBIO MEDIANTE MÉTODOS DE INTELIGENCIA ARTIFICIAL

JORGE LUIS MORÁN LEAL

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL

PROFESOR GUÍA	:	SR. WERNER KRISTJANPOLLER R.
PROFESOR CORREFERENTE	:	SR. RODOLFO SALAZAR A.

AGOSTO 2017

Índice de Contenidos

1. Problema de Investigación **1**
1.1. Objetivos . . . 3
1.1.1. Objetivo General . . . 3
1.1.2. Objetivos Específicos . . . 3

2. Marco Téorico **5**
2.1. Volatilidad Histórica . . . 5
2.2. Mercado de Divisas . . . 5
2.2.1. ¿Qué es el mercado Forex? . . . 5
2.2.2. Indicadores Técnicos . . . 6
2.2.3. Series de Tiempo . . . 8
2.3. Modelos GARCH . . . 9
2.3.1. GARCH . . . 9
2.3.2. EGARCH . . . 9
2.3.3. Distribución Normal . . . 9
2.3.4. Distribución t-Student . . . 10
2.3.5. Distribución GED . . . 10
2.4. Redes Neuronales . . . 11
2.4.1. Artificial Neural Network . . . 11
2.4.2. Algoritmo de entrenamiento . . . 12
2.5. Fuzzy Logic (Lógica Difusa) . . . 13
2.5.1. if-then Rules . . . 13
2.5.2. Fuzzy Sets . . . 14
2.5.3. Funciones de Membresía . . . 14
2.5.4. Fuzzy Interference System . . . 14
2.5.4.1. Sugeno . . . 15

3. Data **17**

4. Metodología **21**
4.1. Pronóstico de Volatilidad . . . 21
4.1.1. Estimación Serie de Retorno y Volatilidad Histórica . . . 21
4.1.2. GARCH/EGARCH . . . 21
4.1.3. Modelos ANN-GARCH/EGARCH . . . 21
4.1.4. ANN-Indicadores Técnicos . . . 22
4.1.5. Fuzzy Interference System . . . 22
4.2. Benchmark . . . 23
4.2.1. Funciones de Pérdida . . . 23
4.2.1.1. MAPE . . . 23
4.2.1.2. MSE . . . 23
4.2.2. Model Confidence Set . . . 23

5. Resultados **25**
5.1. GARCH/EGARCH . . . 25
5.1.1. Funciones de pérdida . . . 25
5.2. ANN-GARCH/EGARCH . . . 27
5.2.1. ANN-GARCH-normal . . . 27
5.2.2. ANN-GARCH-student . . . 28
5.2.3. ANN-GARCH-GED . . . 29
5.2.4. ANN-EGARCH-normal . . . 30
5.2.5. ANN-EGARCH-student . . . 31
5.2.6. ANN-EGARCH-GED . . . 32
5.3. ANN-Technical Analysis(ANN-TA) . . . 33
5.4. Fuzzy Interference System (FIS) . . . 34
5.4.1. FIS . . . 34
5.4.2. FIS-TA . . . 35
5.5. Conclusiones y Recomendaciones . . . 36
5.5.1. Conclusiones . . . 36
5.5.2. Recomendaciones . . . 36

Bibliografía **37**

A. Anexos **39**
A.1. Modelos GARCH/EGARCH . . . 39
A.1.1. Ajuste GARCH-normal . . . 39
A.1.2. Ajuste GARCH-student . . . 40
A.1.3. Ajuste GARCH-GED . . . 40
A.1.4. Ajuste EGARCH-normal . . . 41
A.1.5. Ajuste EGARCH-student . . . 41
A.1.6. Ajuste EGARCH-GED . . . 42
A.2. Benchmark y Model Confidance Set . . . 42

Índice de Tablas

2.1. Resumen de Indicadores Técnicos, Fuente:Elaboración Propia 6

3.1. Resumen Estadístico, Fuente: Elaboración Propia 17

5.1. Funciones de pérdida para modelos GARCH/EGARCH, Fuente: Elaboración Propia 25
5.2. Funciones de pérdida para el modelo ANN-GARCH-normal, Fuente:Elaboración Propia . . 27
5.3. Funciones de pérdida para el modelo ANN-GARCH-student, Fuente:Elaboración Propia . . 28
5.4. Funciones de pérdida para el modelo ANN-GARCH-GED, Fuente:Elaboración Propia . . . 29
5.5. Funciones de pérdida para el modelo ANN-EGARCH-normal, Fuente:Elaboración Propia . 30
5.6. Funciones de pérdida para el modelo ANN-EGARCH-student, Fuente:Elaboración Propia . 31
5.7. Funciones de pérdida para el modelo ANN-EGARCH-GED, Fuente:Elaboración Propia . . . 32
5.8. Funciones de pérdida para el modelo ANN-TA, Fuente:Elaboración Propia 33
5.9. Funciones de pérdida para FIS, Fuente:Elaboración Propia 34
5.10. Funciones de pérdida para FIS+TA, Fuente:Elaboración Propia 35
5.11. Valores recomendados en Substractive Clustering para R_a. Fuente: Elaboración propia. . . . 35
5.12. Benchmark modelos. Fuente: Elaboración propia 36

A.1. Funciones de pérdida y MCS al 1 %, 5 % y 10 %. Fuente: Elaboración propia. 42
A.2. Funciones de pérdida y MCS al 1 %, 5 % y 10 %. Fuente: Elaboración propia. 43
A.3. Funciones de pérdida y MCS al 1 %, 5 % y 10 %. Fuente: Elaboración propia. 44
A.4. Funciones de pérdida y MCS al 1 %, 5 % y 10 %. Fuente: Elaboración propia. 45

Índice de Figuras

2.1. Estructura BPNN, Fuente: Elaboración propia 11
2.2. Diagrama Lógica Difusa, Fuente:Elaboración Propia 13
2.3. Diagrama de flujo, Fuente: MatlabR2017a Documentation 15

3.1. Serie de tipo de cambio USD/EUR histórica 1999-2017, Fuente:Elaboración Propia 18
3.2. Serie de Rentabilidad Logarítmica 1999-2017, Fuente:Elaboración Propia 18
3.3. Serie de volatilidad histórica 1999-2017, Fuente:Elaboración Propia 19

A.1. Ajuste GARCH-normal, Fuente:Elaboración Propia 39
A.2. Ajuste GARCH-student, Fuente:Elaboración Propia 40
A.3. Ajuste GARCH-GED, Fuente:Elaboración Propia 40
A.4. Ajuste EGARCH-normal, Fuente:Elaboración Propia 41
A.5. Ajuste EGARCH-student, Fuente:Elaboración Propia 41
A.6. Ajuste EGARCH-GED, Fuente:Elaboración Propia 42

1 | Problema de Investigación

El mercado Forex (Foreign Exchange Market) es un mercado de intercambio de divisas que tiene alcance en todo el mundo. En el año 2013, el volumen de transacción ascendía a $5.3 trillones, este monto hace que el mercado de intercambio de divisas sea aproximadamete 160 veces más grande que el mercado accionario de Nueva York. En este mercado, el trading se realiza mediante compras y ventas de pares de divisas, entre las cuales se destacan los pares EUR/USD, AUD/USD, GPB/USD, USD/CAD, USD/CHF y USD/JPY.

La estrategia principal en este mercado es comprar a bajo precio y vender a alto precio en un mercado que está abierto 24/7. Dada la alta volatilidad del mercado, es importante monitorearlo constantemente. Es imposible para un humano monitorear un mercado durante 24 horas al día, 5 días a la semana administrando y tomando decisiones racionales de forma manual.

Entre los inversionistas hay concenso general que la volatilidad de tipo de cambio es perjudicial para la economía (Obstfeld y Rogoff (1998)), con un mayor costo de bienestar para los países en desarrollo. Además que incrementa el factor de riesgo de las firmas domésticas , lo que puede implicar un aumento de precios en contra de la prima de riesgo adicional (Giannellis y Papadopoulos (2011)).

Los nuevos avances en machine learning han hecho posible que los inversionistas aprendan de la data histórica del mercado para tener más ganancias en sus transacciones. Yao et al. (2000) empleó indicadores técnicos para capturar las reglas del movimiento de las divisas de las 6 principales monedas en el mercado Forex, probando la eficiencia de cada mercado con el fin de elegir la data para entrenar la red neuronal. Más recientemente, Lahmiri (2017) utilizó la familia de procesos GARCH y EGARCH con diferentes distribuciones para comparar el desempeño de los modelos ANN-GARCH/EGARCH frente a un enfoque de indicadores técnicos, siendo estos últimos los que obtuvieron la mejor performance.

Por otra parte, Hsu et al. (2016) realiza un benchmark de pronósticos de series de tiempo financieras entre Machine Learning, modelos econométricos y análisis técnico, resultando que la metodología de Machine Learning mejora significativamente la precisión del pronóstico. Además, este estudio tomó en cuenta la madurez del mercado y el horizonte del pronóstico.

A pesar de los avances, es necesario buscar una alternativa que permita resolver el problema de forma más eficiente en cuanto a tiempos computacionales y precisión del pronóstico sin necesidad de supuestos o modelos de causalidad además de ser usados en diferentes aplicaciones industriales. Dentro de las investigaciones realizadas con respecto al tipo de cambio, Korol (2014) implementó lógica difusa en el análisis técnico, comparando la efectividad del modelo propuesto frente al modelo ARCH, GARCH ANN y Fuzzy Logic.

Gharleghi et al. (2014) propone pronosticar tipo de cambio en el horizonte de largo y corto plazo utilizando un Sistema Neuro Fuzzy basado en cointegración, el cual es una combinación de una técnica de cointegración; un Sistema de Interferencia Difuso; y una Red Neuronal. Sus resultados revelan que el modelo Neuro-fuzzy basado en cointegración tiene mejor performance que el VECM, logrando pronosticar el tipo de cambio con un alto nivel de precisión.

Hung (2011) presenta un algoritmo iterativo (mediante algoritmos genéticos) para estimar los parámetros de la función de membresía y los modelos GARCH. Los algoritmos genéticos son utilizados para encontrar una solución óptima global con ráida convergencia en el contexto del modelo fuzzy-GARCH. Sus pruebas determinaron que tanto los pronósticos dentro de la muestra como fuera de la muestra mejoran significativamente cuando el modelo GARCH considera efectos de clustering y de pronóstico adaptativo.

Escobar et al. (2013) propone un indicador para análisis técnico basado en lógica difusa, comparándolo con los indicadores técnicos tradicionales como la media móvil, RSI y MACD. Al contrario de los indicadores tradicionales no es un modelo objetivamente matemático, puesto que incorpora característica subjectivas del inversionista como la tendencia al riesgo.

Por lo tanto, es atractivo explorar la posibilidad de utilizar un enfoque basado en indicadores técnicos en conjunto con lógica difusa, dado que estudios recientes (Yin y Yang (2016)) muestran que la utilización de indicadores técnicos resulta en predicciones robustas out-of-sample, el cual es robusto en recesiones y expansiones. Esto último es claro, porque los indicadores técnicos pueden detectar los ciclos económicos de forma efectiva.

1.1. Objetivos

1.1.1. Objetivo General

Pronosticar la volatilidad histórica de la divisa USD/EUR mediante modelos heterocedásticos (GARCH), modelos heterocedásticos híbridos (ANN-GARCH) y sistemas de interferencia difusa (FIS) con el fin de determinar cuál de estos métodos es el que mejor modela y pronostica la volatilidad de este tipo de cambio.

1.1.2. Objetivos Específicos

- Estimar y determinar los indicadores técnicos de FOREX para utilizar como variables predictivas en la red neuronal.
- Encontrar la configuración de parámetros óptima en la red neuronal, tal que se minimice el error del pronóstico tanto con indicadores técnicos, como en los modelos híbridos.
- Comparar el Desempeño de los modelos econométricos frente a los modelos híbridos y el sistema de interferencia difusa propuesto.

2 | Marco Téorico

2.1. Volatilidad Histórica

El comportamiento volátil de los mercados financieros está referido como volatilidad. La volatilidad se ha convertido en un concepto importante en diferentes áreas de ingeniería financiera, como la teoría del portafolio, la gestión del riesgo y el precio de los derivados. Además, en el modelamiento de precios de activos se utiliza la volatilidad como una estimación simple de medida de riesgo. En estadística, la volatilidad usualmente es medida como la desviación estándar o varianza (Daly (2008)).

En este trabajo el activo estudiado son las divisas, de las cuales se detallará en la siguiente sección.

2.2. Mercado de Divisas

2.2.1. ¿Qué es el mercado Forex?

Forex (Foreign Exchange o simplemente FX) es el mercado donde todos los participantes son capaces de comprar, vender, intercambiar y especular en monedas. El mercado Forex está compuesto de bancos, compañías comerciales, Bancos Centrales, entre otras compañías financieras. El mercado de divisas es considerado actualmente el mercado financiero más grande en el mundo, procesando trillones de dólares en transacciones todos los días.

El mercado Forex tiene las siguientes particularidades sobre otros mercados:

1. Al contrario de los mercados de acciones, futuros y opciones, el mercado de divisas no tiene reglas de intercambio ni está controlado por algún cuerpo gubernamental. En efecto, tener información privilegiada podría generar ganancias de mucho dinero sin tener repercusiones legales.
2. No existen comisiones.
3. Es posible comprar las 24 horas del día.
4. No existe límite de cuánto es posible comprar o vender.
5. No existen barreras de salida.

Actualmente existen variadas estrategias de inversión en este mercado, en la próxima sección se revisará una de las más utilizadas basadas en indicadores.

2.2.2. Indicadores Técnicos

El análisis Técnico es un set de herramientas y técnicas usadas para identificar la tendencia del mercado y/o identificar si el mercado está en niveles de sobrecompra y sobreventa.

Los indicadores técnicos consisten en fórmulas matemáticas, las cuales son aplicadas al precio de las series de tiempo para generar otra serie de datos. Los indicadores técnicos pueden clasificarse en tres grupos:

1. Indicadores de Tendencia: se utilizan para detectar las tendencias en los mercados financieros. Este grupo de indicadores es ineficiente para los períodos de equilibrio (Flat) del mercado. Los indicadores de tendencia indican a la dirección del movimiento del precio. La media móvil y el MACD son ejemplos de indicadores de tendencia.
2. Indicadores de Momentum: muestran la variación del precio de un instrumento financiero en un período determinado de tiempo.
3. Indicadores de Volatilidad: están basados en los cambios rápidos de volatilidad en el precio. Las Bandas de Bollinger son ejemplos de estos indicadores.

El análisis técnico busca predecir el precio de un activo basado en el análisis de los precios históricos Pring (2002), asumiendo que las dinámicas de cierto activo están caracterizadas por información interna del mercado, y estas variaciones se repetirán en el futuro.

La siguiente tabla muestra los indicadores utilizados en este estudio y cuáles son sus principales características:

Tabla 2.1: Resumen de Indicadores Técnicos, Fuente:Elaboración Propia

Indicador Técnico	Tipo de Indicador
Bandas de Bollinger	Volatilidad
Momentum	Momentum
Aceleración	Momentum
EMA	Tendencia
RSI	Momentum
MACD	Tendencia

A continuación, se estudiarán en detalle los indicadores a utilizar en este trabajo:

1. Bandas de Bollinger: Propuestas por Bollinger (2001), tiene como principal característica que el espacio entre las bandas varía según la volatilidad de los precios. En momentos de mayor volatilidad en el mercado, las bandas serán más amplias, y durante períodos de baja volatilidad, las bandas se estrechan a fin de contener los precios de las divisas.
   - Middle Band: La banda media corresponde a una media móvil simple, donde T corresponde a 20 períodos:
   $$\frac{\sum_{t=1}^{T} HV_t}{T}$$
   - Upper Band: La banda superior se calcula como la banda media más dos veces la desviación estándar de la media móvil en los T períodos. Con esto, se obtienen resultados 68 % más precisos, evitando el exceso de especulación.
   $$MiddleBand + MiddleBand + [D * \frac{\sum_{t=1}^{T}(HV_t - MiddleBand)^2}{T}]$$
   - Lower Band: Finalmente, la banda inferior se calcula de la misma forma que la banda superior, pero restando los términos calculados en el indicador anterior a la banda media.
   $$MiddleBand - [D * \frac{\sum_{t=1}^{T}(HV_t - MiddleBand)^2}{T}]$$

En donde HV_t es la volatilidad histórica, t es el script de tiempo, T es el número de días seteado en 20, n seteado en 4, $\alpha = \frac{2}{1+k}$ es el factor de suavizamiento, k es el número de observaciones.

El propóstito principal de las Bandas de Bollinger es dar una definición relativa de alto y bajo. Por definición, la volatilidad está alta en la Upper Band y baja en la Lower Band. Esta definición puede ayudar a generar un patrón de reconocimiento, el cual es útil para tomar decisiones de trading sistemático.

2. Momentum: Corresponde a la serie de volatilidad en t menos la serie rezagada en $t = 12$.

$$HV_t - HV_{t-12}$$

3. Aceleración: Similar al caso anterior, se utilizada el indicador de Momentum calculado en tiempo t menos el Momentum rezagado en $t = 12$.

$$Momentum_t - Momentum_{t-12}$$

4. Media Móvil Exponencial (EMA): La Media Móvil Exponencial es un indicador que a diferencia de la media móvil simple, esta le concede mayor importancia a las últimas observaciones en el transcurso del período, es decir que no todos los valores tienen el mismo valor.

$$EMA(k)_{t-1} + \alpha * (HV_t - EMA(k)_{t-1})$$

En cuanto al método del cálculo, para el primer dato se toma el promedio de las primeras observaciones para posteriormente suavizar este valor para la siguiente observación, utilizando el valor de alfa dado para el número de observaciones. Similar a los otros indicadores, HV_t corresponde a la volatilidad histórica y $EMA(k)_{t-1}$ al valor anterior de la media móvil exponencial.

5. MACD: El MACD (Moving Average Convergence/Divergence) es un indicador que mide la diferencia entre dos medias móviles exponenciales, el cual fue propuesto por Appel (1985). El MACD incluye otra EMA, la cual se le conoce como la línea de señal. En otras palabras el MACD(l,s,n) es un indicador donde la serie es una diferencia entre una EMA larga (l) y una EMA corta (s) con el parámetro de la serie (n).

$$MACD(n)_{t-1} + \frac{2}{(n+1)}(Diff_t - MACD(n)_{t-1})$$

En donde $Diff = EMA(12) - EMA(26)$

6. RSI: Fue propuesto por Wilder (1978). Este indicador compara la magnitud de las ganancias recientes con las pérdidas recientes para determinar las condiciones de sobrecompra y sobreventa de una divisa. El RSI se calcula de la siguiente forma:

$$RSI = 100 - \frac{100}{1+RS}$$

En donde RS es:

$$RS = \frac{Media_{subidas}}{Media_{bajada}}$$

En donde $Media_{subidas}$ corresponde a la media de x días en subida (gannacias) y $Media_{bajada}$ corresponde a la media de x días en bajada (pérdidas).

El rango de este indicador oscila entre 0 y 100 %. Si este valor es sobre 70 %, la divisa se considera como sobrecomprada y el RSI genera una señal de venta. En el caso contrario, con un RSI menor a 30 % la divisa se considera en sobrevendida, generando una señal de compra.

2.2.3. Series de Tiempo

Una serie de tiempo es una colección de datos tomada a lo largo del tiempo, es decir datos de corte longitudinal. El objetivo principal del uso de series de tiempo es controlar, predecir y analizar el comportamiento de algún proceso.

Los pronósticos de los métodos de series de tiempo se basan en patrones históricos de los datos, sin ser explicados por una variable exógena además del tiempo. Las mediciones son tomadas en puntos sucesivos o períodos sucesivos (cada hora, día, mes, año o cualquier intervalo regular o irregular).

En la siguiente sección se estudiará uno de los modelos econométricos más populares para capturar la volatilidad de las series de tiempo financieras.

2.3. Modelos GARCH

2.3.1. GARCH

Los modelos más utilizados para estimar la volatilidad son los modelos ARCH (Auto Regressive Conditional Heteroscedasticity) desarrollado por Engle (1982). Desde el desarrollo del modelo ARCH original, se han generado extensiones de este, como el modelo GARCH (Bollerslev (1986)) y el EGARCH (Nelson y Cao (1992)).

El proceso GARCH(p,q) se puede expresar de la siguiente forma:

$$\varepsilon_t = \sigma_t Z_t$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

Donde ω, α_i , β_i, p y q son coeficientes no negativos. Z_t representa una secuencia de variables aleatorias i.i.d (independientes e idénticamente distribuidas) con media cero y varianza unitaria. Por definición, e_t es una secuencia sin correlación serial con media cero y varianza condicional σ_t^2, la cual puede ser no estacionaria.

Además, la estabilidad del proceso GARCH requiere las siguientes condiciones:

$$\alpha_i, \beta_i > 0$$

$$(\sum_{i=1}^{m} \alpha_i + \sum_{j=1}^{n} \beta_j) < 1$$

Gracias a las limitaciones citadas anteriormente, se introduce el modelo EGARCH al estudio de este trabajo.

2.3.2. EGARCH

El modelo EGARCH (Exponential Generalized Autorregresive Conditional Heterocedasticity) fue introducido por Nelson (1991) y Nelson y Cao (1992), encontrando las siguientes limitaciones en los modelos GARCH:

- Las condiciones impuestas sobre los parámetros para asegurar que no sea negativo son violadas en algunas aplicaciones empíricas.
- El modelo GARCH es incapaz de modelar una respuesta asimétrica de la volatilidad ante las subidas y bajadas de la serie. Con el fin de solventar estas deficiencias, Nelson propuso un nuevo modelo GARCH exponencial o EGARCH.

El modelo EGARCH garantiza la no negatividad de la varianza condicional formulando la ecuación de la volatilidad en términos de logaritmo de σ_t^2, mediante una representación lineal del tipo:

$$log(\sigma_t^2) = \omega + \sum_{i=1}^{q} \beta_i \log_{t-i}^2 + \frac{\sum_{i=1}^{p} \alpha_i \left|\varepsilon_{t-i}^2\right|}{(\sigma_{t-i}^2)^{1/2}} + \frac{\sum_{i=1}^{q} \gamma_i \varepsilon_{t-i}}{(\sigma_{t-i}^2)}$$

Además, este modelo no cuenta con restricciones para los parámetros α_i y β_i para asegurar la no negatividad de las varianzas condicionales.

2.3.3. Distribución Normal

Usualmente se asume que las distribuciones condicionales para el error ϵ_t tienen una distribución normal.

La función de densidad de probabilidad para la distribución normal está dada por:

$$f(z) = \frac{1}{2\pi^{1/2}} exp(-\frac{z^2}{2})$$

En investigaciones recientes de Horvath y Sopov (2016) sus resultados sugieren que los valores realizados y los proyectados difieren de forma importante para modelos GARCH con distribución condicional normal, los cuales subestiman el riesgo de las colas. En contraste, el modelo GARCH con distribución condicional t-Student logra capturar las colas de forma más precisa. Por lo tanto, se estudiarán dos alternativas de distribución: t-Student y Generalized Error Distribution (GED).

2.3.4. Distribución t-Student

La función de distribución t-student está dada por:

$$f(z, \mu, \sigma, \nu) = \frac{C}{\sigma^2}(1 + \frac{(z-\mu)^2}{\sigma^2(\nu-2)})$$

donde μ y σ son respectivamente la media y la varianza, y el parámetro C está dado por:

$$C = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi(\nu-2)}\Gamma(\frac{\nu+1}{2})}$$

donde $\Gamma(.)$ es la función gamma y el parámetro $\nu > 2$ es el grado de libertad de las colas de la distribución.

2.3.5. Distribución GED

Finalmente, la función de distribución GED (Generalized Error Distribution) es:

$$f(z_t, \mu, \sigma, \alpha) = \frac{\beta}{2\Gamma(\frac{1}{\beta})} exp(-(\frac{|z_t-\mu|}{\sigma})^\beta)$$

A pesar de que los modelos GARCH son frecuentemente utilizados, estos no logran detectar las relaciones no-lineales que pueden estar presentes en la data. Por lo mismo, en la siguiente sección se presentarán métodos de inteligencia artificial.

2.4. Redes Neuronales

2.4.1. Artificial Neural Network

En general, las series de tiempo financieras asumen una estructura de correlación lineal entre los datos de la serie donde pueden existir relaciones o patrones no lineales en la data, las cuales no pueden ser capturadas por los modelos GARCH. Una red neuronal ANN (Artificial Neural Network) es un modelo computacional que intenta imitar la estructura funcional de los aspectos biológicos de las Redes Neuronales, imitando la habilidad del cerebro humano para procesar data e información para extraer patrones existentes.

Basándose en la estructura del cerebro humano, un set de elementos o neuronas (nodos) están interconectados y organizados en capas. Esas capas pueden ser organizadas en capas de entrada, capas ocultas (medias) y capas de salida.

Una de las mayores ventajas de las Redes Neuronales es que teóricamente son capaces de aproximar una función continua, de forma que el modelo tiene capacidad de adaptación basada en las características de la data (Haofei et al. (2007).

En este estudio se aplicará BPNN (Back Propagation Neural Network), la cual es la red más usada en aplicaciones financieras (Tseng et al. (2008)).

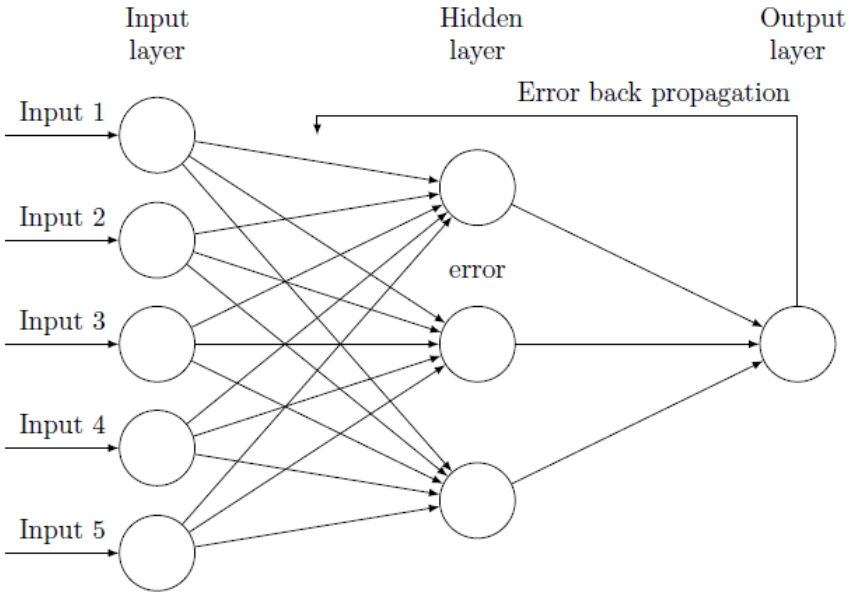

Figura 2.1: Estructura BPNN, Fuente: Elaboración propia

2.4.2. Algoritmo de entrenamiento

En una estructura estándar con una capa de entrada y x variables predictoras, una capa que mapea las entradas y salidas, y una capa de salida con la variable a predecir y. La salida y se calcula como:

$$y_i = f(\sum_{j=1}^{m} x_j w_{ij} + \theta_i)$$

Donde $w_{i,j}$ es el peso conector entre la neurona j a la neurona i, θ denota el bias (parámetro adicional que brinda grados de libertad al modelo) y $f(.)$ es la función de activación usada para controlar la amplitud de la salida (output). En este trabajo la función de activación utilizada es la sigmoide. En cuanto al algoritmo de entrenamiento, se utilizará el algoritmo de Levenberg-Marquardt (Levenberg (1944)), donde los pesos son ajustados como:

$$\Delta w_k = -(J_T^k + \mu I)^{-1} J_k e_k$$

Donde J es la matriz Jacobiana (primeras derivadas), I es la matriz identidad, μ es el parámetro de aprendizaje adaptativo seteado en 0.001 y e es el vector de errores de la red.

A pesar de lograr un mejor desempeño que los modelos econométricos, este approach no es muy práctico en términos de tiempos computacionales. En la sección siguiente se estudiará el modelo propuesto para el tema, basado en lógica difusa.

2.5. Fuzzy Logic (Lógica Difusa)

La interferencia difusa es un método que interpreta los valores de un vector de entrada y basado en un set de reglas (Reglas Difusas/Fuzzy Sets), le asigna valores al vector de salida como se muestra en la siguiente imagen:

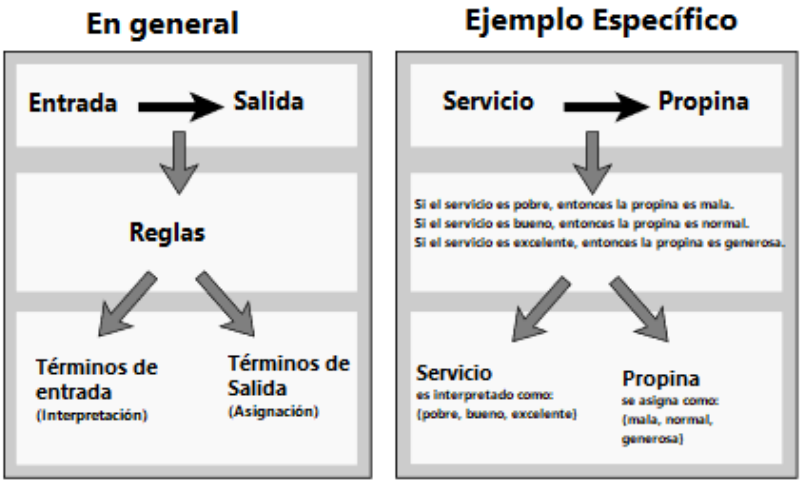

Figura 2.2: Diagrama Lógica Difusa, Fuente:Elaboración Propia

Cualquier afirmación puede ser difusa. La herramienta de razonamiento difuso permite responder una pregunta con tipo de respuesta booleana (sí,no;1,0) con un valor intermedio (sí, no mucho, no). Este tipo de aproximación es lo que generalmente realizan los humanos en su día a día.

Además, el uso de lógica difusa tiene los siguientes beneficios:

- La lógica difusa es conceptualmente fácil de enteder, sus conceptos matemáticos son simples e intuitivos.
- La lógica difusa es tolerante a la data imprecisa.
- La lógica difusa puede modelar funciones no lineales o de complejidad arbitraria.
- La metodología de la lógica difusa no reemplaza los métodos tradicionales, muchas veces incluso facilita su implementación.

Los sets difusos y los operadores difusos son los sujetos y verbos de la lógica difusa. A continuación, se revisarán al detalle cada uno de los conceptos.

Para este capítulo se utilizará el manual de H. (2016) sobre lógica difusa para Matlab.

2.5.1. if-then Rules

Las reglas difusas (if-then) son usadas para formular las afirmaciones como la siguiente:

$$si\ x\ es\ parte\ de\ A,\ entonces\ y\ es\ parte\ de\ B$$

Donde A y B son valores linguísticos definidos por sets difusos en los rangos X e Y, respectivamente. La parte del if de la regla $x\ es\ parte\ de\ A$,es llamada antecedente de o premisa, mientras que *then* es la parte de la regla $y\ es\ parte\ de\ B$ es llamada la consecuencia o conclusión.

2.5.2. Fuzzy Sets

Un set difuso es un set que puede contener elementos solo con un grado parcial de membresía al conjunto, es decir que tienen un porcentaje de pertenencia a este. Al contrario, los conjuntos concretos son aquellos donde los elementos deben pertenecer obligatoriamente a un conjunto.

2.5.3. Funciones de Membresía

Una función de membresía es una curva que define cómo cada elemento del input es mapeado a un valor o grado de membresía entre 0 y 1.

2.5.4. Fuzzy Interference System

Interferencia difusa es el proceso de formular el mapeo desde un input dado hacia un output usando lógica difusa. El mapeo provee una base donde se hacen las decisiones. El proceso de interferencia difusa implica que todos los elementos vistos en las secciones anteriores: funciones de membresía, operadores difusos y las reglas difusas. Existen dos tipos de sistemas de Interferencia Difusa, los cuales varían en cómo se determina el output: el tipo de sistema Mamdani (Mamdani y Assilian (1975)) y el sistema Sugeno (Sugeno y Nishida (1985)).

En este trabajo se estudiará este último para la proyección de volatilidad de tipo de cambio.

2.5.4.1. Sugeno

El método de interferencia difusa Takagi-Sugeno-Kang es similar a Mamdani en varios aspectos. En las primeras dos partes de la interferencia, fuzzificando las entradas y aplicando el operador difuso son las mismas. La principal diferencia entre Mamdani y Sugeno es que las funciones de membresía de salida de Sugeno son lineales o constantes. Una regla típica en el modelo tiene la forma de:

$$si\ el\ input\ es\ x\ y\ el\ input\ 2\ es\ y,\ la\ salida\ nivel\ es\ z = ax + by + c$$
$$para\ un\ modelo\ de\ orden\ 0,\ el\ nivel\ z\ es\ una\ constante$$

Cada una de las reglas posee su propio nivel de salida, z_i, y su grado de pertenencia a la regla para los inputs, w_i. Por ejemplo, para la regla AND con input de $x = 1$ e $y = 2$, el grado de pertenencia es:

$$w_i = AndMethod(F_1(x), F_2(y))$$

Donde F_1, F_2 son las funciones de membresía para los input 1 y 2.

Finalmente, la salida del sistema es la media ponderada de todas las reglas de salida:

$$OutputFinal = \frac{\sum_{i=1}^{N} w_i zi}{\sum_{i=1}^{N} w_i}$$

Donde N es el número de reglas.

A continuación, se muestra un diagrama de flujo que muestra en detalle el funcionamiento del método:

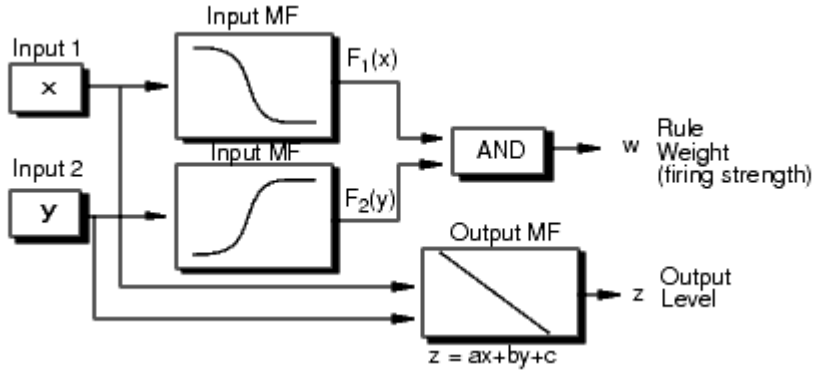

Figura 2.3: Diagrama de flujo, Fuente: MatlabR2017a Documentation

3 | Data

La data histórica utilizada comprende desde 17/02/99 hasta 03/05/17 con un total de 4578 observaciones diarias. El 80 % de las observaciones fue usado como datos de entrenamiento y el 20 % restante como datos de prueba. La base de datos fue proporcionada por el Banco Federal de Saint-Louis. Los modelos GARCH/EGARCH fueron testeados mediante Eviews mientras que los modelos híbridos, los indicadores técnicos y el sistema de interferencia difuso fueron testeados en el software Matlab.

A continuación, la tabla 4.1 muestra un resumen estadístico para las series de tipo de cambio, rentabilidad y volatilidad de USD/EUR:

Tabla 3.1: Resumen Estadístico, Fuente: Elaboración Propia

Serie	Media	Mediana	Máximo	Mínimo	Desv. Estándar	Asimetría	Curtosis	Jarque-Bera	Observaciones
Tipo de Cambio (USD/EUR)	1,210	1,243	1,601	0,827	0,176	-0,301	2,300	163,932	4609
Rentabilidad (Rt)	-5.18E-06	0,000	0,046	-0,030	0,006	0,115	5,161	900,957	4578
Volatilidad (HVt)	3.87E-05	3.40E-05	0,000	3.59E-06	2.75E-05	3,076	17,673	47.950,410	4546

- Tipo de cambio:

  En cuanto a la serie de tipo de cambio, podemos observar una media de 1,210 y una desviación estándar de 0,176. Por otro lado, la Curtosis es menor que 3 (valor en el cual la serie usualmente es Mesocúrtica), indicando que la serie presenta un bajo grado de concentración en los valores centrales de la variable y además, se tiene asimetría es negativa cercana a cero, mostrando que la serie es cercana a una curva simétrica. Finalmente, el test de normalidad de Jarque-Bera muestra que la serie no distribuye normal.

La siguiente figura ilutra el comportamiento histórico del tipo de cambio del par USD/EUR en el período estudiado:

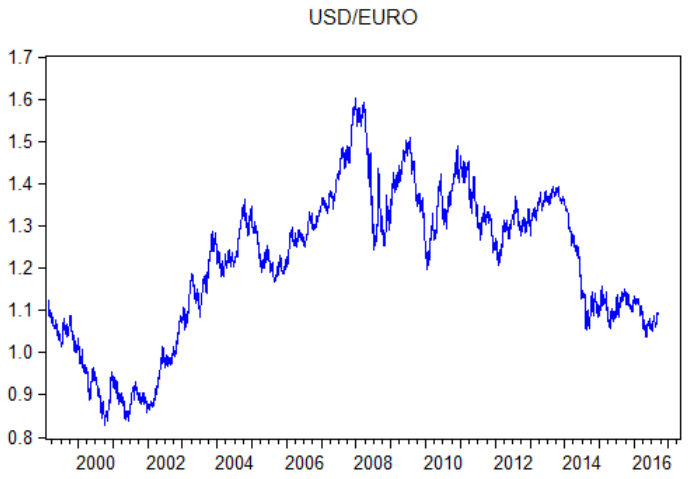

Figura 3.1: Serie de tipo de cambio USD/EUR histórica 1999-2017, Fuente:Elaboración Propia

Continuando con el análisis, a partir de la serie de tipo de cambio se obtiene la rentabilidad logarítmica:

- Rentabilidad Logarítmica

  Para la serie de Rentabilidad, se tiene una media de -5.18E-06 y una desviación estándar de 0,006. En cuanto a las medidas de concentración y asimetría se encuentra un alto grado de concentración a los valores centrales de la variable y una leve asimetría positiva. En cuanto al test de Jarque-Bera, se rechaza la hipóstesis nula de normalidad.

  El siguiente gráfico ilustra el comportamiento de la serie en el período estudiado:

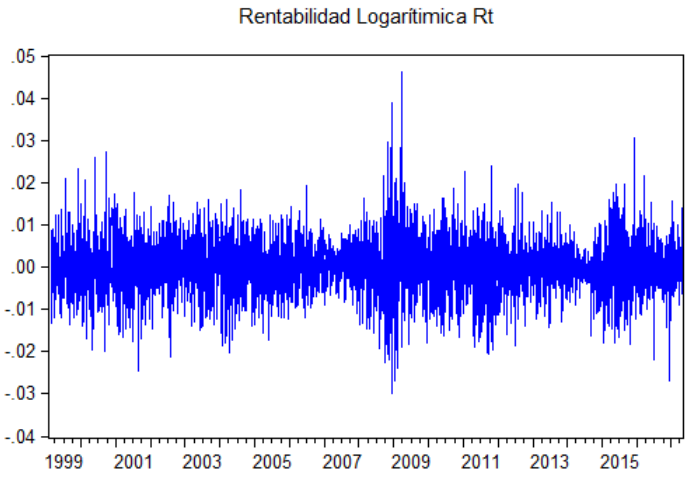

Figura 3.2: Serie de Rentabilidad Logarítmica 1999-2017, Fuente:Elaboración Propia

Finalmente, a partir de la Rentabilidad Logarítmica se genera la serie de Volatilidad Histórica (HV_t)

- Volatilidad Histórica

  En cuanto a la serie de Volatilidad Histórica, se observa una media de 3.87E-05 y una desviación estándar de 2.75E-05. Dentro del período estudiado se encuentran cúmulos de volatilidad, en donde se destacan especialmente los años 2008-2009 debido a la crisis subprime. Con respecto a la concentración, se encuentra una distribución Platicúrtica, en donde los datos están concentrados en los valores centrales de la variable y una clara asimetría positiva. Finalmente, en cuanto al test de normalidad se encontró que se rechaza la hipótesis de normalidad.

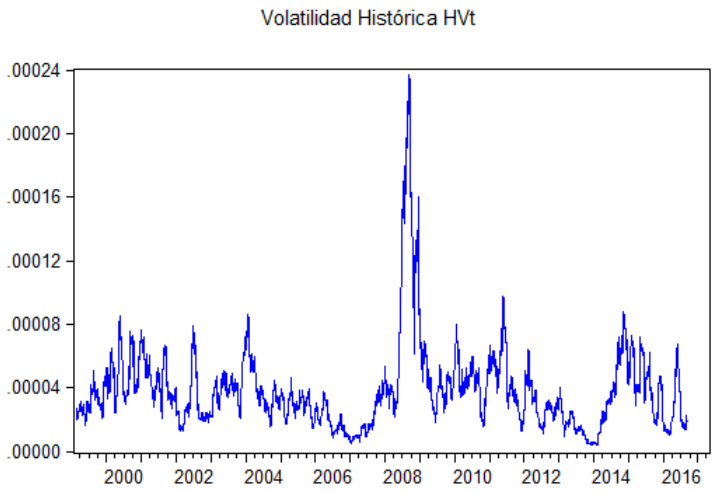

Figura 3.3: Serie de volatilidad histórica 1999-2017, Fuente:Elaboración Propia

A partir de estas tres series se ajustarán los modelos econométricos (GARCH) y se prepararán los input para las redes neuronales y el sistema de interferencia difusa.

4 | Metodología

En esta sección se desarrollarán los cuatro tipos de modelos a comparar. El primer modelo corresponde a un modelo GARCH para pronosticar la volatilidad histórica. El segundo se apoya de los pronósticos del modelo GARCH para usarlos como input en la red neuronal (ANN-GARCH). El tercer modelo corresponde a siete indicadores técnicos de FOREX, los cuales alimentarán la red neuronal para pronosticar el valor futuro de la volatilidad (ANN-TA). Finalmente, se usará un sistema de interferencia difusa alimentado por los rezagos de la serie de volatilidad.

La implementación computacional se realizó en un equipo con las siguientes características técnicas: Intel Core i5-6200U con gráficos Intel HD 520 (2,3 GHz, hasta 2,8 GHz, 3 MB de caché, 2 núcleos)

Para comparar el desempeño de los modelos se utilizarán funciones de pérdida (MAPE y MSE).

4.1. Pronóstico de Volatilidad

4.1.1. Estimación Serie de Retorno y Volatilidad Histórica

La volatilidad histórica se modelará usando la desviación estándar de la variación de precios históricos. La volatilidad histórica (HV) se calculará como la varianza de los retornos logarítmicos de una ventana móvil de 30 días (Yao et al. (2000)) .

$$HV_t = \sqrt{n^{-1} \sum_{i=t}^{t+n} (r_t - r_{prom})^2}$$

Donde r es la primera diferencia del logaritmo del precio, n es el tamaño de la ventana móvil (30 días) y r_{prom} la media de la volatilidad histórica en la ventana.

4.1.2. GARCH/EGARCH

Dado el método de cálculo estimado en la sección anterior, la volatilidad se estimará mediante varios modelos GARCH(1,1)/EGARCH(1,1) usando 3 distribuciones para el error (normal, t-student y GED) usando una ventana móvil de 252 observaciones para el pronóstico de 21 observaciones hacia adelante calculados con Eviews 8.

4.1.3. Modelos ANN-GARCH/EGARCH

Utilizando los resultados de la sección anterior y los primeros 5 rezagos de la serie de volatilidad histórica (HV_t) se alimenta la red neuronal, en la cual se probarán diferentes configuraciones de capas y neuronas (desde 2 a 5 capas y desde 5 a 20 neuronas). El target para este modelo será la serie (HV_{t+21}). La implementación computacional se realizó a través de Matlab R2016a.

4.1.4. ANN-Indicadores Técnicos

Similar al caso anterior, se utilizó como input 8 indicadores técnicos de FOREX además de los 5 primeros rezagos de la serie target (HV_{t+21}).

Se construyeron los siguientes indicadores a partir de la serie HV_t:

1. Bandas de Bollinger (Middle, Lower, Upper bands)
2. Momentum
3. Aceleración
4. Media Móvil Exponencial (EMA)
5. MACD
6. RSI

El método de cálculo y su análisis se encuentran en la sección de Marco Teórico.

4.1.5. Fuzzy Interference System

Similar al caso anterior, se pronosticará la serie de volatilidad histórica a partir de los rezagos de la serie HV_t y la misma matriz con rezagos más los indicadores técnicos calculados en la sección anterior. Ambas configuraciones utilizarán como target a $HV_{(t+21)}$.

La implementación computacional se realizó a través de Matlab R2016a a través del comando genfis2 probando una serie de rangos de clúster entre 0 y 1.

4.2. Benchmark

Para este trabajo se emplearán algunos indicadores de error que permitirán determinar y comparar su desempeño pronosticando la volatilidad histórica del par de divisas.

En cuanto a la selección de los modelos híbridos, se analizarán cada una de las configuraciones por neurona para posteriormente elegir entre aquellas que tuvieron mejor desempeño.

A continuación se estudiarán las funciones de pérdida que se utilizarán para el análisis de los resultados.

4.2.1. Funciones de Pérdida

4.2.1.1. MAPE

$$MAPE = \frac{100}{m} \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{y_i}$$

4.2.1.2. MSE

Se calcula como:

$$MSE = m^{-1} \sum_{i=1}^{m}(y_i - \hat{y}_i)^2$$

La naturaleza de esta función de pérdida es cuadrática, es decir que mientrás mas lejos se encuentre el dato de la media, este será penalizado proporcionalmente al cuadrado. Por lo tanto, los outliers son castigados de forma más fuerte que en una función de pérdida linealmente proporcional

Dentro de las funciones de pérdida estudiadas para analizar los errores se priorizará en primera instancia al MAPE, puesto que provee una comparación entre la volatilidad histórica realizada y la pronosticada en términos porcentuales.

En términos metodológicos, además del MAPE se considerará la complejidad del problema a resolver con respecto a capas y neuronas(en el caso de los modelos ANN-GARCH y ANN-TA), es decir si la diferencia entre las funciones de pérdida es realmente significativa entre uno u otro pronóstico. Para el análisis global se considerarán los tiempos computacionales y el ranking con respecto a las demás funciones de pérdida.

4.2.2. Model Confidence Set

El test para determinar cuál modelo tiene un habilidad superior de predicción (Superior Predictive Ability) consiste en determinar cuál de los modelos (k=1, ..., m) son los mejores en términos de pérdida esperada. La hipótesis nula del test es que la función de pérdida de cada alternativa sea mayor que la del modelo propuesto. Esta prueba fue propuesta por Hansen (2005).

$$H_0 = \mu \leq 0$$

Se tiene además, que $\mu \epsilon \mathbb{R}^m$ y se define μ como:

$$\mu = E(d_t)$$

y $d_t = (d_{1,t}, ...d_{m,t})'$ se define como:

$$d_{k,t} = L(\varepsilon_t, \delta_{0,t-h}) - L(\varepsilon_t, \delta_{k,t-h})$$

En donde m es la cantidad de modelos alternativos a testear, h la la cantidad de períodos, ε la variable de interés, L la función de pérdida asociada a ε y $\delta_{k,t-h}$ el conjunto de posibles reglas de decisión.

5 | Resultados

Con el fin de analizar la robustez de los resultados y buscar la mejor arquitectura de la Red Neuronal, se sensibilizarán los parámetros de el ANN, capas y neuronas. Específicamente las redes serán evaluadas con 5, 10, 15 y 20 neuronas y 2, 3, 4 y 5 capas. En el caso del Sistema de Inteferencia difusa se analizará la influencia de los diferentes radios de cluster.

5.1. GARCH/EGARCH

A continuación, se muestran las funciones de pérdida para la familia de modelos GARCH/EGARCH con las 3 distribuciones estudiadas. Los ajustes de los modelos GARCH/EGARCH se encuentran en el Anexo 1.

5.1.1. Funciones de pérdida

En siguiente tabla se compara el desempeño de la familia de modelos GARCH/EGARCH con tres distribuciones distintas.

Tabla 5.1: Funciones de pérdida para modelos GARCH/EGARCH, Fuente: Elaboración Propia

Modelo	MAPE	MSE	MCS
GARCH-normal	69,37 %	1,4790E-09	1
GARCH-student	162,79 %	4,1934E-09	0,011
GARCH-GED	160,03 %	4,9671E-09	0,012
EGARCH-normal	104,46 %	1,3617E-09	0,037
EGARCH-student	165,39 %	1,4256E-08	0,037
EGARCH-GED	103,18 %	1,2633E-09	0,037

Para el par USD/EUR se observa un mejor desempeño en términos porcetuales del modelo GARCH-normal seguido de los modelos EGARCH (GED y normal). En cuanto a la dispersión (MSE), es posible notar que el el EGARCH-GED y el EGARCH-normal tienen mejor desempeño seguido del GARCH-normal, es decir que los primeros dos modelos funcionan mejor que el GARCH-normal en períodos de alta volatilidad.

Con respecto al test de Hansen, se tiene que el modelo GARCH-normal posee la mayor capacidad predictiva, dado que su P-valor para el test es de 1.

Además, para este par se observa que los modelos EGARCH a pesar de ser más complejos (puesto que toman en cuenta los efectos asimétricos) estos son superados por el modelo GARCH-normal tanto en MAPE como en capacidad predictiva.

Los valores pronosticados en esta sección se utilizarán para alimentar la red neuronal para el modelo híbrido mostrado en la sección siguiente.

5.2. ANN-GARCH/EGARCH

A continuación, se analizarán los modelos híbridos (ANN-GARCH), los cuales usan como target a pronosticar la volatilidad histórica en $t + 21$.

5.2.1. ANN-GARCH-normal

Tabla 5.2: Funciones de pérdida para el modelo ANN-GARCH-normal, Fuente:Elaboración Propia

Modelo		ANN-GARCH-normal			
Neuronas	Capas	MAPE	MSE	MCS	Var.MAPE %
5	2	56,64 %	1,125E-09	0,165	-22,46 %
5	3	56,93 %	1,013E-09	0,087	-21,85 %
5	4	55,62 %	1,071E-09	0,279	-24,73 %
5	5	54,76 %	8,718E-10	1	-26,67 %
10	2	60,66 %	1,214E-09	0	-14,36 %
10	3	61,37 %	1,228E-09	0	-13,04 %
10	4	59,71 %	1,170E-09	0	-16,17 %
10	5	56,40 %	9,594E-10	0,174	-23,00 %
15	2	69,25 %	1,517E-09	0	-0,17 %
15	3	65,33 %	1,523E-09	0	-6,17 %
15	4	63,65 %	1,325E-09	0	-8,98 %
15	5	61,23 %	1,208E-09	0	-13,28 %
20	2	75,85 %	1,960E-09	0	8,54 %
20	3	69,86 %	1,562E-09	0	0,71 %
20	4	66,44 %	1,455E-09	0	-4,40 %
20	5	64,24 %	1,457E-09	0	-7,98 %
GARCH-normal		69,37 %	1,479E-09		

Para el modelo ANN-GARCH-normal la mejor configuración de la red obtenida corresponde a (5 neuronas, 5 capas), la cual obtiene una mejora de un 26,46 % con respecto al GARCH-normal, obteniendo un MAPE de 54,76 % y un MSE de 8,71783E-10.

En relación MCS (con 0.05 de significancia), se obtuvieron como cantidatos a mejores modelos las siguientes configuraciones: (5 capas, 2 neuronas), (5 capas, 3 neuronas), (5 capas, 4 neuronas), (5 capas, 5 neuronas) y (10 capas, 5 neuronas). Con respecto a la selección de arquitectura de la red neuronal, se nota una variación de 1,5 % en términos de MAPE entre las configuraciones (5 neuronas, 5 capas) y (5 neuronas, 4 capas). A pesar que la primera configuración es más compleja y requiere mayor tiempo de procesamiento, esta tiene el mejor rendimiento del set según el test de Hansen dentro del set, con un P-valor de 1.

Para este set se nota que un mayor número de neuronas no aumenta la precisión del pronóstico, a excepción del modelo (10 neuronas, 5 capas), el cual obtiene el tercer lugar en términos de precisión del pronóstico y el segundo en cuanto a MSE.

5.2.2. ANN-GARCH-student

Tabla 5.3: Funciones de pérdida para el modelo ANN-GARCH-student, Fuente:Elaboración Propia

Modelo		ANN-GARCH-student			
Neuronas	Capas	MAPE	MSE	MCS	Var.MAPE %
5	2	56,53 %	1,104E-09	0,132	-187,98 %
5	3	56,71 %	1,187E-09	0,132	-187,04 %
5	4	54,94 %	9,893E-10	0,855	-196,28 %
5	5	54,80 %	9,373E-10	1	-197,08 %
10	2	62,62 %	1,656E-09	0	-159,95 %
10	3	61,44 %	1,455E-09	0	-164,93 %
10	4	58,69 %	1,617E-09	0,001	-177,37 %
10	5	58,03 %	1,234E-09	0,004	-180,53 %
15	2	69,62 %	1,901E-09	0	-133,84 %
15	3	64,10 %	1,782E-09	0	-153,95 %
15	4	64,20 %	1,709E-09	0	-153,58 %
15	5	60,87 %	1,254E-09	0	-167,43 %
20	2	74,69 %	2,654E-09	0	-117,94 %
20	3	68,66 %	1,928E-09	0	-137,10 %
20	4	66,21 %	2,115E-09	0	-145,88 %
20	5	61,48 %	1,489E-09	0	-164,78 %
GARCH-student		162,79 %	4,193E-09		

Para el modelo ANN-GARCH-student la mejor configuración de la red obtenida corresponde a (5 neuronas, 5 capas), la cual mejora un 197,08 % con respecto al GARCH-student, obteniendo un MAPE de 54,80 % y un MSE de 9,37291E-10.

En relación test de Hansen, se obtuvieron como cantidatos a mejores modelos las siguientes configuraciones: (5 capas, 2 neuronas), (5 capas, 3 neuronas), (5 capas, 4 neuronas), (5 capas, 5 neuronas), (10 capas, 4 neuronas) y (10 capas, 5 neuronas). En cuanto a la arquitectura de la red neuronal, se nota que las dos mejores configuraciones (5 neuronas, 5 capas) y (5 neuronas, 4 capas) obtienen p-valores para el test de 1 y 0,855 respectivamente. Para este caso, ambos modelos tienen un desempeño similar en cuanto a funciones de pérdida.

En cuanto al resto de los modelos del set, la variación entre el MAPE con respecto a la mejor configuración, oscila entre un 3,16 % y un -4,70 %. En contraste del MSE, donde estas diferencias llegan a alcanzar un -36,22 % dentro del mismo set, implicando una menor capacidad predictiva en períodos de alta volatilidad.

Similar al caso anterior, una mayor cantidad de neuronas no mejora la precisión del pronóstico, a pesar de que dos configuraciones sí pertenezcan al set, su capacidad predictiva no es comparable con las mejores configuraciones.

5.2.3. ANN-GARCH-GED

Tabla 5.4: Funciones de pérdida para el modelo ANN-GARCH-GED, Fuente:Elaboración Propia

Modelo		ANN-GARCH-GED			
Neuronas	Capas	MAPE	MSE	MCS	Var.MAPE %
5	2	56,42 %	1,161E-09	0,238	-183,64 %
5	3	55,21 %	1,063E-09	0,765	-189,87 %
5	4	55,28 %	1,234E-09	0,765	-189,47 %
5	5	54,75 %	1,011E-09	1	-192,31 %
10	2	61,76 %	1,399E-09	0	-159,10 %
10	3	61,16 %	1,477E-09	0	-161,66 %
10	4	57,95 %	1,273E-09	0,003	-176,13 %
10	5	57,08 %	1,268E-09	0,093	-180,36 %
15	2	70,24 %	2,454E-09	0	-127,83 %
15	3	67,18 %	1,763E-09	0	-138,20 %
15	4	63,03 %	1,895E-09	0	-153,91 %
15	5	60,66 %	1,422E-09	0	-163,82 %
20	2	76,08 %	2,780E-09	0	-110,35 %
20	3	69,49 %	2,062E-09	0	-130,29 %
20	4	66,13 %	1,998E-09	0	-141,99 %
20	5	62,47 %	1,684E-09	0	-156,15 %
GARCH-GED		160,03 %	4,967E-09		

En cuanto al modelo ANN-GARCH-GED, se obtuvo como mejor configuración (5 neuronas, 5 capas), la cual mejora un 160,03 % con respecto al GARCH-student, obteniendo un MAPE de 54,75 % y un MSE de 1,0108E-09.

Al estudiar la capacidad predictiva de los modelos, se encontró que los cantidatos a mejores modelos son las siguientes configuraciones: (5 capas, 2 neuronas), (5 capas, 3 neuronas), (5 capas, 4 neuronas), (5 capas, 5 neuronas), (10 capas, 4 neuronas) y (10 capas, 5 neuronas). En cuanto a la arquitectura de la red neuronal, se tienen tres mejores configuraciones (5 neuronas, 5 capas), (5 neuronas, 3 capas) y (5 neuronas, 4 capas). El primer modelo tiene un p-valor de 1, mientras que las otras dos configuraciones tienen la misma capacidad predictiva al tener un un p-valor de 0,765.

Finalmente, al aumentar el número de neuronas se observa que el desempeño de los pronósticos no mejora de manera significativa.

5.2.4. ANN-EGARCH-normal

Tabla 5.5: Funciones de pérdida para el modelo ANN-EGARCH-normal, Fuente:Elaboración Propia

Modelo		ANN-EGARCH-normal			
Neuronas	Capas	MAPE	MSE	MCS	Var.MAPE %
5	2	56,59 %	1,043E-09	0,005	-84,58 %
5	3	55,61 %	1,010E-09	0,104	-87,85 %
5	4	54,32 %	9,118E-10	0,744	-92,29 %
5	5	54,06 %	9,293E-10	1	-93,21 %
10	2	63,62 %	1,416E-09	0	-64,19 %
10	3	60,35 %	1,179E-09	0	-73,07 %
10	4	58,56 %	1,194E-09	0	-78,37 %
10	5	57,27 %	1,146E-09	0,004	-82,40 %
15	2	66,85 %	1,607E-09	0	-56,25 %
15	3	64,86 %	1,380E-09	0	-61,04 %
15	4	62,27 %	1,278E-09	0	-67,75 %
15	5	59,91 %	1,247E-09	0	-74,35 %
20	2	72,51 %	1,716E-09	0	-44,06 %
20	3	69,18 %	1,867E-09	0	-51,00 %
20	4	64,85 %	1,411E-09	0	-61,08 %
20	5	61,38 %	1,374E-09	0	-70,17 %
EGARCH-normal		104,46 %	1,362E-09		

Con respecto al modelo ANN-EGARCH-normal la mejor configuración obtenida corresponde a (5 neuronas, 5 capas), la cual mejora el MAPE un 93,21 % con respecto al GARCH-normal, obteniendo un MAPE de 54,06 % y un MSE de 9,29271E-10.

Por otra parte, se determinaron los mejores modelos según el test de Hansen resultando como mejores modelos las siguientes configuraciones: (5 capas, 2 neuronas), (5 capas, 3 neuronas), (5 capas, 4 neuronas), (5 capas, 5 neuronas) y (10 capas, 5 neuronas). Además, dentro del set no existen variaciones porcentuales significativas con respecto al MAPE entre modelos, pero sí con respecto al MSE. Esto último es lo que provoca que la capacidad predictiva de el resto de los modelos sea inferior a las mejores configuraciones.

Similar a los casos anteriores, un mayor número de neuronas no mejora la precisión del pronóstico. También se nota que ciertos modelos con 15 y 20 neuronas cuentan con un menor desempeño en términos de MSE que el modelo econométrico.

5.2.5. ANN-EGARCH-student

Tabla 5.6: Funciones de pérdida para el modelo ANN-EGARCH-student, Fuente:Elaboración Propia

Modelo		ANN-EGARCH-student			
Neuronas	Capas	MAPE	MSE	MCS	Var.MAPE %
5	2	57,70 %	1,386E-09	0,017	-186,64 %
5	3	56,47 %	2,278E-09	0,331	-192,88 %
5	4	55,88 %	1,704E-09	0,331	-195,98 %
5	5	55,01 %	1,125E-09	1	-200,64 %
10	2	63,51 %	2,771E-09	0	-160,41 %
10	3	62,90 %	4,151E-09	0	-162,92 %
10	4	58,92 %	1,285E-09	0,008	-180,70 %
10	5	58,91 %	1,693E-09	0,008	-180,76 %
15	2	71,58 %	4,189E-09	0	-131,05 %
15	3	63,37 %	3,227E-09	0	-160,98 %
15	4	62,89 %	2,194E-09	0	-162,97 %
15	5	61,10 %	2,347E-09	0	-170,68 %
20	2	78,62 %	6,570E-09	0	-110,37 %
20	3	72,49 %	1,180E-08	0	-128,16 %
20	4	67,12 %	2,417E-09	0	-146,40 %
20	5	64,01 %	3,551E-09	0	-158,36 %
EGARCH-student		165,39 %	1,426E-08		

En cuanto al modelo ANN-EGARCH-student la mejor configuración obtenida corresponde a (5 neuronas, 5 capas), la cual mejora el MAPE un -200,64 % con respecto al modelo econométrico, resultando en un MAPE de 55,01 % y un MSE de 1,12478E-09.

Analizando la capacidad predictiva de los modelos se encontró que los mejores modelos son: (5 capas, 2 neuronas), (5 capas, 3 neuronas), (5 capas, 4 neuronas), (5 capas, 5 neuronas),(10 capas, 4 neuronas) y (10 capas, 5 neuronas). Similar a modelos anteriores, dentro del set no existen variaciones porcentuales significativas con respecto al MAPE entre modelos, pero sí con respecto al MSE, es decir que en períodos de alta volatilidad su capacidad predictiva disminuye. Estas variaciones oscilan entre 0,6 % y 22,3 %, observandose que la configuración de (10 neuronas, 4 capas) a pesar de tener el peor MAPE del set es la segunda configuración en cuanto a MSE.

5.2.6. ANN-EGARCH-GED

Tabla 5.7: Funciones de pérdida para el modelo ANN-EGARCH-GED, Fuente:Elaboración Propia

Modelo		ANN-EGARCH-GED			
Neuronas	Capas	MAPE	MSE	MCS	Var.MAPE %
5	2	55,01 %	9,479E-10	0,15	-87,58 %
5	3	55,04 %	1,016E-09	0,15	-87,47 %
5	4	55,11 %	9,367E-10	0,15	-87,23 %
5	5	53,61 %	9,013E-10	1	-92,48 %
10	2	63,82 %	1,386E-09	0	-61,69 %
10	3	60,08 %	1,283E-09	0	-71,74 %
10	4	59,11 %	1,162E-09	0	-74,57 %
10	5	57,26 %	1,115E-09	0	-80,21 %
15	2	66,74 %	1,612E-09	0	-54,61 %
15	3	63,69 %	1,309E-09	0	-62,00 %
15	4	60,83 %	1,235E-09	0	-69,61 %
15	5	59,27 %	1,034E-09	0	-74,08 %
20	2	73,64 %	1,816E-09	0	-40,11 %
20	3	69,61 %	1,607E-09	0	-48,23 %
20	4	64,94 %	1,609E-09	0	-58,89 %
20	5	61,14 %	1,389E-09	0	-68,77 %
EGARCH-GED		103,18 %	1,263E-09		

En cuanto al modelo ANN-EGARCH-GED la mejor configuración obtenida corresponde a (5 neuronas, 5 capas), la cual mejora el MAPE un 92,48 % con respecto al EGARCH-GED, resultando en un MAPE de 53,61 % y un MSE de 9,013E-10.

Mediante el test de Hansen se encontró que las configuraciones que pertenecen al MCS son: (5 capas, 2 neuronas), (5 capas, 3 neuronas), (5 capas, 4 neuronas) y (5 capas, 5 neuronas). Se observa que la última configuración planteada tiene un p-valor para el test de 1, siendo estadísticamente la que tiene una mejor capacidad predictiva. Con respecto al resto del set, se observa un desempeño parejo entre modelos con excepción de la configuración (5 neuronas, 4 capas) que a pesar de ser la última en cuanto a MAPE, es la segunda con respecto a MSE. Finalmente, el resto del set tiene la misma capacidad predictiva con un p-valor de 0,15.

5.3. ANN-Technical Analysis(ANN-TA)

Tabla 5.8: Funciones de pérdida para el modelo ANN-TA, Fuente:Elaboración Propia

Modelo		ANN-TA		
Neuronas	Capas	MAPE	MSE	MCS
5	2	43,98 %	6,1445E-10	0,168
5	3	43,10 %	6,54458E-10	0,39
5	4	43,51 %	6,22392E-10	0,269
5	5	42,63 %	5,73707E-10	1
10	2	49,97 %	9,37566E-10	0
10	3	48,68 %	8,14954E-10	0
10	4	47,12 %	7,68315E-10	0
10	5	46,01 %	8,48693E-10	0
15	2	51,55 %	9,71207E-10	0
15	3	48,83 %	8,10045E-10	0
15	4	47,19 %	7,51388E-10	0
15	5	46,65 %	7,33233E-10	0
20	2	62,92 %	1,52764E-09	0
20	3	56,00 %	1,13863E-09	0
20	4	53,18 %	9,86144E-10	0
20	5	50,38 %	9,29659E-10	0

Para el enfoque basado en indicadores técnicos, las configuraciones que pertenecen al MCS con 5 % de significancia son: (5 neuronas, 2 capas), (5 neuronas, 3 capas), (5 neuronas, 4 capas) y (5 neuronas, 5 capas). Analizando los resultados del set, se tiene que el modelo con p-valor igual a 1 corresponde a (5 neuronas, 5 capas), obteniendo un MAPE de 42,63 % y un MSE de 5,73707E-10. Las demás configuraciones de 5 capas y 2,3 y 4 neuronas se encuentran dentro del mismo orden de magnitud de MAPE y MSE.

A pesar que la configuración de (5 capas, 5 neuronas) sea más compleja y que su variación en términos de MAPE no sea significativa frente al resto de modelos del set, esta diferencia si es notoria con respecto al MSE, es decir que el modelo seleccionado tiene mejores pronósticos en períodos de alta volatilidad.

Finalmente, en cuanto las modelos con mayor número de neuronas, se obtuvo que los modelos (15 neuronas, 5 capas) y (10 neuronas, 5 capas) obtuvieron buen rendimiento a pesar de no pertenecer al Model Confidence Set.

Este enfoque tiene la ventaja de no utilizar supuestos sobre distribuciones como en los modelos GARCH y ANN-GARCH, son sencillos de calcular y logran capturar la volatilidad y la tendencia de la divisa estudiada.

5.4. Fuzzy Interference System (FIS)

5.4.1. FIS

A continuación, se muestran las funciones de pérdida para el Sistema de Interferencia Difusa (FIS) de Sugeno utilizando diferentes radios de cluster:

Tabla 5.9: Funciones de pérdida para FIS, Fuente:Elaboración Propia

Radio de Cluster	MAPE	MSE	MCS
0.15	31,17 %	2,6421E-10	0,806
0.2	31,06 %	2,6427E-10	1
0.25	31,16 %	2,6451E-10	0,806
0.35	32,12 %	2,6545E-10	0,287
0.4	32,12 %	2,6554E-10	0,287
0.45	32,28 %	2,6576E-10	0,251
0.5	33,92 %	2,6622E-10	0,11

Para el Sistema de Interferencia Difusa de Sugeno se probaron radios de cluster en el intervalo [0.15,0.5], encontrándose que el radio que entrega los mejores pronósticos corresponde a 0.2, con un MAPE de 31,06 % y un MSE de 2,64267E-10. En cuanto al test de Hansen, este modelo resulta ser el que tiene la mejor capacidad predictiva, con un p-valor igual a 1. Es importante mencionar que todos los modelos testeados pertenecen al Model Confidence Set, notando que hay configuraciones que tienen el mismo p-valor (Radios de 0.15 y 0.25 y 0.35 y 0.4), es decir la misma capacidad predictiva.

En cuanto al rango de cluster, es importante mencionar que al seleccionar un valor pequeño como óptimo (0.2), se producen muchas reglas que al utilizar un radio más grande. A pesar de esto, el valor obtenido está dentro de los rangos aceptados Demirli y Muthukumaran (2000).

Una de las ventajas del FIS es que al ser un enfoque computacional (Model Free) no necesita supuestos de distribución y requiere menos información y tiempos de procesamiento (los cuales son inferiores a un minuto, en comparación al enfoque de redes neuronales que requiere en promedio entre 4 y 6 horas)

En la siguiente sección se reforzará este enfoque utilizando los indicadores técnicos calculados en el modelo ANN-TA.

5.4.2. FIS-TA

Usando la misma base de rezagos anterior, se agregan los indicadores técnicos utilizados en la sección de ANN-TA, resultando:

Tabla 5.10: Funciones de pérdida para FIS+TA, Fuente:Elaboración Propia

Radio de Cluster	MAPE	MSE	MCS
0.3	29,79 %	2,3021E-10	1
0.45	30,20 %	2,3344E-10	0,329
0.5	30,70 %	2,3257E-10	0,217
0.65	30,64 %	2,3014E-10	0,217
0.8	31,51 %	2,4154E-10	0,217
0.95	31,51 %	2,4154E-10	0,217

Como resultado del test de Hansen, se aprecia que todos los modelos pertenecen al Model Confidence Set, teniendo 4 de ellos la misma precisión en los pronósticos (Radios de 0.5, 0.65, 0.8, 0.95) y las dos mejores configuraciones, las cuales tienen órdenes de magnitud similares en MAPE y MSE.

En cuanto a las funciones de pérdida, la mejor configuración corresponde al radio de cluster de 0,3, obteniendo un MAPE de 29,79 % y un MSE de 2,30206E-10. Cabe destacar que esta configuración es superada en términos de MSE por el modelo con radio de cluster de 0,6, obteniendo una variación no significativa de 0,03 % con respecto al primer modelo.

Similar al caso anterior, el radio de cluster que obtiene los mejores resultados en cuanto a MAPE y MSE es el Radio de 0.3. Similar al caso anterior, se encuentra dentro del rango recomendado por Chiu (1994).

Tabla 5.11: Valores recomendados en Substractive Clustering para R_a. Fuente: Elaboración propia.

Autor	Radio de Cluster
Chiu	[0.25,0.50]
Demirli	[0.15,1]

Para ambos casos los tiempos computacionales fueron inferiores a 1 minuto, logrando ser una aproximación sin supuestos (como el caso de los modelos GARCH y los modelos híbridos ANN-GARCH) similar al enfoque ANN-TA.

Finalmente, este enfoque cuenta con las mismas ventajas mencionadas en la sección anterior, además de ser reforzado por los mismos indicadores técnicos utilizados en la seción de ANN-TA. Es posible notar que tanto el enfoque de redes neuronales como el sistema de interferencia difuso ayudan a reconocer y capturar patrones no lineales (que los modelos econométricos no logran capturar) además de mejorar el pronóstico de manera significativa.

La estructura construida a través de genfis2 no realiza ninguna optimización iterativa y puede ser utilizado para preprocesar las reglas a utilizar en ANFIS.

5.5. Conclusiones y Recomendaciones

5.5.1. Conclusiones

A continuación, se muestra un resumen de las mejores configuraciones estudiadas en las secciones anteriores:

Tabla 5.12: Benchmark modelos. Fuente: Elaboración propia

Modelo	Ranking	MCS (5 %)
FIS-TA(0.3)	1	1
FIS (0.2)	2	0,017
ANN-GARCH-student (5,5)	3	0
ANN-TA	4	0
ANN-EGARCH-student (5,5)	5	0
ANN-GARCH-normal (5,5)	6	0
ANN-GARCH-GED (5,5)	7	0
ANN-EGARCH-normal (5,5)	8	0
GARCH-normal	9	0
ANN-EGARCH-GED (5,5)	10	0

A partir de los resultados obtenidos es posible concluir:

- Las redes neuronales ayudan a mejorar la performance del pronóstico de los modelos econométricos (GARCH y EGARCH), siendo esta mejora estadísticamente significativa según el test de Hansen en cada uno de los casos estudiados. En términos de resultados, los modelos econométricos e híbridos que lograron mejor performance corresponden a ANN-EGARCH-GED, ANN-EGARCH-normal, ANN-GARCH-GED y ANN-GARCH-normal.

  En el caso de los modelos econométricos, a pesar que los modelos EGARCH sean más complejos, dado que toman en cuenta los efectos asimétricos de las distribuciones, la precisión de sus pronósticos no logra superar al modelo GARCH-normal.

- El enfoque de indicadores técnicos y el Sistema de Interferencia Difusa cuentan con la ventaja de ser modelos libres de supuestos de distribución, además de ser más simples de entender y cuantificar. A diferencia del primer modelo, el Sistema de Inteferencia Difusa cuenta con bajos tiempos computacionales, dado que no requiere ningún proceso iterativo.

- Dentro del test de Hansen realizado entre todos los pronósticos se encontró que los únicos que pertenecen al set de mejores son todos los modelos FIS-TA y los modelos FIS (con clusters pequeños). Esto deja en evidencia que los indicadores técnicos mejoran la performance del Sistema de Interferencia Difusa, logrando capturar la tendencia y la volatilidad del par USD/EUR.

5.5.2. Recomendaciones

Resulta interesante ampliar este análisis a otras divisas, por ejemplo a los pares USD/JPY y USD/CAD, con el fin de generar una posible estrategia de inversión en el mercado Forex. Referente al Sistema de Interferencia Difusa propuesto en este trabajo, es posible mejorar aún más su desempeño utilizándolo como pre-procesador para ANFIS para determinar las reglas iniciales.

Bibliografía

Appel, Gerald (1985). *The Moving Average Convergence-Divergence Trading Method (Advanced Version)*. Traders Pr. 5

Bollerslev, Tim (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307 – 327. 2.3.1

Bollinger, John A. (2001). *Bollinger on Bollinger Bands*. McGraw-Hill Education. 1

Chiu, Stephen L. (1994). Fuzzy model identification based on cluster estimation. *Journal of Intelligent and Fuzzy Systems*, 2, 267–278. 5.4.2

Daly, Kevin (2008). Financial volatility: Issues and measuring techniques. *Physica A: Statistical Mechanics and its Applications*, 387(11), 2377 – 2393. 2.1

Demirli, Kudret y Muthukumaran, P. (2000). Higher order fuzzy system identification using subtractive clustering. *Journal of Intelligent and Fuzzy Systems*, 9, 129–158. 5.4.1

Engle, Robert F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4), 987–1007. 2.3.1

Escobar, Alejandro; Moreno, Julián; y Múnera, Sebastián (2013). A technical analysis indicator based on fuzzy logic. *Electronic Notes in Theoretical Computer Science*, 292, 27–37. 1

Gharleghi, Behrooz; Shaari, Abu Hassan; y Shafighi, Najla (2014). Predicting exchange rates using a novel "cointegration based neuro-fuzzy system". *International Economics*, 137, 88 – 103. 1

Giannellis, Nikolaos y Papadopoulos, Athanasios P. (2011). What causes exchange rate volatility? evidence from selected emu members and candidates for emu membership countries. *Journal of International Money and Finance*, 30(1), 39 – 61. 1

H., Godfrey (2016). *Fuzzy Logic with MATLAB*. CreateSpace Independent Publishing Platform. 2.5

Hansen, Peter Reinhard (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics*, 23(4), 365–380. 4.2.2

Haofei, Zou; Guoping, Xia; Fangting, Yang; y Han, Yang (2007). A neural network model based on the multi-stage optimization approach for short-term food price forecasting in china. *Expert Systems with Applications*, 33(2), 347 – 356. 2.4.1

Horvath, Roman y Sopov, Boril (2016). GARCH models, tail indexes and error distributions: An empirical investigation. *North American Journal of Economics and Finance*, 37, 1–15. 2.3.3

Hsu, Ming-Wei; Lessmann, Stefan; Sung, Ming-Chien; Ma, Tiejun; y Johnson, Johnnie E.V. (2016). Bridging the divide in financial market forecasting: machine learners vs. financial economists. *Expert Systems with Applications*, 61, 215 – 234. 1

Hung, Jui-Chung (2011). Applying a combined fuzzy systems and garch model to adaptively forecast stock market volatility. *Applied Soft Computing*, 11(5), 3938 – 3945. 1

Korol, Tomasz (2014). A fuzzy logic model for forecasting exchange rates. *Knowledge-Based Systems*, 67, 49 – 60. 1

Lahmiri, Salim (2017). Modeling and predicting historical volatility in exchange rate markets. *Physica A: Statistical Mechanics and its Applications*, 471, 387 – 395. 1

Levenberg, Kenneth (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2), 164–168. 2.4.2

Mamdani, E.H. y Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7(1), 1 – 13. 2.5.4

Nelson, Daniel B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2), 347–370. 2.3.2

Nelson, Daniel B. y Cao, Charles Q. (1992). Inequality constraints in the univariate garch model. *Journal of Business and Economic Statistics*, 10(2), 229–235. 2.3.1, 2.3.2

Obstfeld, Maurice y Rogoff, Kenneth (1998). *Risk and Exchange Rates*. Technical report. 1

Pring, Martin J (2002). *Technical analysis explained: The successful investor's guide to spotting investment trends and turning points*. McGraw-Hill Professional. 2.2.2

Sugeno, M. y Nishida, M. (1985). Fuzzy control of model car. *Fuzzy Sets and Systems*, 16(2), 103 – 113. 2.5.4

Tseng, Chih-Hsiung; Cheng, Sheng-Tzong; Wang, Yi-Hsien; y Peng, Jin-Tang (2008). Artificial neural network model of the hybrid egarch volatility of the taiwan stock index option prices. *Physica A: Statistical Mechanics and its Applications*, 387(13), 3192 – 3200. 2.4.1

Wilder, J. Welles (1978). *New Concepts in Technical Trading Systems*. Trend Research. 6

Yao, Jingtao; Li, Yili; y Tan, Chew Lim (2000). Option price forecasting using neural networks. *Omega*, 28(4), 455 – 466. 1, 4.1.1

Yin, Libo y Yang, Qingyuan (2016). Predicting the oil prices: Do technical indicators help? *Energy Economics*, 56, 338–350. 1

A | Anexos

A.1. Modelos GARCH/EGARCH

A.1.1. Ajuste GARCH-normal

Equation: GARCH_NORMAL Workfile: MEMORIA EXRATE USDEUR...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RT
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 06/15/17 Time: 12:35
Sample: 4298 4549
Included observations: 252
Convergence achieved after 17 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(3)*RESID(-1)^2 + (1 - C(3))*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.000109	0.000275	-0.396490	0.6917
RT(-1)	-0.056844	0.059314	-0.958352	0.3379
Variance Equation				
RESID(-1)^2	0.060690	0.013731	4.419874	0.0000
GARCH(-1)	0.939310	0.013731	68.40705	0.0000

R-squared	0.004819	Mean dependent var	-0.000152
Adjusted R-squared	0.000838	S.D. dependent var	0.005214
S.E. of regression	0.005212	Akaike info criterion	-7.640140
Sum squared resid	0.006792	Schwarz criterion	-7.598123
Log likelihood	965.6577	Hannan-Quinn criter.	-7.623234
Durbin-Watson stat	2.034052		

Figura A.1: Ajuste GARCH-normal, Fuente:Elaboración Propia

A.1.2. Ajuste GARCH-student

Equation: GARCH_STUDENT Workfile: MEMORIA EXRATE USDE...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RT
Method: ML - ARCH (Marquardt) - Student's t distribution
Date: 06/15/17 Time: 15:16
Sample: 4298 4549
Included observations: 252
Convergence achieved after 150 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.000104	0.000291	-0.357571	0.7207
RT(-1)	-0.023504	0.059363	-0.395935	0.6922
Variance Equation				
C	1.62E-06	2.14E-06	0.759900	0.4473
RESID(-1)^2	0.037125	0.039476	0.940443	0.3470
GARCH(-1)	0.902632	0.109133	8.270919	0.0000
T-DIST. DOF	5.411972	1.671152	3.238468	0.0012

R-squared	0.002720	Mean dependent var	-0.000152
Adjusted R-squared	-0.001269	S.D. dependent var	0.005214
S.E. of regression	0.005218	Akaike info criterion	-7.744693
Sum squared resid	0.006806	Schwarz criterion	-7.660659
Log likelihood	981.8313	Hannan-Quinn criter.	-7.710879
Durbin-Watson stat	2.098164		

Figura A.2: Ajuste GARCH-student, Fuente:Elaboración Propia

A.1.3. Ajuste GARCH-GED

Equation: GARCH_GED Workfile: MEMORIA EXRATE USDEURO FIN...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RT
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Date: 06/15/17 Time: 15:22
Sample: 4298 4549
Included observations: 252
Convergence achieved after 18 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.000191	0.000270	-0.707959	0.4790
RT(-1)	-0.059757	0.055768	-1.071539	0.2839
Variance Equation				
C	1.69E-06	2.56E-06	0.661592	0.5082
RESID(-1)^2	0.037840	0.044714	0.846268	0.3974
GARCH(-1)	0.899593	0.128892	6.979418	0.0000
GED PARAMETER	1.174548	0.120991	9.707701	0.0000

R-squared	0.004969	Mean dependent var	-0.000152
Adjusted R-squared	0.000989	S.D. dependent var	0.005214
S.E. of regression	0.005212	Akaike info criterion	-7.743872
Sum squared resid	0.006791	Schwarz criterion	-7.659838
Log likelihood	981.7279	Hannan-Quinn criter.	-7.710058
Durbin-Watson stat	2.028596		

Figura A.3: Ajuste GARCH-GED, Fuente:Elaboración Propia

A.1.4. Ajuste EGARCH-normal

Equation: EGARCH_NORMAL Workfile: MEMORIA EXRATE USD...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RT
Method: ML - ARCH
Date: 06/15/17 Time: 12:50
Sample: 4298 4549
Included observations: 252
Convergence achieved after 130 iterations
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(3) + C(4)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(5)
*RESID(-1)/@SQRT(GARCH(-1)) + C(6)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-8.76E-05	0.000328	-0.266791	0.7896
RT(-1)	-0.059125	0.074463	-0.794017	0.4272
Variance Equation				
C(3)	-16.22439	3.607471	-4.497441	0.0000
C(4)	0.209064	0.110255	1.896189	0.0579
C(5)	0.003524	0.080501	0.043781	0.9651
C(6)	-0.525362	0.343718	-1.528469	0.1264

R-squared	0.004779	Mean dependent var	-0.000152
Adjusted R-squared	0.000798	S.D. dependent var	0.005214
S.E. of regression	0.005212	Akaike info criterion	-7.649261
Sum squared resid	0.006792	Schwarz criterion	-7.565227
Log likelihood	969.8069	Hannan-Quinn criter.	-7.615447
Durbin-Watson stat	2.029456		

Figura A.4: Ajuste EGARCH-normal, Fuente:Elaboración Propia

A.1.5. Ajuste EGARCH-student

Equation: EGARCH_STUDENT Workfile: MEMORIA EXRATE USDEU...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RT
Method: ML - ARCH (Marquardt) - Student's t distribution
Date: 06/15/17 Time: 15:32
Sample: 4298 4549
Included observations: 252
Convergence achieved after 21 iterations
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(3) + C(4)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(5)
*RESID(-1)/@SQRT(GARCH(-1)) + C(6)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-2.87E-05	0.000290	-0.098821	0.9213
RT(-1)	-0.065725	0.060070	-1.094137	0.2739
Variance Equation				
C(3)	-18.16573	2.909129	-6.244390	0.0000
C(4)	0.183301	0.145072	1.263514	0.2064
C(5)	-0.054797	0.083899	-0.653129	0.5137
C(6)	-0.709927	0.280061	-2.534904	0.0112
T-DIST. DOF	5.112081	1.502433	3.402535	0.0007

R-squared	0.004438	Mean dependent var	-0.000152
Adjusted R-squared	0.000456	S.D. dependent var	0.005214
S.E. of regression	0.005213	Akaike info criterion	-7.734774
Sum squared resid	0.006794	Schwarz criterion	-7.636734
Log likelihood	981.5815	Hannan-Quinn criter.	-7.695324
Durbin-Watson stat	2.015830		

Figura A.5: Ajuste EGARCH-student, Fuente:Elaboración Propia

A.1.6. Ajuste EGARCH-GED

Equation: EGARCH_GED Workfile: MEMORIA EXRATE USDEURO FI...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RT
Method: ML - ARCH (Marquardt) - Generalized error distribution (GED)
Date: 06/15/17 Time: 15:41
Sample: 4298 4549
Included observations: 252
Convergence achieved after 23 iterations
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(3) + C(4)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(5)
*RESID(-1)/@SQRT(GARCH(-1)) + C(6)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-9.03E-05	0.000264	-0.342298	0.7321
RT(-1)	-0.091114	0.055840	-1.631685	0.1027
Variance Equation				
C(3)	-18.23027	3.233038	-5.638745	0.0000
C(4)	0.177502	0.155172	1.143907	0.2527
C(5)	-0.052349	0.097327	-0.537872	0.5907
C(6)	-0.716476	0.311650	-2.298979	0.0215
GED PARAMETER	1.143642	0.113144	10.10788	0.0000

R-squared	0.004544	Mean dependent var	-0.000152
Adjusted R-squared	0.000562	S.D. dependent var	0.005214
S.E. of regression	0.005213	Akaike info criterion	-7.733399
Sum squared resid	0.006793	Schwarz criterion	-7.635359
Log likelihood	981.4083	Hannan-Quinn criter.	-7.693950
Durbin-Watson stat	1.968047		

Figura A.6: Ajuste EGARCH-GED, Fuente:Elaboración Propia

A.2. Benchmark y Model Confidance Set

Tabla A.1: Funciones de pérdida y MCS al 1 %, 5 % y 10 %. Fuente: Elaboración propia.

Modelo	Capas	MAPE	MSE	MCS (1 %)	MCS (5 %)	MCS(10 %)
GARCH-normal	-	69,37 %	1,479E-09	0	0	0
GARCH-student	-	162,79 %	4,193E-09	0	0	0
GARCH-GED	-	160,03 %	4,967E-09	0	0	0
EGARCH-normal	-	104,46 %	1,362E-09	0	0	0
EGARCH-student	-	165,39 %	1,426E-08	0	0	0
EGARCH-GED	-	103,18 %	1,263E-09	0	0	0

Tabla A.2: Funciones de pérdida y MCS al 1 %, 5 % y 10 %. Fuente: Elaboración propia.

Modelo	Capas	MAPE	MSE	MCS (1 %)	MCS (5 %)	MCS(10 %)
GARCH-normal (5 neuronas)	2	56,64 %	1,125E-09	0	0	0
GARCH-normal (5 neuronas)	3	56,93 %	1,013E-09	0	0	0
GARCH-normal (5 neuronas)	4	55,62 %	1,071E-09	0	0	0
GARCH-normal (5 neuronas)	5	54,76 %	8,718E-10	0	0	0
GARCH-normal (10 neuronas)	2	60,66 %	1,214E-09	0	0	0
GARCH-normal (10 neuronas)	3	61,37 %	1,228E-09	0	0	0
GARCH-normal (10 neuronas)	4	59,71 %	1,170E-09	0	0	0
GARCH-normal (10 neuronas)	5	56,40 %	9,594E-10	0	0	0
GARCH-normal (15 neuronas)	2	69,25 %	1,517E-09	0	0	0
GARCH-normal (15 neuronas)	3	65,33 %	1,523E-09	0	0	0
GARCH-normal (15 neuronas)	4	63,65 %	1,325E-09	0	0	0
GARCH-normal (15 neuronas)	5	61,23 %	1,208E-09	0	0	0
GARCH-normal (20 neuronas)	2	75,85 %	1,960E-09	0	0	0
GARCH-normal (20 neuronas)	3	69,86 %	1,562E-09	0	0	0
GARCH-normal (20 neuronas)	4	66,44 %	1,455E-09	0	0	0
GARCH-normal (20 neuronas)	5	64,24 %	1,457E-09	0	0	0
GARCH-student (5 neuronas)	2	56,53 %	1,104E-09	0	0	0
GARCH-student (5 neuronas)	3	56,71 %	1,187E-09	0	0	0
GARCH-student (5 neuronas)	4	54,94 %	9,893E-10	0	0	0
GARCH-student (5 neuronas)	5	54,80 %	9,373E-10	0	0	0
GARCH-student (10 neuronas)	2	62,62 %	1,656E-09	0	0	0
GARCH-student (10 neuronas)	3	61,44 %	1,455E-09	0	0	0
GARCH-student (10 neuronas)	4	58,69 %	1,617E-09	0	0	0
GARCH-student (10 neuronas)	5	58,03 %	1,234E-09	0	0	0
GARCH-student (15 neuronas)	2	69,62 %	1,901E-09	0	0	0
GARCH-student (15 neuronas)	3	64,10 %	1,782E-09	0	0	0
GARCH-student (15 neuronas)	4	64,20 %	1,709E-09	0	0	0
GARCH-student (15 neuronas)	5	60,87 %	1,254E-09	0	0	0
GARCH-student (20 neuronas)	2	74,69 %	2,654E-09	0	0	0
GARCH-student (20 neuronas)	3	68,66 %	1,928E-09	0	0	0
GARCH-student (20 neuronas)	4	66,21 %	2,115E-09	0	0	0
GARCH-student (20 neuronas)	5	61,48 %	1,489E-09	0	0	0
GARCH-GED (5 neuronas)	2	56,42 %	1,161E-09	0	0	0
GARCH-GED (5 neuronas)	3	55,21 %	1,063E-09	0	0	0
GARCH-GED (5 neuronas)	4	55,28 %	1,234E-09	0	0	0
GARCH-GED (5 neuronas)	5	54,75 %	1,011E-09	0	0	0
GARCH-GED (10 neuronas)	2	61,76 %	1,399E-09	0	0	0
GARCH-GED (10 neuronas)	3	61,16 %	1,477E-09	0	0	0
GARCH-GED (10 neuronas)	4	57,95 %	1,273E-09	0	0	0
GARCH-GED (10 neuronas)	5	57,08 %	1,268E-09	0	0	0
GARCH-GED (15 neuronas)	2	70,24 %	2,454E-09	0	0	0
GARCH-GED (15 neuronas)	3	67,18 %	1,763E-09	0	0	0
GARCH-GED (15 neuronas)	4	63,03 %	1,895E-09	0	0	0
GARCH-GED (15 neuronas)	5	60,66 %	1,422E-09	0	0	0
GARCH-GED (20 neuronas)	2	76,08 %	2,780E-09	0	0	0
GARCH-GED (20 neuronas)	3	69,49 %	2,062E-09	0	0	0
GARCH-GED (20 neuronas)	4	66,13 %	1,998E-09	0	0	0
GARCH-GED (20 neuronas)	5	62,47 %	1,684E-09	0	0	0

Tabla A.3: Funciones de pérdida y MCS al 1 %, 5 % y 10 %. Fuente: Elaboración propia.

Modelo	Capas	MAPE	MSE	MCS (1 %)	MCS (5 %)	MCS(10 %)
EGARCH-normal (5 neuronas)	2	56,59 %	1,043E-09	0	0	0
EGARCH-normal (5 neuronas)	3	55,61 %	1,010E-09	0	0	0
EGARCH-normal (5 neuronas)	4	54,32 %	9,118E-10	0	0	0
EGARCH-normal (5 neuronas)	5	54,06 %	9,293E-10	0	0	0
EGARCH-normal (10 neuronas)	2	63,62 %	1,416E-09	0	0	0
EGARCH-normal (10 neuronas)	3	60,35 %	1,179E-09	0	0	0
EGARCH-normal (10 neuronas)	4	58,56 %	1,194E-09	0	0	0
EGARCH-normal (10 neuronas)	5	57,27 %	1,146E-09	0	0	0
EGARCH-normal (15 neuronas)	2	66,85 %	1,607E-09	0	0	0
EGARCH-normal (15 neuronas)	3	64,86 %	1,380E-09	0	0	0
EGARCH-normal (15 neuronas)	4	62,27 %	1,278E-09	0	0	0
EGARCH-normal (15 neuronas)	5	59,91 %	1,247E-09	0	0	0
EGARCH-normal (20 neuronas)	2	72,51 %	1,716E-09	0	0	0
EGARCH-normal (20 neuronas)	3	69,18 %	1,867E-09	0	0	0
EGARCH-normal (20 neuronas)	4	64,85 %	1,411E-09	0	0	0
EGARCH-normal (20 neuronas)	5	61,38 %	1,374E-09	0	0	0
EGARCH-student (5 neuronas)	2	57,70 %	1,386E-09	0	0	0
EGARCH-student (5 neuronas)	3	56,47 %	2,278E-09	0	0	0
EGARCH-student (5 neuronas)	4	55,88 %	1,704E-09	0	0	0
EGARCH-student (5 neuronas)	5	55,01 %	1,125E-09	0	0	0
EGARCH-student (10 neuronas)	2	63,51 %	2,771E-09	0	0	0
EGARCH-student (10 neuronas)	3	62,90 %	4,151E-09	0	0	0
EGARCH-student (10 neuronas)	4	58,92 %	1,285E-09	0	0	0
EGARCH-student (10 neuronas)	5	58,91 %	1,693E-09	0	0	0
EGARCH-student (15 neuronas)	2	71,58 %	4,189E-09	0	0	0
EGARCH-student (15 neuronas)	3	63,37 %	3,227E-09	0	0	0
EGARCH-student (15 neuronas)	4	62,89 %	2,194E-09	0	0	0
EGARCH-student (15 neuronas)	5	61,10 %	2,347E-09	0	0	0
EGARCH-student (20 neuronas)	2	78,62 %	6,570E-09	0	0	0
EGARCH-student (20 neuronas)	3	72,49 %	1,180E-08	0	0	0
EGARCH-student (20 neuronas)	4	67,12 %	2,417E-09	0	0	0
EGARCH-student (20 neuronas)	5	64,01 %	3,551E-09	0	0	0
EGARCH-GED (5 neuronas)	2	55,01 %	9,479E-10	0	0	0
EGARCH-GED (5 neuronas)	3	55,04 %	1,016E-09	0	0	0
EGARCH-GED (5 neuronas)	4	55,11 %	9,367E-10	0	0	0
EGARCH-GED (5 neuronas)	5	53,61 %	9,013E-10	0	0	0
EGARCH-GED (10 neuronas)	2	63,82 %	1,386E-09	0	0	0
EGARCH-GED (10 neuronas)	3	60,08 %	1,283E-09	0	0	0
EGARCH-GED (10 neuronas)	4	59,11 %	1,162E-09	0	0	0
EGARCH-GED (10 neuronas)	5	57,26 %	1,115E-09	0	0	0
EGARCH-GED (15 neuronas)	2	66,74 %	1,612E-09	0	0	0
EGARCH-GED (15 neuronas)	3	63,69 %	1,309E-09	0	0	0
EGARCH-GED (15 neuronas)	4	60,83 %	1,235E-09	0	0	0
EGARCH-GED (15 neuronas)	5	59,27 %	1,034E-09	0	0	0
EGARCH-GED (20 neuronas)	2	73,64 %	1,816E-09	0	0	0
EGARCH-GED (20 neuronas)	3	69,61 %	1,607E-09	0	0	0
EGARCH-GED (20 neuronas)	4	64,94 %	1,609E-09	0	0	0
EGARCH-GED (20 neuronas)	5	61,14 %	1,389E-09	0	0	0

Tabla A.4: Funciones de pérdida y MCS al 1 %, 5 % y 10 %. Fuente: Elaboración propia.

Modelo	Capas	MAPE	MSE	MCS (1 %)	MCS (5 %)	MCS(10 %)
FIS (Clusters)	0.15	31,17 %	2,642E-10	0,085	0,116	0,117
FIS (Clusters)	0.2	31,06 %	2,643E-10	0,183	0,194	0,186
FIS (Clusters)	0.25	31,16 %	2,645E-10	0,183	0,194	0,186
FIS (Clusters)	0.35	32,12 %	2,654E-10	0	0,001	0,002
FIS (Clusters)	0.4	32,12 %	2,655E-10	0	0,001	0,002
FIS (Clusters)	0.45	32,28 %	2,658E-10	0	0,001	0,002
FIS (Clusters)	0.5	33,92 %	2,662E-10	0	0,001	0,002
FIS+TA	0.3	29,79 %	2,302E-10	1	1	1
FIS+TA	0.45	30,20 %	2,334E-10	0,345	0,35	0,352
FIS+TA	0.5	30,70 %	2,326E-10	0,187	0,194	0,198
FIS+TA	0.65	30,64 %	2,301E-10	0,187	0,194	0,198
FIS+TA	0.8	31,51 %	2,415E-10	0,085	0,116	0,117
FIS+TA	0.95	31,51 %	2,415E-10	0,183	0,194	0,186
ANN-TA (5 neuronas)	2	43,98 %	6,144E-10	0	0	0
ANN-TA (5 neuronas)	3	43,10 %	6,545E-10	0	0	0
ANN-TA (5 neuronas)	4	43,51 %	6,224E-10	0	0	0
ANN-TA (5 neuronas)	5	42,63 %	5,737E-10	0	0	0
ANN-TA (10 neuronas)	2	49,97 %	9,376E-10	0	0	0
ANN-TA (10 neuronas)	3	48,68 %	8,150E-10	0	0	0
ANN-TA (10 neuronas)	4	47,12 %	7,683E-10	0	0	0
ANN-TA (10 neuronas)	5	46,01 %	8,487E-10	0	0	0
ANN-TA (15 neuronas)	2	51,55 %	9,712E-10	0	0	0
ANN-TA (15 neuronas)	3	48,83 %	8,100E-10	0	0	0
ANN-TA (15 neuronas)	4	47,19 %	7,514E-10	0	0	0
ANN-TA (15 neuronas)	5	46,65 %	7,332E-10	0	0	0
ANN-TA (20 neuronas)	2	62,92 %	1,528E-09	0	0	0
ANN-TA (20 neuronas)	3	56,00 %	1,139E-09	0	0	0
ANN-TA (20 neuronas)	4	53,18 %	9,861E-10	0	0	0
ANN-TA (20 neuronas)	5	50,38 %	9,297E-10	0	0	0