



**UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO**



**CENTRO UNIVERSITARIO UAEM TEXCOCO**

**APLICACIÓN DE LA DESCOMPOSICIÓN EMPÍRICA EN MODOS A LA PREDICCIÓN  
DEL MERCADO BURSÁTIL CON LOS MODELOS DE ARIMA-ARCH Y REDES  
NEURONALES ARTIFICIALES EVOLUTIVAS.**

**T E S I S**

**QUE PARA OBTENER EL GRADO DE:  
MAESTRO EN CIENCIAS DE LA COMPUTACIÓN.**

**PRESENTA:  
LUIS MANUEL LEÓN ANAYA.**

**TUTOR ACADÉMICO:  
DR. VICTOR MANUEL LANDASSURI MORENO.**

**TUTORES ADJUNTOS:  
DR. HÉCTOR RAFAEL OROZCO AGUIRRE.  
DRA. MARICELA QUINTANA LÓPEZ.**

**TEXCOCO, ESTADO DE MÉXICO, JULIO DE 2017.**



## Resumen

El mercado bursátil es un sistema dinámico que se caracteriza por su complejidad, volatilidad, no estacionariedad, irregularidad, pero sobre todo por las repentinas y pronunciadas caídas en los precios. Dadas estas características, y con el fin de contrarrestar las fluctuaciones aparentemente aleatorias, la inherente no linealidad en los datos financieros, y puesto que en muchos de los enfoques tradicionales que abordan la predicción del mercado bursátil en periodos de crisis, estos por lo regular no son capaces de capturar de manera fiable los rasgos distintivos del fenómeno. En esta investigación, se propone como primer paso, descomponer a los indicadores que representan al mercado accionario de los Estados Unidos y México en periodos de crisis, mediante la herramienta llamada Descomposición Empírica en Modos (DEM) que se encarga de descomponer la serie original de los índices accionarios en un número finito de descomposiciones llamadas Funciones de Modo Intrínseco (FMIs) y un elemento residual. A continuación, cada una de las FMIs y el residuo, son pronosticadas individualmente, utilizando por un lado, un modelo paramétrico (Autorregresivo Integrado de Media Móvil-Modelo de Volatilidad Condicional Heterocedástico (ARIMA-ARCH)) y por otro lado, por un modelo no paramétrico Redes Neuronales Artificiales (RNAs), este último es configurado por medio de un algoritmo evolutivo llamado Selección de Características de Programación Evolutiva de Redes Neuronales Artificiales (FS-EPNet). Posteriormente, se adquiere la predicción del modelo paramétrico, mediante la suma de las predicciones resultantes de cada FMI y del residuo, de igual forma se realiza el mismo procedimiento para obtener la predicción final del modelo no paramétrico. Finalmente, las predicciones de los modelos paramétrico y no paramétrico son combinadas mediante un promedio ponderado, para producir una combinación de pronósticos, estas predicciones a su vez son comparadas. Los resultados empíricos obtenidos demuestran que los modelos que colaboraron en conjunción con la técnica de descomposición de señales DEM, tienen una predicción más precisa de la crisis bursátil, a diferencia de los modelos que confeccionaron su pronóstico de manera aislada.

# IMPLEMENTATION OF EMPIRICAL MODE DECOMPOSITION TO STOCK MARKET PREDICTION WITH EVOLUTIONARY ARTIFICIAL NEURAL NETWORKS

## **Abstract**

The stock market is a dynamic system characterized by its complexity, volatility, non-stationarity, irregularity, but mostly by the sudden and pronounced downfalls in prices. Given these characteristics, and to counter the seemingly random fluctuations, the inherent nonlinearity in financial data, and since many of the traditional approaches that address the prediction of the stock market in times of crisis, usually are not able to reliably capture the distinctive features of the phenomenon. In this research, it is proposed as a first step, to decompose the indicators that represent the stock market in the United States and Mexico in times of crisis, through a tool called Empirical Mode Decomposition (EMD) which is responsible for decomposing the original series of stock indexes in a finite number of Intrinsic Mode Functions (IMFs) and a residue. Then each IMFs and residue are individually predicted, firstly using a parametric model (Autoregressive Integrated Moving Average (ARIMA-ARCH)) and secondly, by a non-parametric model (Artificial Neural Networks (ANNs)), the latter is designed by means of an Evolutionary Algorithm (EA) called Feature Selection EPnet (FS-EPNet). Subsequently, the prediction of the parametric model is acquired by summing the predictions resulting from each IMF and residue, similarly the same procedure is performed to obtain the final prediction of the non-parametric model. Finally, predictions of parametric and non-parametric models are combined using a weighted average to produce a combination of forecasts, in turn these predictions are compared. The empirical results show that the models who worked in conjunction with the signals decomposition technique EMD have a more accurate prediction of the market crisis, unlike the models that make its forecast in isolation.

## **Agradecimientos**

En primer lugar me gustaría expresar mi gratitud a mi asesor, el Dr. Víctor Manuel Landassuri Moreno, por haber aceptado dirigir este trabajo, así como sus enseñanzas, apoyo, consejos e infinita paciencia. Al Dr. Héctor Rafael Orozco Aguirre, por transmitirme sus conocimientos y por el tiempo otorgado. Agradezco también a aquellos que me apoyaron en mi formación como estudiante, a la Dra. Maricela Quintana López, Dr. Juan Carlos Baltazar Escalona, Dra. en Ed. Ivonne Rodríguez Pérez, M. en C. Saturnino Job Morales Escobar, Dr. Horacio Alberto García Salas, Dr. José Sergio Ruiz Castilla y al Dr. Joel Ayala de la Vega.

Además quiero agradecer al Consejo de Ciencia y Tecnología (CONACYT), al Consejo Mexiquense de Ciencia y Tecnología (COMECYT). De manera especial quisiera agradecer al Centro Universitario UAEM Valle de México y al Centro Universitario UAEM Texcoco, por su apoyo para la realización de este proyecto, además de permitirme continuar con mis estudios.

Finalmente quiero agradecer al Señor por dejarme saber que él me cuida y me guía.

## **Dedicatoria**

Durante el transcurso de mi vida siempre he tenido a mi lado a dos personas que siempre me han apoyado, todo lo que he realizado y lo que soy se los debo a ellos, y no hay manera alguna en la cual pueda pagarles. Desde lo más profundo de mi corazón quiero expresarles mi agradecimiento, este trabajo está dedicado a mis padres, Carlos y Marina. Para mí es un honor ser su hijo.



# ÍNDICE

Capítulo 1. Introducción .....	1
1.1 Antecedentes .....	1
1.2 Planteamiento del problema.....	2
1.3 Objetivos .....	6
1.3.1 Objetivo general.....	6
1.3.2 Objetivos específicos .....	6
1.4 Hipótesis .....	6
1.5 Justificación.....	6
1.6 Delimitación.....	7
1.7 Metodología .....	8
1.8 Organización de la tesis.....	11
Capítulo 2. Mercado accionario.....	13
2.1 Teoría económica .....	13
2.2 Definición del mercado financiero .....	14
2.3 Función del mercado financiero .....	15
2.3.1 Acción .....	15
2.3.2 Riesgo.....	15
2.3.3 Crack bursátil .....	15
2.3.4 Análisis técnico .....	16
2.3.5 Análisis fundamental.....	17
Capítulo 3. Indicadores bursátiles .....	19
3.1 IPC de la Bolsa Mexicana de Valores .....	19
3.2 Índice Dow Jones de los Estados Unidos .....	19



3.3 Hipótesis de los Mercados Eficientes (HME) .....	19
3.4 Teoría del paseo aleatorio .....	22
3.4.1 Teoría Moderna del Portafolio (TMP).....	23
3.4.2 Modelo de Valoración de Activos Financieros (CAPM).....	24
Capítulo 4. Modelos de predicción .....	27
4.1 Antecedentes .....	27
4.2 Metodología estadística .....	30
4.2.1 Series de tiempo .....	30
4.2.2 Distribución normal .....	34
4.2.2 Prueba de Portmanteau .....	36
4.2.3 Tipos de pronósticos .....	36
4.3 Modelo paramétrico ARIMA-ARCH.....	38
4.3.1 Introducción .....	38
4.3.2 Modelo Autorregresivo Integrado de Media Móvil (ARIMA) .....	39
4.3.3 Metodología Box-Jenkins.....	40
4.3.4 Modelo Estacional Autorregresivo Integrado de Media Móvil .....	42
4.3.5 Modelos de Volatilidad Condicional Heterocedástica .....	44
4.3.5.1 Introducción .....	44
4.3.5.2 Determinación del modelo ARCH(q).....	47
4.4 Modelo no paramétrico.....	49
4.4.1 Redes neuronales artificiales .....	50
4.4.1.1 Introducción .....	50
4.4.1.2 Neurona biológica .....	50
4.4.1.3 Neurona artificial y sus particularidades .....	51

4.4.1.4	Aprendizaje mediante el algoritmo de retropropagación de errores..	61
Capítulo 5.	Algoritmos y técnicas utilizadas.....	67
5.1	Transformada de Hilbert-Huang y descomposición empírica en modos .....	67
5.1.1	Introducción.....	67
5.1.2	Teoría de la descomposición empírica en modos .....	67
5.2	Algoritmos evolutivos .....	72
5.2.1	Antecedentes .....	72
5.2.2	Características de los algoritmos evolutivos .....	73
5.2.3	Representación de los individuos.....	76
5.2.4	Neuroevolución (NE).....	77
5.2.5	Algoritmo Selección de Características de Programación Evolutiva de Redes Neuronales Artificiales (FS-EPNet) .....	80
5.3	Algoritmo de recocido simulado .....	88
5.4	Exponente de Hurst.....	90
5.4.1	Introducción.....	90
5.4.2	Historia del exponente de Hurst .....	91
5.4.3	Interpretación del exponente de Hurst .....	92
5.5	Métodos de combinación de pronósticos .....	92
5.5.1	Introducción.....	93
5.5.2	Teoría de la combinación de pronósticos .....	93
5.5.2.1	Promedio simple.....	95
5.5.2.2	Promedio ponderado.....	96
Capítulo 6.	Estudio empírico.....	97
6.1	Introducción.....	97

6.2 Selección de las muestras.....	98
6.3 Configuración experimental del IPC y DJIA .....	100
6.4 Etapas de experimentación del IPC .....	105
6.4.1 Descomposición del IPC .....	108
6.4.2 Análisis de los datos del IPC con el exponente de Hurst .....	110
6.4.3 Experimentación del IPC.....	111
6.4.3.1 Etapa de experimentación A.....	112
6.4.3.1.1 Escenario I.....	112
6.4.3.1.2 Escenario II.....	113
6.4.3.2 Etapa de experimentación B .....	115
6.4.3.3 Combinación de pronósticos. Etapa C.....	120
6.5 Fases experimentales del DJIA.....	123
6.5.1 Descomposición del DJIA .....	125
6.5.2 Análisis de los datos del DJIA con el exponente de Hurst .....	126
6.5.3 Experimentación del DJIA.....	127
6.5.3.1 Fase experimental A .....	127
6.5.3.2 Fase experimental B .....	131
6.5.3.3 Fase C. Combinación de pronósticos. ....	134
Capítulo 7. Conclusiones y trabajos futuros .....	137
7.1 Conclusiones.....	137
7.2 Recomendaciones para trabajos futuros.....	139
Apéndice .....	141
Bibliografía.....	153
Anexo .....	162

## Lista de figuras

Figura 4. 1: Componentes de una Serie de Tiempo desde la perspectiva clásica. ....	32
Figura 4. 2: Distribución Normal. ....	34
Figura 4. 3: Representación gráfica de los tipos de curvas. ....	35
Figura 4. 4: Metodología Box-Jenkins. ....	42
Figura 4. 5: Ejemplo: Correlograma de una Serie de Tiempo con comportamiento estacional. ....	44
Figura 4. 6: Ejemplo de agrupamiento de la volatilidad. ....	46
Figura 4. 7: Distribución Leptocúrtica. ....	47
Figura 4. 8: La Neurona Biológica. ....	51
Figura 4. 9: Neurona Artificial. ....	51
Figura 4. 10: Arquitectura de un Perceptrón. ....	56
Figura 4. 11: Algoritmo de aprendizaje del Perceptrón. ....	57
Figura 4. 12: Algoritmo de aprendizaje de la red ADALINA. ....	60
Figura 4. 13: Algoritmo de Retropropagación. ....	62
Figura 5. 1: Boceto del proceso de cernido. ....	70
Figura 5. 2: IPC del 17/09/2001 al 13/03/2009 y sus descomposiciones. ....	71
Figura 5. 3: Diagrama básico del funcionamiento del algoritmo FS-EPNet. ....	85
Figura 5. 4: Comparación de Métodos de Selección por Ruleta y Rango. ....	86
Figura 5. 5: Esquema de la mutaciones del algoritmo FS-EPNet. ....	87
Figura 5. 6: Esquema básico del algoritmo de Recocido Simulado. ....	90
Figura 6. 1: Sistema Experimental. ....	104
Figura 6. 2: Etapas de predicción del IPC del 08/07/2002 al 16/12/2008. ....	105
Figura 6. 3: IPC del 08/07/2002 al 13/03/2009, y sus descomposiciones. ....	109
Figura 6. 4: Etapa A del IPC. Escenario I. Periodo del 29/08/2008 al 17/10/2008, 35 días de pronóstico. ....	112
Figura 6. 5: Etapa A del IPC. Escenario II. Periodo del 29/08/2008 al 17/10/2008, 35 días de pronóstico. ....	113
Figura 6. 6: Combinación de pronósticos. Etapa A del IPC. Escenario II. Del 29/08/2008 al 17/10/2008. ....	114
Figura 6. 7: a) Diseño topológico de un GMLP mediante FS-EPNet para pronosticar la FMI <sub>3</sub> del IPC. b) Ejemplo de un MLP con 59 neuronas de entrada, 35 neuronas ocultas. ....	117
Figura 6. 8: IPC del 15/10/2008 al 31/10/2008, 13 días. Predicción del 20/10/2008 al 31/10/2008, ....	118
Figura 6. 9: IPC Etapa B de experimentación, predicción del 20/10/2008 al 16/12/2008, 40 días. ....	119
Figura 6. 10: Etapa C. Combinación de Predicciones del IPC. ....	122

Figura 6. 11: Fases de predicción del DJIA, del 06/01/1989 al 12/19/1997. ....	124
Figura 6. 12: DJIA del 06/01/1989 al 10/23/1997, y sus descomposiciones .....	126
Figura 6. 13: Fase A. Predicción de los modelos no paramétricos del DJIA. Del 09/05/1997 al 10/23/1997. ....	128
Figura 6. 14: Fase A. Predicción DJIA de modelos paramétricos. Del 09/05/1997 al 10/23/1997. ....	128
Figura 6. 15: Fase A. Combinación de pronósticos del DJIA. ....	130
Figura 6. 16: Fase B. Predicción 8 días, DJIA. Del 10/24/1997 al 11/04/1997. ....	132
Figura 6. 17: Fase B. Predicción de modelos no paramétricos a 40 días del DJIA. ...	133
Figura 6. 18: Fase B. Predicción de modelos paramétricos a 40 días del DJIA. ....	133
Figura 6. 19: Fase C. Combinación de pronósticos para el DJIA, 40 días. Del 10/24/1997 al 12/19/1997. ....	135
Figura A. 1: Serie de Tiempo DJIA-FMI <sub>4</sub> . ....	141
Figura A. 2: Histograma del DJIA-FMI <sub>4</sub> .....	142
Figura A. 3: Correlograma del DJIA- FMI <sub>4</sub> . ....	143
Figura A. 4: Estadístico Q de Ljung-Box para DJI-FMI <sub>4</sub> . ....	145
Figura A. 5: Correlograma de los residuales al cuadrado DJIA-FMI <sub>4</sub> . ....	146
Figura A. 6: Correlograma de los residuales y correlograma al cuadrado de los residuales de DJIA-FMI <sub>4</sub> . ....	148
Figura A. 7: Correlograma de DJIA, DJIA-FMI <sub>2</sub> , DJIA-FMI <sub>6</sub> y DJIA-FMI <sub>10</sub> . ....	149
Figura A. 8: RNAs diseñadas por FS-EPNet y los parámetros del modelo ARIMA- ARCH para resolver las series del DJIA, DJIA-FMI <sub>2</sub> , DJIA-FMI <sub>6</sub> y DJIA-FMI <sub>10</sub> . ....	150

## Lista de tablas

Tabla 4. 1: La tabla indica el número de veces que un modelo de pronóstico predice correctamente la dirección de la rentabilidad del índice sobre los 60 períodos de pronóstico fuera de la muestra de enero de 1991 a diciembre de 1995. ....	28
Tabla 4. 2: Cálculo del MSE: dentro y fuera de la muestra. GARCH y RNA .....	29
Tabla 4. 3: Técnicas de Predicción; I) SSP/Estático y II) MSP/MSF. ....	38
Tabla 4. 4: Ejemplo: predicción con el modelo ARMA. ....	40
Tabla 4. 5: Simbología del aprendizaje del Perceptrón. ....	58
Tabla 6. 1: Exponente de Hurst para el IPC y sus descomposiciones. ....	111
Tabla 6. 2: Medición de errores: IPC Etapa A. Escenario I del 29/08/2008 al 17/10/2008. ....	113
Tabla 6. 3: Medición de errores: IPC Etapa A. Escenario II del 29/08/2008 al 17/10/2008. ....	114
Tabla 6. 4: Combinación de pronósticos. Medición de errores: IPC Etapa A. ....	115
Tabla 6. 5: IPC Etapa B. Parámetros de los modelos Paramétrico y No Paramétrico. ...	116
Tabla 6. 6: Medición de errores: IPC Etapa B. 10 días del 20/10/2008 al 31/10/2008 ...	118
Tabla 6. 7: Medición de error: IPC Etapa B, para 13, 17 y 40 días de predicción, del 20/10/2008 al 16/12/2008. ....	120
Tabla 6. 8: Combinación de pronósticos, Etapa C. Medición de errores: predicción del IPC para 5, 11 y 40 días del 20/10/2008 al 16/12/2008 .....	123
Tabla 6. 9: Exponente de Hurst para el DJIA y sus descomposiciones. ....	127
Tabla 6. 10: Error: DJIA Fase A, para 5, 15 y 35 días. ....	129
Tabla 6. 11: Parámetros: modelos Paramétrico y No-Paramétrico. Etapa B, DJIA. ...	131
Tabla 6. 12: Errores: DJIA Fase B, a 8 días del 20/10/2008 al 31/10/2008. ....	132
Tabla 6. 13: Formulación de la Combinación Ponderada del DJIA. ....	134
Tabla 6. 14: Combinación de pronósticos, Etapa C. Medición de errores: predicción del DJIA para 5, 10, 20 y 40 días del 10/24/1997 al 12/19/1997. ....	136
Tabla A. 1: Prueba de raíz unitaria para la serie de DJIA-FMI4. ....	144
Tabla A. 2: Criterios de selección: AIC, BIC y HQC. ....	145
Tabla A. 3: Prueba ARCH a los residuos del modelo ajustado de DJIA-FMI4. ....	147
Tabla A. 4: Resultados de los criterios AIC y BIC para los modelos: ARCH(5), GARCH(1,1), para la serie DJIA-FMI4. ....	147
Tabla A. 5: Prueba ARCH a los residuales del modelo ajustado ARCH(5). ....	148

## Lista de acrónimos

<b>AE</b>	Algoritmo Evolutivo .....	5
<b>AG</b>	Algoritmos Genético .....	5
<b>AIC</b>	Criterio de Información de Akaike .....	101
<b>ADF</b>	Dickey-Fuller Aumentado .....	33
<b>ANOVA</b>	Análisis de Varianza .....	100
<b>AR</b>	Autorregresivo .....	39
<b>ARMA</b>	Autorregresivo de Media Móvil.....	36
<b>ARIMA</b>	Autorregresivo Integrado de Media Móvil.....	11
<b>ARIMA-ARCH</b>	Autorregresivo Integrado de Media Móvil-Modelo de Volatilidad Condiciona l Heterocedástico .....	11
<b>ARIMA-GARCH</b>	Autorregresivo Integrado de Media Móvil-Modelo de Volatilidad Condiciona l Heterocedástico Generalizado .....	93
<b>BIC</b>	Criterio de información Bayesiano .....	101
<b>BMV</b>	Bolsa Mexicana de Valores .....	9
<b>CP</b>	Combinación Ponderada .....	120
<b>BP</b>	Retropropagación de errores .....	27
<b>CAPM</b>	Modelo de Valoración de Activos Financieros.....	24
<b>DEM</b>	Descomposición Empírica en Modos.....	4
<b>DJIA</b>	Dow Jones Industrial Average .....	9
<b>EE</b>	Estrategias Evolutivas .....	74
<b>EPNet</b>	Programación Evolutiva de Redes Neuronales Artificiales .....	80

<b>FS-EPNet</b> Selección de Características de Programación Evolutiva de Redes Neuronales Artificiales .....	5
<b>FAC</b> Función de Autocorrelación .....	41
<b>FACP</b> Función de Autocorrelación Parcial .....	41
<b>FMI</b> Función de Modo Intrínseco .....	9
<b>GARCH</b> Modelo de Volatilidad Condicional Heterocedástico Generalizado .....	29
<b>GMLP</b> Perceptrón Multicapa Generalizado .....	9
<b>HQC</b> Criterio de información de Hannan-Quinn .....	144
<b>HME</b> Hipótesis de los Mercados Eficientes .....	3
<b>IPC</b> Índice de Precios y Cotizaciones.....	9
<b>THH</b> Transformada de Hilbert-Huang.....	67
<b>LMS</b> Least Mean Square.....	60
<b>MBP</b> Modified Back Propagation.....	82
<b>MA</b> Media Móvil .....	39
<b>MLP</b> Perceptrón Multicapa .....	5
<b>MSF</b> Multi-Step-Ahead Forecasts .....	37
<b>MSP</b> Multi-step Prediction .....	37
<b>MSE</b> Mean Squared Error .....	29
<b>NE</b> Neuroevolución .....	76
<b>NEC</b> Neuroevolución Convencional.....	80
<b>PE</b> Programación Evolutiva .....	74
<b>PS</b> Promedio Simple.....	10
<b>PP</b> Promedio Ponderado .....	10



<b>RNA</b> Red Neuronal	Artificial 5	
<b>RNAE</b> Redes Neuronales Artificiales Evolutivas		79
<b>RNAA</b> Red Neuronal Artificial Autorregresiva		93
<b>ST</b> Serie de	Tiempo 5	
<b>SA</b> Recocido Simulado		66
<b>SARIMA</b> Modelo Estacional Autorregresivo Integrado de Media Móvil		43
<b>SSP</b> Single-Step-Ahead Prediction		37
<b>TMP</b> Teoría Moderna del Portafolio		11
<b>VAR</b> Vectores Autorregresivos		28
<b>WTI</b> West Texas Intermediate		93

# Capítulo 1. Introducción

## 1.1 Antecedentes

En las empresas el pronóstico es una herramienta fundamental en la planeación, así como también en muchas ocasiones es la base de la toma de decisiones operativas y estratégicas. Un pronóstico efectivo puede brindar información valiosa de un posible futuro, además el pronóstico puede ser coadyuvante en la disminución de la incertidumbre y en la gestión del riesgo. El comportamiento y funcionamiento de los mercados financieros en las últimas décadas ha sufrido cambios estrepitosos, principalmente por la desregulación de los controles en el flujo financiero y en la fijación del tipo de cambio. Algunas señales que muestran un cambio sustancial son: la refutación de los fundamentos que soportan tanto a la teoría económica como al comportamiento del mercado, así como el aumento sin precedente de la volatilidad de los precios en la mayoría de los mercados.

Cabe destacar que de entre los diferentes mercados, el bursátil suele ser el más atractivo para los inversionistas puesto que existe una gran liquidez, se pueden obtener altas tasas de rendimiento a diferencia de otros instrumentos de inversión, además por el hecho de que hay un gran potencial de ganancia. Dados los argumentos anteriormente expuestos tanto el gobierno por medio del banco central, la industria pública y privada, así como los inversionistas mantienen una estrecha vigilancia en el comportamiento del mercado de valores, y sobre los diferentes sucesos que pudieran incidir sobre él. En los últimos años se ha incrementado de manera notoria el interés por analizar y predecir el comportamiento y el valor futuro de las series de tiempo financieras. Una de las tareas de predicción más complejas en el moderno ámbito financiero es la estimación del acontecer futuro del mercado de valores. Una infinidad de factores que inciden en el mercado de valores como pueden ser; las tasas de interés, sucesos políticos, políticas monetarias, factores tecnológicos, situación económica, expectativas de los inversionistas, etc., conforman una de las actividades con mayor dificultad de predicción. En la actualidad prevalece el debate entre si el mercado de valores es o no predecible. Los mercados financieros y particularmente el mercado de valores se encuentra

inmerso en un ambiente inherentemente plagado de riesgo e incertidumbre que lo hace altamente inestable, la necesidad de contar con herramientas precisas de predicción es mayor, y por lo tanto, es necesario perfeccionar los modelos que han demostrado ser confiables así como ampliar el horizonte en la búsqueda de modelos que se ajusten y comprendan de una manera más fidedigna el complejo funcionamiento del mercado de valores y los entornos en los cuales interactúa. En el entorno económico los índices de los mercados de valores resultan ser de los indicadores más importantes para los diversos agentes económicos, ya que los movimientos de estos indicadores inciden de manera notoria sobre el resto de la economía, y al ser estos los indicadores adelantados de referencia, el prever su devenir es de gran importancia para el ámbito académico pero especialmente en el profesional.

Sin embargo, a pesar de los grandes avances en la computación, en la ciencia estadística, en el análisis del mercado y en la gran amplitud en la disponibilidad de la información, el pronóstico no logra asumir el papel preponderante que se supone debe tener. La investigación tiene como primera premisa refutar la eficiencia del mercado accionario, demostrando de manera empírica que es posible en cierta medida predecir de manera certera su comportamiento.

## **1.2 Planteamiento del problema**

Los mercados financieros son una pieza clave en el desarrollo de la industria y del comercio, su comportamiento incide de manera notoria en la evolución de la economía de un país. Con el devenir de los años, la globalización de los mercados financieros va en aumento, la constante e incesante eliminación de restricciones entre los mercados financieros nacionales con los mercados financieros mundiales tiene como consecuencia que los capitales no tengan fronteras. Dado que el grado de interdependencia entre los mercados financieros es mayor, así también es el riesgo que se asume cuando surge una crisis. Con el fin de insertarse en el mercado global, la economía de muchos países se ha liberalizado y en muchos casos su planeación ha sido deficiente. Esto trae como consecuencia que se encuentren en clara desventaja ante un mercado voraz y competitivo, además de que sean más

propensos a sufrir el contagio de una crisis, y que el nivel de las afectaciones a estas naciones suele ser mayor. En las últimas décadas las caídas o crisis bursátiles se han convertido en un fenómeno con mayor grado de recurrencia. Estas caídas bursátiles son inherentes al sistema: solo es cuestión de que transcurra un determinado tiempo para que vuelva a producirse una de ellas y los efectos que desencadenan dichas caídas suelen ser devastadores, además de que sus consecuencias pueden perdurar por largos periodos.

Por estas razones tanto el gobierno, la industria, el banco central y los inversionistas deben mantener una vigilancia constante al comportamiento del mercado de valores y sobre los diferentes acontecimientos que pudieran repercutir sobre él, tanto de manera negativa como positiva. Por lo descrito anteriormente, es menester de los analistas crear modelos que expliquen las causas y efectos de las crisis. De manera primordial deben situarse en una posición que les permita pronosticar acertadamente y de forma anticipada dichas crisis con el propósito de tomar previsiones y medidas que mermen su impacto.

En lo concerniente a la predictibilidad del mercado bursátil, básicamente existen dos corrientes que se contraponen. La primera (Bachelier, 1900; Cowles, 1933; Fama, 1970; Malkiel, 1992), la que se apoya sobre uno de los paradigmas más controversiales en el ámbito financiero, la Hipótesis de los Mercados Eficientes (HME), la cual fue teóricamente formalizada en el año de 1970 por Eugene Fama en una innovadora e influyente investigación (Fama, 1970). Donde se determinó que no es posible pronosticar la rentabilidad de un activo financiero, en especial aquel activo comercializado en el mercado accionario (ver sección 3.3). Por otro lado, la segunda corriente propone que el mercado de valores no es totalmente eficiente y que efectivamente tiene memoria, es decir que en cierta medida la información histórica del mercado accionario guarda cierta relación con el comportamiento futuro de dicho mercado. De entre algunos autores que fundamentan sus investigaciones tomando como premisa el postulado anterior y aportan evidencia empírica que sustenta la predictibilidad del mercado bursátil, tomando como materia prima la información histórica en forma de series de tiempo

(STs) de variables económicas y financieras, se destacan: Keim y Stambaugh (1986), Lo y MacKinlay (1988), Campbell (1987), Balvers, Cosimano, y McDonald (1990), Schwert (1990), Leung, Daouk y Chen (2000), Zhang (2004), Satchell y Knight (2007), entre otros. Es así que una de las cuestiones que intenta solventar esta investigación, es obtener evidencia empírica que clarifique el comportamiento del mercado accionario con el fin de refutar la HME.

Una gran cantidad de factores afectan el comportamiento del mercado accionario, que lo convierte en un proceso con un dinamismo volátil e inestable, y la forma tradicional de abordar su estudio es a través del supuesto de una relación lineal como una aproximación. Como consecuencia de lo anteriormente expuesto, los datos provenientes del mercado accionario son inherentemente no lineales, se encuentran plagados de ruido y con un alto grado de complejidad por lo que el análisis de este fenómeno por medio de las herramientas tradicionales pudiera estar incompleto. Tomando en consideración las características de los datos, y como consecuencia de la vasta literatura financiera que indica que los índices bursátiles de los mercados de valores representados a través de una ST son de naturaleza cercana a la aleatoriedad o en definitiva son inciertas (Seiler y Rom, 1997; Urrutia, 1995), se propone utilizar un análisis alternativo al tradicional. Por ello en primer lugar se propone descomponer los indicadores con el fin de atenuar el ruido intrínseco contenido en los datos, además se parte del supuesto de que con una desagregación de la serie es posible disminuir la complejidad de estos y encontrar mayor claridad en los datos con el fin de facilitar su análisis y su posterior predicción. El método utilizado para desagregar los datos es la Descomposición Empírica en Modos (DEM) introducida por Huang et al. (1998). La DEM es una técnica guiada por los datos, por lo tanto las series emanadas mediante la descomposición por DEM y que son insertadas al modelo son provenientes de los mismos datos existentes en el modelo y no son datos externos al sistema.

En lo referente a los métodos de pronóstico, estos se pueden dividir en términos generales en dos categorías: en estadísticos y en las técnicas basadas en Inteligencia Artificial (IA). Herramientas como ARIMA, Regresión Múltiple y

Suavizado Exponencial son ejemplos de métodos estadísticos, mientras que los paradigmas de IA incluyen Sistemas Difusos de Inferencia, Algoritmos Genéticos (AGs), Redes Neuronales Artificiales (RNAs), Aprendizaje Automático o Machine Learning (ML), etc. (Zhang et al., 2001).

Los métodos de pronósticos utilizados en esta investigación son: por un lado, con un instrumento de carácter lineal como lo es el modelo paramétrico ARIMA-ARCH y por el otro lado, con una herramienta no paramétrica capaz de modelar relaciones no lineales, las denominadas RNAs. De esta forma se intenta capturar tanto el comportamiento lineal, así como el no lineal, mediante el empleo de ambas técnicas. Las RNAs son una de las herramientas más conocidas para predecir Series de Tiempo (STs) económicas y financieras, así como también, se considera como un aproximador de funciones altamente flexible y universal para el reconocimiento de patrones y como instrumento clasificador (Chen et al., 2009). Por ello, la elección de este modelo, sin embargo, la edificación de un modelo neuronal no es una tarea trivial, de entre los diferentes modelos neuronales se destaca como el modelo más popular el caso del Perceptrón Multicapa (MLP) (Zhang, 2004; Andersson y Palm, 2009), donde es indispensable determinar a priori el número de capas y de neuronas por capa, la tasa de aprendizaje, la inicialización de los pesos y del umbral, las funciones de activación, entre otros parámetros.

Dado que el diseño topológico de una Red Neuronal Artificial (RNA), tiene una gran influencia en el desempeño mostrado y en su capacidad de generalización, se propone utilizar un Algoritmo Evolutivo (AE) que auxilie a configurar la arquitectura lo más cercano posible al óptimo. Por ello, el algoritmo Selección de Características de Programación Evolutiva de Redes Neuronales Artificiales (FS-EPNet) (Landassuri-Moreno y Bullinaria, 2009b; Landassuri-Moreno y Bullinaria, 2011) se utiliza para configurar la arquitectura y los pesos del modelo neuronal elegido. De manera similar a la forma tradicional en cómo se aborda el análisis del mercado accionario, en esta investigación se propone estudiar el comportamiento del mercado accionario por medio los índices bursátiles, ya que estos reflejan el comportamiento del mercado.

### **1.3 Objetivos**

Tomando en cuenta que el mercado bursátil no es un fenómeno determinista y que no es posible predecir su comportamiento con total exactitud. Lo que pretende esta investigación es cumplir con lo siguiente:

#### **1.3.1 Objetivo general**

Proponer un modelo mediante el cual se aumente la probabilidad de pronosticar con un mayor grado de exactitud la evolución del mercado accionario en un periodo de crisis, y con ello poder cuestionar la validez de la Hipótesis del Mercado Eficiente.

#### **1.3.2 Objetivos específicos**

- Diagnosticar si mediante el uso de un modelo no paramétrico (RNA) se obtienen predicciones más precisas, y si estos pueden ser candidatos a substituir a los modelos ARIMA-ARCH o en cambio sirven de complemento a una predicción más robusta mediante la combinación de predicciones.
- Determinar si es beneficioso descomponer una señal con la herramienta DEM, con el fin de alcanzar una comprensión más clara del fenómeno que se analiza y con esto se puedan develar los procesos subyacentes que se encuentran ocultos para los modelos de predicción.
- Evidenciar que para ciertos periodos de tiempo existe memoria en el comportamiento del mercado bursátil y como consecuencia es posible predecirlo.
- Producir una investigación que sirva como precedente en la predicción y modelado del mercado bursátil.
- Determinar en base a la experiencia empírica hasta qué punto el horizonte de predicción es fiable.

### **1.4 Hipótesis**

Si mediante una combinación de predicciones provenientes de un método lineal (ARIMA-ARCH) y de otro no lineal (RNA), obtenemos un pronóstico más preciso, que incluso nos de la capacidad de anticiparnos ante una eventual caída estrepitosa del mercado accionario, o por el contrario es preferible utilizar de manera individual el método de mejor desempeño.

### **1.5 Justificación**

Los mercados de valores desempeñan un papel fundamental en la economía de un país, respaldando sus políticas fiscales y monetarias. Una precipitada caída en el precio de las acciones puede ser causante de una perturbación en la economía de un país, he incluso puede ser causante de colapsar el mercado a nivel global. Un ejemplo de ello, es el memorable crack bursátil de 1929, detonante de la gran depresión de la década de 1930. El efecto nocivo que tuvo en México el crack bursátil de octubre de 1987, trajo como consecuencia, que el peso se devaluara, una incesante fuga de capitales, incremento en los precios y como corolario, un bajo crecimiento de la economía mexicana. Debido a la importancia que tiene para la economía la evolución del mercado de capitales, y siendo que uno de los sucesos que causa mayor impacto en el ámbito financiero y en todos los niveles de la economía, ya sea que se esté explícitamente o no ligado a él, son las reiteradas caídas en el nivel de precios, y los desastrosos cracks bursátiles ya que sus efectos devastadores permean toda la actividad económica del país donde sucede y que incluso las trágicas consecuencias que conlleva se extienden en todo el orbe. Es por ello que es fundamental encontrar señales y contar con herramientas fiables que muestren indicios de una posible manifestación de una crisis financiera futura. De este modo los diferentes participantes del mercado bursátil estarán en la capacidad de tomar medidas anticipadas, y no solo ellos, sino los diferentes agentes económicos que intervienen en la economía, pero sobre todo aquellos funcionarios que tienen en sus manos las decisiones de política económica, que estarán de este modo situados en una posición más adecuada, para una certera toma de decisiones.

## **1.6 Delimitación**

El presente proyecto de investigación se involucra específicamente en el estudio y predicción del mercado accionario mediante sus respectivos índices bursátiles, para el caso de México a través del IPC y para los Estados Unidos mediante el DJIA, en periodos de crisis. Se analizan dos casos, ambos casos son detallados en el apartado de la Metodología.

Por un lado se empleó el modelo estadístico ARIMA y sus modelos derivados, estos



modelos frecuentemente se encuentran asociados a datos de características lineales, como consecuencia, cuando los datos no son estacionarios hay que convertirlos a estacionarios, por ello y por otras adecuaciones que se le tienen que realizar en la mayoría de las veces a los datos, se puede considerar que los datos se tienen que ajustar al modelo, a diferencia de las RNAs, donde los datos modelan y dirigen el proceso. Otra ventaja que ofrecen las RNAs, es que son capaces de manejar procesos no lineales así como también lineales, su amplia difusión y su probada capacidad predictiva para resolver problemas de predicción en diferentes disciplinas como: el mercado de valores (Hamid y Iqbal, 2004), procesos de lluvia-escorrentía (Srinivasulu y Jain, 2006), consumo eléctrico (León-Anaya et al., 2013), series caóticas (Landassuri-Moreno et al., 2011), entre otras. A partir de estas ideas se puede concluir que las RNAs, son herramientas robustas e ideales para la predicción de series financieras.

El algoritmo FS-EPNet se encargó de diseñar el modelo neuronal predictivo para cada una de las ST de los indicadores, y de igual forma configuro un modelo neuronal para cada una de las descomposiciones resultantes de las ST, por otra parte, en el modelo paramétrico de igual forma se propusieron uno o varios modelos para cada una de las ST analizadas de los indicadores, así como también para cada una de las descomposiciones, donde el modelo es estimado en base a las características de la ST. Finalmente, en la fase de combinación de los pronósticos, se combinaron las predicciones de cada modelo, es decir, se combinan solo los datos resultantes de cada modelo, no hay una combinación de metodologías.

## **1.7 Metodología**

El presente trabajo tiene como finalidad el análisis de la evolución del mercado accionario, concretamente en el periodo en el cual ocurre un crisis bursátil, se analizaron dos situaciones de ocurrencia de crisis, por un lado se analiza el mini-crack bursátil que tuvo ocurrencia el lunes 27 de octubre de 1997, el cual tuvo una caída porcentual del 7.18% en una sola sesión. Para ello se analizar el índice llamado Dow Jones Industrial Average (DJIA). Por otra parte y en otro periodo de tiempo se analiza el mercado de un país emergente, en este caso la Bolsa Mexicana

de Valores (BMV) representada por el IPC. Donde el día 22 de octubre de 2008 el IPC muestra un retroceso importante en las cotizaciones bursátiles. Mediante las diferentes herramientas expuestas a lo largo del trabajo se pretende prever o anticipar dichos periodos de crisis. En cuanto a la parte experimental el primer paso que tomamos es descomponer las Serie de Tiempo (ST) de los índices por medio de Descomposición Empírica en Modos (DEM) el resultado que se obtiene al descomponer una ST son un conjunto de descomposiciones denominadas Funciones de Modo Intrínseco (FMIs) y un elemento residual, esto con el propósito de obtener más información del fenómeno y de aminorar la complejidad de la serie original. Por medio del exponente de Hurst se analizará el Índice de Precios y Cotizaciones (IPC) y el Dow Jones Industrial Average (DJIA) y sus correspondientes FMIs y el residuo, para evaluar su complejidad y su grado de aleatoriedad. Para la predicción por un lado se utilizan los modelos paramétricos ARIMA-ARCH, la familia de los modelos ARCH son los modelos que con mayor frecuencia se utilizan para Serie de Tiempo (ST) de tipo financieras debido a la efectividad que han mostrado (Engle, 1982; Bollerslev, 1986).

Por otro lado, el método no paramétrico de elección son las Redes Neuronales Artificiales (RNAs), ya que se consideran capaces de aproximar cualquier función, por ello son considerados como aproximadores universales (Zhang et al., 1998; Moreno, 2006). De entre los múltiples modelos neuronales artificiales que existen el Perceptrón Multicapa (MLP) es una técnica que ha sido utilizada de manera amplia para la predicción y que ha demostrado su efectividad en el manejo de series de tiempo con diferentes características. Por las causas mencionadas anteriormente, se eligió utilizar una ampliación del MLP, el modelo elegido es el Perceptrón Multicapa Generalizado (GMLP), dicho modelo permite arquitecturas más generales y potencialmente más poderosas que el MLP (Landassuri-Moreno, 2012). Cabe destacar que diseñar la arquitectura óptima de una RNA no es una labor trivial, y que existe evidencia empírica que demuestra que la capacidad de procesar información por una RNA se encuentra en función de su arquitectura (Yao y Liu, 1997b). Para subsanar la problemática que implica la edificación de una RNA se propone utilizar un Algoritmo Evolutivo (AE), una herramienta basada en la

evolución biológica la cual fue propuesta inicialmente por Yao y Liu (1997b) y que posteriormente fue extendida (Landassuri-Moreno y Bullinaria, 2009b; Landassuri-Moreno y Bullinaria, 2011). Este algoritmo denominado Selección de Características de Programación Evolutiva de Redes Neuronales Artificiales (FS-EPNet) será el encargado de configurar de manera automática los parámetros variables de cada una de las RNAs que se utilizarán para predecir las distintas ST empleadas en la investigación.

Como anteriormente se había mencionado el propósito fundamental de esta investigación es encontrar si alguno de los modelos predictivos aquí propuestos, es capaz de predecir de manera precisa el movimiento del mercado financiero. Al obtener los diferentes pronósticos emanados de los diferentes modelos, se realizan mediciones del rendimiento de cada uno de ellos con el fin de compararlos y comprobar la efectividad de cada uno de ellos, y así develar quien captura de manera más fidedigna el comportamiento de las ST financieras.

Pero más allá de realizar el análisis anteriormente descrito que podría denominarse un análisis clásico donde el mejor modelo prevalece y el otro es descartado por completo, lo que se pretende en esta investigación es encontrar si una combinación de pronósticos emanados cada uno de ellos de diferentes paradigmas es capaz de mejorar sustancialmente al mejor modelo individual. A partir de la obra pionera de Bates y Granger (1969), se ha popularizado la combinación de pronósticos como técnica que coadyuva a mejorar la precisión y a robustecer la predicción sobre la ofrecida por los modelos individuales. Puesto que el entorno donde confluyen los diferentes mercados financieros es altamente inestable y siempre cambiante, se presupone que no hay un modelo individual que pueda ser capaz de generar los resultados más precisos en todas las situaciones y ante cualquier circunstancia. En esta investigación, se aplicarán dos de los métodos más comunes que reporta la literatura sobre combinación de pronósticos, como lo son: el método del promedio simple (PS) y la técnica del promedio ponderado (PP). Los dos métodos se basan en la asignación de una ponderación de peso a cada modelo individual. El promedio simple asigna un peso igualitario a cada modelo, por el otro lado, el promedio

ponderado, calcula el peso de cada modelo tomando en cuenta el desempeño de los modelos individuales en base a predicciones del pasado.

## **1.8 Organización de la tesis**

El presente documento está organizado de acuerdo al siguiente esquema:

- En el capítulo 2, se introducen algunos de los conceptos teóricos de los mercados accionarios que son utilizados en el transcurso del trabajo.
- En el capítulo 3, se exponen los indicadores bursátiles (IPC y DJIA) sobre los cuales descansa esta investigación. Se presenta una de las teorías (HME) más representativas que intenta explicar el comportamiento del mercado bursátil. Así como también la Teoría Moderna del Portafolio (TMP) la cual le ha proporcionado cimientos a ciertas corrientes contemporáneas de economía financiera. Además se presenta una de las herramientas (CAPM) más populares en el ámbito financiero.
- Las características y los conceptos más relevantes de los modelos predictivos utilizados en la presente investigación se introducen en el capítulo 4, como lo son: el modelo paramétrico (Autorregresivo Integrado de Media Móvil-Modelo de Volatilidad Condicional Heterocedástico (ARIMA-ARCH)) y modelo no paramétrico (Redes Neuronales Artificiales).
- El capítulo 5 presenta los conceptos claves de las herramientas de apoyo en el estudio y en el tratamiento de los datos, así como en la configuración de los modelos de predicción.
- En el capítulo 6 se analizarán las descomposiciones (FMI y el residuo) derivadas de los índices que son el objeto de análisis, así como también las ST que representan a los indicadores. Se abordara el desarrollo experimental de los modelos utilizados en el pronóstico. En primera instancia se indicaran las muestras de datos que se analizan a través del presente capítulo, en segundo lugar se descompondrán los indicadores y se analizaran mediante el coeficiente de Hurst. Posteriormente se realizaran las consabidas predicciones, para cada indicador, así como también las predicciones de cada una de las descomposiciones provenientes de dichos indicadores (IPC y DJIA), en las

muestras fuera del pronóstico para cada caso. Finalmente, se realizara la combinación de las predicciones emanadas del modelo paramétrico y del modelo no paramétrico. A lo largo, de las diferentes secciones de pronóstico (pronóstico fuera de la muestra y combinación de predicciones) se realizaran las comparaciones de los modelos predictivos utilizados en este trabajo.

- Finalmente en el capítulo 7, se exponen las conclusiones y los trabajos futuros.

## **Capítulo 2. Mercado accionario**

### **2.1 Teoría económica**

La teoría económica tiene como finalidad tratar de explicar en base a modelos o hipótesis las diferentes relaciones entre los agentes económicos, así como el funcionamiento de la economía. Los modelos económicos intentan representar de manera sintetizada y matemática la realidad. Para probar estos modelos la economía emplea instrumentos como las matemáticas y la estadística implementada a la econometría. Habitualmente las herramientas empleadas para el estudio experimental de las ST económicas se basan en supuestos de una relación lineal como método de aproximación. Por un lado, su principal fortaleza es la facilidad en su interpretación, por otro lado, su cálculo computacional no es tan demandante, consecuencia de ello es la popularidad a favor de ellas.

No obstante, en la mayoría de los casos la conducta económica es de carácter no lineal, discontinuo y complejo. Diversos estudios han demostrado que las series de datos económicos y particularmente financieros son esencialmente no lineales (Carles, 1996; Lorenzo Valdés, 2005; Herrera, 2008; Balacco, 2012), bajo este supuesto es imprescindible buscar herramientas alternativas para solventar las limitaciones de los modelos tradicionalmente utilizados. Como muchas otras ciencias la investigación económica se ha alimentado de otras disciplinas, por ejemplo: la teoría neoclásica se encuentra en gran parte basada en la física (Carles, 1996), la teoría del comportamiento del consumidor y del productor por la psicología, de la matemática y la física surge la teoría del caos que explica los intrincados y complejos mercados de valores (Monroy Olivares, 1997; Balacco, 2012; Herrera, 2008).

Derivado de todo lo anterior se deduce que es necesario utilizar modelos alternos a los que típicamente son utilizados, en base a las características de no linealidad en las ST, y dada las restricciones que supone un modelado paramétrico tradicional (posibilidad de resultados alterados), es que se propone la utilización de un modelo no paramétrico como lo son las Redes Neuronales Artificiales ya que son modelos de naturaleza no lineal y que a diferencia de los modelos tradicionales (regresión

lineal, ARIMA) estos no asumen la hipótesis de normalidad en la distribución de los errores. En base a lo anterior se presupone que el modelo no paramétrico será capaz de develar el comportamiento no aleatorio y por lo tanto predecible, que no pudiera captar el modelo paramétrico, debido a su capacidad de aprender las relaciones no lineales (relaciones ocultas en los modelos lineales) entre los datos. De esta forma es que el presente capítulo tiene como objetivo introducir los aspectos básicos y generales del mercado accionario, así como dar a conocer el cometido del índice Dow Jones y del índice IPC de la bolsa mexicana de valores. Así como también exponer la teoría clásica de los modelos de predicción de los mercados accionarios.

## **2.2 Definición del mercado financiero**

El mercado financiero se encuentra compuesto por: Mercado de deuda, Mercado de derivados, Mercado cambiario y Mercado accionario, este último es nuestro objeto de análisis, dicho ente es un mercado de carácter público que puede ser físico o virtual donde concurren los inversionistas (demandantes) de forma individual o en conjunto, que buscan obtener un beneficio invirtiendo su capital monetario, así como también los intermediarios y por el otro lado se encuentran las empresas emisoras (oferentes) que buscan obtener fondos para el desarrollo de sus empresas, en este espacio se pueden realizar operaciones de compra y venta de títulos accionarios emitidos por las empresas. Hay dos tipos de mercados:

1. Mercado primario: ocurre en el momento en que se emiten y se ponen a la venta por primera vez los títulos accionarios. En él participan las instituciones o empresas emisoras de las acciones, los intermediarios colocadores y los primeros compradores de los activos emitidos. El precio es fijado en base a la capacidad y potencial de la empresa emisora.
2. Mercado secundario: es cuando las acciones ya habían sido previamente compradas por medio de una adquisición pública o privada. En este mercado no intervienen las empresas emisoras y por lo tanto las transacciones que de ahí emanan no redundan en ninguna ganancia o recursos a las empresas emisoras, los participantes son los inversionistas que transfieren por medio

de las transacciones de compraventa los títulos accionarios. El precio se establece de acuerdo a la libre demanda y oferta.

## **2.3 Función del mercado financiero**

La función del mercado accionario es encauzar los bienes de los inversionistas como una fuente de suministro de capital a largo plazo, hacia los diferentes emisores para que ellos a su vez lo canalicen a sus operaciones de inversión y financiamiento.

### **2.3.1 Acción**

Se le nombra acción a cada una de las pequeñas fracciones en las cuales se encuentra dividido el capital de una empresa. Es decir si se posee una acción de determinada empresa, se tiene un título que otorga la pertenencia de una porción minúscula de la sociedad. El propietario de las acciones tiene derecho solicitar una parte de los dividendos de la empresa, y si además dicho individuo es dueño de una cantidad cuantiosas de acciones, es acreedor a ser partícipe en las decisiones de negocio de la sociedad y también tiene como deber en caso de quiebra, el de apoyar con su capital invertido a los demás acreedores del negocio. Cuando una compañía quiere expandirse o necesita capital para operar su negocio, se emiten acciones para poder allegarse de la liquidez necesaria para hacer frente a dichas necesidades. Estas acciones tiene la facilidad de ser comercializadas en el mercado, esta característica dota al vendedor de la acción de liquidez inmediata, así como también hace que la propiedad de la acción pueda ser fácilmente intercambiable entre inversionistas.

### **2.3.2 Riesgo**

Es la incertidumbre asociada con cualquier inversión sobre una posible pérdida. La situación factible de que el retorno real de la inversión sea menor al beneficio esperado, incluso el posible riesgo de perder la mayoría o hasta la totalidad de la correspondiente inversión original.

### **2.3.3 Crack bursátil**

Frecuentemente se considera a un crack bursátil como un suceso aleatorio y de dimensiones catastróficas, ya que sus efectos negativos pueden dejar secuelas



por largos periodos de tiempo. La caída bursátil es un súbito declive en el precio de las acciones o de un indicador en un determinado periodo de tiempo. Bajo el supuesto de la teoría económica de que el precio de un bien se rige bajo la ley de la oferta y de la demanda, y puesto que el precio del valor de un activo financiero es el reflejo de una incesante oleada de buenas y malas noticias, las cuales a su vez son interpretadas tanto por inversores como por analistas, y ante un cúmulo de noticias ominosas estas pueden ser causantes de detonar la pérdida de confianza de los inversores, y estos ante un temor exagerado, donde el miedo los domina colocan demasiadas órdenes de venta al mismo tiempo, que se manifiesta en una significativa caída porcentual en el índice bursátil.

Los posibles orígenes de un crack bursátil se engloban en 2 clasificaciones: a) Crack Exógeno; cuando el origen de un crack se encuentra relacionado por una cantidad significativa de información dañina, debido a eventos externos como el intento de golpe de estado en la Unión Soviética el 21 de agosto de 1991, o el ocurrido el 11 de septiembre de 2011. b) Crack Endógeno; este suceso tiene su causa germinal en una burbuja especulativa. Una burbuja especulativa se puede definir; es cuando el precio de un activo o de un índice aumenta significativamente, y se encuentra sobrevaluado, como resultado de una retroalimentación positiva (Jacobsson, 2009). Un crack a menudo resulta en un mercado bajista o con una tendencia a la baja con una correspondiente disminución en los precios, durante este proceso la mayor parte de los inversionistas entran en pánico y tratan de vender la totalidad de sus inversiones e incluso las rematan, esta fase de descenso puede prolongarse por un lapso de tiempo que puede durar semanas, meses e inclusive prolongarse por años.

#### **2.3.4 Análisis técnico**

Se denomina análisis técnico al conjunto de herramientas y procedimientos que utilizan los datos históricos para predecir la cotización futura del precio de los activos financieros tasados en un mercado de valores. Este análisis parte en dos categorías por un lado se utiliza el enfoque por medio del estudio exclusivamente de gráficas, que se le llama análisis gráfico o análisis chartista, pero dada la

subjetividad que este tipo de análisis pudiera tener, es aconsejable combinar con la otra categoría denominada simplemente análisis técnico que utiliza herramientas estadísticas, matemáticas y modelos econométricos. La combinación de ambas técnicas pretende eliminar la subjetividad de solo utilizar gráficas para explicar y predecir el fenómeno de estudio en cuestión, es usual que los especialistas utilicen ambas técnicas enriqueciendo así el análisis. En conclusión el enfoque moderno engloba estas dos categorías y cuando se hable de análisis técnico se estará refiriendo a la combinación de ambas categorías.

### **2.3.5 Análisis fundamental**

Esta técnica intenta hallar el valor verdadero de los títulos accionarios, al cual se le denomina valor fundamental. Partiendo del supuesto que el valor real o intrínseco de un activo financiero depende de tres componentes económicos que son: estudio de la empresa, análisis de la industria y los factores macroeconómicos y políticos. Tomando como base estos tres factores el analista fabrica indicadores de desempeño sobre el riesgo, utilidades, flujo de efectivo estimado, tasas de interés, y demás herramientas que pudieran ser útiles para poder determinar el valor fundamental de las acciones.

Una vez que el analista ha obtenido el valor fundamental del activo lo compara con el precio por acción tasado por el mercado en ese instante, si el precio de la acción se encuentra por debajo de su valor fundamental, es decir, se encuentra infravalorado se debe de comprar dicho título ya que se presupone aumentara su valor en el largo plazo, por el contrario si el precio del título se encuentra sobrevalorado por lo tanto se aconseja vender, puesto que en un futuro próximo el precio de la acción bajara.



## **Capítulo 3. Indicadores bursátiles**

### **3.1 IPC de la Bolsa Mexicana de Valores**

Definición de Índice: Un índice accionario es un indicador que tiene como función exponer el comportamiento de un grupo de acciones. Como tal los índices son herramientas estadísticas que sirven para medir y mostrar las fluctuaciones o movimientos de una o más variables que guardan relación en el transcurso del tiempo. El índice más significativo y representativo de la Bolsa Mexicana de Valores (BMV) es el Índice de Precios y Cotizaciones también conocido como el IPC o IPyC. El IPC está compuesto de las 35 acciones con el grado más alto de bursatilidad (liquidez del título) del mercado accionario mexicano y es el indicador más utilizado por inversionistas nacionales y extranjeros. La muestra de los elementos que la conforman es inspeccionada anualmente en febrero y de manera usual se sustituye a un conjunto de cinco a seis acciones.

### **3.2 Índice Dow Jones de los Estados Unidos**

El Dow Jones Industrial Average por sus siglas en inglés (DJIA) y popularmente conocido como Dow Jones, en español conocido como el Índice de Promedio Industrial Dow Jones. El índice hizo su aparición el 26 de Mayo de 1896, fue creado por Charles Henry Dow editor financiero de The Wall Street Journal. Al principio solo evaluaba el desempeño de 12 empresas del sector industrial estadounidense. En el año de 1916 el índice aglutino a 20 empresas, y en 1928 creció a 30 empresas grandes y reconocidas este número se mantiene hasta el día de hoy, además de ser el índice bursátil más referenciado globalmente. Es un indicador que trata de reflejar la conducta de una gran variedad de acciones ya que las empresas tasadas pertenecen a diferentes industrias como son: servicios, telecomunicaciones, minoristas, servicios financieros, empresas de entretenimiento, farmacéuticas, entre otras. El mercado estadounidense es analizado y medido habitualmente a través de este indicador (GuzmánPlata, 2008).

### **3.3 Hipótesis de los Mercados Eficientes (HME)**

Se califica como un mercado eficiente aquel que manifiesta en el valor de los activos

toda la información disponible de su pasado así como de su posible evolución, además de que sus precios se ajustan de manera inmediata a la información más reciente. Bajo tres supuestos se edifica la HME:

- a. Existe una gran cantidad de inversores en el mercado, donde de manera individual estos inversores analizan y calculan el precio de los títulos accionarios sin tomar en cuenta a los otros competidores o analistas.
- b. La información más novedosa que se incorpora al mercado aparece de forma aleatoria, no hay patrones, ni normas definidas.
- c. La velocidad en el ajuste en los precios de los activos es inmediata a la nueva información, lo anterior surge como consecuencia de que exista una enorme cantidad de inversionistas que tratan de maximizar sus beneficios.

Eugene Fama se encargó de popularizar y desarrollar la HME en su trabajo doctoral (Fama, 1965), donde comienza sentando las bases, argumentando que en cualquier mercado existen una gran cantidad de inversores bien informados, y que el valor presente de los activos refleja toda la información disponible. En el año de 1970 las ideas antes expuestas evolucionaron y maduraron en su clásico trabajo (Fama, 1970), en el expone la teoría así como la evidencia empírica de la hipótesis dotando a la HME de una estructura enriquecida y refinada, donde introduce las definiciones de las tres formas del mercado eficiente: Débil, Semifuerte y Fuerte. La HME se presenta en tres versiones atendiendo al tipo de información que se examine:

- Hipótesis de los Mercados Eficientes (HME) en su forma débil: esta hipótesis postula que la información pasada ya se encuentra reflejada en el valor de los títulos, tal conocimiento es plasmado a través de las series históricas de precios, esto se traduce en: en el mercado accionario, no hay forma o modo de encontrar alguna estrategia basada en los datos históricos que pueda servir para obtener un rendimiento superior al del promedio del mercado, sobre el precio y el retorno de las acciones, ya que todos los inversionistas que concurren en el mercado son capaces de descifrar y comprender cualquier indicio o señal que pudieran mostrar las series históricas, y si se llegara a obtener un beneficio más alto que los del resto de los participantes en el mercado examinando únicamente la información pasada sería exclusivamente por cuestiones de azar. La variación en los precios

es azarosa, ya que los elementos que afectan a los precios son los sucesos o noticias desconocidas y estas ocurren aleatoriamente. En consecuencia la predicción más certera para tasar el precio de un activo en un futuro es emplear el valor presente. Esto lleva a concluir que no hay manera de beneficiarse mediante el análisis técnico.

- Hipótesis de los Mercados Eficientes (HME) en su forma semifuerte: a esta modalidad se le incluye la información de su forma previa, y se le agrega la información pública sobre la empresa como puede ser la situación financiera, situación contable y en el caso del mercado accionario, la valoración en libros de las acciones, el cálculo de los dividendos por acción, la correlación entre el precio de un título accionario y la utilidad por acción, y el de su entorno que incida sobre el valor de los títulos, que vendrían siendo las noticias sobre política, economía y cualquiera que guarde una relación intrínseca con la empresa. Ni con el análisis técnico ni con el análisis fundamental se podría encontrar capacitado para lograr un rendimiento mayor al promedio del mercado. La única herramienta válida para poder alcanzar una ganancia mayor al promedio es mediante la obtención y uso de información privilegiada.
- Hipótesis de los Mercados Eficientes (HME) en su forma fuerte: esta última versión indica que en el precio de los activos se encuentra plasmada la totalidad de la información crucial que sea tanto del pasado, como privada y pública. No hay fórmula que ponga al inversionista en ventaja sobre los demás competidores para ser capaz de lograr una rentabilidad superior al promedio del mercado. En este caso no existe la información privilegiada. Por lo tanto, esta versión es ficticia ya que esto implicaría que es un mercado de competencia perfecta.

La versión semifuerte de la Hipótesis de los Mercados Eficientes (HME) es la más ampliamente aceptada por la comunidad de especialistas y por el círculo de inversionistas (Balacco, 2012). Por lo tanto, cuando se hace alusión a la HME se debe inferir que se refiere a la forma semifuerte. En este caso debe entenderse que las variaciones en los precios de los activos son aleatorias ya que en su valoración incluyen toda la información pública relevante.

### **3.4 Teoría del paseo aleatorio**

El paseo aleatorio afirma que los cambios en los precios de las acciones siguen una trayectoria aleatoria en el tiempo y éstas pueden ser descritas por un proceso aleatorio, donde las fluctuaciones de los precios del pasado no ejercen influencia alguna sobre los precios actuales y futuros, por lo tanto resulta imposible predecir el movimiento de los precios. La teoría supone, que si se emplean herramientas como el análisis técnico, esto no conducirá a obtener ganancias mayores, a las que en promedio obtienen los demás inversionistas. Por lo tanto, está teoría implica que una serie de cambios en el precio de las acciones no tiene memoria, que no se puede predecir los precios futuros del mercado sobre la base del pasado histórico del comportamiento de los precios.

Dado que el valor intrínseco de una acción se estima en base a un análisis fundamental sobre el nivel de desarrollo que se supone puede alcanzar la empresa emisora en un futuro, y puesto que la nueva información sobre la empresa se vuelve disponible de manera inmediata, los inversionistas actualizarán las estimaciones previstas sobre los rendimientos futuros de la empresa, esta nueva estimación afectará el valor intrínseco de la acción. Es así, que la introducción de información novedosa provoca un cambio en el valor actual de la acción. Lo que a su vez implica que en cualquier momento, el valor actual de una acción y su precio en el mercado representa el valor más cercano al valor intrínseco estimado en base a toda la información disponible. En otras palabras, un mercado es eficiente si el precio de los activos que se comercializan en este mercado, se ajusta a la nueva información de manera pronta y expedita, y a su vez el precio del activo debe de reflejar toda esta información reciente, la cual debe estar a disposición de cualquier participante. Es amplia la literatura que emplea al modelo del paseo aleatorio como instrumento empírico para desarrollar la HME (Fama, 1965b; Samuelson, 1965; Fama, 1970; Malkiel, 1973), bajo la siguiente premisa: la variación en el precio de los activos en el día presente es causado por noticias o eventos no esperados de hoy, y no por eventos que ocurrieron con anterioridad, por lo tanto, las ganancias obtenidas el día de hoy son independientes de los eventos pasados. Partiendo del

supuesto de independencia en los cambios en el precio de los activos o de las ganancias obtenidas, la evolución en la cotización de los activos se ajusta a la de una distribución normal. Esta forma de representación empírica de la HME mediante el paseo aleatorio, es la más restrictiva y condicionante ya que los incrementos  $\eta$  o perturbaciones deben ser independientes e idénticamente distribuidas (i.i.d.), y tener la característica de ser ruido blanco, con media  $\mu$  cero y varianza constante  $\sigma^2$ ; de tal forma que la dinámica del precio de un activo (P), donde la  $\mu$  representa el cambio esperado en el precio del activo, y esta se puede describir en función de la siguiente ecuación (Balacco, 2012):

$$P_t = \mu + P_{t-1} + \eta_t \quad \eta \approx (i.i.d.)(0, \sigma^2) \quad (3.1)$$

Sin embargo, para la HME no es indispensable ser representada de manera empírica bajo los supuestos del paseo aleatorio, no obstante, paseo aleatorio si supone HME en su forma fuerte, aunque, cabe destacar que existen otras dos versiones de HME, que no son tan restrictivas. Como lo expuso Fama (1965a), donde precisa que la HME puede ser representada bajo el modelo de martingala, el cual es un proceso estocástico que no necesita de independencia ni de estar idénticamente distribuido, esto trae como consecuencia que se extienda el panorama a los participantes del mercado de títulos.

### 3.4.1 Teoría Moderna del Portafolio (TMP)

Por otro lado, previo a las contribuciones antes mencionadas, en la década de los 50, se desarrolló una teoría que ha llegado a ser uno de los pilares fundamentales en el ámbito del mercado de capitales. En el año de 1952 el economista Harry Markowitz a través de su investigación pionera (Markowitz, 1952), sentó las bases de la Teoría Moderna del Portafolio (TMP). La TMP propone un enfoque opuesto al tradicional para la selección de activos. La TMP plantea que los participantes del mercado deben abordar la comprensión del mercado en su conjunto, analizando los rasgos del retorno y del riesgo de manera global, en vez de estudiar los activos de manera individual. En la TMP la selección de los activos así como la gestión del riesgo son los elementos básicos de esta teoría, por ello las inversiones son especificadas estadísticamente en términos de la tasa de retorno esperado en el



largo plazo y de la volatilidad esperada en el corto plazo. La volatilidad se trata como un factor de riesgo, y la cartera se conforma en virtud de la tolerancia al riesgo de cada inversor en particular, tras elegir el máximo nivel de retorno disponible para el nivel de riesgo escogido. En su modelo, Markowitz, dice que los inversionistas tienen una conducta racional a la hora de seleccionar su cartera de inversión y por lo tanto siempre buscan obtener la máxima rentabilidad sin tener que asumir un nivel de riesgo más alto que el estrictamente necesario. Muestra también, como hacer una cartera óptima disminuyendo el riesgo de manera que el rendimiento no se vea afectado. Para poder integrar una cartera de inversión equilibrada lo más importante es la diversificación, ya que de esta forma se reduce la variación de los precios. La idea de la cartera es, entonces, diversificar las inversiones en diferentes mercados y plazos para así disminuir las fluctuaciones en la rentabilidad total de la cartera y por lo tanto también del riesgo (Velasco Robles, 2008).

#### **3.4.2 Modelo de Valoración de Activos Financieros (CAPM)**

Sharpe, Littner y Mossin trabajando al mismo tiempo pero por separado, retomaron las contribuciones realizadas por Markowitz, relacionando a la HME con la TMP, con el fin de desarrollar el modelo de Valoración de Activos Financieros (CAPM) (por sus siglas en inglés, Capital Asset Pricing Model). Donde se introduce el empleo del cálculo de probabilidades con el propósito de analizar el mercado de capitales. El CAPM es un modelo que relaciona el riesgo y la rentabilidad de un activo de manera lineal. Dicho modelo es utilizado para valorar el rendimiento que es posible esperar dado el nivel de riesgo de la inversión. En CAPM el riesgo de un activo se calcula dividiéndolo en riesgo no sistemático y riesgo sistemático. El primero de ellos es el riesgo intrínseco del activo, al cual también se le conoce como riesgo diversificable, ya que como su nombre lo indica es posible atenuar el riesgo mediante la diversificación, para poder llevar acabo la posible minimización del riesgo es necesario poseer activos de diferentes empresas y ramos, así como también detentar bonos gubernamentales libres de riesgo, entre otros. Al segundo tipo de riesgo también se le conoce como riesgo de mercado el cual es propio e incide en la totalidad del mercado y no solo de manera particular. Basándose en el supuesto de que los mercados financieros son eficientes y que los participantes

como conjunto se conducen de forma eficiente diversificando el riesgo, esto da como resultado, que el principal riesgo asociado a una acción es su riesgo sistemático, ya que no es posible eliminarlo. Cuanto mayor es la beta (medida de riesgo sistémico) de una acción, mayor es el riesgo relevante de ese activo, y una mayor rentabilidad es exigida. Si se supone que el riesgo no sistemático es diversificado y por ende este es eliminado, la tasa de rendimiento requerida para la acción  $j$  es (Van Horne y Wachowicz, 1998):

$$\bar{R}_j = R_f + (\bar{R}_m - R_f) \beta_j \quad (3.2)$$

Donde  $R_f$  es la tasa libre de riesgo,  $\bar{R}_m$  es el retorno esperado para el portafolio de mercado, y  $\beta_j$  es el coeficiente beta para el activo  $j$ . Finalmente, el trabajo realizado por Elton et al., (2014), es una referencia imprescindible si se quiere profundizar en los temas anteriormente expuestos.



## Capítulo 4. Modelos de predicción

### 4.1 Antecedentes

En el ámbito de la economía, se han tratado de solucionar una gran variedad de problemas tanto de clasificación como de aproximación de funciones por medio de las RNAs. Las RNAs como herramientas clasificadoras se han utilizado en la predicción del fracaso empresarial (Trippi y Turban, 1996; Bell, 1997; Wheelock y Wilson, 2000; Rodríguez, Piñeiro y de Llano, 2014; Ravi-sankar y Ravi, 2009), en la clasificación de obligaciones (Surkan y Singleton, 1990; Flórez López, 2007), en la identificación de los segmentos de mercado (van Wezel et al., 1996; Cao et al., 2005; Ertay y cekyay, 2005).

El uso de la Red Neuronal Artificial (RNA) en el estudio de los mercados financieros no es un tema tan novedoso, se considera al trabajo propuesto por White (1988) como el primer estudio en el ámbito financiero que emplea como modelo predictivo a las RNAs, en este trabajo Halbert White emplea un modelo neuronal tipo MLP con un método de aprendizaje de retropropagación de errores (BP). En dicho estudio el MLP es utilizado análogamente a un método de análisis técnico, donde la variable explicativa es ella misma rezagada en el tiempo, es decir, utiliza un modelo univariante que no incorpora información adicional más que la inherente a la serie. En el experimento se comienza con una regresión lineal simple sobre los registros históricos de la cotización de IBM. Donde la evidencia demuestra que ciertamente el mercado es eficiente, y por ello esta sugiere la imposibilidad de predecirlo. Pero si se toma en cuenta que los modelos neuronales son no lineales y que son capaces de integrar una gran cantidad de datos, cabría la posibilidad de un mejor rendimiento que el modelo lineal. El MLP fue entrenado con un conjunto de datos derivados de los retornos accionarios de IBM por 1,000 días. En la primera fase del pronóstico, con los datos de entrenamiento el modelo neuronal pronóstico notablemente mejor que el modelo lineal, desafortunadamente con un par de colecciones de datos desconocidos provenientes de antes y de después del conjunto de entrenamiento, el rendimiento no fue mejor que el obtenido por el azar. Leung et al. (2000), compara la capacidad predictiva de varias técnicas como son: el Suavizado Exponencial

Adaptativo (SEA), Función de Transferencia Multivariada (FTM), Vectores Autorregresivos (VAR) con filtro de Kalman (VARK), y la red neuronal multicapa feedforward (MLP), donde se analizan 3 indicadores financieros; S&P500 (USA), FTSE (Reino Unido) y Nikkei225 (Japón). En el estudio se emplean los datos mensuales del periodo de enero de 1967 a diciembre de 1990, siendo la suma total de 228 observaciones, el pronóstico es realizado de enero de 1991 y concluye en diciembre de 1995, esto da un total de 60 períodos mensuales. El citado estudio maneja un enfoque analítico alternativo al tradicional, donde se menciona que en la mayoría de estos modelos propuestos, toman como base la precisión del pronóstico del valor subyacente del índice bursátil o de su retorno. Donde el grado de exactitud y la aptitud de ciertos pronósticos son medidos mediante el cálculo de las desviaciones de las estimaciones de los valores observados. En cambio en este estudio se estima la predictibilidad de la dirección/signo del movimiento en el índice, ya que se destaca que una predicción con un error mínimo no se traduce necesariamente en una ganancia de capital.

Tabla 4. 1: La tabla indica el número de veces que un modelo de pronóstico predice correctamente la dirección de la rentabilidad del índice sobre los 60 períodos de pronóstico fuera de la muestra de enero de 1991 a diciembre de 1995.

	US S&P500		UK FTSE100		Japan Nikkei225		Total	
	Número	Porcentaje	Número	Porcentaje	Número	Porcentaje	Aciertos	Porcentaje
Suavizado exponencial adaptativo	29	0.48	33	0.55	38	0.63	100	56 %
Función de transferencia multivariada	32	0.53	32	0.53	35	0.55	99	55 %
VAR con filtro de Kalman	32	0.53	34	0.56	35	0.58	101	56 %
Red neuronal multicapa feedforward	38	0.63	30	0.50	36	0.60	104	<b>58 %</b>

Fuente: Adaptado de Leung et al. (2000).

Los resultados que arroja la tabla 4.1 sobre el número de aciertos, demuestran que el modelo neuronal es el que tiene un rendimiento ligeramente mejor de entre todos los modelos propuestos, aunque cabe destacar que en general todos los modelos presentan un desempeño deficiente muy cercano al que se obtendría con decisiones al azar. Por medio del índice Dow Jones Riga Stock Exchange (RSE) el mercado de valores de Letonia es analizado por Zorin y Borisov (2002). En esta

investigación, se evalúa la predicción de los modelos de RNAs y ARIMA; los resultados indican que la RNA supera al modelo ARIMA. Además, la investigación concluyó que la predicción del indicador (RSE) es aceptable por un período de dos semanas.

Por otra parte, el mercado financiero mexicano también ha sido analizado por varios autores (Mendoza-Velázquez y Gómez-Gil, 2011; Guzmán Plata et al., 2007; Cruz Torres, 2007; García Salgado y Morales Castro, 2012), donde se destaca la investigación realizada por Gómez et al. (2011), en dicho estudio se compararon dos modelos uno paramétrico el Modelo de Volatilidad Condicional Heterocedástico Generalizado (GARCH) y otro no paramétrico el modelo del Perceptrón Multicapa (MLP), sobre los índices accionarios IPC de México y el DJIA de los Estados Unidos. Los datos que fueron sujetos de análisis son los precios de cierre del intervalo entre el 7 de junio de 2010 al 6 de enero de 2011, que suman un total de 150 observaciones diarias para ambos indicadores (DJIA y IPC) y realiza un pronóstico del 7 al 20 de enero de 2011. Se utilizó un modelo neuronal con una arquitectura para ambos casos de (6-10-1) y el modelo GARCH(0,1) de igual forma, y como medida de desempeño Mean Squared Error (MSE).

Tabla 4. 2: Cálculo del MSE: dentro y fuera de la muestra. GARCH y RNA

IPC			DJIA		
OBSERVACIÓN	GARCH	RNA	OBSERVACIÓN	GARCH	RNA
dentro	0.008	0.025	dentro	0.009	0.018
fuera	0.003	0.021	fuera	0.006	0.016
1	0.003	0.017	1	0.006	0.013
2	0.003	0.016	2	0.008	0.012
3	0.003	0.018	3	0.007	0.014
4	0.004	0.021	4	0.006	0.016
5	0.004	0.020	5	0.005	0.016

Fuente: Gómez et al. (2011).

En la tabla 4.2 se muestran los diferentes resultados de dicha investigación, donde los resultados muestran que el modelo paramétrico GARCH(0,1) tiene un ajuste ligeramente mejor a los datos que el modelo neuronal. Cabe destacar que resulta un poco difícil determinar si la evidencia mostrada en los diferentes documentos

anteriormente mencionados pueden llevar a conclusiones absolutas, ya que los productos emanados de los diversos investigadores carecen de homogeneidad entre estos, el rasgo que en mayor medida marca la heterogeneidad es la notable disparidad en las medidas de desempeño utilizadas, a causa de que no se emplea una medida de referencia estandarizada. Aunado a lo anteriormente descrito y que en menor medida pudiera incidir sobre los resultados, son las diferentes circunstancias sobre las que se desempeña el investigador, así como las herramientas utilizadas, y los diversos procedimientos entre otros elementos o particularidades. Es por ello que los resultados empíricos de estos estudios no los podemos interpretar como evidencia concluyente ni tratarlos con rigurosidad.

## **4.2 Metodología estadística**

### **4.2.1 Series de tiempo**

Meza Carvajalino (2007, p.117) señala que “el concepto de series de tiempo es utilizado para denotar una sucesión de datos ordenados en función, como su nombre lo indica, del tiempo”. En otras palabras una serie de tiempo es una sucesión de observaciones o conjunto histórico de datos ordenados de manera cronológica. Por ejemplo, precios de un artículo, índice de inflación, tasas de empleo y desempleo, lluvias en sucesivos días, etc. A la evolución de un conjunto de datos que tiene un patrón de comportamiento invariable se le nombra serie determinista. Una serie no determinista es aquella en donde los valores posteriores o futuros se determinan en base a una distribución de probabilidad. Una gran parte de las series de tiempo económicas son estocásticas, por consiguiente, los valores futuros de la serie dependen o se determinan solo en cierta medida de los valores de su propio pasado, ya que también el componente de aleatoriedad es inherente a las series, esto lleva a inferir que es casi imposible obtener predicciones totalmente exactas. El análisis de las series de tiempo persigue los siguientes objetivos:

- Lograr una descripción detallada, clara y precisa del suceso generador de la ST.
- Diagnosticar y construir un modelo que represente de la mejor forma posible la evolución de la serie de datos.
- A partir de la serie de datos de que se disponen pronosticar valores futuros o

pasados desconocidos, con el mayor grado de exactitud posible.

Un enfoque al análisis de series de tiempo comprende un intento de identificar los factores que ejercen influencia sobre cada uno de los valores periódicos de una serie. Este procedimiento de identificación se denomina descomposición. La serie de tiempo se divide en los diferentes componentes que la comprenden con el objeto de analizar por separado cada uno y comprender con mayor precisión el proceso generador de la serie, para que a su vez los datos históricos sean proyectados en el corto, mediano y largo plazo. Son cuatro los componentes que una serie de tiempo que a continuación se mencionan:

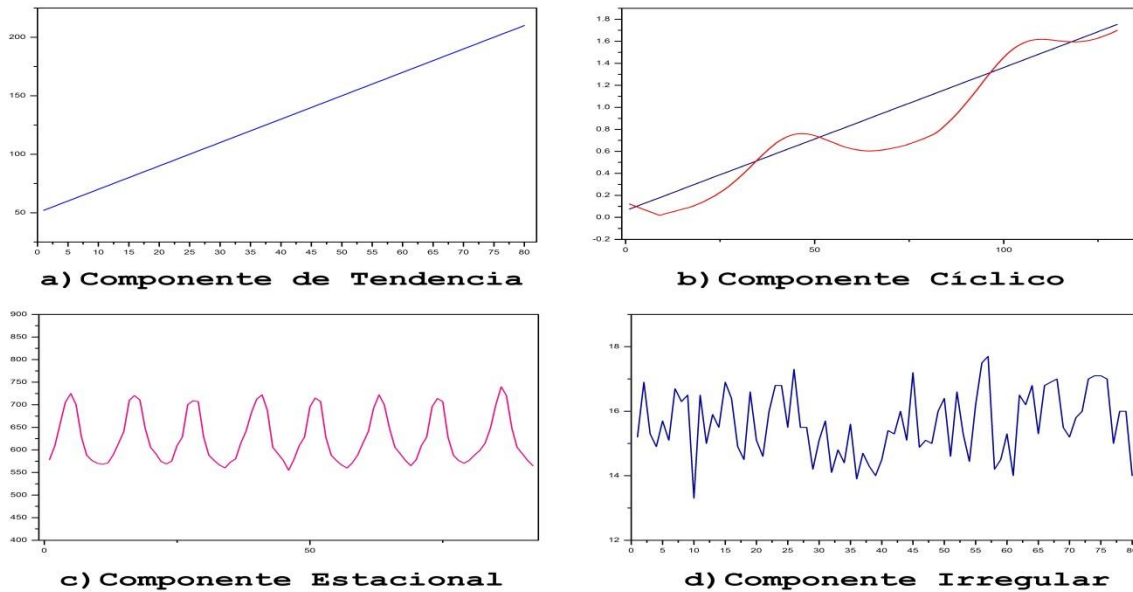
1. Componente de tendencia: es el comportamiento que indica la trayectoria o dirección en la cual una serie manifiesta dirigirse en el largo plazo, las más frecuentes son; lineales, logarítmicas y exponenciales. Algunos fenómenos que impactan en las series históricas económicas son; inflación, crecimiento poblacional, cambios en los precios y cambios tecnológicos.
2. Componente Cíclico: es el patrón de larga duración que oscila ondulatoriamente alrededor de la tendencia, su movimiento puede ser periódico o estocástico. Es necesario restarle a la tendencia el patrón cíclico para obtenerlo, pero dado que en muchas ocasiones esta separación no es posible se adopta como solución incluir a los dos componentes en un solo patrón.
3. Componente Estacional: las variaciones estacionales son patrones repetibles en un rango igual o menor a un año estas fluctuaciones se localiza en series clasificadas semanalmente, mensualmente, bimestralmente, trimestralmente, etc. La evolución cambiante del patrón estacional se produce recurrentemente a través del tiempo. Su comportamiento es reiterativo año tras año, y su desplazamiento o movimiento concluye dentro del transcurso de un año.
4. Componente Irregular: son fluctuaciones que duran un corto intervalo de tiempo que no presentan un comportamiento sistemático ni periódico, y por lo mismo son impredecibles. Se produce de manera ocasional afectando de manera casual a la variable que se analiza y de forma no continua.

La figura 4.1 ilustra los componentes que integran una ST desde la perspectiva clásica, en la figura 4.1a se muestra una tendencia lineal ascendente, la figura 4.1b



exhibe el componente cíclico montado a una tendencia positiva, en la figura 4.1c es mostrado el componente estacional y finalmente en la figura 4.1d se presenta el componente irregular.

Figura 4. 1: Componentes de una Serie de Tiempo desde la perspectiva clásica.



Fuente: Elaboración propia.

- **Series estacionarias y no estacionarias:** los datos de una serie temporal estacionaria deben oscilar o fluctuar en el tiempo alrededor del eje horizontal, esto quiere decir que el promedio de la serie debe de ser constante independientemente del tiempo. Generalmente las STs económicas suelen ser no estacionarias, ya que presentan patrones estacionales y de tendencia ya sea creciente o decreciente que causan que la media y la varianza de las series sean inconstantes, y no sigan una distribución normal. Por medio de los métodos gráficos muchas de las veces es posible inferir si una serie de datos es no estacionaria, se parte primero de una inspección visual de la gráfica de la ST y del diagrama de autocorrelación, la gráfica de autocorrelación de una serie de datos no estacionarios decrece lentamente en el tiempo a diferencia de una serie estacionaria donde la gráfica decrece de una manera rápida. Una técnica más robusta para la detección de la no estacionariedad en las series es el análisis o test de raíces unitarias entre los más populares están las pruebas de Dickey-Fuller Aumentado (ADF), Dickey-Fuller GLS (DFGLS), Phillips-Perron (PP) y

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) (Meza Carvajalino, 2007). La herramienta fundamental de los métodos AR y MA es la regresión y al emplear el método de mínimos cuadrados a series de datos no estacionarios, puede conducir a descubrir parámetros incorrectos sobre las relaciones entre las variables, además bajo esta circunstancia es común encontrar relaciones significativas entre las variables cuando en realidad no existen. Se denomina regresión espuria cuando ocurre este suceso (Granger y Newbold, 1974). Antes de empezar a construir los modelos se debe remover la no estacionariedad en los datos para evitar estimaciones erróneas, para poder satisfacer las condiciones que imponen estos modelos existen diferentes técnicas para transformar los datos de entre los más populares son:

- a) Si la serie aumenta con el tiempo sería evidente que no se tiene una media constante, y si hay evidencia que se está ante una tendencia determinista, se podría realizar una regresión ya sea lineal, logarítmica, exponencial, etc., y de esta forma convertirla en una serie que si sea estacionaria. Posteriormente la serie alterada se examinaría, se determinarían sus patrones para luego pronosticar sus futuros valores. Para finalizar se tendría que realizar la transformación inversa, es decir adicionarle la tendencia que se le había sustraído con anterioridad.

$$Y = a + bx + e$$

- b) Cuando los datos no muestran una tendencia determinista sino una tendencia estocástica es necesario transformarla mediante la técnica de diferenciación (también llamada de periodo a periodo o corta). Se les llama series integradas a las series que son alteradas mediante la diferenciación y que las convierte en series estacionarias. Dependiendo del número de diferenciaciones que se le aplique a la serie se le dará el distintivo de integrada en orden  $n$ , donde  $n$  es el número de diferenciaciones que se le tuvieron que aplicar a la serie para convertirla en una serie estacionaria, por ejemplo; si a la serie se le aplica una diferencia será una serie integrada en orden uno y así consecutivamente, cabe destacar que las series económicas y financieras rara vez se le aplica más de una diferencia (Fabris, 2009).

$$Y_t = Y_t - Y_{t-1}$$

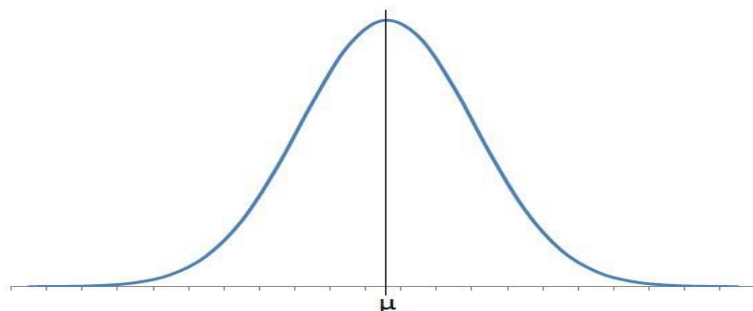
- c) Las series macroeconómicas regularmente exhiben heterocedasticidad (la varianza es inconstante) (Fabris, 2009), para suavizar la serie se estila utilizar las transformaciones Box y Cox, que consiste aplicarle logaritmos a la serie.
- d) Cuando el patrón estacional se manifiesta con mucha insistencia, y con la diferenciación corta no es suficiente, se le aplica una diferenciación estacional o larga, dependiendo del comportamiento del componente estacional es que se le aplica la diferencia. Una serie de datos que presente un comportamiento estacional cada 12 periodos se diferenciara de la siguiente manera.

$$Y_t = Y_t - Y_{t-12}$$

#### 4.2.2 Distribución normal

La distribución de probabilidad más destacada y con un mayor grado de popularidad es la distribución normal, puesto que a través de ella se explica de forma aproximada una gran cantidad de sucesos que acontecen en el ámbito de la industria, en las diferentes ramas de la investigación, así como también muchos fenómenos naturales son descritos por medio de la distribución normal.

Figura 4. 2: Distribución Normal.



Fuente: Elaboración propia.

La distribución normal fue propuesta inicialmente por Abraham de Moivre en el año 1733. También se le conoce como la campana de Gauss en honor de Carl Friedrich Gauss, ya que él fue pionero en la aplicación de esta técnica para el estudio de datos astronómicos. Es por medio de la estimación de dos parámetros estadísticos que se estima la distribución normal: el promedio o la media  $\mu$ , donde la cima tiene su punto más alto, y la desviación estándar  $\sigma$ , que muestra la dispersión. La curva

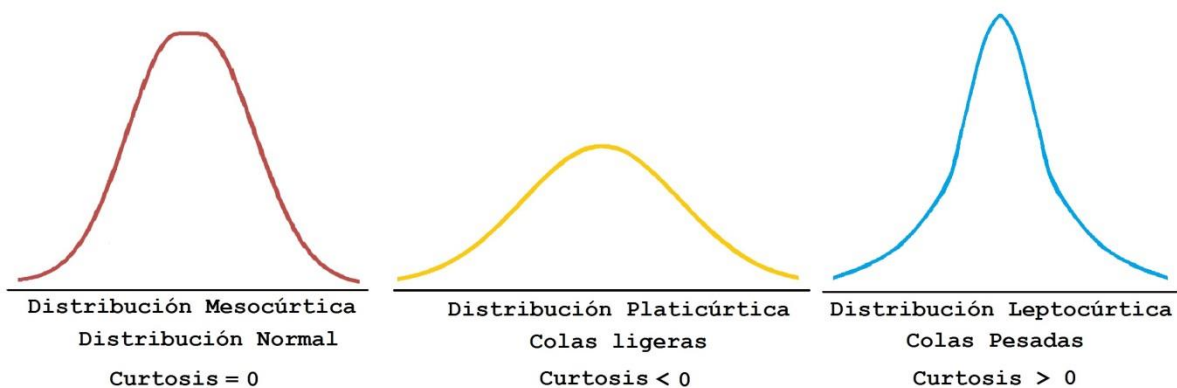
normal tiene la forma de campana y es simétrica como la que se presenta en la figura 4.2.

**Prueba Jarque–Bera.** Una de las primeras pruebas que se realiza sobre los datos es verificar la normalidad de los datos, una de las pruebas más conocidas es la Prueba Jarque-Bera. La prueba de Jarque-Bera basa su análisis en la asimetría y curtosis de la densidad empírica. La prueba estadística de Jarque-Bera se define:

$$JB = \frac{N}{6} \left( s^2 + \frac{1}{4} (K - 3)^2 \right)$$

Donde el número de observaciones es  $N$ , la asimetría de la muestra es  $S$ , y la curtosis de la muestra es  $K$ . La asimetría proporciona una medida de la simetría de las observaciones alrededor de la media. Una distribución normal, tiene una asimetría de valor 0. Una distribución con sesgo a la derecha tiene asimetría positiva y una distribución con sesgo a la izquierda tiene sesgo negativo.

Figura 4. 3: Representación gráfica de los tipos de curvas.



Fuente: Elaboración propia.

Mediante el coeficiente de curtosis se mide el grado de concentración de los valores de los datos alrededor de la media, que de igual forma se comparan con la distribución normal, y en base a esta medición permite conocer si la distribución de los datos es leptocúrtica (mayor concentración en torno a la media), mesocúrtica (similar a la distribución normal) o platicúrtica (mínimo aglomeramiento de datos alrededor de la media). El coeficiente de curtosis para una Distribución Normal es tres. La figura 4.3 muestra los tres tipos de curvas de manera gráfica.

#### 4.2.2 Prueba de Portmanteau

La prueba de Ljung-Box es una herramienta que se emplea para verificar que los residuales de un modelo ARIMA que se ha propuesto sean similares al ruido blanco, es decir, se comprueba que los residuos son independientes. De manera general la prueba de Ljung-Box se define así:

$H_0$ : El modelo se encuentra ajustado correctamente.

$H_a$ : El modelo exhibe deficiencia en el ajuste.

$$Q = n(n + 2) \sum_{k=1}^m \frac{\hat{r}_k^2}{n - k}$$

Donde  $n$  es la longitud de la ST,  $\hat{r}_k$  es la autocorrelación estimada de la serie en el rezago  $k$ , por último  $m$  es el número de rezagos que se están verificando. El rechazo de la hipótesis nula en la prueba de Ljung-Box ocurre en caso de:

$$Q > X_{1-\alpha}^2$$

Donde  $X_{1-\alpha}^2$ ,  $h$  es el valor tabla de la distribución de chi-cuadrado con  $h$  grados de libertad y con un nivel de significancia  $\alpha$ . Dado que la prueba es empleada sobre los residuos, los grados de libertad deben considerar los parámetros estimados del modelo a fin de que  $h = m - p - q$ , donde  $p$  y  $q$  indican el número de parámetros del modelo Autorregresivo de Media Móvil (ARMA).

#### 4.2.3 Tipos de pronósticos

Comúnmente las pruebas estadísticas para determinar la precisión del desempeño predictivo de un modelo propuesto, se realizan dividiendo el conjunto de datos en dos grupos de datos históricos. El primer subconjunto de datos se le denomina periodo de estimación dentro de la muestra y el segundo subconjunto periodo de estimación fuera de la muestra.

- a) Pronóstico dentro de la muestra: este pronóstico se genera a partir del conjunto de datos que se han utilizado previamente para estimar los parámetros del modelo. Dada la situación se espera que el pronóstico tenga un buen ajuste a los datos. En esta fase experimental se utiliza por parte de los modelos

paramétricos el denominado pronóstico estático el cual también es conocido como el pronóstico a 1-paso por delante o a un periodo, donde se utilizan los valores reales o actuales en lugar de los valores pronosticados. El término estático proviene del software Eviews. El modelo no paramétrico utiliza una técnica similar llamada Single-Step-Ahead Prediction (SSP) (Landassuri-Moreno, Bustillo-Hernández, Carbajal-Hernández, & Fernández, 2013). La técnica SSP de manera análoga proyecta un solo valor en el futuro donde no se utiliza una retroalimentación para seguir con las sucesivas predicciones, al igual que la predicción estática, esta herramienta utiliza los datos actuales para continuar con los pronósticos. En la etapa experimental no se utilizan estas técnicas, ya que se concluyó basado en experimentaciones previas, que estas no reportan utilidad para los fines requeridos en la presente investigación.

b) Pronóstico fuera de la muestra MSF y MSP: El Pronóstico fuera de la muestra en teoría debe de ser una herramienta para medir la precisión de los modelos predictivos de forma imparcial ya que prescinden de los ajustes y de los filtros que previamente habían sido utilizados en el pronóstico dentro de la muestra. Los datos que se utilizan en este análisis son de un rango diferente a los que fueron utilizados previamente para ajustar o calibrar el modelo (Pronóstico dentro de la muestra). Los modelos paramétricos utilizan el pronóstico llamado Multiple-Step-Ahead Forecast (MSF) (Brooks, 2008), también denominado pronóstico dinámico por el software utilizado (Eviews), esta técnica calcula los pronósticos posteriores al primer periodo en la muestra mediante el uso de los valores previamente pronosticados de la variable, es decir, la predicción se realiza tomando en consideración solamente las observaciones que inicialmente se utilizaron para estimar los parámetros. El modelo análogo en los modelos no paramétricos es la técnica llamada Multi-step Prediction (MSP) (Landassuri-Moreno et al., 2012), las predicciones varios pasos adelante son aquellas generadas para n-pasos adelante, donde el horizonte o distancia de la predicción viene a ser delimitada por medio de los n-periodos. A continuación, se muestra en la tabla 4.3 las técnicas de predicción.

Tabla 4. 3: Técnicas de Predicción; I) SSP/Estático y II) MSP/MSF.

I) Pronóstico	Entradas	II) Pronóstico	Entradas
$Y_{t+1}$	$X_t, X_{t-1}, X_{t-2}$	$Y_{t+1}$	$X_t, X_{t-1}, X_{t-2}$
$Y_{t+2}$	$X_{t+1}, X_t, X_{t-1}$	$Y_{t+2}$	$Y_{t+1}, X_t, X_{t-1}$
$Y_{t+3}$	$X_{t+2}, X_{t+1}, X_t$	$Y_{t+3}$	$Y_{t+1}, Y_{t+2}, X_t$
$Y_{t+4}$	$X_{t+3}, X_{t+2}, X_{t+1}$	$Y_{t+4}$	$Y_{t+1}, Y_{t+2}, Y_{t+3}$

Fuente: Adaptado de Landassuri-Moreno et al. (2012).

La tabla 4.3, muestra de una manera simplificada cómo se calculan las predicciones, para las técnicas empleadas cotidianamente. La primera columna así como la tercera indican el número de predicción estimada, en la segunda y cuarta columna se muestran los datos empleados (entrada) para realizar dichas predicciones. Donde ambos métodos se pueden apreciar como una ventana con un movimiento continuo a un paso, para cada predicción estimada. Lo que distingue a cada método es el conjunto de datos que utilizan, los métodos SSP/Estático emplean los datos originales de la serie para realizar la estimación del pronóstico. En cambio las técnicas MSP/Dinámico en un principio utilizan los datos originales para estimar la primera predicción, pero conforme avanzan las predicciones se mezclan datos originales con predicciones, hasta finalmente realizar pronósticos tomando en cuenta solamente valores predichos.

### 4.3 Modelo paramétrico ARIMA-ARCH

#### 4.3.1 Introducción

Los modelos paramétricos parten de suponer un modelo estadístico donde los datos se encuentran distribuidos de una forma normal, las inferencias sobre los datos se hacen en función de la estimación de los parámetros (ejemplos: promedio, desviación estándar, asimetría, etc.). El grado de relación entre dos variables se mide por medio del coeficiente de correlación, el rango en el cual fluctúa esta relación varía entre  $+1$  que indica una relación positiva total, ya sea que aumente o disminuya una variable la otra también lo hará en el mismo sentido, por el contrario si la relación es de  $-1$  es una relación inversa o contraria en esta relación si la variable aumenta la otra disminuye y si esta disminuye la otra aumenta. Dos variables no se

encuentran relacionadas cuando el coeficiente es igual a cero, es decir si una de las variables disminuye o aumenta ese movimiento no tendrá impacto sobre la otra. En los modelos univariantes la autocorrelación juega un papel preponderante para determinar los patrones elementales y definir el modelo que comprenda y se ajuste a los datos. El coeficiente de correlación es un factor análogo al coeficiente de autocorrelación, la diferencia radica en que este último explica la relación entre los valores de la misma variable, rezagadas en diferentes fases o periodos del tiempo.

#### 4.3.2 Modelo Autorregresivo Integrado de Media Móvil (ARIMA)

El modelo Autorregresivo (AR) fue propuesto por George Yule (1927), Evgeny Evgenievich Slutsky retomo las ideas propuestas por Yule (Yule, 1921; Yule, 1926) y creo el modelo de Media Móvil (MA) (Slutsky, 1927). En el modelo AR las variables independientes o que explican el modelo, es la misma serie rezagada en el tiempo.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (4.1)$$

Los coeficientes que señalan el grado de influencia de los retardos en el momento  $t$  están representados por  $\phi$ , estos últimos se estiman por medio de la autocorrelación de una forma similar a la regresión.

$Y_t$  = Es la variable dependiente.

$Y_{t-1}, Y_{t-2} \dots Y_{t-n}$  = Son las variables independientes, son valores de la misma serie en periodos anteriores.

$\varepsilon_t$  = Perturbación o shock aleatorio que no explica el modelo.

Por otro lado, el proceso de MA realiza la estimación en función de los errores o perturbaciones del pasado como variables independientes o explicativas.

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (4.2)$$

La ecuación (4.2) es similar a la ecuación (4.1), la diferencia estriba en que los modelos AR utilizan como variable explicativa el valor de la misma variable a diferencia del modelo MA que construye el modelo en base a los errores o perturbaciones del pasado de la variable misma.

$Y_t$  = es la variable dependiente

$\varepsilon_t$  = los errores o perturbaciones aleatorias

$\theta_t$  = el ponderador que indica la persistencia de la perturbación



El modelo que combina ambas ecuaciones es llamado modelo mixto, ya que es posible que este describa de una manera más adecuada el proceso que se está analizando.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_n Y_{t-n} + \varepsilon_t - \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (4.3)$$

A continuación se realizara un ejemplo somero con fines didácticos, con el fin de clarificar un poco más la construcción de los modelos. Una pequeña parte de la ST manchas solares será analizada, donde los valores de  $\phi$  y  $\theta$  son ficticios, de forma real los valores de las ponderaciones son estimados por medio de la regresión que de igual forma estima el signo de los diferentes coeficientes que componen al modelo. La tabla 4.4 muestra una situación ficticia y se presenta la construcción de un modelo mixto Autorregresivo de Media Móvil (ARMA).

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \theta_4 \varepsilon_{t-4}$$

Tabla 4. 4: Ejemplo: predicción con el modelo ARMA.

Período	Manchas Solares	Predicción Manchas Solares	Error o Shock	Ponderación $\phi$	Ponderación $\theta$
t-5	81.6	80	1.6	-	-
t-4	82.8	83	-0.2	-	0.2
t-3	84.1	84	0.1	-0.1	0.5
t-2	86.3	85	1.3	0.4	0.6
t-1	87.8	86	1.8	0.7	0.3

Fuente: Elaboración propia.

El modelo que se estima es de la forma ARMA(p,q), si p=3 (valor de ponderación de  $\phi_1 = 0.7$ ,  $\phi_2 = 0.4$ ,  $\phi_3 = -0.1$ ) y q=4 (valor de ponderación de  $\theta_1 = 0.3$ ,  $\theta_2 = 0.6$ ,  $\theta_3 = -0.5$ ,  $\theta_4 = 0.2$ ).

El modelo ARMA(3,4), quedaría de la siguiente forma:

$$Y_t = 0.7(87.8) + 0.4(86.3) + [-0.1(84.1)] + \varepsilon_t + 0.3(1.8) + 0.6(1.3) + [-0.5(0.1)] + 0.2(-0.2)$$

$$Y_t = 88.8 + \varepsilon_t$$

### 4.3.3 Metodología Box-Jenkins

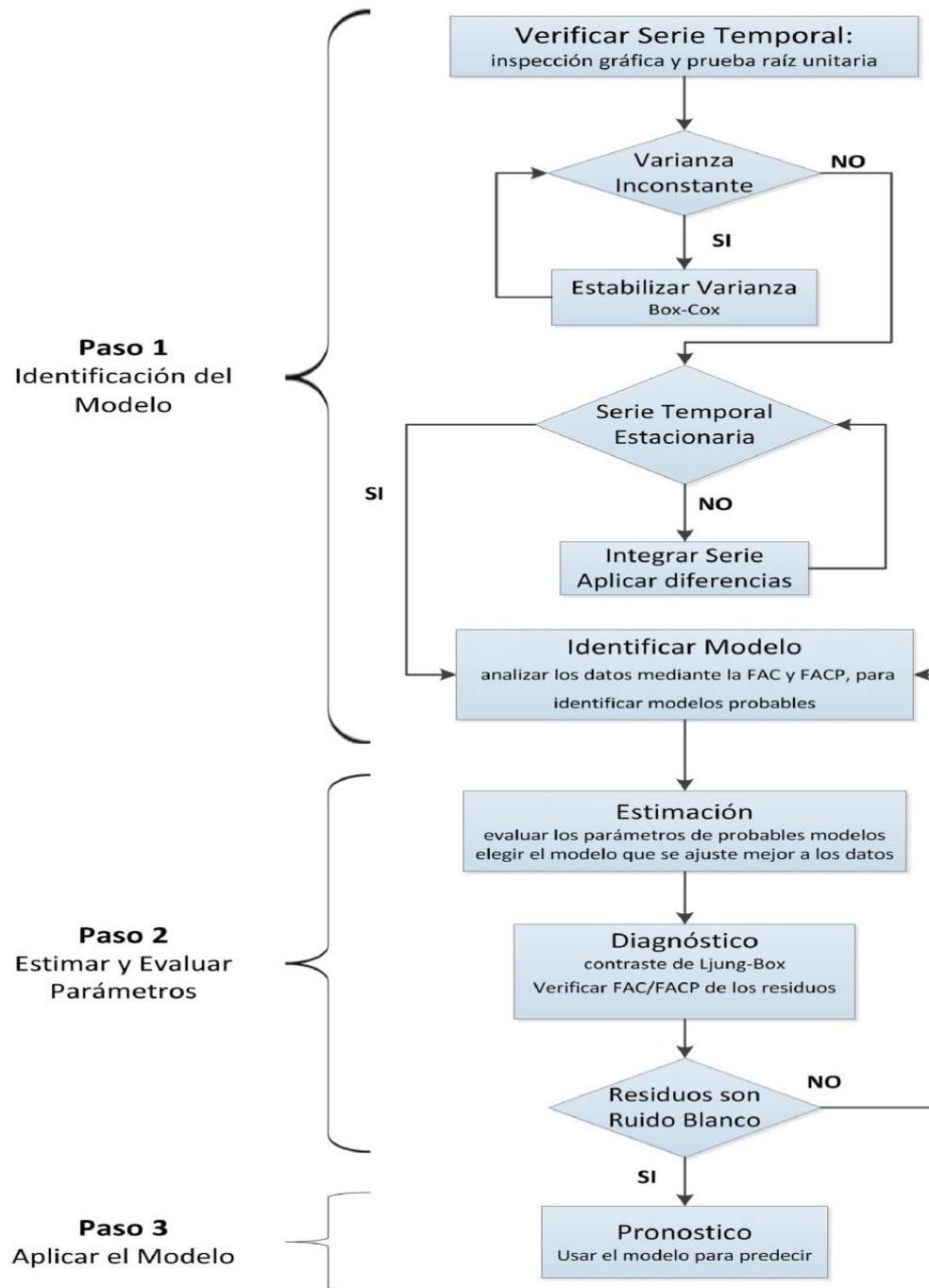
En la década de los años setenta del siglo pasado George E. P. Box y Gwilym M. Jenkins publican la obra Time Series Analysis: Forecasting and Control donde proponen un estudio de las series de tiempo donde solo se involucra una variable a diferencia del análisis causal tradicional, donde el proceso se examina por medio de una o más variables que explican el fenómeno. El modelador no determina ni

presupone la estructura del modelo a priori, en cambio los datos son los que conducen la metodología. La modelación y posterior predicción de una variable está en función de sus valores pasados sin necesidad que otra variable explique o determine el fenómeno, de este modo la evolución de la ST está definida por su propio pasado más el movimiento de una perturbación aleatoria que no incluye el modelo al ser eventos fortuitos o accidentales (Fabris, 2009). De manera análoga a la metodología Box-Jenkins también se le conoce como modelo ARIMA y su ecuación se estructura mediante tres componentes: Autorregresivo (AR)/Integrado (I)/ Media Móvil (MA). La metodología Box-Jenkins hace referencia a un grupo de técnicas o procedimientos de identificación, estimación, ajuste y comprobación de los modelos ARIMA. En general la metodología Box-Jenkins está compuesta por cuatro fases:

1. Identificación: Se debe de definir las modificaciones necesarias para volver la serie en un proceso estacionario, en otras palabras integrar la serie de tiempo aplicando diferenciación u otro método que ajuste nuestros datos, esto da por consecuencia encontrar el grado u orden de integración (I). Determinar los órdenes de los componentes (p) autorregresivo AR, (q) promedio móvil MA. Se seleccionan uno o más modelos tentativos que posteriormente serán analizados, estimados y comparados para evaluar cuál es el modelo que representa la evolución de la serie de la manera más apta. El grado de los parámetros se determinara en base a la función de Autocorrelación (FAC) y la función de Autocorrelación Parcial (FACP). En la mayoría de los casos se debe de elegir el modelo más parsimonioso, es decir, que tenga menos coeficientes en AR y MA.
2. Estimación: la estimación consiste en un procedimiento iterativo donde se estiman los valores de los parámetros de cada modelo o modelos identificados en la anterior etapa.
3. Evaluación: en la etapa de Evaluación se verifica que el modelo que se haya seleccionado sea el adecuado. Se deberán realizar y analizar pruebas a los residuos para constatar si no existen patrones que no se hubieran incluido en el modelo elegido.
4. Pronóstico: en base a los parámetros del modelo que se ha elegido se estimará la predicción, se recomienda que el número de periodos a predecir no sea mayor a cinco (Meza Carvajalino, 2007).

En la figura 4.4 se esquematizan las etapas de la metodología Box-Jenkins.

Figura 4. 4: Metodología Box-Jenkins.



Fuente: Adaptado de (Makridakis, Wheelwright y Hyndman, 1988).

#### 4.3.4 Modelo Estacional Autorregresivo Integrado de Media Móvil

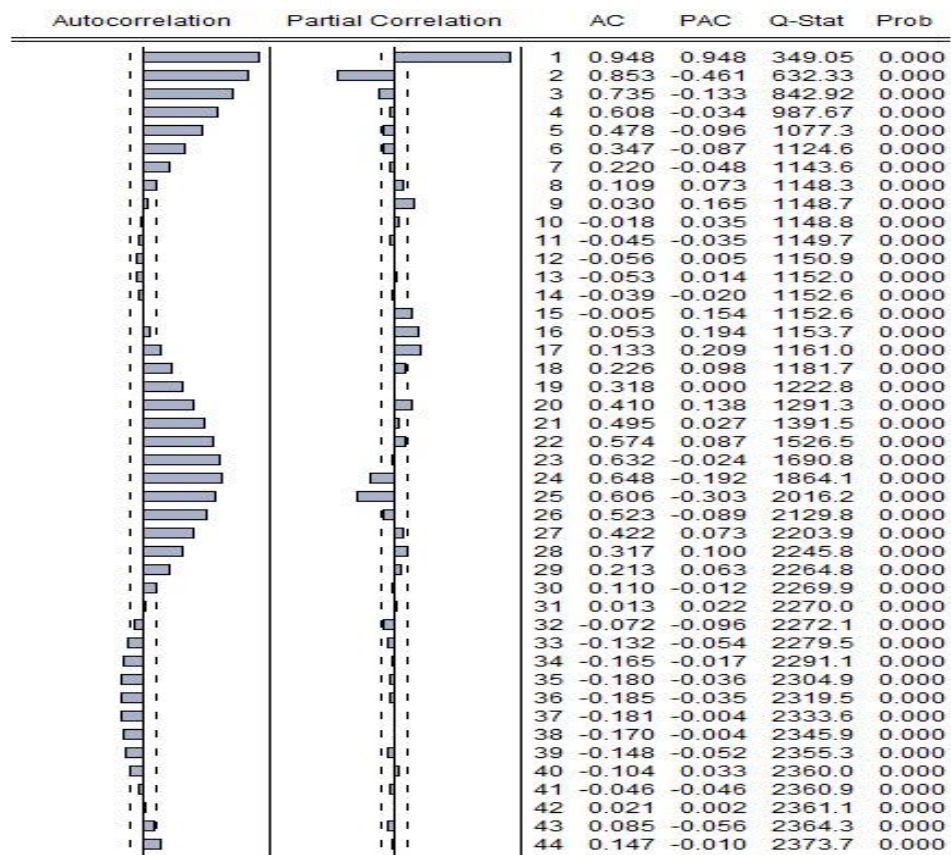
El patrón estacional se presenta en una ST cuando en ésta inciden factores estacionales, como pueden ser los días de la semana, circunstancias mensuales, cuatrimestrales entre otras. La estacionalidad estable es un patrón regular y

repetitivo de períodos fijos y conocidos. El modelo  $ARIMA(p,d,q)$  resulta de la generalización del modelo  $ARMA(p,q)$ , donde el término agregado  $d$  se refiere al número de diferencias necesarias para estacionarizar el proceso, esto con el propósito de ampliar el modelo y hacerlo capaz de manejar series no estacionarias.

De igual forma, el Modelo Estacional Autorregresivo Integrado de Media Móvil (SARIMA) es una ampliación del modelo  $ARIMA(p,d,q)$ , que se utiliza en el caso de que la ST contenga un persistente comportamiento estacional. El modelo SARIMA puede contener en su estructura una parte no estacional, al igual que el modelo ARIMA, dado que puede tener los factores típicos de un modelo ARIMA como lo son el AR, el factor MA y un orden de diferenciación. En el Modelo Estacional Autorregresivo Integrado de Media Móvil  $(SARIMA)(p,d,q)(P,D,Q)_S$ , el primer paréntesis hace referencia a la parte regular de la ST, y el segundo paréntesis se encuentra relacionado con la parte estacional de la ST. Donde  $P$  = número de términos autorregresivos estacionales (SAR),  $D$  = número de diferencias estacionales,  $Q$  = número de términos de promedio móvil estacionales (SMA), y el subíndice  $S$  = se refiere a la periodicidad (días, meses, trimestre, etc.) de la serie.

Como es habitual la primera aproximación para detectar estacionalidad en una ST es mediante un reconocimiento visual del gráfico. Mediante el cual es factible que se advierta los posibles periodos estacionales. Mediante el correlograma se visualiza la Autocorrelación (FAC) y la Autocorrelación Parcial (FACP), dichos coeficientes deben ser significativamente diferentes de 0, usualmente se pueden apreciar periodos repetitivos de rápido crecimiento para posteriormente de igual forma decaer como se aprecia en la figura 4.5, donde se puede apreciar una forma similar a la de un abanico. Tanto la FAC como la FACP, pueden intercambiar roles en el comportamiento estacional.

Figura 4. 5: Ejemplo: Correlograma de una Serie de Tiempo con comportamiento estacional.



Fuente: Elaboración propia.

Típicamente se debe de diferenciar la ST, hay que tener cuidado en la elección de la diferenciación donde se puede tener: diferencia no estacional; 0 o 1 y diferencia estacional; 0 o 1. No se recomiendan más de 2 diferenciaciones en el modelo. La estimación y posterior verificación de los parámetros (P,D,Q), se realizan de manera semejante a los de la parte regular (p,d,q), pero en este caso se debe tomar en cuenta la periodicidad implícita o explícita de la ST.

#### 4.3.5 Modelos de Volatilidad Condicional Heterocedástica

##### 4.3.5.1 Introducción

La flexibilización, la eliminación de restricciones, así como la eliminación de impedimentos a la entrada y salida de capitales en los diferentes mercados financieros han motivado un incremento sustancial en la inestabilidad de los mercados que se traduce en una creciente volatilidad, como tal, en las pasadas décadas el estudio y análisis de dicho fenómeno ha venido a cobrar un especial

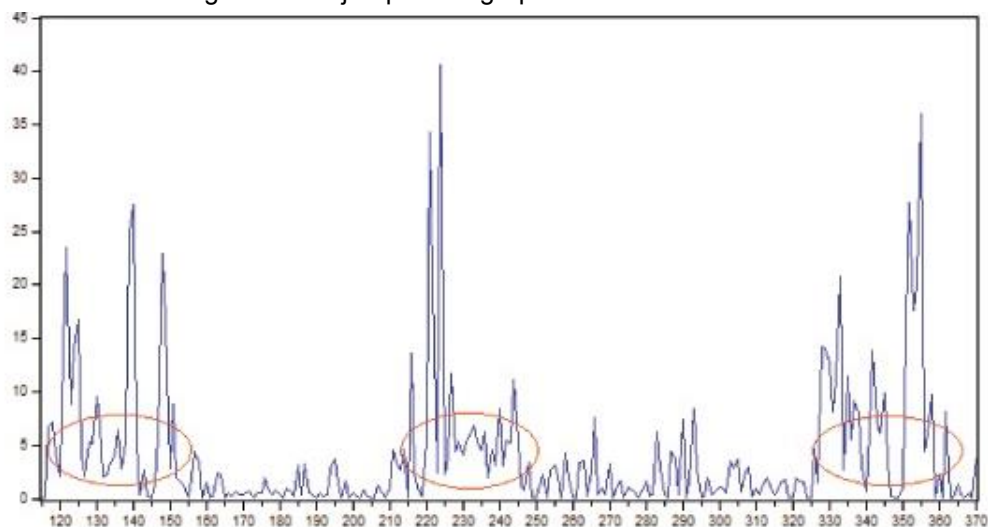
interés tanto en el ámbito académico como en el profesional. La entidad emisora del instrumento financiero presta atención en el comportamiento del activo durante la duración de la actividad o vida del papel emitido, por ello basa su interés en el valor de la media y varianza no condicional o incondicional, que basa su estudio en la conducta probabilística de forma individual partiendo de la distribución marginal de la variable de interés. Por el contrario el inversor al asumir el riesgo del papel estará interesado en la rentabilidad del instrumento en el corto plazo, como tal, el inversionista debe estudiar la varianza y media condicional, en la búsqueda del mayor rendimiento (determinar el momento óptimo para: comprar, vender o retener) del activo, gestionando el riesgo, por lo tanto el inversor deberá utilizar un pronóstico condicional esto significa; una proyección hacia el futuro que se encuentra sujeta o condicionada a la información acumulada o disponible hasta dicho momento, ya sea que este se encuentre situado en un momento de baja o de alta volatilidad, se debe tomar en cuenta los periodos más cercanos o inmediatos de tranquilidad o de intranquilidad según sea el caso.

Los modelos clásicos de naturaleza lineal como son la regresión lineal y el modelo ARIMA, le confieren a sus parámetros linealidad p.e. al asumir que al igual que la media la varianza es constante (homocedasticidad), y si esta no lo es la transforman en una varianza lineal por diferentes técnicas como pueden ser las de la familia Box-Cox, pero como tal, las series de tiempo financieras son en su gran mayoría series no lineales (Brooks, 2008; Balacco, 2012). Dada las características de no linealidad y varianza heteroscedástica (no constante) de las series financieras que surge como consecuencia o en respuesta de una interacción no lineal de los diferentes agentes económicos (Brooks, 2008), es evidente que los modelos clásicos dado los supuestos en los que fundamenta su análisis provoca diversas inconsistencias estadísticas dando como resultado incapacidad en la captación de la dinámica del sistema que subyace en las series financieras. Como respuesta a lo anteriormente descrito a principios de la década de los ochenta surge una modelización que trata de resolver la volatilidad de los mercados financieros, de esta forma surgen los modelos de volatilidad condicional heteroscedástica que conceptualmente son idóneos para tratar la inestabilidad de la varianza en el transcurso del tiempo. Es

así que en el año de 1982 Robert Engle de la mano de su obra seminal (Engle, 1982) introduce los modelos de volatilidad condicional (ARCH) marcando así el inicio de ellos, en esencia el modelado del proceso toma en cuenta los cambios significativos en la varianza. El conjunto o familia de los modelos ARCH, tratan de captar los rasgos más representativos de las series financieras que se encuentran con mayor reiteración, sus características se presentan a continuación.

- Agrupamiento de la volatilidad: la aglomeración de la volatilidad también conocido en inglés como clúster de volatilidad, es un rasgo o particularidad que se presenta especialmente en el mercado accionario, por lo regular la volatilidad se manifiesta en ráfagas, es decir, que grandes cambios vienen seguidos de grandes cambios y cambios menores vienen seguidos de cambios menores, en ambos casos estos cambios pueden ser positivos o negativos. La figura 4.6 ilustra el comportamiento de los retornos al cuadrado del DJIA en un periodo histórico escogido al azar donde se aprecia de forma bastante clara como se aglomera y se dispersa la volatilidad. Se han marcado en círculos rojos los periodos de alta volatilidad donde el mercado muestra intranquilidad y sus valores son altamente fluctuantes a diferencia de los periodos de tranquilidad donde las oscilaciones son menores. La figura 4.6 es una serie con rasgos de agrupamiento de volatilidad.

Figura 4. 6: Ejemplo de agrupamiento de la volatilidad.



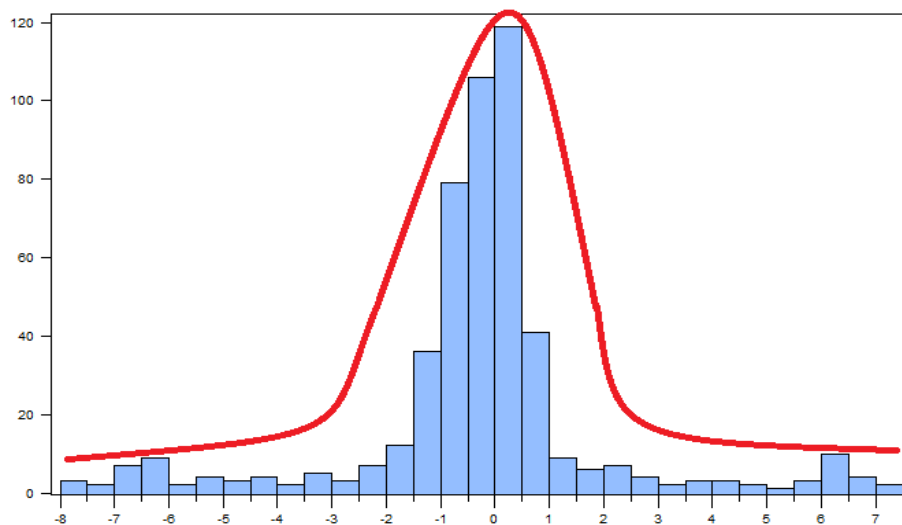
Fuente: Elaboración propia.

- El efecto de Apalancamiento. En los mercados financieros de forma habitual los

rendimientos de las acciones tienden a estar correlacionados de forma negativa con los cambios en la volatilidad, en otras palabras, si los rendimientos son más pequeños de lo esperado (malas noticias) impactan de una forma mayor, traduciéndose en una mayor volatilidad, a diferencia de las buenas noticias (mayor de lo esperado) donde la volatilidad tiende a disminuir.

- **Distribución Leptocúrtica.** La distribución de las series la mayoría de las veces presenta una distribución simétrica, similar a una distribución normal, aunque las colas de su distribución tienden a ser más anchas que la normal, y una curtosis de tipo leptocúrtica, esto se explica por la cantidad masiva de valores extremos. Se denomina coeficiente de curtosis a la medida que nos indica la altura y el grado de afilamiento del pico o punto más alto de una distribución, con respecto a una medida estándar. Dicho patrón de medida estándar, es la distribución normal que tiene una forma de campana. En la figura 4.7 se muestra una distribución de datos con características leptocúrticas.

Figura 4. 7: Distribución Leptocúrtica.



Fuente: Elaboración propia.

#### 4.3.5.2 Determinación del modelo ARCH(q)

Como antes ya se había descrito la característica más emblemática de las series financieras son; los cambios desmesurados de los valores de las series seguidos de cambios de gran magnitud que son los periodos de alta volatilidad, después de



un tiempo indeterminado estos exabruptos retornan la serie a la normalidad es decir los valores tienden a regresar a la media donde la serie se estabiliza estos periodos son denominados de baja volatilidad ya que los cambios son pequeños y estos son seguidos por cambios reducidos, este comportamiento variante demuestra que la serie en las perturbaciones contiene autocorrelación o se encuentra serialmente correlacionada. Este proceso presenta una especie de heterocedasticidad concurrente en donde la varianza de la perturbación de la regresión obedece o es dependiente de la volatilidad de las perturbaciones en el pasado inmediato. El modelo ARCH(q) presupone que la varianza depende de las perturbaciones pasadas, asemejándose a un proceso MA ya que la varianza condicional  $\sigma_t^2$  depende de los valores de los rezagos de los residuales al cuadrado, la ecuación tiene la estructura siguiente:

$$\epsilon_t = \mu_t z_t \quad (4.4)$$

$\epsilon_t$  = termino de error

$\mu_t$  = desviación estándar

$z_t$  = proceso de ruido blanco

Donde el ruido blanco se denota:

$$z = a_t ; a_t \sim iid \quad N(0, 1)$$

El ruido blanco es homocedástico ya que la varianza tiene un valor de uno y es constante. El orden q del modelo ARCH(q) hace alusión al número de periodos rezagados.

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_1 \epsilon_{t-2}^2 + \dots + \alpha_p \epsilon_{t-p}^2 \quad (4.5)$$

Donde  $\alpha_0 > 0, 0 \geq \alpha_i \leq 1$

para que el proceso que se modela sea heterocedástico este se obtiene multiplicando a  $z_t$  por un coeficiente de  $\epsilon$  donde la varianza debe ser positiva y los rezagos más recientes impactan de una forma mayor, si el valor de  $\epsilon$  es alto un valor cercano a la unidad, la varianza de la siguiente observación condicionada a este valor será también alta. La condición  $|\alpha_i| < 1$  garantiza la estabilidad del

modelo evitando que la  $\sigma^2$  no sea explosiva. Para poder calificar a un modelo ARCH como apto este debe de cumplir ciertos requisitos o acatar ciertas restricciones. Es usual utilizar muchos rezagos para lograr adecuar un modelo ARCH, esta característica es indeseable ya que reduce los grados de libertad a la regresión (Fabris, 2009).

Con la aparición del modelo GARCH se subsanaron estas cuestiones que limitaban al modelo ARCH. En el año 1986 Tim Bollerslev (Bollerslev, 1986) plantea el modelo GARCH (Generalized Autorregresive Conditional Heterocedasticity) con la encomienda de extender y flexibilizar los requisitos o restricciones que imponía el modelo ARCH. El modelo GARCH(p,q) se asemeja al modelo ARMA, ya que es un modelo compuesto donde se incluyen de igual forma tanto los retardos de  $\sigma_t^2$  como de  $\varepsilon_t^2$  de forma similar al modelo ARMA. El parámetro p indica el número de retardos de la variable ( $\varepsilon_t^2$ ), y el parámetro q muestra la cantidad de rezagos de la varianza condicional ( $\sigma_t^2$ ) (Fabris, 2009). El ordenamiento o armazón de un modelo GARCH(p,q) es como sigue:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2 \quad (4.6)$$

El modelo univariante GARCH(1,1), es muy popular y es ampliamente utilizado en la práctica, ya que dicho modelo puede captar de manera fehaciente la volatilidad que generalmente se presenta en las series financieras (Ludlow, 2006). Tomando como base los modelos ARCH y GARCH se han diseñado y creado una serie o conjunto de modelos opcionales que varían su estructura en harás de captar el dinamismo de la volatilidad, pero siempre manteniendo la esencia de los modelos en los cuales se basaron (ARCH y GARCH), de los cuales destacan IGARCH, ARCH-M, EGARCH, TARCH (Ludlow, 2006).

Cabe destacar que el modelado ARCH del proceso heterocedástico puede elaborarse mediante dos vertientes como son: el estudio causal y mediante la metodología Box-Jenkins, lo cual hace a dicho modelado una herramienta flexible y robusta para el analista.

#### **4.4 Modelo no paramétrico**

## **4.4.1 Redes neuronales artificiales**

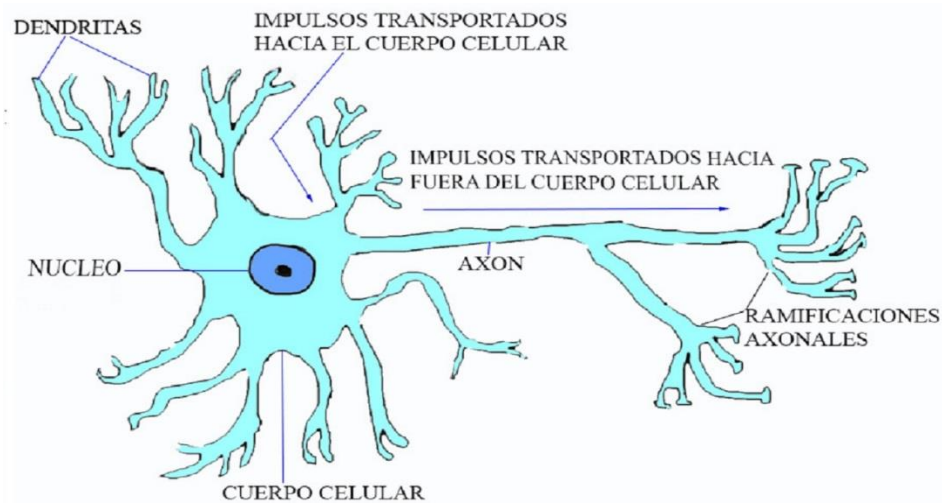
### **4.4.1.1 Introducción**

En los últimos años se ha popularizado el empleo de los diversos modelos de Redes Neuronales Artificiales (RNAs) como herramienta de análisis en diversos campos como son la medicina, finanzas, ingeniería, y telecomunicaciones entre muchas otras, y también en la predicción de series de tiempo. Man, Tang y Kwong (1999) sostienen que el hecho de esta gran aceptación y del amplio uso en diversas áreas es debido principalmente a la emulación del comportamiento de la naturaleza del cerebro humano y al hecho de que su estructura puede ser matemáticamente formulada (p. 155).

### **4.4.1.2 Neurona biológica**

La neurona se encuentra compuesta por los mismos componentes que las células biológicas. Lo que distingue a la neurona de las demás células vivas es que posee la facultad o capacidad de comunicación. La neurona es la unidad de procesamiento fundamental del sistema nervioso. Las neuronas se encuentran vinculadas entre ellas, por medio de conexiones de entrada y de salida. De forma general, una neurona se encuentra formada por las dendritas, por un cuerpo celular llamado soma de forma esférica, los cuales reciben impulsos de entrada, a través de las dendritas que a su vez lo transmite al soma, el cual se encarga de unir, conjuntar e integrar dichas señales para luego reenviarla a otras neuronas por medio de una prolongación del cuerpo celular denominado axón. El axón a través de sus ramificaciones se encarga de trasladar las señales, hasta sus botones terminales, los cuales tienen el cometido de repartir la señal a un nuevo grupo de neuronas. En oposición a las computadoras que operan principalmente de forma secuencial (procesamiento de tareas de una a la vez), el sistema nervioso es masivamente paralelo (procesamiento de varias tareas al mismo tiempo). Se estima que el sistema nervioso humano contiene alrededor de cien mil millones de elementos de procesamiento llamados neuronas, donde cada uno se encuentra conectado a miles de elementos análogos.

Figura 4. 8: La Neurona Biológica.

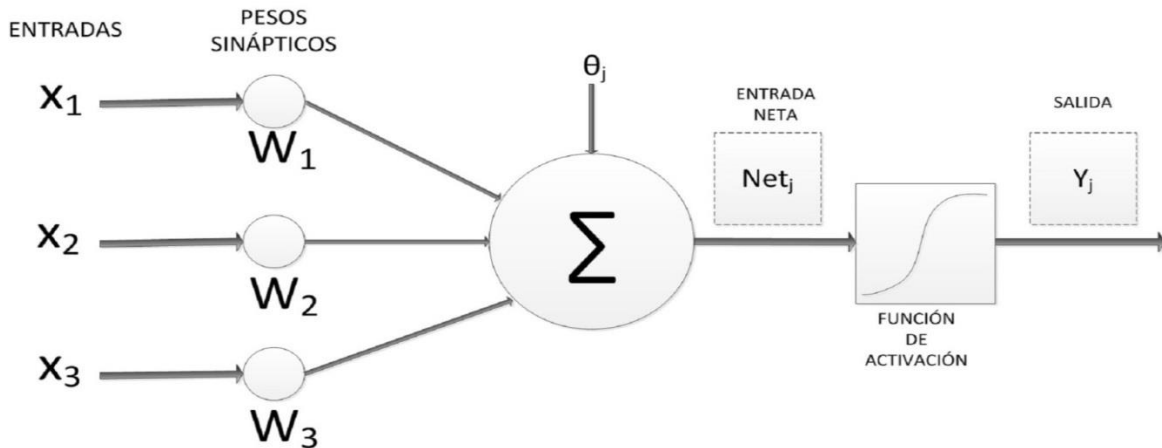


Fuente: Elaboración propia.

#### 4.4.1.3 Neurona artificial y sus particularidades

Las redes neuronales artificiales se diseñaron para intentar imitar el funcionamiento de las redes neuronales biológicas que se encuentran en el cerebro. El elemento básico y fundamental es la neurona artificial o también denominada procesador elemental, las cuales pueden ser vistas como unidades simples que reciben un estímulo procedente de un vector de entrada procedente de un agente externo o mediante una señal obtenida por otras neuronas con las que tenga conexión, para luego realizar cálculos y producir una respuesta en forma de salida. Los elementos básicos que integran una neurona se pueden apreciar en la figura 4.9 y se enlistan a continuación

Figura 4. 9: Neurona Artificial.



Fuente: Elaboración propia.

1. Pesos sinápticos: a través de una etapa o proceso de aprendizaje es que la red consigue el conocimiento el cual es acumulado en las conexiones que hay entre las neuronas denominadas como pesos sinápticos. Los pesos son valores numéricos que tienden a variar durante la fase de entrenamiento de la red neuronal, esta modificación en los pesos da como resultado que la información se encuentre contenida en ellos. Un elemento que se le agrega al conjunto de pesos es el umbral o bias ( $\theta$ ) que dependiendo de la función de activación que se utilice puede emplearse como un valor de un peso adicional, y en otros modelos como su nombre lo indica cumple la función de umbral de la neurona no activándola o activándola siempre y cuando la suma ponderada de las entradas sea igual o mayor que dicho umbral.
2. Variables de entrada y salida: las señales de entrada y salida pueden tomar diferentes valores que dependiendo de la utilización y del modelo de red a emplear pueden ser de carácter binario o continuo. Las neuronas de la capa de entrada reciben impulsos provenientes del exterior, las señales de salida de la capa de entrada son entradas a las neuronas de la capa que le precede.
3. Función de Propagación: la regla de propagación determina los preceptos y condiciones bajo las cuales se debe combinar las variables de entrada y los pesos de las conexiones que convergen en la neurona. De manera similar a la forma en como recibe información la neurona biológica, la neurona artificial recibe un cúmulo de entradas o señales provenientes del exterior o de otras neuronas con las cuales se encuentra vinculada, para el modelo que se ilustra en la figura 4.9, se aprecia que las señales que recibe la neurona se definen mediante un vector de entradas  $X = [x_1, x_2, x_3]$ . Este vector de entradas que recibe la neurona es transformada por medio de un vector ( $W$ ) de pesos que puede incrementar o debilitar dicha señal, se le nombra como entrada total o neta a la suma de las entradas modificadas, los pesos sinápticos interpretan un papel similar a la función que desempeña la sinapsis biológica. Los pesos sinápticos determinan la magnitud de la relación recíproca entre la neurona presináptica ( $j$ ) y la postsináptica ( $i$ ). La función que utiliza de forma más común

es la de tipo lineal y se fundamenta en la suma ponderada de las entradas con los pesos con los que guarda relación.

$X$  = Vector de Entradas

$W$  = Vector de Pesos

$\theta$  = Umbral o Bias

$$Net_j(t) = \sum_j W_{ij}X_j + \theta_j \quad (4.7)$$

4. Función de activación: esta función establece el estado de activación actual mediante la combinación del potencial postsináptico  $Net_j$ , el cual es obtenido a partir de la función de propagación, con el estado inicial de la neurona ( $a_j(t-1)$ ), para elaborar un nuevo estado de activación conforme a los datos más recientes ( $a_j(t)$ ).

$a_j(t-1)$  = Estado inicial de la neurona.

$Net_j(t)$  = Entrada Neta.

$t$  = Iteración.

$$a_j(t) = \mathcal{F}[a_j(t-1), Net_j(t)] \quad (4.8)$$

En la función de activación se valora o evalúa la entrada neta y se obtiene una señal de salida.

5. Función de Salida: toda neurona está ligada a una función de salida ( $y_j(t)$ ) que se determina en base al estado de activación actual  $a_j(t) = \mathcal{F}(Net_j(t))$ . Es usual que tanto la función de activación como la de salida coincidan, es decir, la función de salida es la misma o es una identidad de la función de activación.

$$y_j(t) = \mathcal{F}(a_j(t)) = a_j(t) \quad (4.9)$$

Esta circunstancia ocurre en los modelos neuronales que se utilizan de forma habitual como la ADALINA y el Perceptrón Multicapa.

Algunos de los rasgos distintivos de las redes neuronales artificiales se enlistan a continuación:

1. Generalización: al concepto de generalización de una red se entiende como la habilidad de responder de la red adecuadamente a datos o eventos que no se

utilizaron en la etapa de aprendizaje, es decir, la capacidad de extender una solución correcta ante patrones que no fueron incluidos en la fase de entrenamiento. La capacidad de generalización de una red está condicionada por tres factores que son: a) grado de dificultad del problema que se intenta resolver, b) la topología o arquitectura de la red y c) la cantidad de ejemplos de entrenamiento (Haykin, 1999).

2. Aprendizaje: las Redes Neuronales Artificiales son capaces de aprender cuando se le suministra un conjunto de ejemplos que representen de manera adecuada la cuestión o el problema al que se intenta dar solución. La calidad del aprendizaje está determinada por la calidad de la información que se le suministra a la red. La capacidad de aprendizaje es una de las características distintivas que poseen las RNAs dotándola de una capacidad de adaptación ante nuevos eventos cada vez que se le proporcione los datos suficientes y adecuados a la red para entrenarse.
3. Flexibilidad: de manera automática la red es capaz de ajustarse a un nuevo entorno sin necesidad de utilizar instrucciones preprogramadas.
4. Cálculo Asociativo: es capaz de realizar una determinada tarea de una manera distribuida y ejecutar múltiples operaciones en paralelo.
5. Robustez y maleabilidad: la red tiene la capacidad para manejar una gran variedad de datos provenientes de naturaleza distinta. Así como también la red puede ser apta en la operación de información inconsistente, difusa, y ruidosa.

Cabe recalcar que las RNAs pueden ser clasificadas por medio de su arquitectura. De esta forma, a su planeación, estructuración y distribución es a lo que se le nombra como arquitectura de una red neuronal. Donde los nodos o neuronas se encuentran conectados por medio de la sinapsis, la conducta de la red es determinada por la estructura de las conexiones sinápticas. El algoritmo que se emplee para el aprendizaje guarda una estrecha relación con la arquitectura que se disponga de la RNA. Las neuronas se encuentran distribuidas dentro de la red formando niveles o capas y conforme a la estructura de las capas se clasifican en:

- Redes Monocapa. En esta arquitectura solo se dispone de una capa de neuronas para realizar el procesamiento.

- Redes Multicapa. La organización de la red neuronal está compuesta de varias capas que son: entrada, oculta(s) y salida. La forma mediante la cual se puede distinguir a que capa pertenece una neurona artificial es verificando la procedencia de las señales que recibe así como la dirección o rumbo hacia donde dirige las señales que emite.

De igual forma las RNAs se pueden clasificar mediante el direccionamiento de las señales. Por ello, la forma en que las neuronas se conectan entre ellas se encuentra vinculado al modo en como fluye la información y se clasifican en:

- Redes Unidireccionales (feedforward). El flujo de información se desplaza en un solo sentido, partiendo de las neuronas de entrada en dirección a las de salida.
- Redes Recurrentes (feedback). La información fluye en cualquier sentido entre las capas. Las salidas de las neuronas pueden ser entradas de otras neuronas que estén situadas en la misma capa o en una capa anterior.

De igual manera, es posible clasificar las RNAs mediante el aprendizaje.

- Aprendizaje supervisado en RNAs. Este tipo de aprendizaje consiste en que a la red se le muestran un grupo de patrones de entrada y otro conjunto de patrones de salida ideales u objetivos que son deseables para las entradas, donde los pesos son modificados en función del error cometido y de forma iterativa se ajustan dichos pesos hasta lograr obtener la salida que se aspira lograr.
- Aprendizaje no supervisado en RNAs. Este aprendizaje se caracteriza por no presentar un conjunto de patrones deseados, en este caso únicamente se muestran un grupo de patrones de entrada, y la red por si misma tendrá la misión de encontrar los rasgos característicos de los diferentes patrones y clasificarlos de acuerdo a los atributos que guardan en común los patrones.

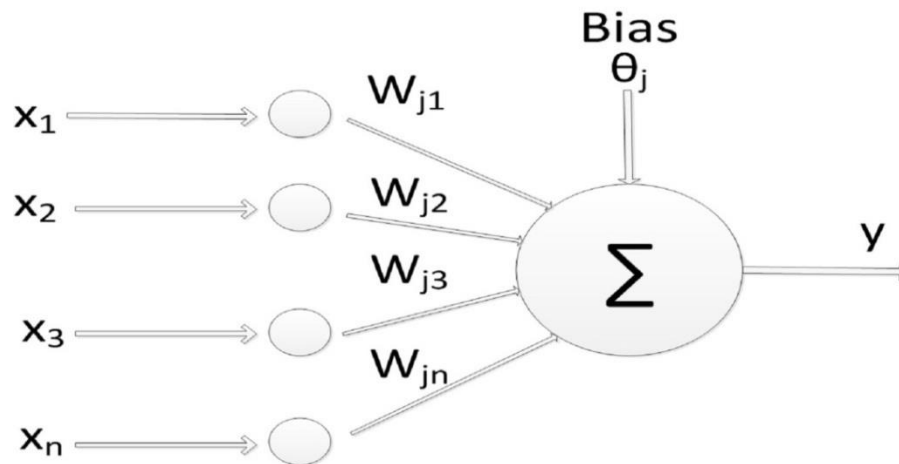
Existe una gran diversidad de tipos de redes neuronales artificiales. A continuación se enlistan algunos de los modelos de RNAs más populares.

1. Perceptrón Simple. Tomando como fundamentos los prototipos de neuronas biológicas de McCulloch y Pitts y los principios del aprendizaje Hebbiano, Frank Rosenblatt concibió en 1957 el algoritmo del perceptrón. En el perceptrón la información fluye desde la capa de entrada hacia la capa de salida, es decir, las



conexiones son en un sólo sentido, y utiliza el aprendizaje supervisado.

Figura 4. 10: Arquitectura de un Perceptrón



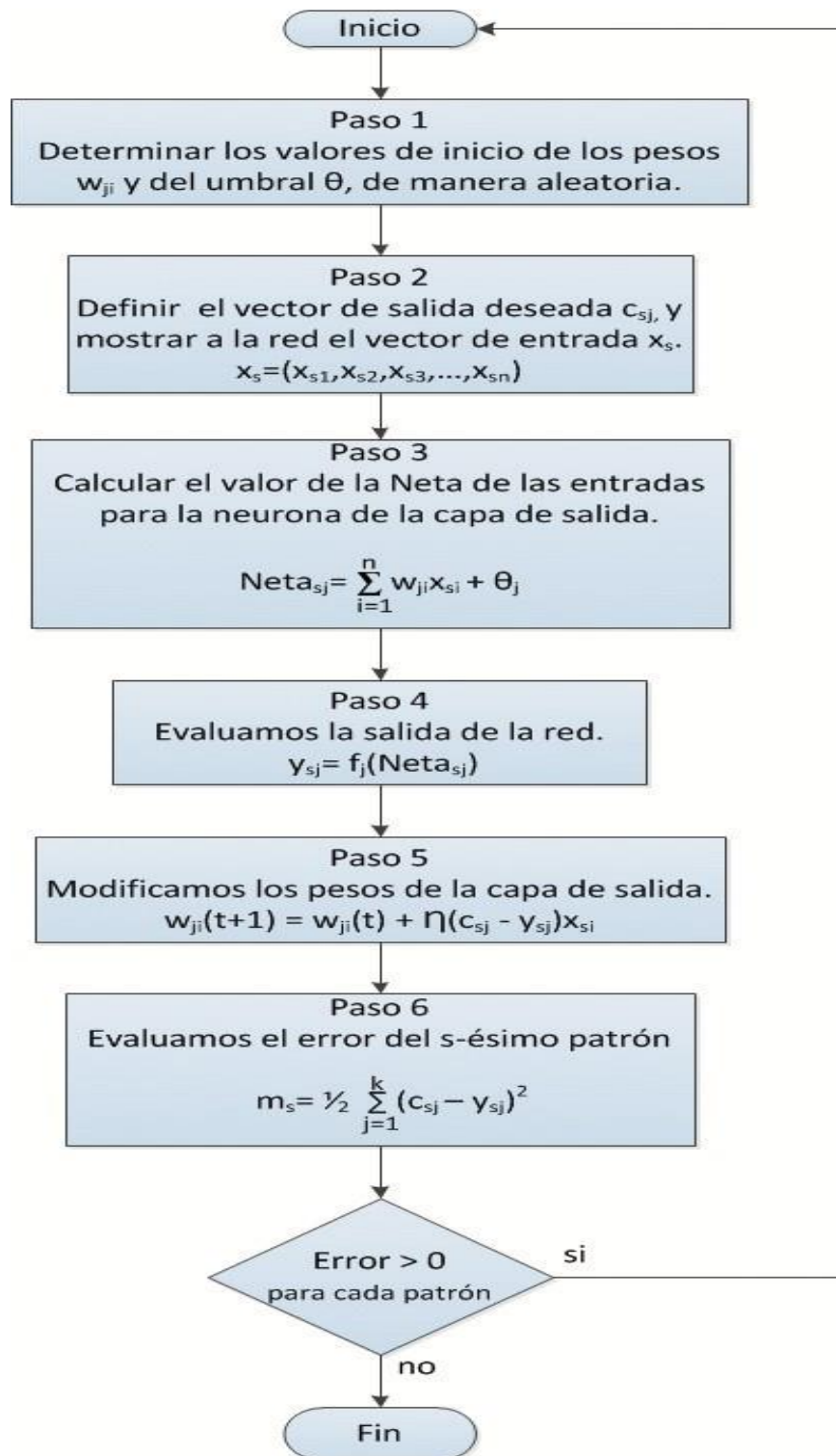
Fuente: Elaboración propia.

Su estructura consta de dos capas de neuronas, como consecuencia de que en la capa de entrada no se efectúa ningún tipo de procesamiento a la señal introducida, y debido a que el procesamiento solo se realiza en la capa de salida se le considera como una red monocapa. La neurona de salida se encuentra totalmente conectada con las neuronas de entrada como se puede apreciar en la figura 4.10. La salida del perceptrón son valores binarios o bipolares. En el caso del perceptrón los valores de las entradas son ponderados, y cuando la salida presente el valor de 1, se dice que el perceptrón ha sido activado. Si la suma es superior a un umbral  $\theta$ , entonces se activa el perceptrón, en el caso contrario, donde la suma no supere el umbral el perceptrón presenta como salida de  $f(x)$  el valor de 0 o de -1, según sea el caso. En el perceptrón se calcula su salida basándose en el patrón actual de entrada (un vector de valores  $x_1 \dots x_n$ ) y por los pesos (un vector de valores  $w_1 \dots w_n$ ), para luego actualizar el valor de los pesos, en base al grado que difiere la salida actual  $Y$  de la salida deseada  $C$ .

$$Y_{pj} = f\left(\sum_{i=1}^n W_{ji}X_{pi}\right) \text{ donde } f(\theta) = \begin{cases} 1, & \text{si } \theta > 0 \\ 0, & \text{si } \theta < 0 \end{cases} \quad (4.10)$$

Los pasos del algoritmo de aprendizaje del perceptrón se encuentran representados en la figura 4.10.

Figura 4. 11: Algoritmo de aprendizaje del Perceptrón.



Fuente: Elaboración propia.

En la tabla 4.5 se enlistan los símbolos utilizados y su correspondiente descripción.

Tabla 4. 5: Simbología del aprendizaje del Perceptrón.

Valor del peso del vínculo o conexión proveniente de la i-ésima capa de entrada hacia la j-ésima capa de salida o de procesamiento.	$W_{ji}$
La cantidad de neuronas en la capa de entrada.	$N$
Valor del umbral perteneciente a la j-ésima neurona de la capa de salida.	$\theta_j$
Tasa de aprendizaje.	$H$
Estimación del valor de error en el s-ésimo patrón de aprendizaje.	$m_s$
Para la neurona de la capa de salida valor de salida deseado.	$C_{sj}$
Valor actual de la j-ésima capa de salida.	$Y_{sj}$
Valor actual del cálculo de la Neta de la capa de salida o procesamiento.	$Net_{sj}$
Valor de la i-ésima entrada relacionada al s-ésimo patrón.	$X_{si}$
Número de iteración en el cálculo del error.	$K$

Fuente: Elaboración propia.

2. Perceptrón Multicapa (MLP). Rosenblatt demostró que el modelo del perceptrón que propuso es capaz de converger las funciones que son linealmente separables, no así aquellas que no son linealmente separables. Esta limitante del perceptrón simple fue demostrada por Marvin Minsky y Seymour Papert en su libro Perceptrons (1969), donde manifestaron que un perceptrón no tiene la capacidad de aprender una función como la XOR. Como respuesta a lo anteriormente descrito, surge como una propuesta para subsanar aquellos limitantes que le confieren al perceptrón simple, el Perceptrón Multicapa (MLP). Convirtiéndose este modelo en uno de los más populares en la actualidad. El MLP se fundamenta en el perceptrón simple, siendo una de las diferencias más notable de entre ellos la adición de una o más capas de elementos de procesamiento (capa oculta), en oposición a solo una capa del perceptrón simple. La función de activación del perceptrón simple en la capa de procesamiento es una combinación lineal de las entradas y de sus correspondientes pesos, así como de un dispositivo de umbral, este último componente permite activar o no al elemento de procesamiento, mediante una función de escalón.

La inclusión de una o más capas a la estructura del perceptrón no tendría un mayor impacto, si esta no hubiera sufrido cambios a su vez, en la función de activación, pasando de una función lineal a una no lineal como lo es la función sigmoide o la

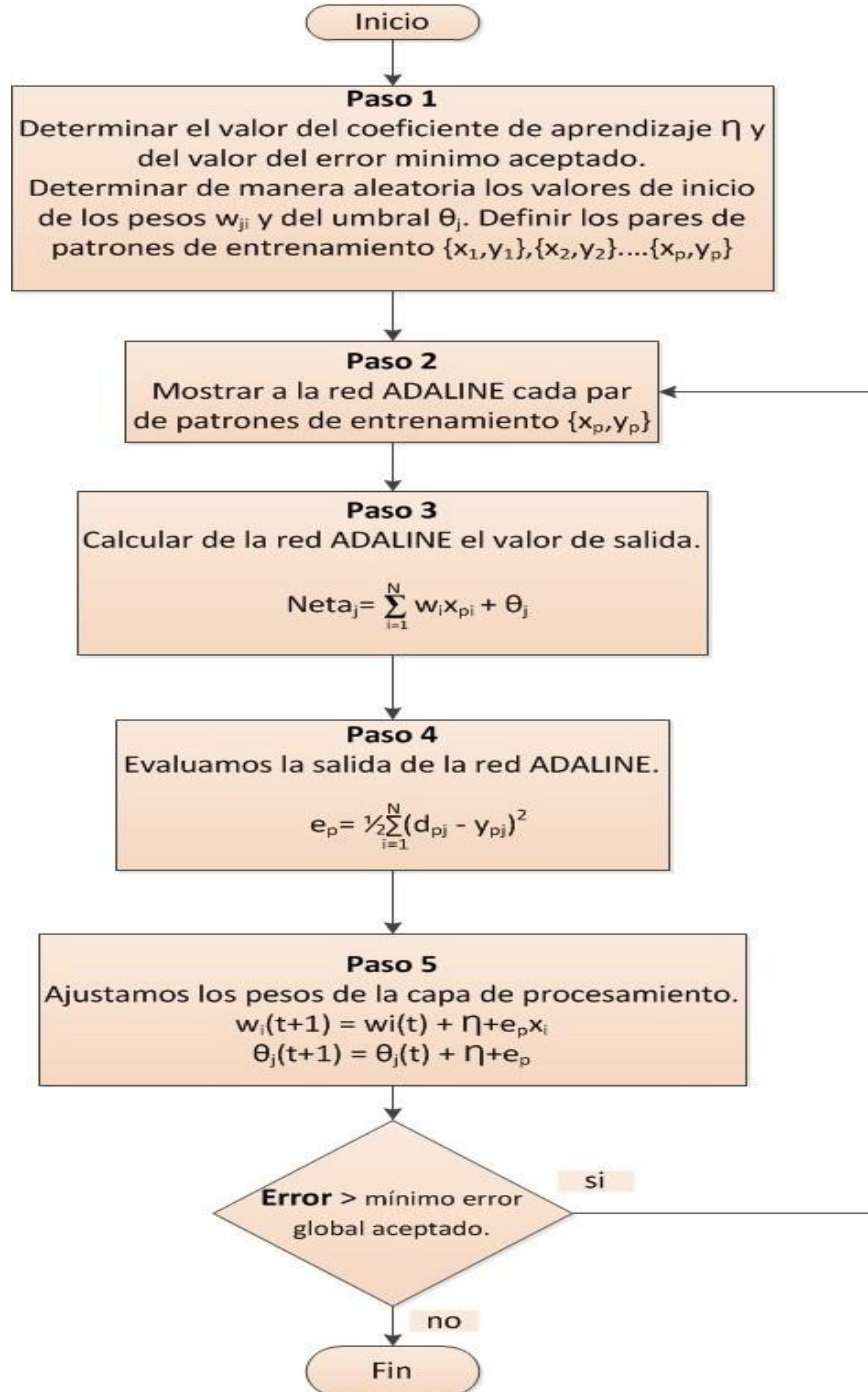
tangente hiperbólica, siendo estas de las más populares en los modelos MLP. El hecho de introducir una función de carácter no lineal en los modelos neuronales así como también el ser apto en el manejo de valores reales, lo dota de la capacidad de poder manejar información no lineal, lo cual es habitual en los diferentes ambientes o situaciones del mundo real. Por consiguiente, a lo anteriormente descrito sobre el MLP se podría decir que estas incorporaciones al modelo le confieren la capacidad de ser un modelador de funciones universal.

3. Perceptrón Multicapa Generalizado (GMLP). Cabe destacar que el modelo utilizado por el algoritmo FS-EPNet es el Perceptrón Multicapa Generalizado (GMLP), este modelo se configura mediante una capa de entrada, una capa de salida y un conjunto de neuronas ocultas que se pueden interconectar entre ellas, así como también se admiten conexiones entre cualquier neurona anterior a una delantera. De esta forma se pueden modelar RNA con neuronas de entrada conectadas a neuronas de salida, además de que es factible conectar neuronas de salida con neuronas de salida. Dada la flexibilidad que se le proporciona a una RNA del tipo GMLP se pueden edificar redes muy complejas pero a la vez muy compactas que contengan un número menor de neuronas, a diferencia de las basadas en un típico MLP con una o varias capas ocultas.

4. ADALINA. Inicialmente el modelo ADALINA era la abreviatura del término Neurona Lineal Adaptativa (Adaptive Neural Network), para después convertirse en la abreviatura de la expresión Elemento Lineal Adaptativo (Adaptive Linear Element). Como consecuencia de la mala reputación que causó en los sistemas neuronales en la década de los sesenta el trabajo de Minsky y Papert, por ello optaron por el cambio en el término. La función de transferencia de tipo lineal, el algoritmo de aprendizaje, así como la tarea que desempeña el elemento umbral (bias), son las principales diferencias entre el modelo ADALINA y el perceptrón. El modelo MADALINE (Múltiples ADALINA) es el resultado de la combinación de múltiples sistemas ADALINA, donde la red MADALINE puede tener varias neuronas de salida a diferencia de la red ADALINA que solo posee una neurona de salida. Aprendizaje de la red ADALINA. La regla delta es el mecanismo de aprendizaje utilizado en la red ADALINA fue desarrollado por Widrow y Hoff,

también llamado el método o regla del mínimo error cuadrado medio (Least Mean Square (LMS)), es una de las reglas de aprendizaje más popular.

Figura 4. 12: Algoritmo de aprendizaje de la red ADALINA.



Fuente: Elaboración propia.

- Mediante el método LMS se intenta encontrar el vector  $W$  de pesos deseados, que se encuentra relacionado con un conjunto de vectores de entradas  $(x_1, x_2, \dots, x_p)$ ,

y el correspondiente valor de salida deseado  $d_p$ , donde se compara el valor de salida con la respuesta deseada. Si la diferencia es igual a 0, no existe aprendizaje, si ocurre lo contrario, los pesos se ajustan para reducir esta diferencia. En otras palabras lo que pretende realizar el algoritmo, es la modificación o ajuste en los valores de los pesos cada vez que se produce una combinación de entrada-salida con el objeto de reducir la diferencia entre la salida deseada y la actual, dicho cálculo se realiza para cada patrón. Apoyándose en el método de descenso de gradiente, se obtiene el criterio mediante el cual se ajustan los pesos, hasta lograr el encontrar un mínimo de una función. La forma en que funciona, es comenzando con una solución inicial estimada y se toma el gradiente de la función en ese punto. Se da un paso de la solución en el sentido negativo de la pendiente y se repite el proceso.

- $d_i$  = Para el  $i$ -ésimo patrón, valor de salida deseado.
- $y_i$  = Valor de salida actual del  $i$ -ésimo patrón del modelo ADALINE.
- $i$  = El patrón  $i$ -ésimo de la colección de patrones de entrenamiento.

$$E = \frac{1}{2} \left( \sum_{i=1}^i d_i - y_i \right)^2 \quad (4.11)$$

El algoritmo es iterado hasta converger al punto donde el gradiente es igual a cero o cercano a este. En la figura: 4.12 se muestran los pasos del algoritmo de entrenamiento de una red ADALINA.

#### 4.4.1.4 Aprendizaje mediante el algoritmo de retropropagación de errores

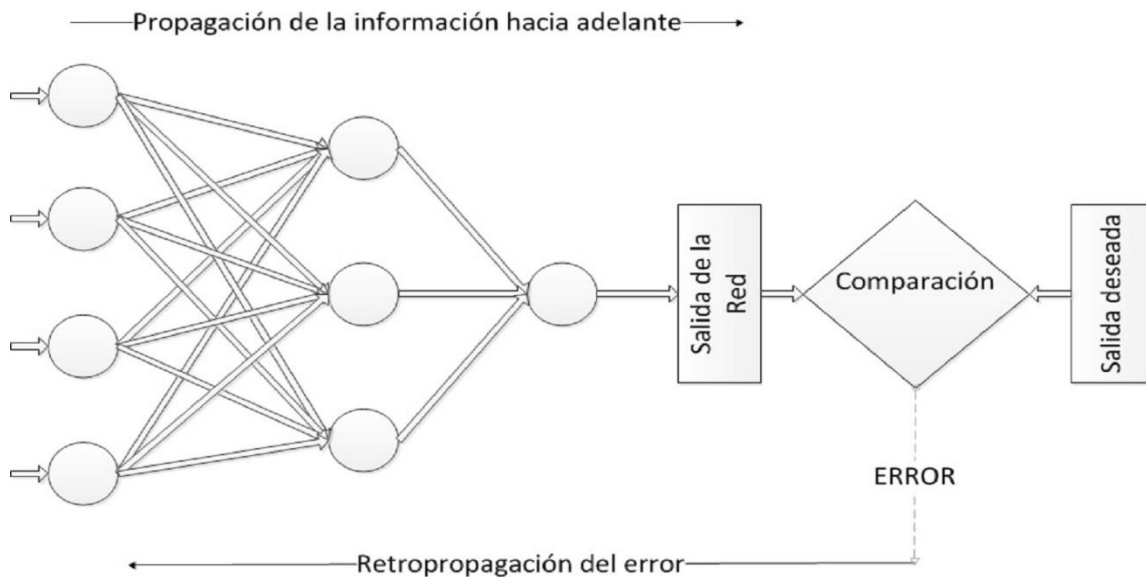
La Regla Delta Generalizada o algoritmo Backpropagation es una generalización del algoritmo LMS, mediante un aprendizaje supervisado tiene como propósito desarrollar un método capaz de proporcionar al algoritmo la capacidad de extender el aprendizaje a una red neuronal multicapa de tipo feedforward o de propagación hacia delante, de manera que la red pueda ser entrenada para capturar de manera adecuada al conjunto de pares de patrones de entrada-salida.

Para alcanzar los valores óptimos del conjunto de pesos, se utiliza el método del gradiente descendente, mediante una exploración a lo largo de la superficie del

error. La medición del error se calcula mediante la diferencia cuadrática entre salida real obtenida en la capa de salida de la red y la salida deseada, a partir de un patrón de entrada proveniente del conjunto del par de patrones de entrada-salida. La salida actual se calcula utilizando la configuración actual de los pesos en todas las capas.

Para lograr obtener los pesos óptimos es necesario ajustarlos, mediante una búsqueda del descenso del gradiente a través de la superficie total del error. Un aprendizaje ideal sería: que en un número reducido de iteraciones se lograra el ajuste óptimo en los pesos de la red para cada presentación de un patrón par de entrada-salida, pero esto no es posible, ya que no es posible calcular el error en la capa oculta.

Figura 4. 13: Algoritmo de Retropropagación.



Fuente: Elaboración propia.

El algoritmo de retropropagación es un proceso de carácter supervisado, donde el aprendizaje se produce en cada ciclo o época (cada vez que a la red se le presenta un nuevo patrón de entrada) comenzando con la entrada de la información en la capa de entrada hacia la capa intermedia, y luego ser transformada mediante una función de activación para fluir hacia adelante hasta llegar a la capa de salida, y posteriormente propagar el error hacia atrás con el fin de ajustar los pesos. Este algoritmo se puede dividir en dos fases, como se aprecia en la figura 4.13.

Fase de aprendizaje hacia adelante. Paso 1. Determinar los valores iniciales de los pesos y de los umbrales aleatoriamente. A la capa de entrada se le presenta un conjunto de patrones de entrada.

$$X_p = (X_{p1}, X_{p2}, \dots, X_{pN})$$

La información fluye de la entrada hacia la capa intermedia y se pondera a través de los pesos  $W_{ji}$ , para luego realizar el cálculo correspondiente de la Entrada Neta. En el Perceptrón básico el valor de  $\theta$  se utilizaba como un umbral de disparo que determinaba el estado inhibitorio o excitatorio en la salida del elemento de procesamiento, a diferencia del anterior uso, en el MLP este elemento se utiliza como un peso adicional que tiene como propósito ayudar a la red neuronal en el proceso de aprendizaje de los patrones. El cálculo de la Neta o de la entrada total se calcula mediante la siguiente formula:

$$Net_{pj} = \sum_{i=1}^N W_{ji}(t)X_{pi}(t) + \theta_j \quad (4.12)$$

Donde la (t) indica la iteración.

Paso 2. Se calcula el valor de salida  $y_j^p$  del elemento de procesamiento j de la capa oculta mediante la aplicación de una función de activación  $f_j$  a su correspondiente entrada neta.

$$Y_{pj} = f_j(Net_{pj}) \quad (4.13)$$

Paso 3. Los valores de salida de la capa oculta se convierten en las señales de entrada de la capa de salida, para luego ser utilizados en el cálculo del valor de la entrada neta de la k-ésima neurona de la capa de salida.

$$Net_{pk} = \sum_{j=1}^H W_{kj}(t)X_{pj}(t) + \theta_k \quad (4.14)$$

Paso 4. Este es el último paso de la propagación hacia adelante de la señal. Proveniente de la señal de estimulación de entrada, se calcula el valor de salida de la red neuronal por medio de la función de activación de la k-ésima neurona de la capa de salida.



$$Y_{pk} = f_k(Net_{pk}) \quad (4.15)$$

Fase de aprendizaje hacia atrás.

Paso 5. El aprendizaje con el algoritmo de retropropagación de errores tiene como finalidad minimizar el error en la comparación de la salida producida por la red y de la salida deseada, donde la salida deseada es un conjunto de patrones determinados por un supervisor o usuario, la cual es contrastada con la señal de salida. La colección de datos deseados es:

$$d_p = (d_{p1}, d_{p2}, \dots, d_{pM})$$

Donde  $d_p$  es el p-ésimo valor de salida deseado. Retomando la fórmula de la regla delta utilizada en la red ADALINA, donde  $e_p$  se definía como el error de solamente un elemento de procesamiento:

$$e_{pk} = (d_{pk} - y_{pk}) \quad (4.16)$$

A diferencia de la regla delta, en el algoritmo de retropropagación se toma en cuenta todas las neuronas, sumando la contribución al error de cada una de ellas, a partir del aporte de cada uno de los elementos de procesamiento obtenemos un error global que se define como:

$$E = \sum_{p=1}^q E_p \quad (4.17)$$

Donde q es la cantidad de patrones de entrenamiento.

El patrón p es mostrado ante la salida deseada  $d_{pk}$  de la neurona k. Se debe considerar que  $E_p$  es función de todos los umbrales o bias y de los pesos sinápticos de la red, donde la derivada parcial de  $E_p$  es igual al vector gradiente de  $E_p$  correspondiente a cada umbral y a cada uno de los pesos sinápticos de la red, en el cual el gradiente es la base para realizar el ajuste de los pesos.

Se calcula el término de error de la capa de salida.

$$\delta_{pk} = -\frac{\partial E_p}{\partial(Net_{pk})} = (d_{pk} - y_{pk}) \frac{\partial f_k(Net_{pk})}{\partial(Net_{pk})} \quad (4.18)$$

Paso 6. Se determina el error en la capa oculta.

$$\delta_{pj} = f'_j(Net_{pj}) \sum_{k=1}^M \delta_{pk} W_{kj} \quad (4.19)$$

Paso 7. Se ajustan cada uno de los bias y de los pesos pertenecientes a las conexiones entre la capa de salida y de la capa oculta.

$$W_{kj}(t + 1) = W_{kj}(t) + \eta \delta_{pk} y_{pj} \quad (4.20)$$

$$\theta_{pk}(t + 1) = \theta_{pk}(t) + \eta \delta_{pk} \quad (4.21)$$

Paso 8. Se actualizan cada uno de los pesos y de los bias de las conexiones que ligan a los elementos de procesamiento entre la capa oculta y la capa de entrada. Con una tasa de aprendizaje  $\eta$ .

$$W_{ji}(t + 1) = W_{ji}(t) + \eta \delta_{pj} x_{pi} \quad (4.22)$$

$$\theta_{pj}(t + 1) = \theta_{pj}(t) + \eta \delta_{pj} \quad (4.23)$$

Paso 9. Se calcula la función de error global, la cuál es la que se intenta disminuir para cada patrón  $p$ -ésimo es la siguiente:

$$E_p = \frac{1}{2} \sum_{k=1}^M (d_{pk} - y_{pk})^2 \quad (4.24)$$

El error global muestra la capacidad adaptativa que tiene la red, y en base a ello se puede establecer si la red está aprendiendo de manera correcta. Además se puede evaluar la condición de terminación o de error mínimo propuesto, en base al cálculo de este error global, cabe aclarar que este término no incide en la actualización de los pesos de la red. Una de las principales desventajas del algoritmo de retropropagación es que el proceso de convergencia puede ser muy tardado, entre otra de las desventajas es la posibilidad de caer en sobre-aprendizaje o sobre-ajuste a los datos de ejemplo suministrados, impactando de forma negativa en la

capacidad de generalización de la red.

Un rasgo distintivo del algoritmo FS-EPNet es la proclividad a impedir el sobre-entrenamiento de en una red, dado que cada red que es seleccionada, se entrena parcialmente con un número reducido de épocas de entrenamiento, además de la aplicación continua de mutaciones en la red, que se manifiesta en una red con modificaciones sistemáticas. Lo anterior, se ve reflejado en las constantes alteraciones topológicas de la red, aunado a ello, FS-EPNet emplea el algoritmo de Recocido Simulado o Simulated Annealing (SA), para que la red escape de los mínimos locales (Plagiannakos et al., 1999; Sexton et al., 1999) en los cuales pudiera quedar atrapado. Tomando en cuenta estas características del algoritmo FS-EPNet, se puede inferir que existen pocas probabilidades de que una red se sobreentrene bajo este esquema.

Cabe destacar, que existen varias técnicas para evitar el sobre-entrenamiento (overfitting), de entre ellas se destaca la herramienta denominada parada temprana o *early-stopping*, que se emplea durante la fase de entrenamiento, donde el funcionamiento del conjunto de validación es supervisado. En la etapa de entrenamiento inicial de la red, el error cometido por esta tiende a disminuir en ambos conjuntos, tanto en el conjunto de entrenamiento como en el conjunto de validación, sin embargo, cuando la red comienza a sobre-ajustarse a los datos, usualmente el error cometido por la red en el conjunto de validación tiende a aumentar.

## **Capítulo 5. Algoritmos y técnicas utilizadas**

### **5.1 Transformada de Hilbert-Huang y descomposición empírica en modos**

#### **5.1.1 Introducción**

Se denomina transformada de Hilbert-Huang, a la mezcla de la técnica de Descomposición Empírica en Modos (DEM) y al análisis espectral de Hilbert. Los métodos Transformada de Hilbert-Huang (THH) y DEM en un principio se utilizaron para el análisis de las olas marítimas, fue propuesto por Huang et al. (1998), se ha demostrado de forma experimental su eficiencia en diversas áreas científicas (Huang y Wu, 2008; Moreno, 2006). Este método matemático en primera instancia consta en descomponer una señal en un conjunto determinado de las denominadas Funciones de Modo Intrínseco (FMIs), es decir mediante un algoritmo adaptable y guiado por los datos, analiza y separa de forma local una señal o ST de forma gradual extrae desde la oscilación más rápida hasta la más lenta, dicho procedimiento lo realiza mediante la técnica llamada DEM para posteriormente aplicar a cada una de ellas la transformada de Hilbert THH.

#### **5.1.2 Teoría de la descomposición empírica en modos**

A diferencia de la transformada de Fourier, la DEM-THH es una técnica para analizar procesos reales o naturales que en su mayoría son señales no lineales y no estacionarias, esta capacidad le confiere una utilidad invaluable para el análisis y estudio de las ST económicas y financieras. El corazón de esta técnica es la DEM que consiste en un proceso de eliminación de ruido de una señal para adquirir los componentes oscilantes, dicho procedimiento es denominado algoritmo de cernido o tamizado. Como anteriormente se mencionó, los componentes que son adquiridos mediante la filtración de la señal original se les nombra como FMIs, dicho apelativo es utilizado porque representa el modo de oscilación embebido en los datos.

Para que una señal depurada sea considerada una FMI debe de cumplir con dos requisitos: a) el total de extremos y la cantidad de cruces por cero debe de ser iguales o distinto a lo sumo por uno, b) El promedio o media de la envolvente determinado por el mínimo local y máximo local, en cualquier lugar del espacio es

igual a cero. La FMI al cumplir estas condiciones asegura que esta solo implicada en un sólo modo de oscilación, dando como resultado que no se admitan ondas complejas incrustadas en la señal. El proceso de cernido es descrito a continuación. El proceso de filtrado o cernido se le aplica a la señal o ST para obtener las FMIs de forma iterativa por medio de los siguientes pasos:

- 1) Determinar el total de máximos y mínimos locales de la señal de entrada o ST  $x(t)$ .
- 2) Obtener la envolvente superior  $e_{sup}(t)$  mediante la interpolación de los máximos locales. Obtener la envolvente inferior  $e_{inf}(t)$  mediante la interpolación de los mínimos locales.
- 3) Calcular el promedio de la envolvente (i-ésimo) superior y de la envolvente inferior.

$$m_i(t) = \frac{e_{sup} + e_{inf}(t)}{2} \quad (5.1)$$

- 4) Se obtiene el componente que aspira a ser FMI mediante la eliminación de la media a la señal de entrada.

$$h_1 = x(t) - m_1 \quad (5.2)$$

El componente residual  $h_1$  que se obtiene es un aspirante a ser una FMI y como tal debe tratarse como un prototipo de una FMI verificando si cumple con los requisitos a convertirse en una FMI, en caso de que  $h_1$  no alcance a ser una FMI se repiten los pasos del (1) al (4) en el lugar de  $x(t)$  se toma al elemento  $h_1$  como la señal de entrada.

$$h_{11} = h_1 - m_{11} \quad (5.3)$$

El proceso de cernido se repite n-veces hasta que  $h_{1j}$  logre cumplir con la condición de ser una FMI.

$$h_{1j} = h_{1(j-1)} - m_{1j} \quad (5.4)$$

A partir de los datos se ha obtenido el primer componente FMI el cual contiene la amplitud más fina es decir el componente con el periodo más corto de la ST y se denota:

$$c_1 = h_{1j} \quad (5.5)$$

Las reglas o criterios de paro en el proceso de cernido determinan el momento en el cual el proceso ha sido completado y se ha adquirido una nueva FMI, es decir estos criterios marcan el número de iteraciones para obtener una FMI. Por la vía de la manipulación o cambio en el criterio de paro del proceso de cernido es posible adquirir diferentes cantidades de FMIs de la señal original o ST. El criterio más común es el propuesto por Huang et al. (1998), en donde se evalúa la semejanza entre dos elementos residuales (prototipos de FMI) de iteraciones sucesivas, examinando la cercanía entre los dos componentes midiendo la desviación que hay entre ellas por medio de la siguiente definición: la diferencia cuadrática  $DC_j$  entre dos prototipos de FMI sucesivos  $h_{j-1}(t)$  y  $h_j(t)$ :

$$DC_j = \sum_{t=1}^T \left[ \frac{(h_{1(j-1)}(t) - h_{1j}(t))^2}{h_{1(j-1)}^2(t)} \right] < U \quad (5.6)$$

Para determinar si un elemento residual lograr ser una FMI se utiliza un coeficiente denominado umbral  $U$ , el cual es fijado de forma arbitraria, si el valor de  $U$  es muy pequeño entonces obtendremos un mayor cantidad de FMIs, por el contrario al determinar un valor muy alto de  $U$  se obtiene un número menor de FMIs. Una vez que se ha obtenido el componente  $c_1$  se separa del resto de los datos:

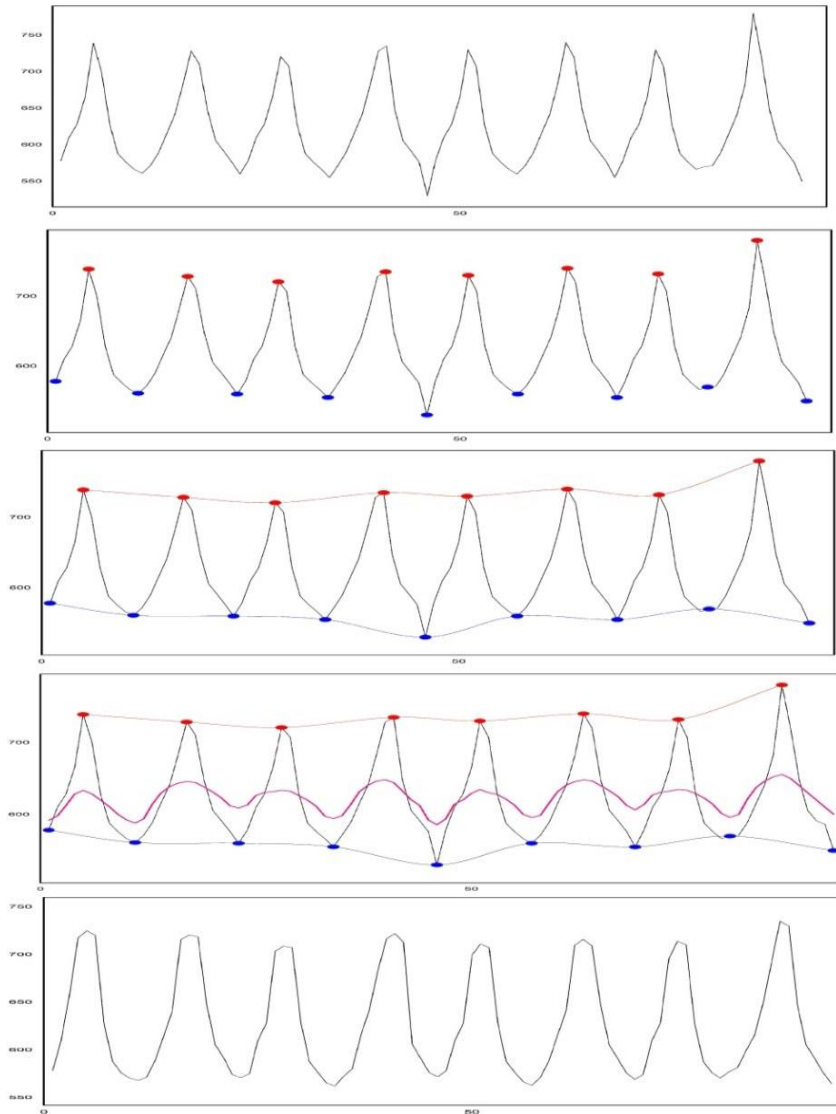
$$r_1 = x(t) - c_1 \quad (5.7)$$

Ya que seguramente el componente  $r_1$  aun incluye largos periodos de variación en los datos, es decir la ST todavía incluye información que se puede aprovechar, se debe volver a repetir el proceso de cernido en donde  $r_1$  será el dato de entrada y se busca obtener un nuevo componente residual  $r_n$  (Moreno, 2006). El método se repite  $r_2 = r_1 - c_2, \dots, r_n = r_{n-1} - c_n$ , hasta que el componente  $r_n$  se convierte en una función monótona o una constante donde no se pueden hallar más FMIs. El producto que se obtiene al finalizar la experimentación con DEM y mediante el cual se puede volver a recobrar la señal de entrada o ST  $x(t)$  es:

$$x(t) = \sum_{j=1}^j IMF_j(t) + r_n(t) \quad (5.8)$$

Donde  $j$  es la suma total de FMIs y  $r_n(t)$  es el elemento residual final o la función monótona que indica la terminación del procedimiento. En la figura 5.1 se presenta un breve esquema del proceso de cernido de una ST.

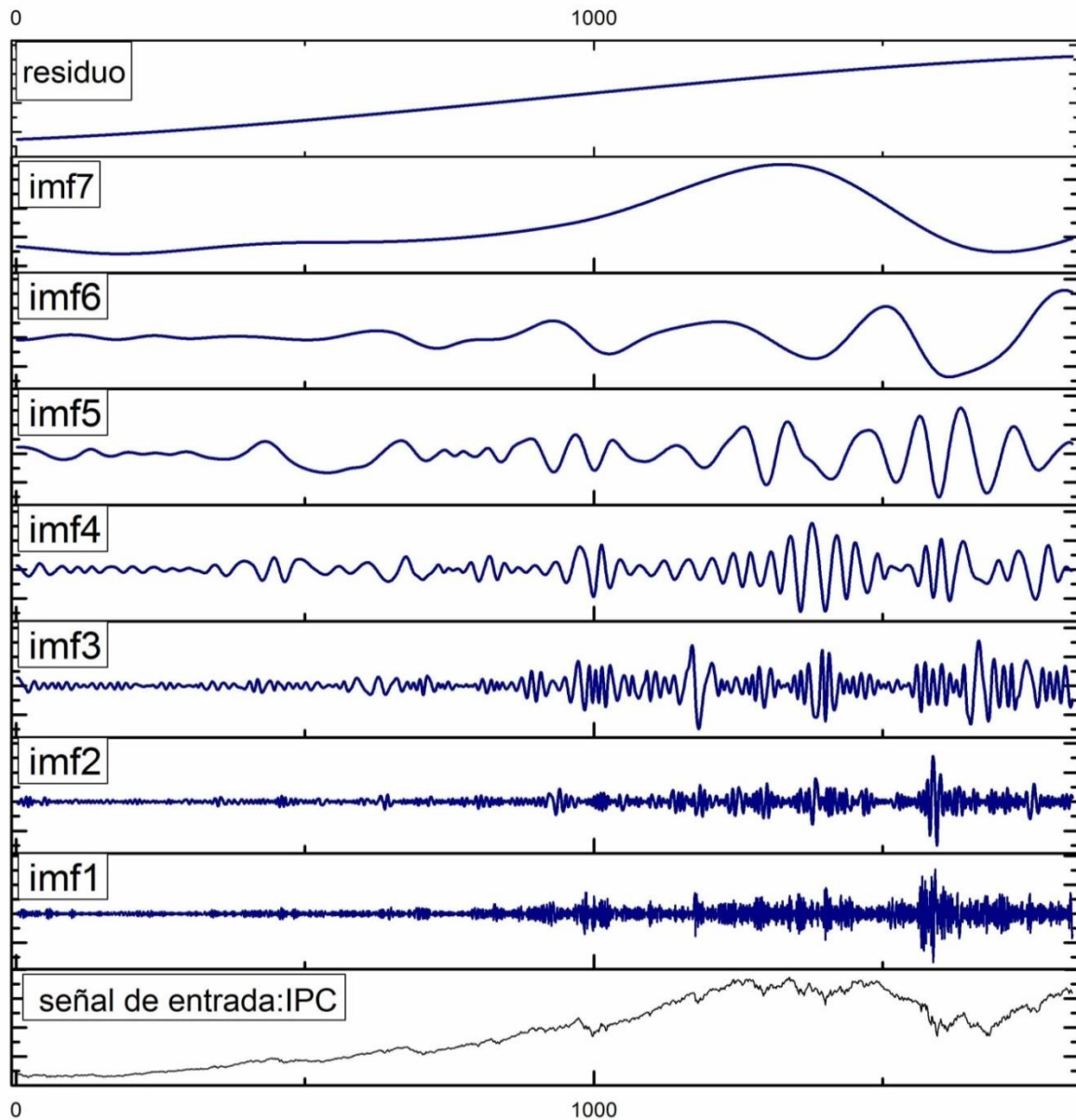
Figura 5. 1: Boceto del proceso de cernido.



Fuente: Elaboración propia.

Varios factores inciden en la cantidad de FMIs que se adquieren al finalizar la experimentación como son: el criterio de paro para obtener una FMI, la longitud de datos en la ST y la complejidad de los datos que se analizan entre otras causas. Se debe tomar en cuenta que no todas las FMIs tienen el mismo valor de utilidad. La figura 5.2 representa la ST y las FMIs que se obtuvieron por descomposición a través de DEM, los datos que son sujetos al proceso de cernido son del IPC.

Figura 5. 2: IPC del 17/09/2001 al 13/03/2009 y sus descomposiciones.



Fuente: Elaboración propia.

En la investigación realizada por Rilling et al. (2003) se utiliza un umbral  $U = 0.05$  para determinar si un prototipo de una FMI alcanza a ser una FMI, de la cual se retoma como guía para las experimentaciones realizadas en la presente investigación. En la figura 5.2 se muestra claramente como las oscilaciones de las FMIs obtenidas comienzan con una alta frecuencia y conforme avanza el proceso esta va decayendo a frecuencias más bajas hasta llegar a obtener la señal monótona que marca el fin del procedimiento. Una vez que se han adquirido los



componentes FMI y el residuo por medio de la DEM, se procede con la segunda parte de la metodología, mediante un análisis armónico fundamentado en la transformada de Hilbert se obtiene la amplitud y la frecuencia instantánea en función del tiempo de que cada una de las FMIs (Blázquez et al., 2010).

## **5.2 Algoritmos evolutivos**

### **5.2.1 Antecedentes**

La teoría evolutiva de Darwin; afirma que todos los seres vivos se encuentran relacionados, que la descendencia parte de un ancestro en común, es decir, que tanto los árboles, las aves, los peces, los vegetales, así como los humanos, y en general toda la vida guarda una conexión. El principio de la selección natural es un proceso en el cuál los organismos se transforman en el transcurso del tiempo, y esto da como resultado los cambios en los rasgos físicos o en el funcionamiento. Estos cambios permiten la posibilidad que un individuo se adapte de manera más adecuada al entorno que le rodea, así como lo dota de una mayor capacidad de sobrevivir y por lo tanto una mayor probabilidad de reproducirse y obtener descendientes, estos descendientes a su vez se benefician de aquellos rasgos superiores que le fueron heredados.

Con el paso del tiempo estas modificaciones serán extensivas a los demás individuos de la población. La microevolución enuncia que con el devenir de varias generaciones es posible que la selección natural altere de forma reducida alguna especie, como por ejemplo ocasionando que una población de individuos cambie el tamaño de sus dientes o el color de su piel. Con el cumplimiento del tiempo necesario y con la cantidad de alteraciones adecuadas, la selección natural es capaz de crear nuevas especies, como lo es transformar a los dinosaurios en aves, a los lobos en perros domésticos y al simio en humano. Se denominan mutaciones a aquellas alteraciones físicas y de funcionamiento que hacen posible que la selección natural pueda acontecer a nivel genético. Esas mutaciones pueden ser causadas por factores exógenos (factores ambientales como la luz solar, la radiación, etc.) o por factores endógenos (errores durante la replicación del ADN pueden conducir a cambios genéticos). Generalmente estas mutaciones suelen

dañar o ser indiferentes, pero excepcionalmente estas alteraciones resultan favorables al individuo. Si este cambio resulta beneficioso para el individuo lo más probable es que en la siguiente generación sea más frecuente y se propague a la población de la cual forma parte. A raíz de la lectura de la obra “Ensayo sobre el principio de la población” escrita por el clérigo y economista inglés Thomas Robert Malthus que trata sobre demografía, Darwin concluye que en condiciones favorables todas las poblaciones tanto de animales como de plantas tienen una posibilidad similar de incrementar su número a no ser que exista escasez de alimentos, agua y de cualquier otro recurso esencial para sobrevivir, así como de la aparición de enfermedades y de los ataques constantes de los depredadores.

En base a estos factores Darwin determina que cuando los recursos ya no son suficientes los individuos que tienen una menor adaptación en una población son los que tienen una mayor probabilidad de fallecer, y que por el contrario aquellos individuos más adaptados tienen una ventaja selectiva y una posibilidad de subsistencia superior, por consiguiente una preeminencia a transmitir sus características o rasgos a la generación siguiente.

### **5.2.2 Características de los algoritmos evolutivos**

Los Algoritmos Evolutivos (AEs) parten de las ideas de la teoría Darwiniana de la evolución biológica como un proceso de optimización. Los AEs son métodos de resolución de problemas que combinan la búsqueda aleatoria y la búsqueda dirigida, esta característica les proporciona la capacidad de poder alcanzar cualquier zona del espacio de búsqueda. A pesar que los AEs comparten similitudes con la evolución natural, también existen notables diferencias, una de ellos es que mientras la evolución natural es un proceso no guiado y no limitado hacia un punto final, los AEs se diferencian en que requieren determinar una función de aptitud bien definida, esto a su vez se traduce en una búsqueda guiada y que finalizara una vez que se haya alcanzado un número predeterminado de generaciones o se consiga encontrar un individuo lo suficientemente adaptado. Otra diferencia notable es que mientras que en la evolución biológica probablemente se necesiten millones de años para que una especie produzca cambios notables, en la evolución artificial se

requiere converger hacia una solución en un periodo de tiempo mucho menor. Los AEs organizan un proceso de aprendizaje colectivo dentro de una población de individuos, donde cada uno de estos individuos además de representar un punto en el espacio de solución, pueden ser candidatos potenciales a ser la solución a un determinado problema.

Se parte de una población inicializada aleatoriamente de individuos (soluciones candidatas), para luego evolucionar sucesivamente hacia regiones que prometan soluciones más óptimas en el espacio de búsqueda a través de la selección, combinación y mutación, es decir, que por medio de estos tres principios básicos, en el ciclo evolutivo se generan cambios subsecuentes en los individuos que forman parte del conjunto poblacional. El proceso selectivo tiende a favorecer a los individuos de mayor calidad (mejor adaptados) a reproducirse con mayor frecuencia, a diferencia de los individuos con peor calidad que a consecuencia de un menor grado de adaptación sus probabilidades de producir descendencia son mínimas. El proceso de recombinación o cruce permite que la combinación de la información genética de los padres sea transmitido hacia sus descendientes, y es a través de las mutaciones que se insertan cambios o novedades en el material genético de los individuos de la población.

Este sistema evolutivo forma parte de los AEs más populares; Programación Evolutiva (PE) (Fogel, Owens & Walsh, 1966; Saravanan y Fogel, 1995): su operador principal es la mutación y no utiliza el operador de cruzamiento (recombinación), Estrategias Evolutivas (EE) (Schwefel, 1981; Schwefel, 1993): la mutación es el operador primordial y el operador de cruce posee un rol secundario; la solución es codificada directamente por sus valores reales, Algoritmos Genéticos (AGs) (Holland, 1975): el cruzamiento (recombinación) destaca como el operador de búsqueda más importante y la mutación es aplicada esporádicamente, ya que tiene una probabilidad muy baja de ocurrencia y por lo tanto representa el papel de un operador secundario. Usualmente utilizan una representación binaria. Un ejemplo básico de los pasos que sigue un AE se describe a continuación:

- 1) Inicialización:

Generar una población inicial de individuos ya sea aleatoriamente o con información previa.

2) Evaluar la adaptación:

Medir la capacidad del desempeño de cada individuo para determinar el grado de adaptación.

3) Selección:

Selecciona un grupo de individuos de la población, a los cuales se les aplicará el operador de Mutación y/o Cruzamiento.

4) Reproducción:

Cruce (opcional). Realizar el cruzamiento entre los individuos seleccionados para generar nuevos descendientes.

Mutación (probabilístico). En base a un cálculo probabilístico emplear el operador de mutación a los descendientes.

5) Reemplazo:

Substituir la antigua población con los nuevos descendientes.

6) Finalización:

Si se ha cumplido un criterio de terminación, como puede ser encontrar un individuo como solución favorable o se ha alcanzado el número máximo de generaciones que se determinó. En caso contrario, sino no se ha logrado concretar el criterio de terminación ir al paso 2.

Como se puede apreciar los AEs muestran un esquema común en su estructuración, el cuál puede aplicarse a una gran cantidad de problemas diferentes, que favorecen el planteamiento y la ejecución del problema que se pretenda abordar. Sin embargo, aunque en apariencia podría parecer sencilla la implementación de este simple esquema de un algoritmo evolutivo aquí representado en la resolución de un problema, esto no suele ser tan sencillo, ya que existen todavía varios aspectos por solventar para poner en funcionamiento de manera efectiva un Algoritmo Evolutivo (Ashlock, 2010, p. 15):

- ¿Qué estructura de datos aplicar?
- ¿Qué función de desempeño se va a utilizar?
- ¿Qué operadores de cruzamiento y mutación se emplearan?

- ¿Cómo seleccionar a los padres de la población?, ¿Cómo se van a insertar a los descendientes en la población?
- ¿Qué condición de terminación detendrá el algoritmo?

### 5.2.3 Representación de los individuos

Un tema central cuando se emplean AEs para la resolución de un problema, es como representar a los individuos de la población, como establecer un vínculo o una estructura que conecte el entorno del problema real o base que se pretende resolver con el espacio de búsqueda de solución en el cual se desarrollara la evolución del algoritmo. En el caso de la Neuroevolución (NE), una población de codificaciones genéticas de RNAs son evolucionadas con el fin de encontrar una red que resuelva la tarea dada. Cada codificación en la población (un genotipo) se elige a su vez y se decodifica en la correspondiente RNA (un fenotipo).

Esta red se emplea en una tarea y su desempeño es medido a través del tiempo, obteniendo un valor de aptitud para el correspondiente genotipo (Sammut y Webb, 2011). Cabe destacar que en esta etapa, un asunto de gran relevancia es determinar la cantidad de información del sistema que será codificada hacia el genotipo. En los AEs generalmente se utilizan dos métodos de codificación para la búsqueda de arquitecturas de RNAs: por un lado en la codificación directa, todos los rasgos de la topología de la RNA son codificados en el individuo, por el otro lado en la codificación indirecta, una especificación compacta de la RNA es evolucionada. El vocabulario empleado en los AEs es una fusión entre el léxico propio de la biología y de la terminología utilizada en las Ciencias de la Computación. Por ello se enlistan algunas definiciones básicas.

- Genotipo: es la estructura genética de un organismo, incluye su material genético completo, es la información heredable.
- Individuo: en los AEs, un individuo es un elemento singular de la población. También se denomina solución candidata, fenotipo, y en ellos está contenido su cromosoma, cada uno de estos individuos representan una posible solución al problema planteado (un solo punto en el espacio de búsqueda).
- Fenotipo: la expresión funcional del genotipo en un entorno específico. Es la

manifestación física del organismo.

- Cromosoma: es la estructura de datos empleada para almacenar cada individuo de la población en la evolución.
- Función de Aptitud o Desempeño: es una técnica para determinar la calidad de los diferentes individuos de la población durante la evolución.
- Codificación Directa: en esta codificación se establece cada uno de los parámetros de la RNA, en consecuencia el fenotipo es el mismo que el genotipo, por ello no hay necesidad de codificar al fenotipo en su correspondiente genotipo.
- Codificación Indirecta: Este tipo de representación suele tener una confección más elaborada y se suele codificar solamente algunas propiedades de la topología neuronal, lo cual puede impactar significativamente en la longitud del resultado de la codificación expresada en el cromosoma.

#### **5.2.4 Neuroevolución (NE)**

Como ya se ha indicado en el capítulo anterior, la efectividad como técnica en la resolución de problemas de las RNAs ha sido ampliamente demostrado, no obstante, todavía existen varios temas por solucionar, así como en cualquier modelo neuronal, sobre la elección de los parámetros que tienen que determinarse previamente. Por ejemplo, el Perceptrón Multicapa (MLP) en combinación con el algoritmo de aprendizaje de Retropropagación de errores (BP), es uno de los modelos de Red Neuronal Artificial (RNA) que con mayor frecuencia se emplea, en el hay que elegir el número de capas y de neuronas por capa, la tasa de aprendizaje, la inicialización de los pesos y del umbral, las funciones de activación, entre otros parámetros. Pese a que en los últimos tiempos se ha incrementado significativamente la investigación en esta área, y se han logrado destacados avances tanto teóricos como empíricos, el problema de diseñar la arquitectura de una RNA cercana al óptimo para una aplicación sigue sin resolverse (Yao y Liu, 1997b, p. 1). Existe una base teórica y práctica limitada para apoyar la selección sistemática de los parámetros de la RNA a través de todo el proceso de desarrollo y de la fase de entrenamiento de una RNA. A causa de esto los parámetros de la RNA son normalmente establecidos en base a la experiencia previa del modelador

o experto mediante el procedimiento de prueba y de error, el cuál suele consumir mucho tiempo. De esta forma los valores óptimos de los parámetros de la RNA que permitan lograr la mejor calidad de la RNA no se encuentran garantizados (Ortiz-Rodríguez et al., 2013).

El diseño arquitectónico de una RNA se ha pretendido resolver desde diferentes enfoques, pero prácticamente se podrían englobar en tres modelos de aprendizaje estructural: métodos destructivos, métodos constructivos y la aplicación de un Algoritmo Evolutivo (AE) que evolucione a la Red Neuronal Artificial (RNA). A los algoritmos de poda de una RNA se les llama también métodos destructivos, inicialmente se entrena una red de una dimensiones mayores a las que proporcionalmente se requiere, para posteriormente comenzar a eliminar componentes que se estiman innecesarios como son las conexiones y los nodos. La principal virtud de esta metodología, es que usualmente se logran construir redes de tamaño pequeño, dando como resultado redes más rápidas y su formulación puede ser más sencilla de llevar a cabo (Reed y Marks, 1999; Castellano et al., 1997; Yao, 1999). Sin embargo, la principal desventaja a la que se enfrenta este tipo de métodos, es la elección de que parámetro debe ser eliminado, en base al crédito asignado (castigo y premio) al nodo o conexión evaluado en función de la salida obtenida.

En cambio los métodos constructivos pretenden adecuar la estructura neuronal al problema que se trata de resolver, en primera instancia se comienza con una red pequeña que es entrenada hasta alcanzar el punto de mayor capacidad de aprendizaje para esa estructura neuronal inicial, una vez llegado hasta este punto se empiezan a insertar nuevos elementos al sistema neuronal como son; capas, nodos y conexiones, hasta lograr converger a una solución óptima. Sus principales desventajas son la propensión a crear redes muy grandes, a un sobreentrenamiento y sobreajuste en los datos, y a quedar estancadas en mínimos locales (Angeline et al., 1994; Tetko y Villa, 1997; Riley et al., 2010). En ambos métodos se suele utilizar el método de descenso de gradiente, en el cuál el aprendizaje puede quedarse atrapado en un mínimo local, aunado a lo anterior es posible que no se alcance a

encontrar el modelo estructural más adecuado al problema en el espacio de búsqueda de posibles estructuras. Para resolver estos problemas, distintos investigadores han encauzado sus esfuerzos a la tarea de investigar métodos que permitan resolver esta clase de problemas. Es así que en años recientes ha aumentado el interés de utilizar a los AEs, como una posible respuesta que permita superar dichos problemas. A pesar, de que los AEs y las RNAs se consideran como métodos de búsqueda de solución de problemas generales, la evidencia empírica ha demostrado que la combinación de ambas estrategias es una forma efectiva de fortalecer y mejorar el desempeño de la solución sobre el generado por los modelos de manera individual (Yao, 1999; Mizuta et al., 2001; Bullinaria, 2005; Moreno, 2006).

El mejor desempeño mostrado por la combinación de ambos modelos, se debe a que por un lado los AEs realizan una búsqueda que permite alcanzar regiones cercanas al mínimo global y de esta manera evitar los mínimos locales, por el otro lado los algoritmos de entrenamiento de descenso de gradiente necesitan un mayor tiempo para alcanzar la vecindad de una solución óptima, pero cuando la alcanzan lo hace de manera más precisa (Koehn, 1994).

Es decir, que a través de un trabajo colaborativo de ambos modelos, en primera instancia los AEs son capaces de realizar una búsqueda global, ubicando a la RNA de manera aproximada al óptimo global, para posteriormente en base a un algoritmo de entrenamiento como lo es el BP o alguna de sus variantes encontrar de manera más exacta el mínimo local, logrando así alcanzar el punto óptimo de manera eficiente y eficaz (Belew et al., 1990; Renders y Flasse, 1996; Castillo et al., 2001). La Neuroevolución (NE) es un método de aprendizaje de máquina o automático que mediante el empleo de Algoritmos Evolutivos (AEs) se evolucionan RNAs, los AEs son las herramientas encargadas de buscar los parámetros óptimos de la red que maximicen la función objetivo. Este método de aprendizaje también se le conoce con el término Redes Neuronales Artificiales Evolutivas (RNAEs) y ha sido utilizado de manera exitosa en diferentes estudios (Yao., 1999; Bullinaria, 2006; Bullinaria, 2005; Landassuri-Moreno y Bullinaria, 2009a; León-Anaya et al., 2013).



Generalmente la evolución se ha introducido en las RNAs en tres categorías: en el peso de las conexiones, la arquitectura, y en las reglas de aprendizaje. El método de aprendizaje paramétrico más simple y que con mayor frecuencia se aplica es la Neuroevolución Convencional (NEC), esta técnica de aprendizaje neuronal se avoca solamente a la tarea de evolucionar los pesos entre las conexiones neuronales y la arquitectura neuronal se mantiene fija. El proceso de generación de descendencia es realizado a través de los operadores de cruzamiento y mutación, estos operadores pueden ser utilizados de manera individual o en combinación o incluso omitirse alguno de ellos, todo depende del modelo evolutivo empleado. Sin embargo, un hecho importante y reconocido indica, que la elección de una determinada topología de una RNA en particular impacta significativamente el desempeño o la capacidad de resolver un problema, por ello es deseable ser capaz de poder optimizar automáticamente el procedimiento de diseño de una arquitectura para una aplicación en particular (Melin y Castillo, 2005; Floreano et al., 2008).

El mayor inconveniente de evolucionar arquitecturas neuronales es el problema de la permutación (Hancock, 1992; Radcliffe, 1993) o el problema de las convenciones competitivas (Hancock; 1992; Schaffer et al., 1992), este problema se presenta cuando un AE es el encargado de elaborar la arquitectura neuronal a través de su genotipo, puesto que existe la posibilidad de que redes topológicamente diferentes sean funcionalmente equivalentes, y que al recombinarse produzcan descendientes no útiles, es decir, que fenotipos estructuralmente diferentes pueden ser similares en su desempeño, pero su representación genotípica suele ser muy diferente. Ya que este problema se presenta al recombinar dos genotipos, muchos expertos han optado por solo usar el operador de mutación para evolucionar topologías, de igual forma en esta investigación se empleara un algoritmo que no utiliza el operador de cruzamiento como lo es FS-EPNet.

### **5.2.5 Algoritmo Selección de Características de Programación Evolutiva de Redes Neuronales Artificiales (FS-EPNet)**

El algoritmo de Programación Evolutiva de Redes Neuronales Artificiales (EPNet)

fue formulado, diseñado y empleado empíricamente por Yao y Liu (1997b), este ha sido extendido por Landassuri-Moreno y Bullinaria (2009a) y se le ha llamado Selección de Características de Programación Evolutiva de Redes Neuronales Artificiales (FS-EPNet). En la propuesta inicial del algoritmo EPNet se empleaba una composición fija de entradas durante todo el desarrollo de la fase evolutiva, a pesar de que se ha demostrado su efectividad para resolver problemas de predicción (Yao y Liu, 1997a; Yao y Liu, 1997b; Landassuri-Moreno y Bullinaria, 2009b), se realizó una ampliación al método, con la inserción del algoritmo de selección de características, que tiene como precepto fundamental reducir la longitud del sistema con el fin de favorecer el rendimiento del modelo, esta selección de características es usada para optimizar el número de entradas de la red y los retardos o espacios entre las entradas. Los cuales si no se evolucionan, se tienen que conocer a priori, o bien, aplicar un método matemático para determinar el número de entradas, así como la cantidad de rezagos.

Al diseñar una RNA con entradas fijas, se parte del supuesto de que es posible que se inserten entradas innecesarias en la red neuronal, ya que estas pueden contener información redundante e irrelevante, y de igual forma los datos que son sujetos de análisis pueden contener ruido e información sin sentido. De esta forma el objetivo del proceso de selección de características es conservar solo la información adecuada o destacada y desechar aquella información que carezca de importancia y sea considerada recurrente. El algoritmo de selección de características, mediante la identificación de las características más sobresalientes para el aprendizaje, enfoca al algoritmo de aprendizaje en los aspectos más relevantes de los datos para su análisis y la predicción futura (Hall, 1999). Por otra parte, si no se utiliza el algoritmo de selección de características en las entradas, es posible que al tener entradas excesivas, esto se traduzca en una RNA de mayores dimensiones, por ende el tiempo requerido en la fase de entrenamiento será mayor, así como también se introduce ruido innecesario a la RNA que generalmente se manifiesta en un detrimento en la precisión del resultado (Landassuri-Moreno y Bullinaria, 2009a). Por ello, se aspira a que mediante el proceso de selección de características, este algoritmo conduzca a determinar la cantidad óptima de

entradas, rezagos y conexiones, que permita lograr un aprendizaje que converja con una mayor rapidez, de igual forma que el modelo sea capaz de comprender de una manera más adecuada el comportamiento del proceso subyacente generador de datos, y finalmente se obtengan los beneficios de una predicción más precisa.

El algoritmo FS-EPNet se considera un método híbrido utilizado para evolucionar RNAs de propagación hacia adelante, ya que combina la búsqueda estructural evolutiva con métodos de descenso del gradiente como el algoritmo de Retropropagación de errores (BP) para el aprendizaje en el entrenamiento de los pesos entre los nodos. Para superar los problemas que se presentan en los AGs con el operador de cruzamiento, FS-EPNet se fundamenta en el algoritmo de la Programación Evolutiva (PE) para evolucionar arquitecturas, solo empleando los operadores de selección y mutación, siendo este último el único operador a través del cual se recorre el espacio de búsqueda de las topologías neuronales. El algoritmo fluctúa alternativamente entre los cambios en la composición estructural de la RNA y la modificación de los pesos mediante el entrenamiento, de esta manera se pretende no romper los vínculos entre los padres y sus descendientes, además de lograr una búsqueda eficiente.

Dado que el algoritmo FS-EPNet se construyó bajo las mismas bases en las cuales fue diseñado e implementado EPNet, y con el fin de no agregarle mayor complejidad al modelo, la evolución de las entradas opera de igual forma a como se realiza en su algoritmo predecesor EPNet, es decir, el modelo trabaja de igual forma sobre las entradas como sobre cualquier nodo de la RNA. En el proceso evolutivo se involucran nueve operadores de mutación: entrenamiento híbrido (se utiliza un algoritmo de Modified Back-Propagation (MBP) y el algoritmo de Recocido Simulado (SA)), eliminación de nodo oculto, eliminación de conexión, eliminación de entrada, eliminación de rezago, adición de: nodo, conexión, rezago y entrada. En el algoritmo FS-EPNet, al conjunto de datos que se analiza se le divide en cuatro subconjuntos. El primer subconjunto es el conjunto de entrenamiento, el cual se utiliza para realizar la actualización de los pesos de la red y del umbral (bias), mediante el entrenamiento híbrido. El segundo subconjunto es el conjunto de

validación, que en primera instancia se puede emplear para evolucionar la tasa de aprendizaje, por otro lado, también se emplea para evitar un sobreajuste en el aprendizaje. El tercer subconjunto es el conjunto interno de prueba de FS-EPNet, que simula una predicción real para obtener el grado de adaptación de la red. Finalmente, el cuarto subconjunto de datos es el conjunto final de prueba, el cual es utilizado para medir la capacidad de generalización de los individuos una vez finalizado el proceso evolutivo del algoritmo.

Mediante un proceso gradual se aplican una sucesión de nueve mutaciones, destacando la importancia en el ordenamiento de las mutaciones con el propósito de impulsar la simplicidad o parsimonia de la RNA. El entrenamiento híbrido es la mutación que se encarga de modificar los pesos entre los nodos de la RNA, este aprendizaje consiste por un lado, en un entrenamiento con el MBP con una tasa de aprendizaje adaptativa, por el otro lado se utiliza como herramienta de apoyo al algoritmo de SA para posibilitar el escape de los mínimos locales. Mientras que las otras ocho mutaciones restantes afectan a la arquitectura neuronal. El algoritmo realiza la eliminación de un nodo aleatoriamente, sin embargo, la agregación y eliminación de una conexión se crea probabilísticamente en función de la importancia de la conexión en la RNA (Yao y Liu, 1997b).

Uno de los rasgos distintivos del algoritmo y como anteriormente se había mencionado, es que por medio de la PE se pretende proteger el vínculo entre el ancestro y el descendiente, por ello, cuando se agrega un nodo este se realiza mediante la división de uno ya existente, y de igual forma cuando tiene lugar de manera exitosa una de estas mutaciones de carácter topológico, se realiza un entrenamiento parcial con el fin de disminuir la brecha en el vínculo funcional entre padre e hijo ocasionado por la mutación. A continuación se muestra un esquema que describe a grandes rasgos los pasos del algoritmo (Riessen et al., 1997).

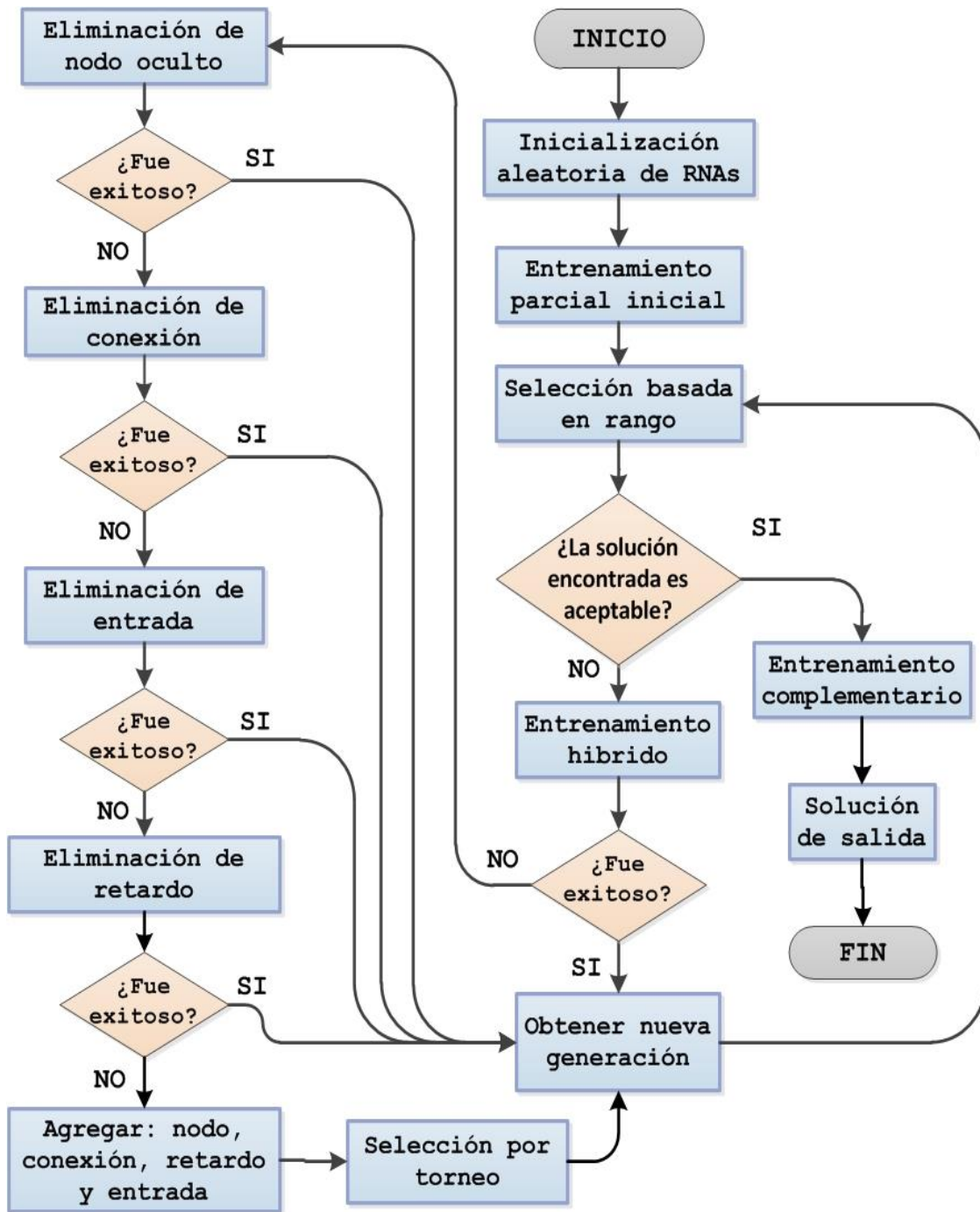
- a) Generación de la Población. El usuario establece los parámetros y se genera una población inicial de RNAs. El conjunto de datos se divide en 4 subgrupos: un conjunto de entrenamiento, un conjunto de validación, un conjunto interno de prueba y un conjunto final de prueba. Los conjuntos de validación, el

interno de prueba y el final de prueba, permanecen ocultos durante la fase de entrenamiento.

- b) Selección. Mediante el operador de selección se clasifican de manera jerárquica cada una de las RNAs basándose en el rendimiento obtenido con el subgrupo de prueba, proporcionándonos una medida de aptitud. Se determina una probabilidad a cada RNA en función de su jerarquía.
- c) Mutación. El algoritmo FS-EPNet al estar basado en técnicas de Programación Evolutiva (PE) utiliza solamente el operador de mutación durante la fase de entrenamiento. Durante el proceso evolutivo, de manera simultánea se actualizan los pesos de las conexiones y se modifica la topología de la RNA mediante la eliminación de las conexiones entre nodos, nodos, o incrementando elementos como adición de conexiones y de nodos. Con el fin de buscar redes más compactas se enfatiza en el ordenamiento de la aplicación de las diferentes mutaciones, es decir, se prioriza la eliminación de elementos antes que la agregación de estos.
- d) Reemplazo. El individuo de la población que es reemplazado puede ser tanto el padre como el individuo con la función de desempeño más baja. Si durante el entrenamiento de los pesos el individuo mejora significativamente su función de desempeño, el padre de este es substituido por su descendiente, de esta forma se pretende mantener el vínculo funcional entre generaciones. Por otro lado, si el entrenamiento estructural es beneficioso entonces el individuo con el menor grado de adaptación de la población es reemplazado, de esta forma el vínculo funcional entre padre y descendiente es desechado.
- e) Terminación de la prueba. Se verifica si el número requerido de generaciones y ejecuciones independientes se ha completado.

La Figura 5.3 representa el esquema general del funcionamiento del algoritmo FS-EPNet.

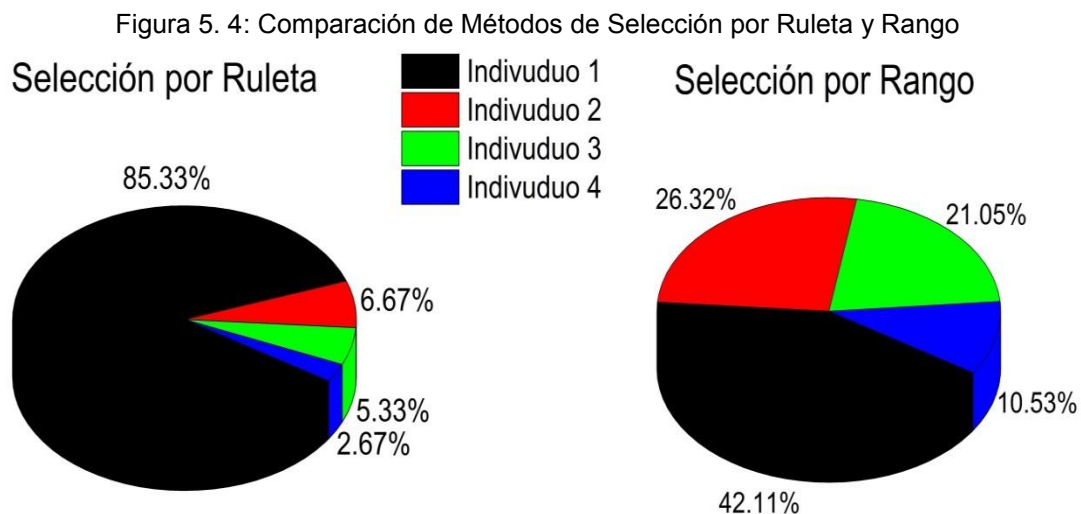
Figura 5. 3: Diagrama básico del funcionamiento del algoritmo FS-EPNet.



Fuente: Elaboración propia.

Un esquema de codificación directa se utiliza en FS-EPNet para representar las topologías de las RNAs, así como de los pesos y de los umbrales. Se emplea la función de transferencia sigmoide, la elección de los individuos se realiza mediante el método de selección por Rango, en este método los individuos de la población

son clasificados o rankeados numéricamente en base al grado de adaptabilidad alcanzado, este método alienta la diversidad, ya que a diferencia del método de selección por ruleta (método más común) permite que los individuos con un notable menor grado de adaptabilidad puedan ser seleccionados, en la figura 5.3 se muestra un ejemplo: si empleamos el método de Selección por Ruleta y el mejor individuo tiene un grado de adaptabilidad del 85.33% de toda la rueda de la ruleta, entonces los otros individuos tendrán posibilidades mínimas de ser seleccionados.

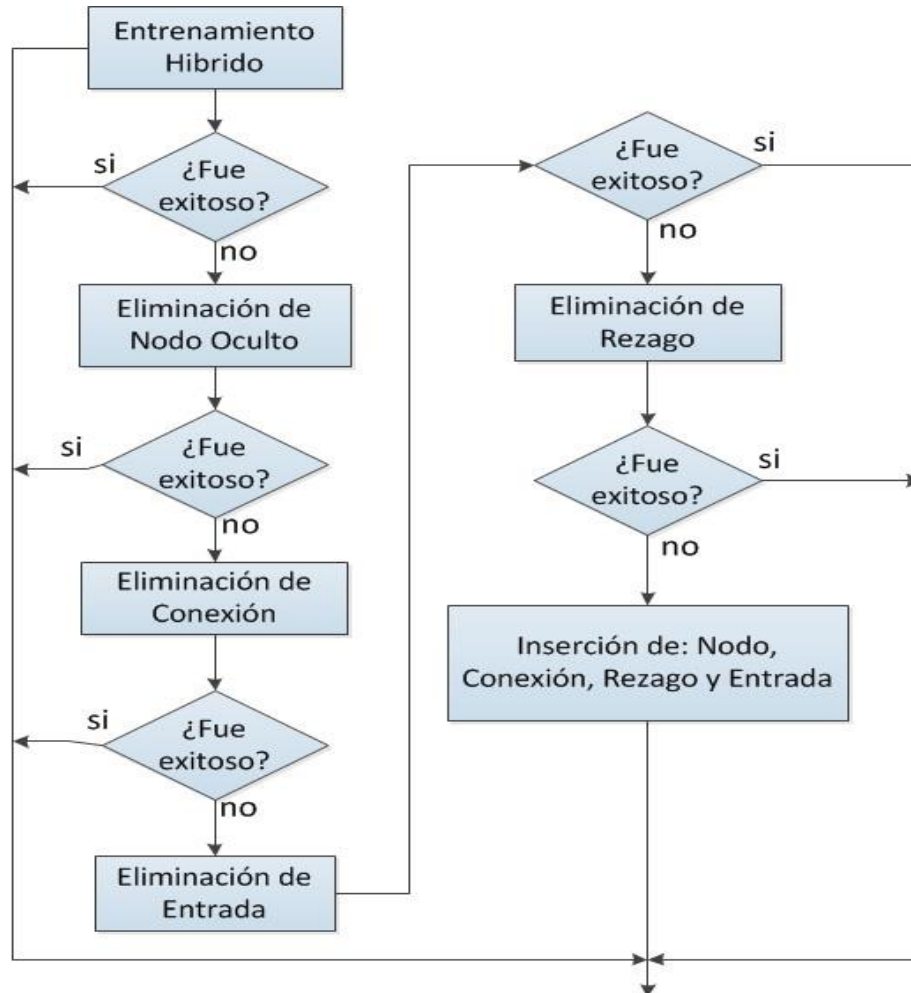


Fuente: Elaboración propia.

En cambio si se utiliza la selección por rango los individuos con una menor adaptación tienen mayores probabilidades de ser elegidos. FS-EPNet es un algoritmo de estado estacionario (Steady-State), donde se reemplazan una cantidad mínima de individuos a diferencia de los algoritmos generacionales que en cada generación por lo regular reemplazan toda la población. La teoría de Lamarck manifiesta que las características modificadas o adquiridas de un individuo durante su vida, se les pueden heredar a sus descendientes, de esta forma, los descendientes tendrán las mismas habilidades que poseían sus ancestros. En FS-EPNet la evolución de las poblaciones de RNAs, se realiza a través de la herencia Lamarckiana, de esta forma, tanto los pesos como las topologías que se aprendieron en una generación, se heredan en la siguiente generación. Es decir, en el caso de FS-EPNet al seleccionar un individuo, como anteriormente se mencionó, la primera mutación que se realiza para generar descendencia es el entrenamiento

hibrido, donde al individuo se le aplica un entrenamiento parcial o el algoritmo de SA.

Figura 5. 5: Esquema de la mutaciones del algoritmo FS-EPNet.



Fuente: Elaboración propia.

Por lo tanto, se puede apreciar como FS-EPNet trata de dirigir el proceso para converger hacia el nivel óptimo mediante una evolución Lamarckiana. Cabe destacar, que la literatura reporta (Whitley et al.; 1994; Castillo Valdivieso et al., 2006) que el uso de la estrategia Lamarckiana ofrece una convergencia más rápida hacia la solución, que si se usara el efecto Baldwin. Sin embargo, la evolución Lamarckiana tiene una probabilidad mayor de quedar atrapado en un óptimo local, a diferencia del efecto Baldwin. Por ello, en FS-EPNet la ocurrencia de este problema se trata de impedir mediante la aplicación del algoritmo de SA. En contraste, en la evolución Baldwiniana este tipo de mecanismo hereditario no tiene



cabida, a pesar de eso, se considera que la descendencia tendrá una mayor capacidad de aprendizaje para adquirir nuevas habilidades.

En el caso de utilizar una evolución Baldwiniana, los pesos aprendidos por padres no son insertados en sus descendientes, por lo tanto, el nuevo individuo tendría que ser entrenado de nueva cuenta por completo, y por consiguiente la evolución Baldwiniana requerirá de una mucha mayor cantidad de generaciones para converger al óptimo. En la figura 5.4 se presenta el esquema de la mutaciones del algoritmo FS-EPNet, como se puede apreciar el algoritmo hace énfasis en una composición compacta de la RNA, descartando elementos antes de adicionarlos.

### **5.3 Algoritmo de recocido simulado**

El recocido simulado (SA) es un algoritmo mediante el cual se pretende encontrar la mejor solución a un problema de optimización, aunque esta no sea una resolución perfecta. Existen algunos problemas de optimización donde hay un extenso espacio de soluciones posibles, que incluso para el poder computacional moderno puede resultar demasiado intenso y poco práctico recorrer todas las soluciones posibles a considerar. El algoritmo SA es una analogía al proceso metalúrgico de recocido, mediante el recocido se mejoran las propiedades del metal. El recocido es un proceso de calentar el metal levemente por encima de la temperatura crítica del metal y permitir que se enfríe gradualmente para poder establecerse en su estado energético más bajo. Las propiedades físicas del metal son alteradas debido a los cambios en su estructura interna, dado el procedimiento de recocido que radica en calentar y enfriar dicho material.

En consecuencia en el algoritmo de recocido simulado se ocupa una variable de temperatura para simular este proceso de calentamiento. Al inicializar el proceso de recocido la variable de temperatura se establece alta y lentamente se va enfriando conforme el algoritmo es iterado. Para evitar que el algoritmo quede atrapado en un óptimo local durante su ejecución, el algoritmo acepta soluciones de calidad inferior con mayor frecuencia a la solución actual, mientras la variable de temperatura es alta. Conforme la temperatura desciende así también la posibilidad de aceptar

soluciones inferiores, de esta forma el algoritmo es dirigido hacia una región del espacio de búsqueda donde se presupone se pueda hallar una solución muy cercana a la óptima. Para los problemas vastos y complejos que contienen múltiples óptimos locales, el algoritmo de SA los resuelve mediante el procedimiento de enfriamiento gradual que le permite eludir cualquier óptimo local.

Paso 1. Se genera aleatoriamente la solución inicial y se evalúa mediante una función objetivo.

Paso 2. Establecer la temperatura inicial.

Paso 3. Se produce una solución aleatoria proveniente del vecindario mediante la alteración de la solución actual.

Paso 4. Evaluar la nueva solución.

Paso 5. Comparar las soluciones.

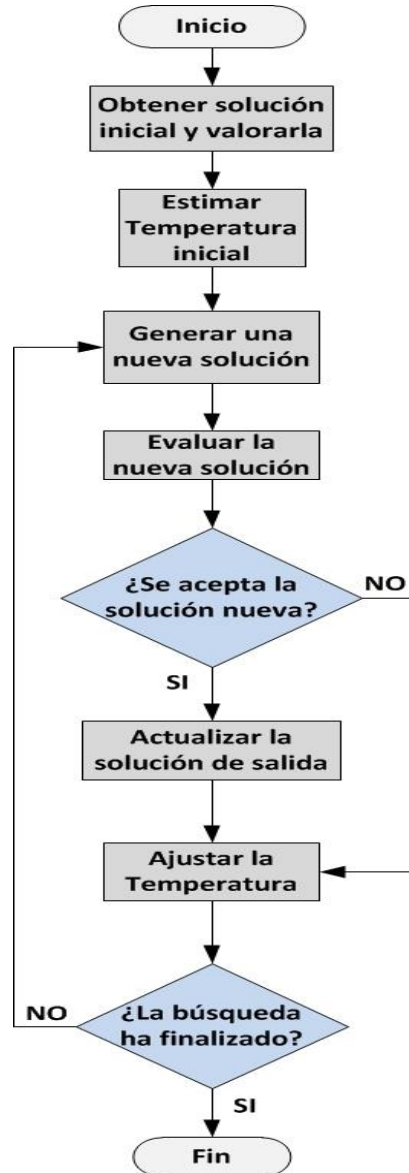
Si la solución nueva tiene una mejor evaluación que la solución anterior se reemplaza incondicionalmente, de esta forma la nueva solución es la base para la siguiente iteración, además esto significa que el algoritmo se está acercando a un óptimo. Sin embargo, si la nueva solución es inferior que la solución actual, se deben de considerar dos factores. En primera instancia se juzga en que grado es peor la nueva solución. En segundo lugar, se calcula que tan elevada es la temperatura actual del procedimiento. El algoritmo trata en lo posible de evitar soluciones peores. Sin embargo, cuando la temperatura es alta aumenta la probabilidad de que el sistema acepte soluciones con una calidad inferior.

Paso 6. Por último, se disminuye la temperatura y se continúa iterando el algoritmo hasta que la búsqueda finalice.

Por otro lado, en lo referente a la evolución de las RNAs el algoritmo SA aparece en la fase de entrenamiento híbrido en el algoritmo evolutivo FS-EPNet. Donde es posible destacar que el método SA en comparación con el algoritmo de retropropagación de errores (BP), necesita de un mayor número de épocas o iteraciones para converger hacia un entrenamiento exitoso (Sekhar Behera & Chattopadhyay, 2012). La inclusión de SA en el algoritmo FS-EPNet en la etapa de entrenamiento híbrido le dota a FS-EPNet la capacidad de escapar de los óptimos locales, marcando diferencia con BP que tiene cierta tendencia a quedar atrapado

en los óptimos locales. La figura 5.6 muestra un esquema general del funcionamiento del algoritmo SA.

Figura 5. 6: Esquema básico del algoritmo de Recocido Simulado.



Fuente: Adaptado de (Sekhar Behera & Chattopadhyay, 2012).

## 5.4 Exponente de Hurst

### 5.4.1 Introducción

Una herramienta que sirve para medir la independencia de una serie de tiempo es el coeficiente o exponente de Hurst, en otras palabras dicho instrumento es útil para verificar si hay presencia de memoria en una serie de tiempo. El exponente de Hurst

es un método estadístico creado por Harold Edwin Hurst, que en primera instancia fue de utilidad para determinar los rasgos distintivos del comportamiento fluvial del río Nilo. Tiempo después los planteamientos e ideas de Hurst son rescatados por el matemático Benoit B. Mandelbrot para posteriormente generalizarlas y que denominaría análisis de rango reescalado (R/S). El exponente de Hurst es un método matemático que ha sido aplicado a diferentes ramas de la ciencia empezando por el fenómeno hidrológico (Valle et al., 2011), en el ámbito financiero es común su empleo dadas las características de los datos que se involucran en su análisis (Peters, 1989; Mandelbrot, 2010; Sierra, 2012), en economía (Valdez-Cepeda, 2000; Matamoros, 2004), en el análisis médico (Morales-Matamoros et al., 2009; Aparicio et al., 2014), entre otras disciplinas donde se ha utilizado.

#### **5.4.2 Historia del exponente de Hurst**

Harold Edwin Hurst (1880-1978) fue un ingeniero hidráulico que trabaja en el río Nilo a principios del siglo XX, a la edad de veintiséis años se fue a Egipto en un inicio por un periodo corto de tiempo pero la estancia se prolongó por 62 años, tiempo suficiente para estudiar a fondo y exhaustivamente la conducta o proceder del río Nilo, recorriéndolo de diferentes formas y en innumerables ocasiones, compilando los datos relevantes que guardaban relación con el variante estado de sus caudales y corrientes. Al ser el Nilo un río que representa gran trascendencia para la nación egipcia desde tiempos inmemoriales, siendo proveedor de abundancia y a la vez causante de desgracias debido a su comportamiento impredecible hasta antes del trabajo puntual que realizó Hurst.

Con el acceso a los datos históricos extensos correspondientes a 800 años sobre los flujos anuales, y con un análisis minucioso en todo lo relativo al Nilo. En dicho estudio encontró que existían ciertas conductas repetitivas y que el comportamiento del Nilo no era del todo azaroso como se suponía, donde los años lluviosos venían seguidos de años que de igual forma se presentaban con un abundante nivel de agua, de igual forma años de sequía venían precedidos de años con un grado menor de lluvias. Es así que llegó a la conclusión que el comportamiento del Nilo era persistente. A la medición de la persistencia de las tendencias de sequías y

abundancia que se presentaban en el Nilo, se le ha bautizado con el nombre de Coeficiente o Exponente de Hurst o simplemente  $H$ , que en base a un cálculo determina la persistencia de una serie de tiempo.

### 5.4.3 Interpretación del exponente de Hurst

El rango en el que se encuentra acotado el exponente de Hurst es entre cero y uno:

- Si el valor de  $H=0.50$  o muy cercano a este número, la serie es aleatoria hay una correlación con un valor de 0, es decir los datos guardan independencia, un decremento o incremento en la serie no se encuentra relacionado. No existe relación entre un dato y otro.
- Un exponente en el rango de  $0 < H < 0.50$  indica un comportamiento antipersistente. Donde un periodo de crecimiento es seguido de otro periodo de decrecimiento, de igual forma un periodo de decrecimiento le sigue uno de crecimiento, la correlación de los datos tiene un valor negativo. Variaciones negativas y positivas en los periodos, comportamiento cíclico. Memoria a corto plazo.
- Cuando el exponente se encuentra ubicado entre  $0.5 < H < 1$  la serie de tiempo es persistente, los datos muestran correlación positiva. Hay una alta probabilidad de que al incremento en los datos le siga un aumento, de manera análoga las fluctuaciones decrecientes le siguen disminuciones. Un comportamiento totalmente determinista en los datos viene dado cuando el exponente ha obtenido un valor de  $H=1$ . Memoria a largo plazo.

Tomando como herramienta el exponente de Hurst para medir la complejidad del fenómeno que se analiza, se puede plantear que modelo de predicción se adecua mejor a los datos. Un exponente de Hurst cercano a la unidad representa un sistema cuya conducta es más fácil de entender y por lo tanto de predecir, en cambio un exponente de Hurst con un valor por debajo de 0.5 implicaría una complejidad mayor, donde la tarea de predicción tendera a ser más compleja.

## 5.5 Métodos de combinación de pronósticos

### **5.5.1 Introducción**

El utilizar combinación de estimaciones no es un tópico novedoso, hay antecedentes que Laplace considero utilizar la combinación de estimaciones de los coeficientes de regresión (Clemen, 1989). Posteriormente en 1936 Edgerton y Kolbe utilizan un criterio de optimización donde minimizan la suma de los cuadrados de las diferencias, hallando así una combinación de estimaciones óptima (Edgerton y Kolbe, 1936; Clemen, 1989).

De manera consensuada académicos y profesionales consideran que en el año 1969 se marca el comienzo formal de la combinación de pronósticos, derivado del trabajo seminal de Bates y Granger donde desarrollan y analizan una serie de técnicas combinatorias para un pronóstico puntual (Bates y Granger, 1969; Clemen, 1989), ya que fueron los primeros en desarrollar analíticamente y de forma particular un modelo general para la combinación de predicciones que fueran aplicables a situaciones del mundo real (Clemen, 1989).

Gonzales Martínez y Luna Acevedo (2011) basándose en la metodología Bates-Granger combinan un modelo Red Neuronal Artificial Autorregresiva (RNAA) y un modelo Autorregresivo Integrado de Media Móvil-Modelo de Volatilidad Condicional Heterocedástico Generalizado (ARIMA-GARCH) buscando obtener una predicción más precisa del precio futuro del petróleo West Texas Intermediate (WTI).

Zhang (2003) utiliza un método híbrido que se divide en dos partes donde el primer paso del modelo se analiza y extrapola mediante un modelo ARIMA y en el segundo paso el resultado de los residuales del modelo ARIMA son la entrada a una RNA que de igual manera analiza y predice, donde finalmente se combinan ambos resultados, en dicho estudio se analizan tres conjuntos de datos ampliamente conocidos como son: las manchas solares de Wolf, datos del lince canadiense, y el tipo de cambio libra esterlina=dólar donde se demuestra la efectividad del método híbrido.

### **5.5.2 Teoría de la combinación de pronósticos**

Generalmente en el proceso de toma de decisiones el utilizar pronósticos ha venido a ser una parte esencial, es así, que aquellos analistas que se encargan de realizar

proyecciones disponen de una múltiple variedad de métodos de predicción. En otras palabras, para proyectar una variable existen diferentes enfoques y formas para poder generarlos, es decir, que para una misma variable pueden existir varios pronósticos, y es una práctica habitual que después de realizar varias predicciones del mismo fenómeno se comparen y se escoja aquel que se ajuste mejor a los datos o sea más preciso.

Sin embargo, persisten varias cuestiones que tienen que solventarse como son: pudiera ser que el modelo seleccionado tenga un buen desempeño en el pronóstico dentro de la muestra pero en cambio la proyección generada fuera de la muestra sea ineficiente, ya sea por un sobre-ajuste a los datos dentro de la muestra, rupturas o cambios estructurales del fenómeno, inconsistencias y sesgo al momento de crear el modelo así como también error u omisión al escoger los parámetros del modelo de predicción. Este problema pudiera ser subsanado al utilizar una combinación de diferentes métodos de predicción.

Tambien es posible que cada uno de estos métodos predictivos fuera capaz de capturar los diferentes patrones que subyacen en el fenómeno que se analiza. Otro enfoque toma como premisa la teoría moderna del portafolio donde lo esencial es la diversificación de inversiones para disminuir el riesgo, en este caso al utilizar una variedad de métodos de predicción y combinarlos se acrecentaría la probabilidad de que el pronóstico final fuera más certero. Existen métodos que son capaces de adecuarse de manera más oportuna o rápida a los cambios estructurales a diferencia de aquellos que sus parámetros se ajustan de una forma más lenta. Por otro lado existen quienes afirman que si un método necesita ser combinado es a causa de una deficiente especificación de sus parámetros. Otra cuestión a tomar en consideración es que si se agregan varios pronósticos individuales a nuestro modelo sin tener la certeza de la trascendencia de cada uno de ellos nos podría conducir a una perdida en la eficiencia de nuestro modelo así como también una notable disminución en la precisión de nuestra predicción. Cabe recalcar que la evidencia empírica muestra resultados mixtos, no obstante en la mayoría de los casos se fomentan los beneficios que otorga la combinación de técnicas logrando como

resultado una proyección más precisa.

Existen diversos estudios (Adams, 1978; Makridakis et al., 1982; Bessler y Chamberlain, 1987; Zhang, 2003; Gonzales Martínez y Luna Acevedo, 2011) donde se ha demostrado que utilizar alguna técnica de combinación se traduce en un desempeño superior al mejor modelo individual. En la literatura sobre el tema de combinación de predicciones se puede hallar una gran variedad de técnicas. Habitualmente la combinación de pronósticos tiene como tarea fundamental hallar los pesos o ponderaciones adecuadas u óptimas para cada modelo. Algunos de los modelos más populares son: el promedio simple (media, media truncada, mediana), media ponderada, regresión múltiple (Krishnamurti et al., 1999), agrupamientos o clúster (Aiolfi y Timmermann, 2006), modelo bayesiano (Garratt, Lee, Pesaran & Shin, 2003), así como también modelos híbridos: ARIMA-RNA (Zhang, 2003) y Encuestas-Series de Tiempo (Aiolfi et al., 2010), entre otros.

En la mayoría de ellos se ha encontrado evidencia que los métodos de combinación más sencillos como por ejemplo un promedio simple (PS) son tan efectivos he incluso superiores en la mayoría de los casos a los métodos más sofisticados y refinados de combinación (Clemen, 1989; Timmermann, 2006). Estas premisas pueden llevar a la conclusión que en general los analistas prefieren un modelo que tenga un buen desempeño y en esencia o por naturaleza sea sencillo e intuitivo, a uno con un rendimiento similar pero con exceso de refinamiento y complejidad. El diseño general de un estimación de un pronóstico combinado  $M_{Cj}$  se puede determinar a partir de  $n$  modelos que se encuentren involucrados para el  $j$ -ésimo periodo, se encuentra especificada como la función  $F()$  de los  $n$  modelos individuales:

$$M_{Cj} = F(M_{1j}, M_{2j}, M_{3j}, \dots, M_{nj}) \quad (5.9)$$

Dependiendo de la técnica de combinación empleada es que se define la función  $F()$ , donde  $M_{ij}$  es el modelo empleado por el  $i$ -ésimo modelo para el  $j$ -ésimo período.

### 5.5.2.1 Promedio simple



Cuando no hay certeza cuál de los modelos de predicción es más preciso y buscando estabilidad en nuestro modelo, el utilizar un promedio simple puede ser la respuesta.

$$M_{Cj} = \frac{1}{n} \sum_{i=1}^n M_{i,j} \quad (5.10)$$

### 5.5.2.2 Promedio ponderado

Dada la situación de que algún método de nuestro conjunto de elección se comporte mejor o sea superior en rendimiento que otro, es decir cuando un modelo tiene una mayor precisión que otro modelo podemos obtener una mejora substancial al utilizar esta herramienta, puesto que un promedio simple en ocasiones puede ser deficiente y el modelo con peor desempeño puede contaminar nuestro modelo. La combinación de diferentes modelos ponderados propuesta por Granger y Ramanathan (1984), es una de las más populares, y se expresa en la siguiente ecuación:

$$M_{Cj} = \sum_{i=1}^n W_i M_{i,j} + e_j \quad (5.11)$$

La ponderación  $W_j$  que se le asigna al  $i$ -ésimo modelo  $M_{ij}$  y la  $e_j$  es el error asociado al ajuste de la combinación.

## **Capítulo 6. Estudio empírico**

### **6.1 Introducción**

En la primera sección del presente capítulo se presentan las muestras que se analizaron, así como una breve descripción de estas, y algunas cuestiones que se tienen que tomar en cuenta al elegir las. A continuación, se descomponen los indicadores mediante la herramienta DEM, y estas descomposiciones extraídas de los indicadores al igual que estos, son analizadas en primera instancia mediante el coeficiente de Hurst para determinar su complejidad y empezar a discernir qué estrategia se puede implementar. Posteriormente, se realiza la parte experimental, la cual se encuentra dividida en tres etapas A, B y C.

En primer lugar se lleva a cabo el modelado y predicción del índice IPC. En la etapa A se plantean dos escenarios, en ambos escenarios se analiza y pronostica el mismo periodo de tiempo, la totalidad de la muestra para esta etapa es 1589 días. Para estimar los parámetros de los diferentes modelos predictivos se utilizan 1554 datos, y se pronostican 35 días. La diferencia entre el primer y segundo escenario consiste en que en el primer escenario se plantea la posibilidad de que el analista o modelador incurra en el error de introducir la totalidad de los datos a la herramienta de descomposición de señales DEM, es decir, que DEM descomponga los 1589 datos disponibles para esta fase y con las resultantes FMIs y residuo realice las estimaciones de los parámetros de los modelos paramétricos y no paramétricos. Ya que de esta forma los resultados se encontrarían sesgados de manera positiva. En el segundo escenario de la fase A se realiza la descomposición del IPC de forma adecuada, ya que las FMIs y el residuo que se obtienen surgen de la descomposición de 1554 datos y se deja a un lado los restantes 35 días que se pretenden predecir. Es así que las predicciones resultantes del primer y segundo escenario son comparadas en el segundo escenario, con el fin de verificar como las predicciones del primer escenario se ajustan de manera más favorable a los datos a predecir, ya que los parámetros de los modelos de predicción se encuentran inducidos. En la etapa B se elabora la predicción en el periodo donde tuvo ocurrencia una caída bursátil, se recorren 35 días para realizar la estimación y el

entrenamiento de los modelos paramétricos y no paramétricos respectivamente. Finalmente se ejecuta la fase de combinación de pronósticos (Etapa C). En las etapas B y C se estiman los parámetros de los diferentes modelos predictivos con la misma cantidad de datos, así como también se pronostica el mismo periodo de tiempo. El segundo caso, el análisis y predicción del DJIA se confecciona de manera similar al experimento anterior (IPC), sin embargo se omite el primer escenario de la fase A.

A pesar de que la presente investigación se basó en un marco general de análisis y modelado, se optó por mezclar la sección de desarrollo y resultados en uno solo, por un lado, con el propósito de que la exposición resulte más clara para el lector. Por otro lado, el desarrollo de los modelos se encuentra supeditado en muchos de los casos a los resultados obtenidos en etapas previas, estos resultados a su vez conducen a seguir un determinado camino, es decir, las fases en el desarrollo de los diferentes modelos se encuentran íntimamente ligados unos con otros. Por último en el apartado APÉNDICE se presenta la edificación de un modelo ARIMA-ARCH aplicado a una de las descomposiciones proveniente de la serie del DJIA.

## **6.2 Selección de las muestras**

Al momento de seleccionar el tamaño de la muestra, puede surgir la interrogante, de cuál es la cantidad correcta de datos a ser ingresados al modelo. Para el caso de las series financieras, algunos pudieran pensar que no es trascendental trabajar con datos de dos o más años ya que el dinamismo del fenómeno es altamente cambiante y que los datos antiguos no son relevantes. Esta deducción como tal es errónea, ya que lo que manifiesta dicha deducción es una carencia en el entendimiento de lo que realiza un adecuado modelo predictivo. Puesto que un modelo correctamente especificado deberá ser capaz de captar los cambios en los patrones a lo largo del tiempo, y no solo si este ha sufrido cambios, sino comprender la evolución futura del fenómeno.

Por ello es necesario pensar que se necesita de una cantidad considerable de datos, ya que con una mayor cantidad de datos es posible determinar con un grado mayor de exactitud los parámetros y la estructura del modelo predictivo. Por ejemplo: un

modelo SARIMA con los parámetros  $(0, 1, 1)(0, 1, 1)_{12}$  necesita disponer de por lo menos 16 datos para una predicción a un paso (Hyndman y Kostenko, 2007), para el caso de la presente investigación se pretende predecir a 35 y 40 pasos, es por ello que se requiere de una mayor cantidad de datos que solo los 16 necesarios para un periodo.

Por otro lado cuando se diseña una RNA existen varias cuestiones por resolver como pudiera ser el número de capas, la cantidad de neuronas por capa, las funciones de transferencia, etc. Aparte de lo anteriormente mencionado, en el modelo no paramétrico es de suma importancia la cantidad de datos, ya que una considerable cantidad de datos de entrenamiento típicamente se traduce en una predicción más precisa (Leondes, 2002). Una red de grandes dimensiones requiere de un gran conjunto de datos de entrenamiento, ya que es proclive a sobreentrenarse con un pequeño número de ejemplos de entrenamiento (Müller et al., 1995). No se deberá usar una red con más parámetros que el número de datos disponibles (Lawrence et al., 1996). Algunas de las interrogantes como son: el número de neuronas, la cantidad de conexiones, entre otras, se encarga de resolverlas el algoritmo evolutivo FS-EPNet. Pero en cuanto a estimar la óptima cantidad de datos, estará en función del grado de ruido presente en los datos, la complejidad del fenómeno, así como también para que una red generalice eficientemente se necesita proporcionarle un conjunto de datos lo suficientemente completo, donde cada grupo de datos este representado de manera adecuada; ya que cada subconjunto del conjunto de datos tiene propensión o inclinación hacia algún o algunos patrones en específico. Otra problemática de las ST económicas es la escasa cantidad de datos, que muchas veces limitan el análisis del fenómeno, y restringen a la variabilidad en la elección del modelo predictivo.

Como consecuencia de tener una cantidad limitada de datos en las ST económicas, es la red la que se trata de ajustar al tamaño del conjunto de datos mediante el algoritmo FS-EPNet. El algoritmo FS-EPNet se encuentra basado en técnicas de computación evolutiva, las cuales a su vez son consideradas como métodos de búsqueda heurísticos (Michalewicz, 1996). Por lo cual se realizaron 12 ejecuciones

independientes del modelo FS-EPNet/RNA para cada una de las FMI's y el residuo provenientes de la ST del DJIA, ya que la estructura arquitectónica de la RNA es construida heurísticamente, el resultado de dicha topología suele ser diferente para cada iteración.

Lo mencionado anteriormente se traduce en resultados diferentes para cada caso, y que para fines prácticos del modelo significa una predicción diferente para cada una de las Redes Neuronales Artificiales Evolucionadas. Donde dichos pronósticos deben de ser homogéneos, esta semejanza en los resultados se consigue al encontrar una formula eficiente en la estimación de los parámetros como son: el número adecuado de generaciones del algoritmo FS-EPNet, la cantidad de épocas de entrenamiento de la RNA con el algoritmo BP, y la cantidad de datos suficientes. Puesto que las ST financieras en su gran mayoría no presentan normalidad y específicamente las ST referentes al mercado de valores tienen como uno de sus rasgos distintivos la varianza heteroscedástica, las pruebas paramétricas, no se deben utilizar en este caso, dado que arrojarían resultados ambiguos. Por ello, se propuso utilizar un método no paramétrico para el análisis de la varianza, el método elegido es la prueba de Friedman que es semejante a la prueba del Análisis de Varianza (ANOVA). La prueba de Friedman es una prueba que se utiliza para probar la diferencia entre varias muestras relacionadas, mediante el paquete estadístico SPSS se realizaron las pruebas de homogeneidad.

Hipótesis Nula y Alternativa:

$H_0$  = es que no existen diferencias entre las variables.

$H_1$  = al menos dos de las variables son significativamente diferentes una de la otra.

Donde el nivel de significancia estadística que se obtuvo fue del 0.001, siendo este inferior al nivel de significancia que se establece usualmente en las ciencias sociales del 0.05, pero en este caso como consecuencia de las divergencias en los resultados, y a causa de la complejidad del fenómeno, esta situación se podría considerar aceptable.

### **6.3 Configuración experimental del IPC y DJIA**

Esta investigación se dividió en tres etapas. En todas las etapas, los modelos

paramétricos y no paramétricos emplearon una técnica similar de predicción a múltiples pasos. Se utilizó la técnica de predicción MSP (MSF) tanto para el modelo no paramétrico como para el modelo paramétrico, el horizonte de predicción en la primera etapa es de 35 puntos, en la segunda y tercera etapa la predicción es de 40 puntos. Bajo la estrategia de Box-Jenkins se utilizaron tres variantes: el primero de ellos modelo ARIMA, el más general y con un uso más amplio, basa su predicción en una combinación lineal de sus valores pasados y una conjunción de errores o innovaciones; el modelo SARIMA (ARIMA estacional) se utiliza cuando el proceso analizado posee un comportamiento estacional; finalmente la familia de los modelos ARCH que se utilizan cuando el proceso presenta una consistente invariabilidad en las perturbaciones, estos modelos también se conocen como modelos de Heterocedasticidad Condicional Autorregresiva. Para utilizar el modelo paramétrico es necesario que los datos introducidos en el modelo sean estacionarios.

Las series del IPC y del DJIA presentan un proceso no estacionario y como tal, algunas descomposiciones presentan rasgos de no estacionariedad. El orden de diferenciación de las series de tiempo para convertirlas en un proceso estacionario fue determinado por la prueba de raíz unitaria (Brooks, 2008). La experimentación comenzó con la normalización de la serie en un intervalo entre [-1 a 1], para después descomponer la ST mediante la técnica DEM, utilizando el software matemático Matlab y la función propuesta por Rilling et al. (2003). A través del procedimiento propuesto por Box y Jenkins (1976), se estimaron los parámetros ARIMA-ARCH, los criterios utilizados para identificar el mejor modelo son: Akaike (AIC) y el criterio de Información Bayesiano (BIC).

La transformación de los datos originales a un proceso lineal en el modelo no paramétrico no es necesario que se realice, teniendo en cuenta esto, los datos se introdujeron sin filtro ni alteración y sólo aplicando la normalización de los datos anteriormente señalada. Para los experimentos del modelo no paramétrico se llevaron a cabo 12 ejecuciones independientes. Los datos provenientes de las series del IPC y DJIA se dividieron en cuatro subconjuntos:

- 1) Conjunto de entrenamiento: se utiliza para llevar a cabo la tarea de aprendizaje

con el algoritmo de Retropropagación de errores modificado (MBP) o la técnica SA.

- 2) Conjunto de Validación: este conjunto se emplea para evitar el sobre-ajuste en el proceso de aprendizaje.
- 3) Conjunto interior de Prueba: se utiliza para simular una predicción (MSP) real (predicción fuera de la muestra) con el fin de medir la aptitud de las RNAs.
- 4) Conjunto Final de Prueba: una vez concluido el proceso evolutivo, este conjunto es utilizado para evaluar el rendimiento de los diferentes individuos pertenecientes a la población final, para luego ser clasificados.

Para producir las predicciones del modelo DEM/FS-EPNet/RNA, se promedió a los cinco individuos mejor clasificados tomando como base el criterio de desempeño NRMSE en cada descomposición para ser parte de la reconstrucción. Es decir, de entre 12 ejecuciones independientes se eligieron los resultados provenientes de las cinco redes neuronales que obtuvieron el error menor (NRMSE) en la predicción del conjunto final de prueba para cada FMI y el residuo. El resultado de la predicción de estas cinco RNAs elegidas a su vez se promedió para obtener una sola predicción por cada descomposición. Posteriormente, para reconstruir una predicción final del modelo DEM/FS-EPNet/RNA, el promedio resultante de cada FMI y el residuo se sumó para adquirir una sola predicción. Por otro lado, los pronósticos generados mediante el modelo FS-EPNet/RNA son el resultante de promediar los pronósticos generados mediante 12 ejecuciones independientes tomando como entrada la serie original del indicador analizado en cuestión. La configuración de los parámetros del algoritmo FS-EPNet se basó en la estrategia utilizada por Landassuri-Moreno (2012). Algunos parámetros se determinaron previamente antes de que comenzara el proceso evolutivo. Los parámetros establecidos para esta investigación fueron los mismos para todos los modelos: tasa de aprendizaje 0.153, las mutaciones de los nodos ocultos y las conexiones entre 1 y 3, las mutaciones de entrada y de los retardos entre 1 y 2, el tamaño poblacional de 30 individuos, con 1400 generaciones esperadas en el proceso evolutivo. No se emplearon criterios de interrupción temprana (early stopping) con el fin de evitar un sobreentrenamiento de las RNAs, puesto que en base a experimentos previos, se infirió que no era necesario, dada la

complejidad de la tarea a resolver, y también derivado de las pruebas preliminares se estableció el número de épocas de entrenamiento parcial durante la evolución en 1200 y de 3600 al finalizar el proceso evolutivo.

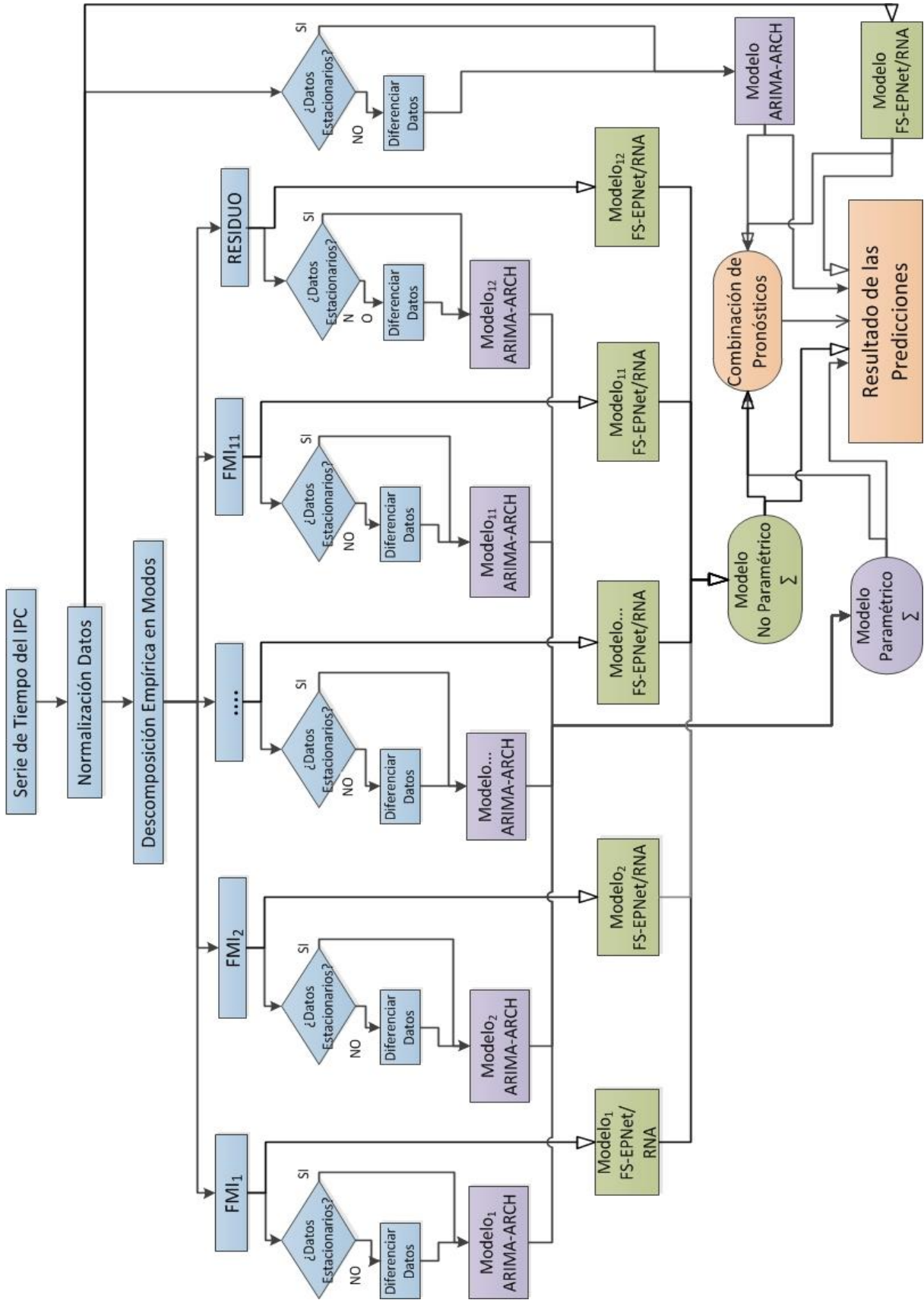
En la parte experimental se utilizaron cuatro sistemas, se diseñaron dos modelos que trabajaron con la serie original de los indicadores (IPC y DJIA) y los otros dos sistemas operaron con las FMIs y el residuo de las series provenientes del IPC y del DJIA. Cada uno de estos modelos trabajo de manera aislada, es decir, no se mezclaron metodologías (paramétricas y no paramétricas) ni datos hasta llegar a la etapa de combinación de predicciones. Los modelos que trabajaron con la serie original, en primera instancia fueron el modelo paramétrico ARIMA-ARCH y de igual forma, se diseñó mediante el algoritmo FS-EPNet un modelo de RNA. A los dos modelos se les introdujo la serie original del indicador, y se encargaron de producir cada uno su pronóstico respectivo. Posteriormente, los otros dos modelos tanto el paramétrico como el no paramétrico, que operaron con las FMIs y el residuo trabajaron de manera independiente. Por un lado, se edificó para cada una de las descomposiciones un modelo ARIMA-ARCH. De igual forma, para cada FMI y para el componente residual se diseñó una RNA, donde cada una de las 12 descomposiciones fueron pronosticadas de manera individual en cada modelo (paramétrico y no paramétrico), finalmente se reconstruyo una predicción para cada modelo, mediante la suma de cada una de las descomposiciones pronosticada, de este modo cada modelo produjo su propia predicción. De esta forma se obtuvieron cuatro pronósticos uno por cada sistema, que a continuación se presentan:

- |                   |                     |
|-------------------|---------------------|
| 1) ARIMA-ARCH     | 2) FS-EPNet/RNA     |
| 3) DEM/ARIMA-ARCH | 4) DEM/FS-EPNet/RNA |

Por último se combinan las predicciones del modelo paramétrico y no paramétrico para crear un pronóstico combinado. La figura 6.1 muestra el diagrama general de la experimentación realizada en esta investigación.



Figura 6. 1: Sistema Experimental.



Fuente: Elaboración propia.

Con el propósito de evaluar y comparar el rendimiento predictivo de los diferentes modelos, se utilizaron diferentes métricas de error: Error Cuadrático Medio (MSE), Root Mean Squared Error (RMSE) (Ardalani-Farsa y Zolfaghari, 2010) y Normalized Root Mean Squared Error (RMSE) (Landassuri-Morenoy Bullinaria, 2009a).

#### 6.4 Etapas de experimentación del IPC

Esta etapa de investigación se encuentra implicada específicamente en la predicción de la BMV de México, en un periodo de crisis. La muestra de datos analizada es la correspondiente al precio de cierre del IPC para el período comprendido entre el 8 de julio de 2002 y el 16 de diciembre de 2008; la cantidad de datos contabilizó un total de 1629 datos del IPC. La figura 6.2 representa las diferentes fases de experimentación realizados en la experimentación del IPC.

Figura 6. 2: Etapas de predicción del IPC del 08/07/2002 al 16/12/2008.



Fuente: Elaboración propia.

- A) Esta fase se encuentra dividida en dos partes y se propusieron dos escenarios, en ambos escenarios se pronosticó el mismo lapso de tiempo (se emplearon los mismos datos).

En el primer escenario de la fase A, se plantea la posibilidad de que el analista descomponga con la técnica DEM la totalidad de los datos (1589 datos), en otras palabras, todos los datos disponibles en esta etapa se descomponen en conjunto para estimar los parámetros de los modelos de predicción (representado en color negro en la figura 6.2, conjunto de datos dentro de la muestra) al igual que los datos que se pretenden predecir (representado en color rojo en la figura 6.2, conjunto de datos fuera de la muestra). De esta forma es posible medir el rendimiento de los modelos predictivos en cada descomposición. Sin embargo, los modelos estarían sesgados, puesto que los parámetros estimados para realizar las predicciones por los modelos paramétricos y no paramétricos se encontrarían inducidos. Puesto que el algoritmo DEM es guiado y dependiente de los datos que son introducidos en él, si se descompone la totalidad de la muestra (conjunto de datos para estimar parámetros y conjunto de datos a pronosticar) se estaría incorporando en las descomposiciones (FMI y el residuo) información futura y como consecuencia los resultados se encontrarían alterados, produciéndose así una predicción anormalmente favorable. Además, esta dependencia de DEM hacia los datos se refleja cuando se descomponen diferentes muestras de una misma serie, ya sea con una cantidad mayor o menor de datos e incluso con la misma cantidad de datos, pero en otro periodo de tiempo, y dada la variabilidad y complejidad, así como también los rasgos de no estacionariedad y volatilidad que presentan las series financieras, es muy probable que DEM no produzca la misma cantidad de FMIs, al igual que la estructura de cada descomposición, que podría variar notablemente. Gracias a esto las predicciones resultantes de los modelos predictivos propuestos en esta investigación (DEM/ARIMA-ARCH y DEM/RNA/FS-EPNet) solo pueden ser analizados y evaluados de una manera confiable, tomando solo en cuenta la adición de todas las descomposiciones pronosticadas y compararlas con los datos reales. Por ello, se plantea un segundo escenario, donde la descomposición de los datos mediante DEM

se realiza de manera correcta. La cual consiste en introducir a DEM sólo 1554 datos (conjunto de datos dentro de la muestra) para descomponer, y con las resultantes FMI y residuo estimar los parámetros de los modelos de predicción, y los restantes 35 datos a pronosticar son separados (conjunto de datos fuera de la muestra).

En esta primera fase el modelo no paramétrico fue entrenado con un total de 1549 datos, del 07/08/2002 al 28/08/2008, este conjunto de datos se encuentra representado en la figura 6.2 en color negro. Mientras que para el modelo paramétrico se estimaron sus parámetros con un total de 350 datos, del 04/12/2007 al 28/08/2008. Esta disminución en la muestra para estimar los modelos paramétricos es a consecuencia de que en base a experimentaciones previas con dichos modelos se obtuvo un rendimiento más favorable de dicho modelo con una cantidad menor de datos.

El horizonte de predicción abarca del martes 29 de Agosto de 2008 al viernes 17 de octubre de 2008, los cuales se encuentran ilustrados en la figura 6.2 en color rojo. Por último, en esta etapa A se presenta una combinación de pronósticos con una igualdad en la ponderación. Esta etapa sirve para adquirir los pesos de los modelos de predicción para la etapa C, donde se combinan los modelos paramétricos y no paramétricos. Puesto que el periodo de tiempo de mayor interés en esta investigación es cuando tiene ocurrencia la caída bursátil del indicador, esta etapa A vendría siendo una fase donde se prueba el funcionamiento y se comprueban las diferentes técnicas de predicción, así como la técnica de descomposición DEM. Es así que esta etapa se podría considerar como el periodo dentro de la muestra con la salvedad de que para realizar la predicción de esta etapa, a los modelos de predicción se les oculta el periodo de datos a pronosticar, ya que el empleo de técnicas como SSP para el modelo no paramétrico y la predicción estática o predicción dentro de la muestra para el modelo paramétrico no resultan de utilidad para esta investigación.

B) En la segunda fase de la experimentación el periodo de predicción

transcurre del lunes 20 octubre de 2008 al martes 16 diciembre de 2008; estos datos se representan en color azul en la figura 6.2. El día de la crisis bursátil fue el miércoles 22 de octubre, con una caída del 7.54%. En este periodo de experimentación, los datos fueron desplazados hacia adelante 35 días o puntos de pronóstico a partir de los utilizados en la fase anterior, para estimar los parámetros de ambos modelos de predicción.

Los parámetros del modelo de ARIMA-ARCH fueron estimados con los datos que abarcan el período comprendido entre 01/06/2007 al 10/17/2008, un total de 350 datos. El periodo de la muestra empleado para el entrenamiento de las RNAs fue del 07/08/2002 hasta 10/17/2008, resultando en un total de 1589 observaciones, en esta fase, al igual que en la tercera, los datos utilizados para las estimaciones están representados en la figura 6.2, a través de la adición de los datos de la serie representados en la figura 6.2 en color negro y rojo.

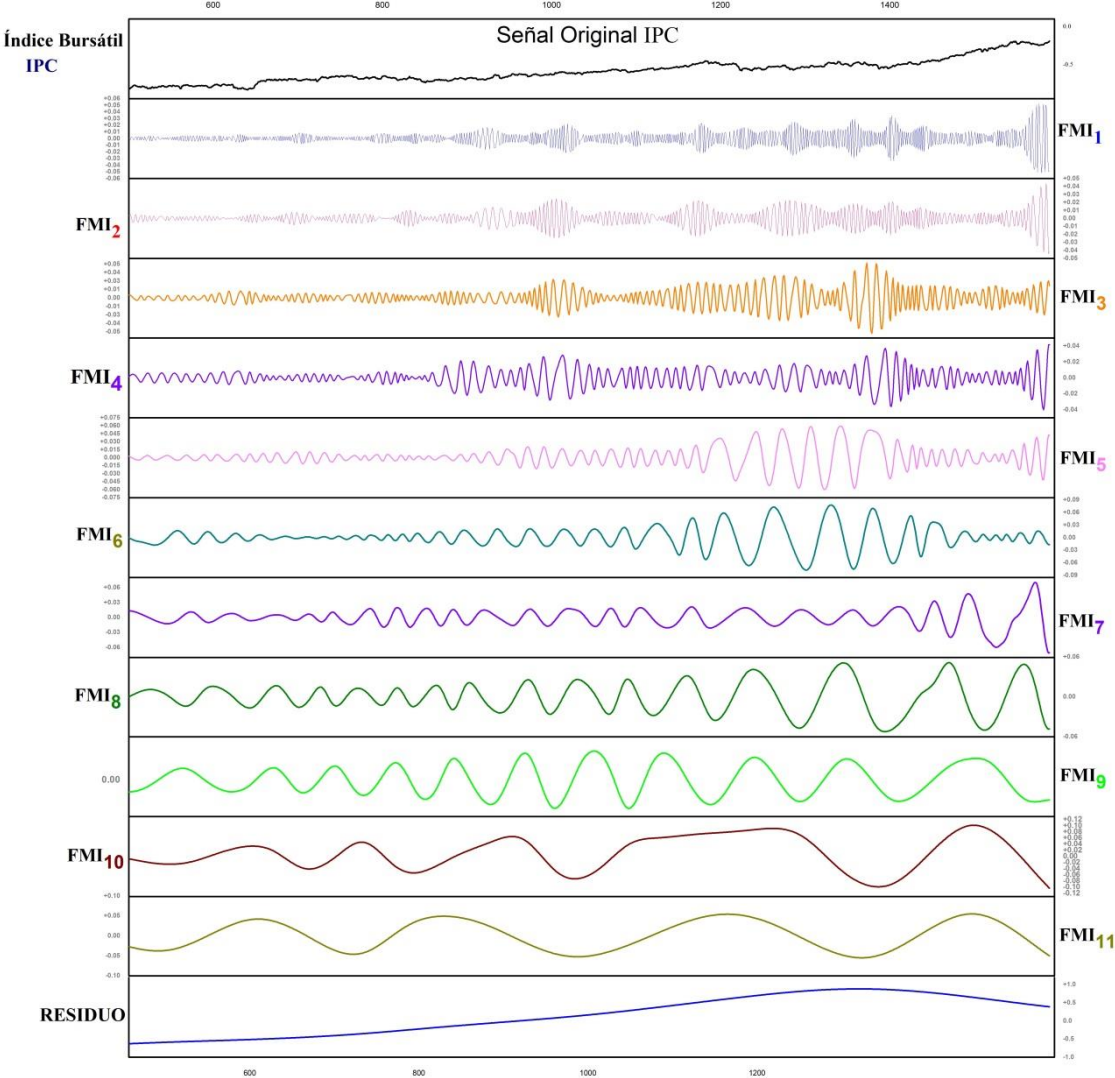
- C) En la fase final, las predicciones resultantes de los modelos paramétricos y no paramétricos se utilizaron para realizar una predicción final combinada, con el objetivo de obtener una predicción más precisa, en esta fase se analiza el mismo período de tiempo que en la experimentación anterior (Etapa B).

#### **6.4.1 Descomposición del IPC**

El enfoque clásico de descomposición y en general los métodos de descomposición surgen como una respuesta para tratar de entender y de analizar con un nivel de complejidad menor el fenómeno en cuestión, es un desglose o subdivisión de la serie. Donde se tratan de aislar los componentes de la serie para que su estudio sea más simple. En esta investigación se cuestionan los fundamentos de la descomposición clásica, ya que los componentes que la conforman y en los cuales se descomponen (tendencia + estacionalidad + ciclo + irregular) son elementos teóricos y como tal el análisis se encuentra sujeto a ciertas restricciones, es decir, los datos que se analizan se tienen que ajustar al modelo propuesto. La descomposición de una ST mediante un enfoque clásico solo tiene aplicación con fines didácticos ya que en

la práctica descomponer una ST para predecir por separado cada elemento para luego ser unido no suele dar buenos resultados (Hanke y Reitsch, 1986). El proceso generador de datos bajo el que se construyen los indicadores de los diferentes mercados de valores, implican una gran complejidad puesto que en él influyen muchos factores tanto cuantitativos como cualitativos. Los índices representan la interacción y combinación de diversos procesos, hay muchas variables que inciden en el mercado de valores, factores económicos, financieros, políticos, decisiones racionales e irracionales de los inversionistas.

Figura 6. 3: IPC del 08/07/2002 al 13/03/2009, y sus descomposiciones.



Fuente: Elaboración propia.

Como anteriormente se mencionó la muestra de datos a la cual se le aplica la

técnica DEM es la correspondiente al precio de cierre en el periodo que transcurre entre el 8 de julio de 2002 al 16 de diciembre de 2008 en su modalidad diaria, la suma de datos es de 1629 datos del Índice de Precios y Cotizaciones (IPC) de la Bolsa Mexicana de Valores (BMV). La figura 6.3 muestra la serie original del IPC, así como las 12 descomposiciones resultantes de aplicar la técnica DEM a dicha serie, de la descomposición se obtuvieron 11 FMIs y un elemento residual. La figura 6.3 ilustra como la primera FMI se distingue por contener oscilaciones rápidas y conforme avanza el proceso de cernido, las FMIs que se van adquiriendo incluyen oscilaciones gradualmente más lentas hasta que se alcanza el residuo que tiene como característica esencial ser una función monótona.

#### **6.4.2 Análisis de los datos del IPC con el exponente de Hurst**

Existe una gran cantidad de métodos de predicción algunos de ellos fueron creados específicamente para capturar algún patrón en particular que caractericé a una serie de tiempo como por ejemplo un modelo SARIMA se especializa en ST donde el patrón o rasgo distintivo es un proceso estacional, otras herramientas son capaces de manejar ST con diversas características de una manera adecuada como son las RNAs, ya que dicho modelo es adaptativo y con la disposición de aprender y puede ser utilizado en el pronóstico de casi cualquier serie de tiempo. Bajo el contexto del mercado de capitales cuando una serie presenta características de ser un paseo aleatorio, se recomienda utilizar otra estrategia como los que utilizan el cálculo de probabilidades. Es por ello que un modelo predictivo que se basa en información de su pasado encuentra su capacidad limitada a diversos factores entre ellos están el desconocimiento o conocimiento limitado de la estructura del fenómeno a predecir, así como también una deficiente especificación en los parámetros del sistema que se analiza y el ingrediente principal los datos que se pretenden analizar.

Lo anteriormente mencionado viene a colación puesto que el exponente de Hurst resulta una herramienta útil para obtener más información del fenómeno, ya que mediante está es posible comprender la estructura que contiene y se determina el grado de complejidad al que se puede enfrentar, y con ello utilizar la herramienta de

predicción que mejor se ajuste a las circunstancias y a los datos. Una herramienta que auxilia en el reconocimiento y caracterización del dinamismo de una ST es el exponente de Hurst, dotando de la capacidad de discernir si una ST es aleatoria o si no lo es y muestra rasgos característicos que ayudan en la descripción del comportamiento del mercado financiero. La información que se analiza por medio del exponente de Hurst es el IPC en el periodo del 22/08/2008 al 16/12/2008, con sus respectivas descomposiciones, ya que es el periodo utilizado para realizar las predicciones. Lo anteriormente descrito es ilustrado en la tabla 6.1.

Tabla 6. 1: Exponente de Hurst para el IPC y sus descomposiciones.

Índice de Precios y Cotizaciones (IPC)		
IPC	H = 0.4943	Antipersistente
FMI <sub>1</sub>	H = 0.0569	Antipersistente
FMI <sub>2</sub>	H = 0.1447	Antipersistente
FMI <sub>3</sub>	H = 0.3674	Antipersistente
FMI <sub>4</sub>	H = 0.5491	persistente
FMI <sub>5</sub>	H = 0.609	persistente
FMI <sub>6</sub>	H = 0.9124	persistente
FMI <sub>7</sub>	H = 0.896	persistente
FMI <sub>8</sub>	H = 0.9368	persistente
FMI <sub>9</sub>	H = 0.853	persistente
FMI <sub>10</sub>	H = 0.8129	persistente
FMI <sub>11</sub>	H = 0.7571	persistente
Residuo	H = 0.8652	persistente

Fuente: Elaboración propia.

Evaluando H del IPC, el de sus correspondientes FMIs y el elemento residual, es posible inferir que el IPC muestra un comportamiento complejo de predecir, que es antipersistente, que existe un movimiento discontinuo de lapsos positivos y negativos. Aunado a lo anteriormente mencionado, de todos los elementos analizados en la tabla 6.1, la señal o la ST que más se aproxima a un proceso aleatorio es la ST del IPC. En cambio las FMIs y el elemento residual pertenecientes al IPC presentan en general una conducta que perdura en el largo plazo, de los 12 elementos nueve de ellos son persistentes y tres son antipersistentes.

#### 6.4.3 Experimentación del IPC



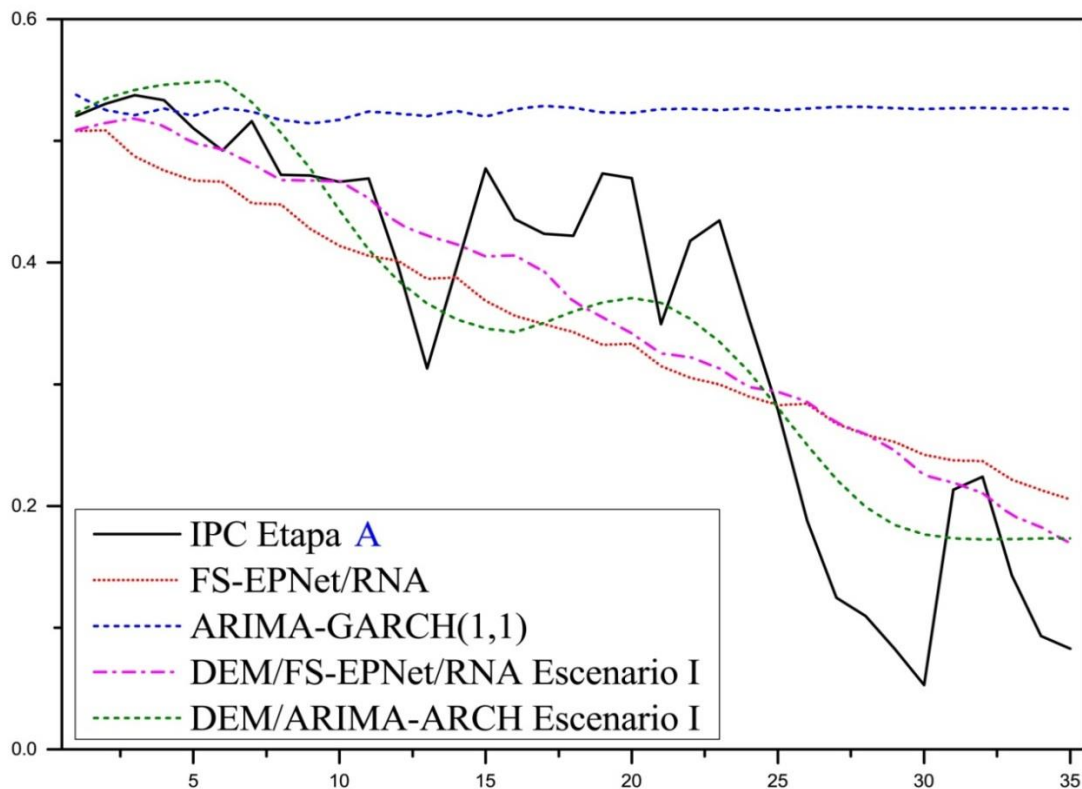
### 6.4.3.1 Etapa de experimentación A

La etapa A, cubre un periodo de pronóstico fuera de la muestra del 29 agosto del 2008 al 17 de octubre de 2008.

#### 6.4.3.1.1 Escenario I

En el primer escenario se comparan los cuatro modelos anteriormente descritos. En la figura 6.4 se muestra el desempeño de los diferentes modelos propuestos en el primer escenario.

Figura 6. 4: Etapa A del IPC. Escenario I. Periodo del 29/08/2008 al 17/10/2008, 35 días de pronóstico.



Fuente: Elaboración propia.

La figura 6.4 ilustra como el modelo paramétrico ARIMA-GARCH (1,1), es el único modelo con un desempeño deficiente de entre todos los modelos propuestos en este escenario. Es visualmente evidente que los otros tres modelos tienen un rendimiento similar. La tabla 6.2 confirma lo que la figura 6.4 ilustra, que los modelos mejoran su rendimiento, utilizando la técnica DEM. El modelo paramétrico DEM/ARIMA-ARCH, es el modelo que se beneficia en mayor medida con la incorporación del empleo de las FMIs y el residuo del IPC para realizar su predicción.

Sin embargo el modelo DEM/FS-EPNet/RNA es el modelo con un error menor.

Tabla 6. 2: Medición de errores: IPC Etapa A. Escenario I del 29/08/2008 al 17/10/2008.

CRITERIO	ARIMA-GARCH(1,1)	FS-EPNet/RNA	DEM/ARIMA-ARCH	DEM/FS-EPNet/RNA
MSE	0.05321	0.00816	0.00605	<b>0.00441</b>
RMSE	0.23067	0.09036	0.07779	<b>0.0664</b>
NRMSE	1.475	0.57783	0.4974	<b>0.42461</b>

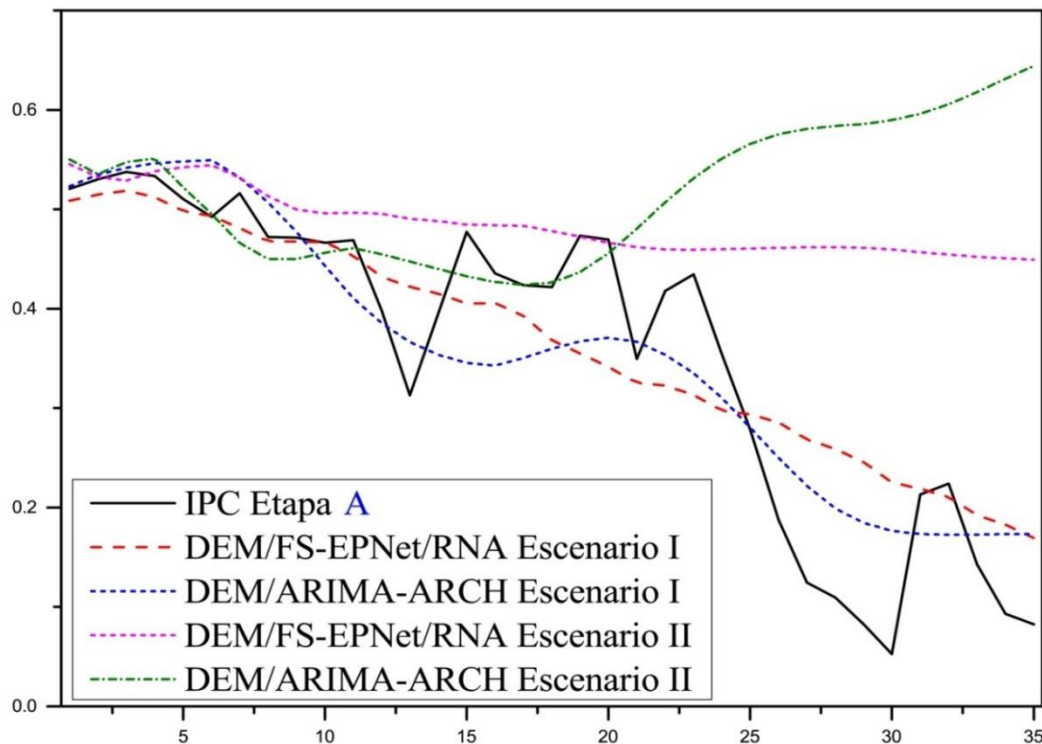
Fuente: Elaboración propia.

#### 6.4.3.1.2 Escenario II

El segundo escenario cubre el mismo periodo de predicción que el Escenario I. Los dos objetivos que se pretenden resolver en este escenario son:

- Determinar si existe diferencia significativa en el enfoque de descomposición en cuanto al periodo en que los datos son descompuestos mediante la técnica DEM.
- Establecer basado en el desempeño de los de los modelos de predicción, los pesos en las predicciones de los modelos que son empleados en la última etapa de pronóstico (la combinación ponderada de las predicciones).

Figura 6. 5: Etapa A del IPC. Escenario II. Periodo del 29/08/2008 al 17/10/2008, 35 días de pronóstico.



Fuente: Elaboración propia.

La figura 6.5 muestra el comportamiento de los diferentes modelos de predicción

propuestos en el presente escenario, así como también la serie original del IPC. A través de una inspección visual a la figura 6.5, es claramente evidente que todos los modelos tienen un rendimiento similar en un lapso de aproximadamente 20 días. De igual forma, la tabla 6.3 indica y reafirma lo que es evidente en la figura 6.5, en cuanto al rendimiento en general para los 35 periodos de los modelos predictivos provenientes del Escenario I, donde los criterios de error indican un rendimiento muy superior sobre los modelos predictivos del escenario II.

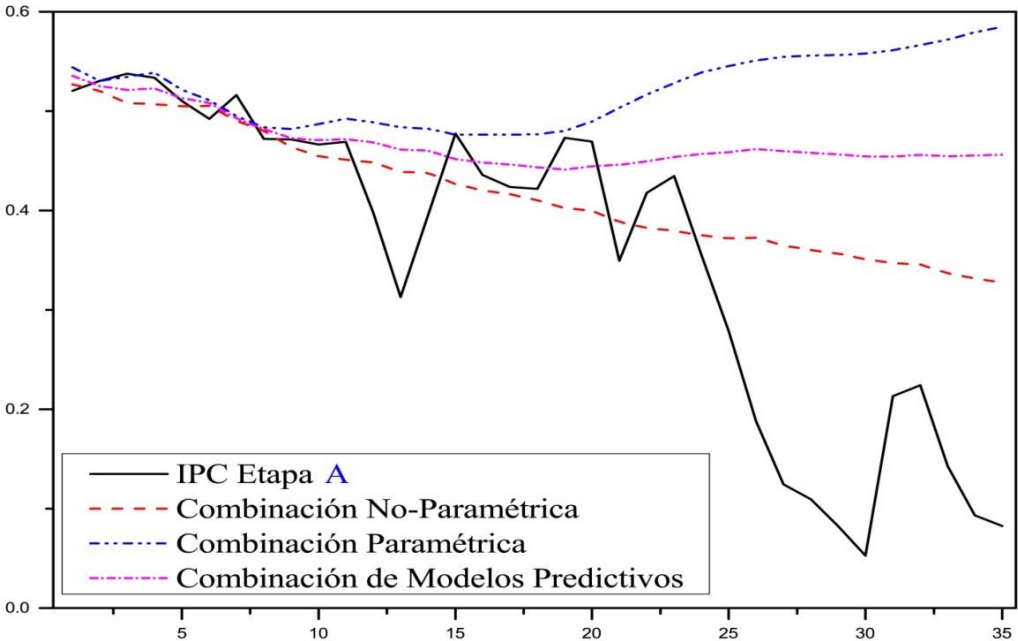
Tabla 6. 3: Medición de errores: IPC Etapa A. Escenario II del 29/08/2008 al 17/10/2008.

CRITERIO	DEM/ARIMA-ARCH Escenario I	DEM/FS-EPNet/RNA Escenario I	DEM/ARIMA-ARCH Escenario II	DEM/FS-EPNet/RNA Escenario II
MSE	<b>0.00605</b>	<b>0.00441</b>	0.03482	0.06957
RMSE	<b>0.07779</b>	<b>0.06640</b>	0.18660	0.26377
NRMSE	<b>0.49740</b>	<b>0.42461</b>	1.19318	1.68665

Fuente: Elaboración propia.

Por lo tanto, la evidencia empírica confirma la suposición de que utilizar la técnica DEM para descomponer la totalidad de los datos (muestra para construir los modelos y muestra para probar los modelos predictivos), conduce hacia pronósticos con un mayor grado de exactitud, creando de esta forma falsas expectativas sobre la calidad y el desempeño sobresaliente de los modelos predictivos.

Figura 6. 6: Combinación de pronósticos. Etapa A del IPC. Escenario II. Del 29/08/2008 al 17/10/2008.



Fuente: Elaboración propia.

Por último, con el fin de probar la posibilidad de que una combinación de pronósticos

puede conducir a un rendimiento superior en el pronóstico final, en esta etapa se propone una combinación igualitaria de pesos de los modelos (promedio simple), excluyendo los modelos provenientes del Escenario I que trabajaron en conjunción con la técnica de DEM. La figura 6.6 muestra un esquema de combinación de tres modelos predictivos. La primera mezcla es la Combinación No Paramétrica que incluye al modelos FS-EPNet/RNA proveniente del primer escenario y al modelo DEM/FS-EPNet/RNA del segundo escenario. El segundo modelo combina al modelo paramétrico utilizado en el primer escenario ARIMA-GARCH(1,1) y al modelo DEM/ARIMA-ARCH del segundo escenario, el resultado de esta combinación produce la Combinación Paramétrica. La tercera y última combinación, llamada Combinación de Modelos Predictivos, incluye la Combinación No-Paramétrica y la Combinación Paramétrica.

La tabla 6.4 muestra el comportamiento de las predicciones en un esquema de combinación con pesos igualitarios. La tabla 6.4 indica, que para los 11 periodos la predicción con un mejor desempeño es la Combinación de Modelos Predictivos. Sin embargo para los 35 periodos la Combinación No-Paramétrica es la predicción que posee el error más bajo en la Etapa A de esta investigación.

Tabla 6. 4: Combinación de pronósticos. Medición de errores: IPC Etapa A.

<b>11 Periodos de Predicción</b>			
CRITERIO	Combinación No-Paramétrica	Combinación Paramétrica	Combinación de Modelos Predictivos
MSE	0.000285	0.000217	<b>0.000156</b>
RMSE	0.016879	0.014747	<b>0.012503</b>
NRMSE	0.114893	0.10645	<b>0.080829</b>
<b>35 Periodos de Predicción</b>			
CRITERIO	Combinación No-Paramétrica	Combinación Paramétrica	Combinación de Modelos Predictivos
MSE	<b>0.015916</b>	0.060290	0.033705
RMSE	<b>0.126158</b>	0.245541	0.183589
NRMSE	<b>0.806689</b>	1.570050	1.173912

Fuente: Elaboración propia.

#### 6.4.3.2 Etapa de experimentación B

En esta etapa, la predicción comienza el lunes 20 octubre de 2008, dos días antes de la caída de la bolsa que aconteció el miércoles 22 de octubre de 2008, hasta el miércoles 13 de marzo de 2009 un total de 40 puntos de pronóstico. La tabla 6.5, muestra los parámetros utilizados para cada descomposición por los modelos paramétrico y no paramétrico.

Tabla 6. 5: IPC Etapa B. Parámetros de los modelos Paramétrico y No Paramétrico.

Modelo No-	Parámetros del Modelo	Modelo Paramétrico	Parámetros del Modelo
FMI <sub>1</sub>			
FS-EPNet/RNA	Entradas:6; Ocultas:5; Conexiones:14; Retardos:6;	ARIMA-GARCH(1,1)	p=AR(1)AR(2),d=0,q=0
FMI <sub>2</sub>			
FS-EPNet/RNA	Entradas:3; Ocultas:8; Conexiones:24;Retardos:2;	ARIMA-ARCH(2)	p=AR(6), d=0,q=MA(1)MA(2)MA(3) MA(4)MA(5)MA(6)
FMI <sub>3</sub>			
FS-EPNet/RNA	Entradas:2; Ocultas:6; Conexiones:18; Retardos:1;	ARIMA-GARCH(1,1)	p=AR(1)AR(2)AR(4)AR(5), d=0,q=MA(1)MA(2)MA(3)
FMI <sub>4</sub>			
FS-EPNet/RNA	Entradas:3; Ocultas:5; Conexiones:16; Retardos:1;	ARIMA-GARCH(1,1)	p=AR(1)AR(3)AR(4)AR(5)AR(6), d=0,q=MA(1)MA(3)MA(4)
FMI <sub>5</sub>			
FS-EPNet/RNA	Entradas:4; Ocultas:16; Conexiones:70; Retardos:7;	SARIMA	(p=AR(3)AR(6)AR(9),d=0,q=MA(1)MA(2) MA(3))x(P=SAR(1)SAR(2),D=0,Q=SMA(2))
FMI <sub>6</sub>			
FS-EPNet/RNA	Entradas:2; Ocultas:12; Conexiones:48; Retardos:2;	ARMA	p=AR(1)AR(2)AR(3)AR(5)AR(7), d=0,q=MA(1)MA(2)MA(7)MA(8)
FMI <sub>7</sub>			
FS-EPNet/RNA	Entradas:2; Ocultas:15; Conexiones:59; Retardos:4;	ARIMA-ARCH(3)	p=AR(1)AR(7), d=0,q=MA(1)MA(2)MA(3)
FMI <sub>8</sub>			
FS-EPNet/RNA	Entradas:2; Ocultas:20; Conexiones:78; Retardos:2;	SARIMA	(p=AR(1)AR(2)AR(4),d=0,q=MA(3)MA(4) MA(19))x(P=SAR(1),D=1,Q=SMA(1))
FMI <sub>9</sub>			
FS-EPNet/RNA	Entradas:2; Ocultas:9; Conexiones:33; Retardos:4;	ARIMA	p=AR(1)AR(2)AR(3),d=1, q=MA(1)MA(2)
FMI <sub>10</sub>			
FS-EPNet/RNA	Entradas:3; Ocultas:8; Conexiones:24; Retardos:4;	ARIMA	p=AR(1)AR(2) AR(3)AR(5) AR(7)AR(8),d=1,q=MA(1)MA(4)
FMI <sub>11</sub>			
FS-EPNet/RNA	Entradas:4; Ocultas:17; Conexiones:80; Retardos:5;	SARIMA	(p=AR(1)AR(3)AR(7),d=1,q=MA(1) MA(2) MA(3))x(P=SAR(1),D=0,Q=0)
Residuo			
FS-EPNet/RNA	Entradas:6; Ocultas:14; Conexiones:60; Retardos:5;	ARIMA	p=AR(1)AR(4),d=2, q=MA(1)MA(2)MA(3)MA(4)

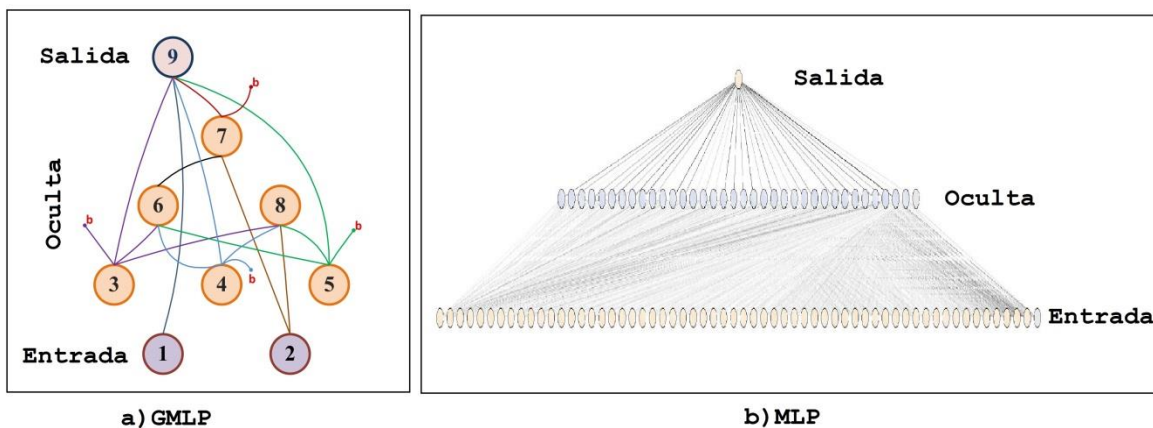
Fuente: Elaboración propia.

La tabla 6.5 muestra que las descomposiciones FMI<sub>1</sub>, FMI<sub>2</sub>, FMI<sub>3</sub>, FMI<sub>4</sub> y FMI<sub>7</sub> se encuentran definidas y regidas por un proceso ARCH. Por otro lado, el modelo SARIMA se emplea en la FMI<sub>5</sub>, FMI<sub>8</sub> y la FMI<sub>11</sub>, esto debido a que sus procesos

tienen características de estacionalidad. La FMI<sub>6</sub> es conducida por un modelo ARMA debido a que no se le aplicó diferenciación, y en última instancia las FMI<sub>9</sub>, FMI<sub>10</sub> y el residuo se modelaron mediante el modelo ARIMA.

Por otra parte, la experimentación del modelo no paramétrico demuestra la alta flexibilidad que poseen las RNAs para manejar las cambiantes facetas y capturar los patrones subyacentes de las diferentes descomposiciones derivados de la ST del IPC. Además, el algoritmo FS-EPNet demuestra su utilidad, al producir redes muy compactas. Como el diseñado para pronosticar la FMI<sub>3</sub>, con tan sólo dos entradas, seis neuronas en la capa oculta y 18 conexiones, dicha edificación topológica se muestra en la figura 6.7a. La red diseñada de mayores dimensiones, es la responsable de predecir la FMI<sub>8</sub>, con dos entradas, 21 neuronas ocultas y 88 conexiones. Otras investigaciones como la mencionada anteriormente en el cuarto capítulo en el apartado de los antecedentes, en el trabajo realizado por Zorin y Borisov (2002), se empleó una arquitectura neuronal con: 59 neuronas de entrada y 35 neuronas ocultas, una ejemplificación de las dimensiones que tendría una red de tal envergadura se exhibe en la figura 6.7b.

Figura 6. 7: a) Diseño topológico de un GMLP mediante FS-EPNet para pronosticar la FMI<sub>3</sub> del IPC. b) Ejemplo de un MLP con 59 neuronas de entrada, 35 neuronas ocultas.

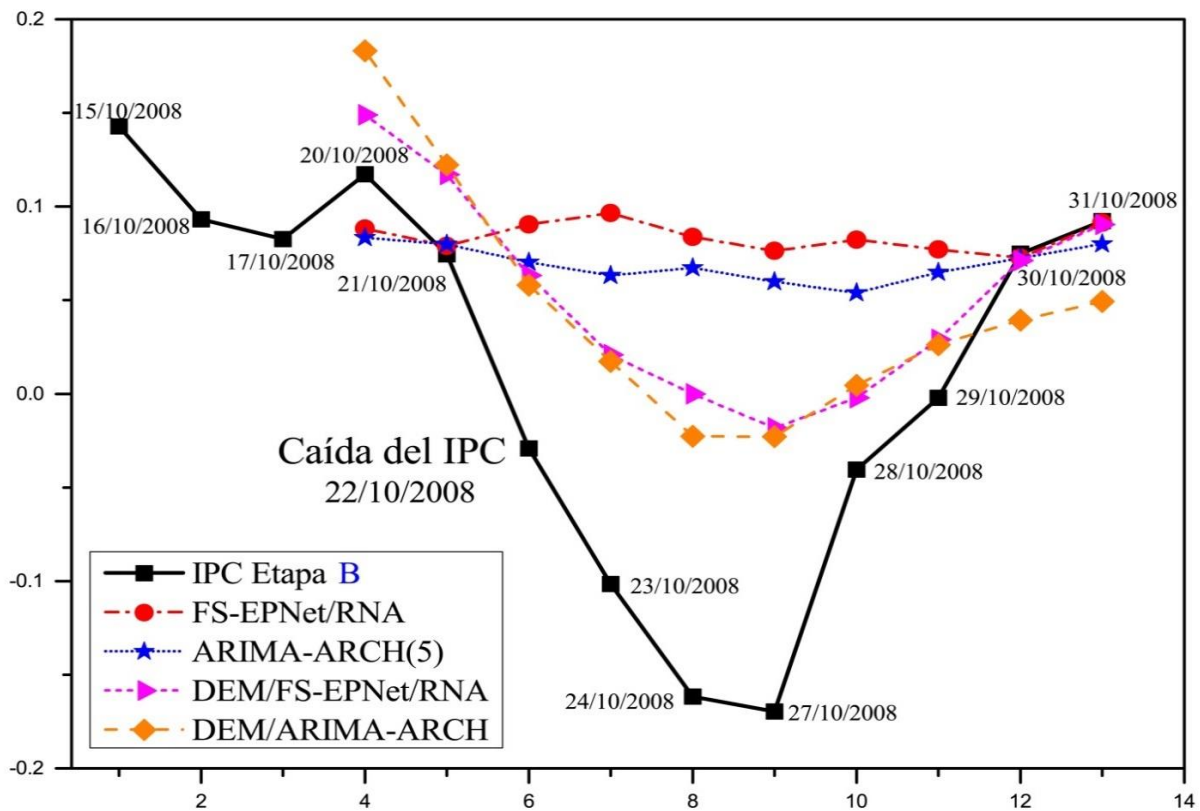


Fuente: Elaboración propia.

Como se puede apreciar hay una gran diferencia entre ambos diseños arquitectónicos, a pesar de que no resuelven la misma tarea estas dos RNAs, esta comparación puede otorgar un atisbo de las dimensiones tan compactas que puede lograr diseñar el algoritmo FS-EPNet. La figura 6.8 muestra en detalle el periodo de

tiempo del 15 al 31 octubre de 2008, un intervalo de 13 días, donde se muestra el comportamiento del IPC, así como también se muestra la evolución de las diferentes predicciones de los modelos paramétricos y no paramétricos del 20/10/2008 al 31/10/2008, las predicciones abarcan un período de 10 días.

Figura 6. 8: IPC del 15/10/2008 al 31/10/2008, 13 días. Predicción del 20/10/2008 al 31/10/2008,



Fuente: Elaboración propia.

La tabla 6.6 muestra el rendimiento de las predicciones de los diferentes modelos paramétricos y no paramétricos del 20/10/2008 al 31/10/2008, un periodo de 10 días. Los modelos que trabajaron en conjunción con la técnica DEM son aquellos que muestran evidencia de un mejor rendimiento, y donde, por el contrario, el modelo que tiene el menor rendimiento es el modelo no paramétrico FS-EPNet/RNA, que basó su predicción sólo con los datos provenientes de la ST del IPC.

Tabla 6. 6: Medición de errores: IPC Etapa B. 10 días del 20/10/2008 al 31/10/2008

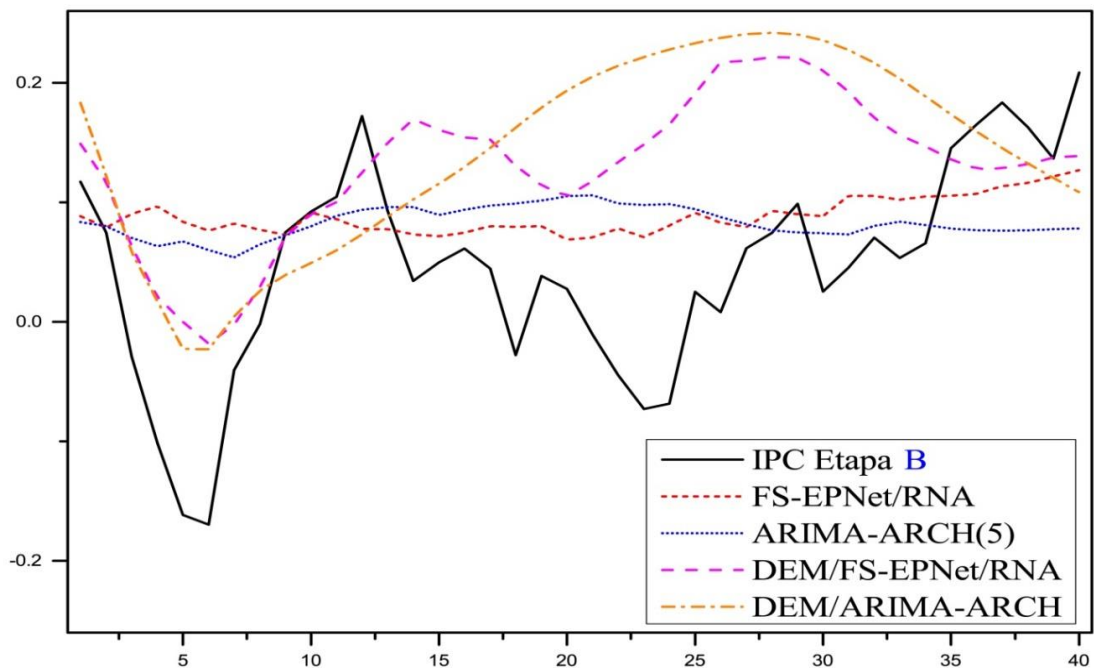
CRITERIO	ARIMA-ARCH	FS-EPNet/RNA	DEM/ARIMA-ARCH	DEM/FS-EPNet/RNA
----------	------------	--------------	----------------	------------------

MSE	0.01569	0.01965	<b>0.00751</b>	0.00778
RMSE	0.12525	0.14017	<b>0.08668</b>	0.08820
NRMSE	1.08163	1.21051	<b>0.74851</b>	0.76169

Fuente: Elaboración propia

La figura 6.8 muestra la evolución del IPC y el comportamiento de las predicciones de los modelos paramétricos y no paramétricos que produjeron su predicción solamente con la serie del IPC, así como también los modelos que trabajaron de manera conjunta con la herramienta DEM, para un periodo de 40 puntos de pronóstico. La caída del IPC se produjo el 22 de octubre de 2008. La predicción comienza 2 días antes de la caída del mercado de valores antes mencionada y continua hasta 37 días posteriores.

Figura 6. 9: IPC Etapa B de experimentación, predicción del 20/10/2008 al 16/12/2008, 40 días.



Fuente: Elaboración propia.

La tabla 6.7 muestra el rendimiento de los modelos de predicción en diferentes puntos de predicción. En primer lugar a los 13 puntos, donde los modelos que utilizan la técnica DEM demuestran una amplia superioridad sobre los modelos que produjeron su predicción solamente con los datos del IPC originales. A los 17 pasos de predicción continua prevaleciendo la superioridad de los modelos DEM/ARIMA-



ARCH y DEM/FS-EPNet/RNA sobre los modelos que solamente basaron su predicción con la ST del IPC. La etapa 40 es el último punto de predicción, y la formula FS-EPNet/RNA es el modelo con un mejor rendimiento en general, de entre todos los modelos propuestos en esta etapa.

Tabla 6. 7: Medición de error: IPC Etapa B, para 13, 17 y 40 días de predicción, del 20/10/2008 al 16/12/2008.

<b>13 Periodos de Predicción</b>				
CRITERIO	ARIMA-ARCH(5)	FS-EPNet/RNA	DEM/ARIMA-ARCH	DEM/FS-EPNet/RNA
MSE	0.012559	0.015843	0.006688	<b>0.006410</b>
RMSE	0.112069	0.125869	0.08178	<b>0.080063</b>
NRMSE	1.022310	1.148196	0.746004	<b>0.730347</b>
<b>17 Periodos de Predicción</b>				
CRITERIO	ARIMA-ARCH(5)	FS-EPNet/RNA	DEM/ARIMA-ARCH	DEM/FS-EPNet/RNA
MSE	0.010145	0.012317	<b>0.006515</b>	0.007893
RMSE	0.100726	0.110984	<b>0.080715</b>	0.088844
NRMSE	1.049348	1.156214	<b>0.840878</b>	0.925560
<b>40 Periodos de Predicción</b>				
CRITERIO	ARIMA-ARCH(5)	FS-EPNet/RNA	DEM/ARIMA-ARCH	DEM/FS-EPNet/RNA
MSE	0.008952	<b>0.008278</b>	0.020205	0.013704
RMSE	0.094615	<b>0.090985</b>	0.142145	0.117064
NRMSE	1.091708	<b>1.049819</b>	1.640108	1.350718

Fuente: Elaboración propia.

#### 6.4.3.3 Combinación de pronósticos. Etapa C.

Los métodos de combinación propuestos y utilizados en esta etapa son: la Combinación Ponderada (CP) y el Promedio Simple (PS). La CP se basa en la estrategia propuesta por Granger y Ramanatha (1984), donde los modelos utilizan un sistema de ponderación de combinaciones con pesos muy cercanos a la igualdad y se encuentra restringida la suma de pesos de los modelos a combinar a la unidad. Por lo tanto, la medida que se eligió fue el periodo de 11 pasos hacia adelante del criterio NRMSE, ya que en dicho periodo el error de ambos modelos es muy cercano uno del otro. Las predicciones que se incluyen en la combinación ponderada son los derivados de los modelos paramétricos y no paramétricos que trabajaron en conjunto con la herramienta DEM. Para llevar a cabo la combinación, los pesos se

determinaron con base en el rendimiento de los modelos en el segundo escenario de la etapa A. La fórmula que se utilizó para determinar la ponderación de pesos es la siguiente:

EGNRMSE = Error Global NRMSE

MPNRMSE = Modelo Paramétrico–NRMSE

MNPNRMSE = Modelo No Paramétrico–NRMSE

$$EGNRMSE = MPNRMSE + MNPNRMSE \quad (6.1)$$

El valor del criterio de error NRMSE a 11 pasos en la fase A del modelo paramétrico (DEM/ARIMA-ARCH) fue de 0.14576 y para el modelo no paramétrico (DEM/FS-EPNet/RNA) fue de 0.19139. Por lo tanto, el error global se determina de la siguiente forma:

$$EGNRMSE = 0.14576 + 0.19139$$

$$EGNRMSE = 0.33716$$

PMP = Peso Modelo Paramétrico

$$PMP = 1 - \frac{MPNRMSE}{EGNRMSE} \quad (6.2)$$

El peso del Modelo Paramétrico se determina:

$$PMP = 1 - \frac{0.14576}{0.33716}$$

PMNP = El peso del Modelo No Paramétrico se precisa:

$$PMP = 1 - \frac{MNPNRMSE}{EGNRMSE} \quad (6.3)$$

El peso del Modelo No Paramétrico se determina:

$$PMNP = 1 - \frac{0.19139}{0.33716}$$

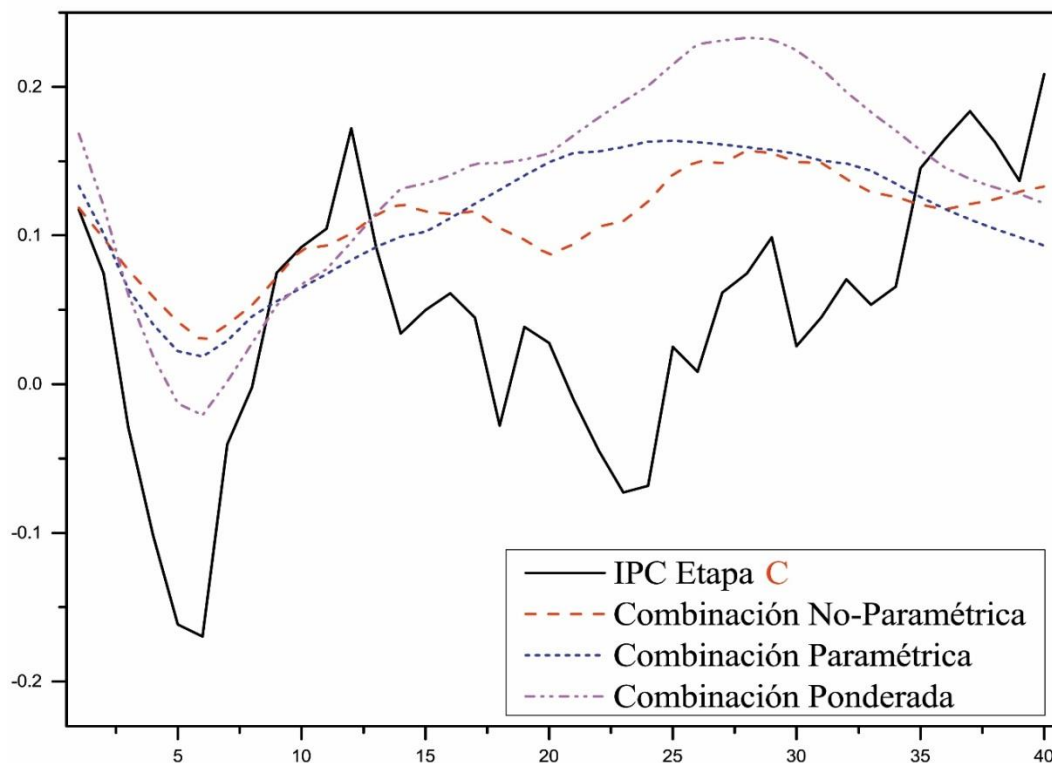
CP = Combinación Ponderada

$$CP = PMP + PMNP \quad (6.4)$$

De esta forma, cada una de las predicciones de la fórmula DEM/ARIMA-ARCH se multiplicó por 0.56766 y para la formula DEM/FS-EPNet/RNA cada pronóstico se multiplicó por 0.43234. Posteriormente ambas predicciones ponderadas son

adicionadas y se obtiene como resultado la Combinación Ponderada (CP). Adicionalmente se propone una combinación de predicciones con pesos igualitarios (PS). Por un lado, se realizó la Combinación No-Paramétrica que incluye a los modelos FS-EPNet/RNA y DEM/FS-EPNet/RNA. Por otro lado, la Combinación Paramétrica es una mezcla de los modelos ARIMA-ARCH(5) y DEM/ARIMA-ARCH. La figura 6.9 ilustra el comportamiento de los modelos: Combinación No-Paramétrica, Combinación Paramétrica y Combinación Ponderada, así como también la ST del IPC, en el mismo periodo experimental llevado a cabo anteriormente del IPC (Etapa B).

Figura 6. 10: Etapa C. Combinación de Predicciones del IPC.



Fuente: Elaboración propia.

A partir de una inspección visual de la figura 6.9 se puede apreciar que la predicción producida por la Combinación Ponderada por aproximadamente 11 días es el modelo que se ajusta mejor al periodo de la caída bursátil. La tabla 6.8 confirma lo anteriormente ilustrado en la figura 6.9, ya que muestra el rendimiento de los modelos predictivos para 5, 11 y 40 periodos de predicción. También, la tabla 6.8 expone como el modelo que mejor se ajusta a los datos en los primeros periodos

es el modelo de Combinación Ponderada. Sin embargo, para los 40 puntos de predicción, el error acumulado por la Combinación Ponderada es el mayor de entre los modelos expuestos en la Etapa C.

Tabla 6. 8: Combinación de pronósticos, Etapa C. Medición de errores: predicción del IPC para 5, 11 y 40 días del 20/10/2008 al 16/12/2008

<b>5 Periodos de Predicción</b>			
CRITERIO	Combinación No Paramétrica	Combinación Paramétrica	Combinación Ponderada
MSE	0.01577	0.01272	<b>0.00985</b>
RMSE	0.12559	0.112764	<b>0.099261</b>
NRMSE	1.02180	0.91743	<b>0.80757</b>
<b>11 Periodos de Predicción</b>			
CRITERIO	Combinación No Paramétrica	Combinación Paramétrica	Combinación Ponderada
MSE	0.01164	0.00984	<b>0.00690</b>
RMSE	0.10788	0.09917	<b>0.08307</b>
NRMSE	0.96426	0.88645	<b>0.74252</b>
<b>40 Periodos de Predicción</b>			
CRITERIO	Combinación No Paramétrica	Combinación Paramétrica	Combinación Ponderada
MSE	<b>0.00967</b>	0.01235	0.01695
RMSE	<b>0.09833</b>	0.11114	0.13021
NRMSE	<b>1.13459</b>	1.28232	1.50240

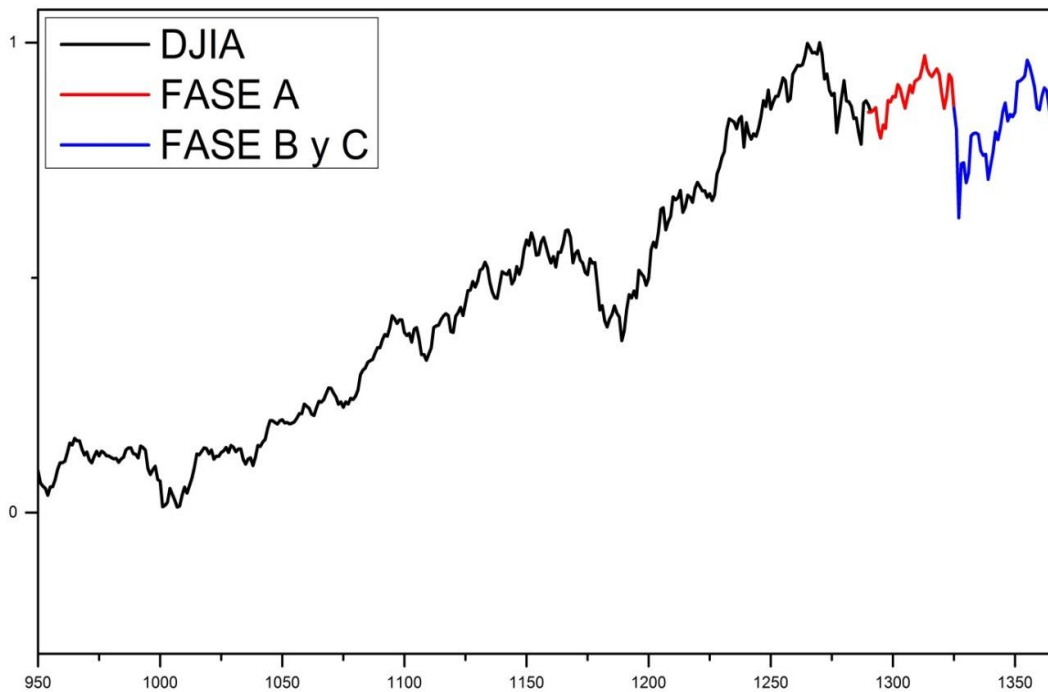
Fuente: Elaboración propia.

## 6.5 Fases experimentales del DJIA

El segundo experimento de la presente investigación trata el caso del índice bursátil del DJIA de los Estados Unidos, en un periodo donde tuvo ocurrencia un mini-crack bursátil. Este segundo estudio mantiene una estructura experimental similar al caso anterior, la experimentación del IPC. Se emplearon las mismas técnicas de predicción (MSP y MSF), así como se plantean tres etapas. La diferencia estriba en que se omite el Escenario I de la primera fase. Donde el método de descomposición DEM, se le aplicaba a la totalidad de la muestra disponible. Dado que queda demostrado que las descomposiciones sufren de perturbaciones si se descompone la totalidad de los datos y como consecuencia hay una alteración favorable en el resultado de las predicciones. La muestra de datos analizada es la correspondiente

al precio de cierre del DJIA en su forma diaria, para el lapso de tiempo que abarca del 1 de junio de 1989 al 19 de diciembre de 1997, con un total de 2165 datos del DJIA. El lunes 27 de octubre de 1997 el índice DJIA iniciaba operaciones con 7715.41 para después sufrir un descenso de 554.26 puntos, cerrando la jornada en 7161.15 puntos. La Figura 6.10 ilustra las diferentes fases de experimentación del índice DJIA.

Figura 6. 11: Fases de predicción del DJIA, del 06/01/1989 al 12/19/1997.



Fuente: Elaboración propia.

- A) La etapa A tiene como objetivo analizar el desempeño de los diferentes modelos propuestos en este estudio con el fin de obtener los pesos de los modelos en la etapa de combinación de pronósticos (fase C). En la etapa A, el modelo no paramétrico fue entrenado con un total de 2090 datos, del 06/01/1989 al 09/04/1997, los datos anteriormente mencionados se encuentran ilustrados en la figura 6.10 en color negro. Por otro lado, para el modelo paramétrico se estimaron sus parámetros con un total de 1100 datos, del 10/20/1993 al 09/04/1997. En esta etapa la predicción comprende del viernes 5 de septiembre de 1997 al jueves 23 de octubre de 1997, este conjunto de datos se presenta en la figura 6.10 en color rojo.

B) En la etapa B de experimentación la predicción cubre el periodo que abarca del viernes 24 de octubre de 1997 al viernes 19 de diciembre de 1997; dicho conjunto de datos se encuentran ilustrados en color azul en la figura 6.10. El crack bursátil aconteció el lunes 27 de octubre, con una caída porcentual del 7.18%. Al igual que en la experimentación anterior (IPC) los datos son adelantados 35 días (pronostico fuera de la muestra, etapa A), de la etapa anterior para calcular los parámetros de los modelos predictivos. Los modelos no paramétricos emplearon para su entrenamiento un total de 2125 datos, dicho lapso de tiempo comprende el periodo del 06/01/1989 al 10/23/1997.

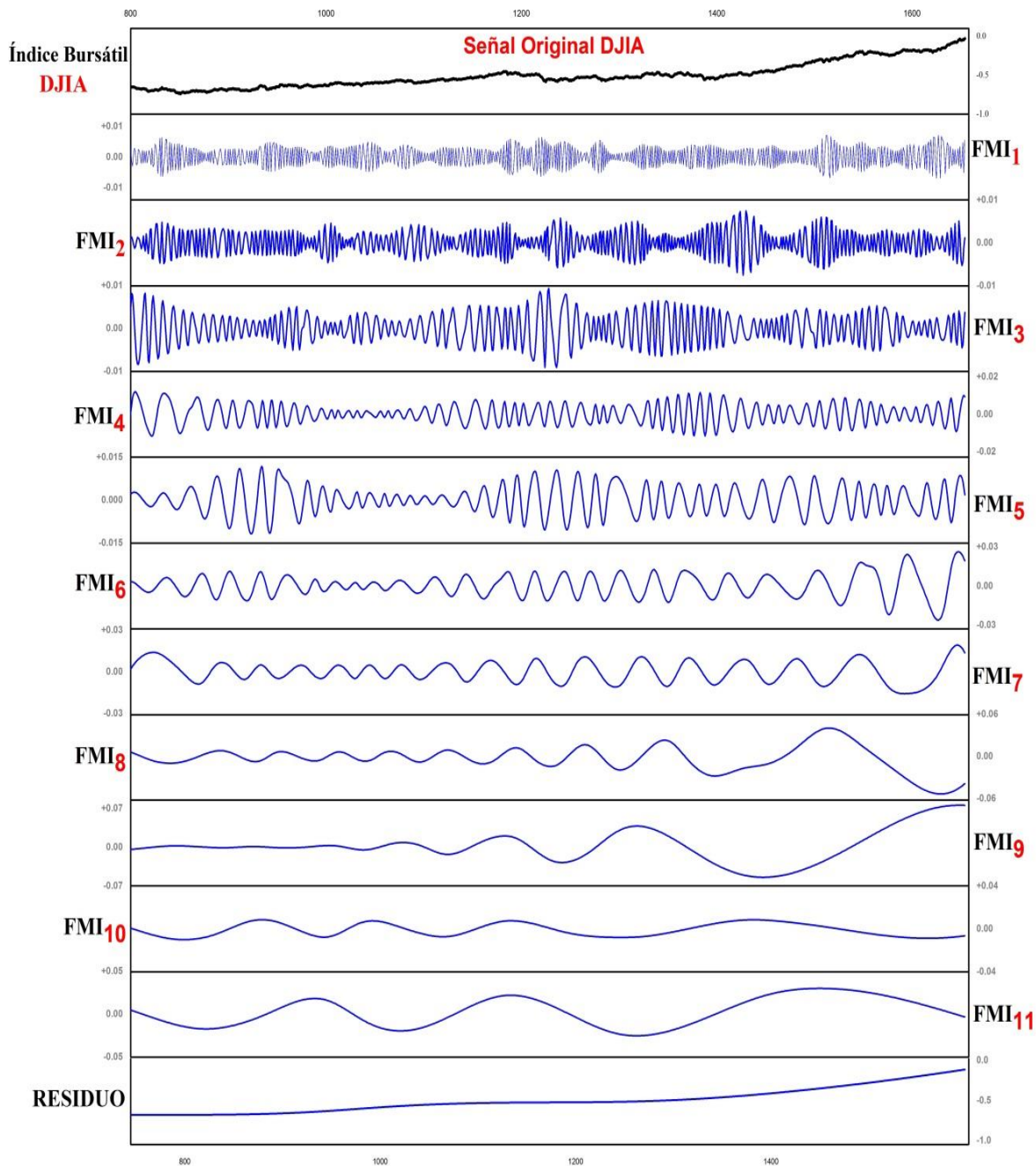
En contraste, para estimar los parámetros de los modelos ARIMA-ARCH se empleó el periodo que abarca del 6/21/1993 al 10/23/1997, registrando un total de 1100 datos. Tanto en esta etapa así como también en la etapa C, el conjunto de datos empleados para realizar las estimaciones de los diferentes parámetros de los modelos, se encuentran ilustrados en la figura 6.10, mediante la suma de los datos de la Serie de Tiempo (ST) escenificados en la figura 6.10 en color negro y rojo.

C) En la etapa C, las predicciones emanadas de la formula paramétrica DEM/ARIMA-ARCH y del modelo no paramétrico DEM/RNA/FS-EPNet se combinan para realizar una combinación de pronósticos. En esta etapa se analiza el mismo lapso de tiempo que en la etapa B.

### **6.5.1 Descomposición del DJIA**

Para obtener las descomposiciones de la ST del DJIA se introdujeron a la herramienta DEM un total de 2090 datos para la primera etapa, y un total de 2125 datos para las etapas B y C. Correspondientes al precio de cierre del DJIA en su modalidad diaria. Como consecuencia de aplicar la herramienta DEM a la ST del DJIA, se obtuvieron 11 FMIs y un residuo, las descomposiciones así como la serie original son ilustradas en la figura 6.11.

Figura 6. 12: DJIA del 06/01/1989 al 10/23/1997, y sus descomposiciones



Fuente: Elaboración propia.

### 6.5.2 Análisis de los datos del DJIA con el exponente de Hurst

La tabla 6.9 muestra el resultado de la aplicación del coeficiente de Hurst a la serie original del DJIA, así como a sus respectivas descomposiciones. Se analizaron la totalidad de los datos de la muestra, es decir para cada descomposición así como para la serie original del DJIA se emplearon los 2165 datos. La serie original del

DJIA muestra un coeficiente de H que se muestra como un proceso muy cercano a la aleatoriedad. De la primera FMI a la cuarta muestran un comportamiento antipersistente, por el contrario desde la quinta FMI hasta el elemento residual muestran un comportamiento persistente. De esta manera, el 66.66% de las descomposiciones muestran que existe persistencia en los datos.

Tabla 6. 9: Exponente de Hurst para el DJIA y sus descomposiciones.

Dow Jones Industrial Average (DJIA)		
DJIA	H = 0.4867	antipersistente
FMI1	H = 0.0668	antipersistente
FMI2	H = 0.0076	antipersistente
FMI3	H = 0.1618	antipersistente
FMI4	H = 0.4967	antipersistente
FMI5	H = 0.7097	persistente
FMI6	H = 0.8949	persistente
FMI7	H = 0.9505	persistente
FMI8	H = 0.9601	persistente
FMI9	H = 0.9639	persistente
FMI10	H = 0.9901	persistente
FMI11	H = 0.9733	persistente
Residuo	H = 0.9879	persistente

Fuente: Elaboración propia.

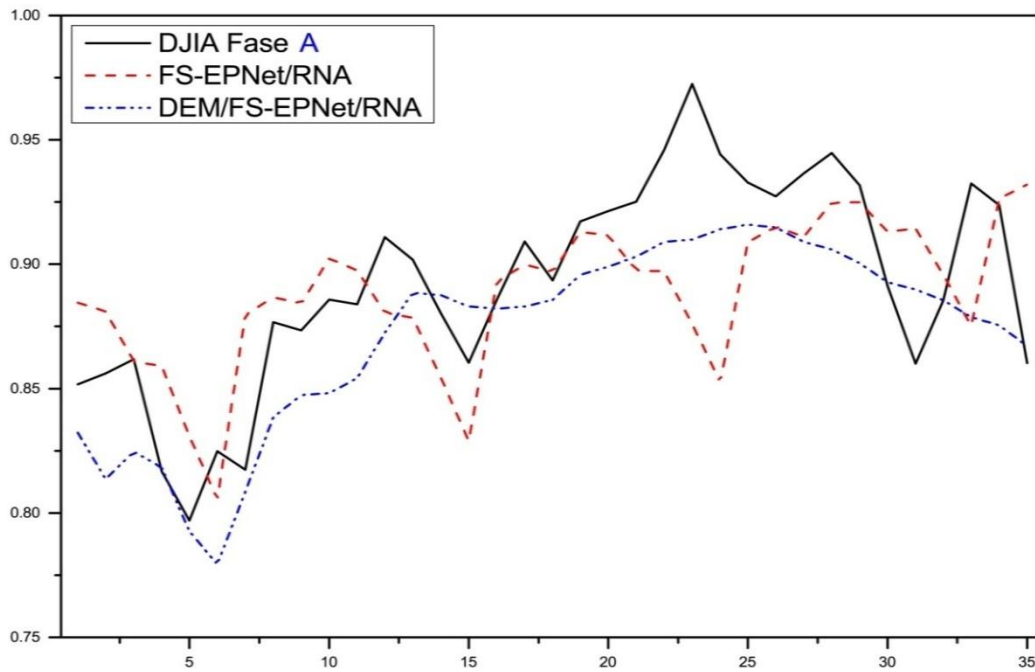
### 6.5.3 Experimentación del DJIA

#### 6.5.3.1 Fase experimental A

La fase A, parte de una muestra de 2090 observaciones del precio histórico del DJIA, para estimar los parámetros de los modelos paramétricos y no paramétricos. El horizonte para realizar la predicción es 35 periodos. Por cuestiones de apreciación se ilustran el comportamiento de los diferentes modelos predictivos de forma separada y la serie original del DJIA. La primera figura 6.12 de esta fase, muestra el desempeño de los modelos no paramétricos.



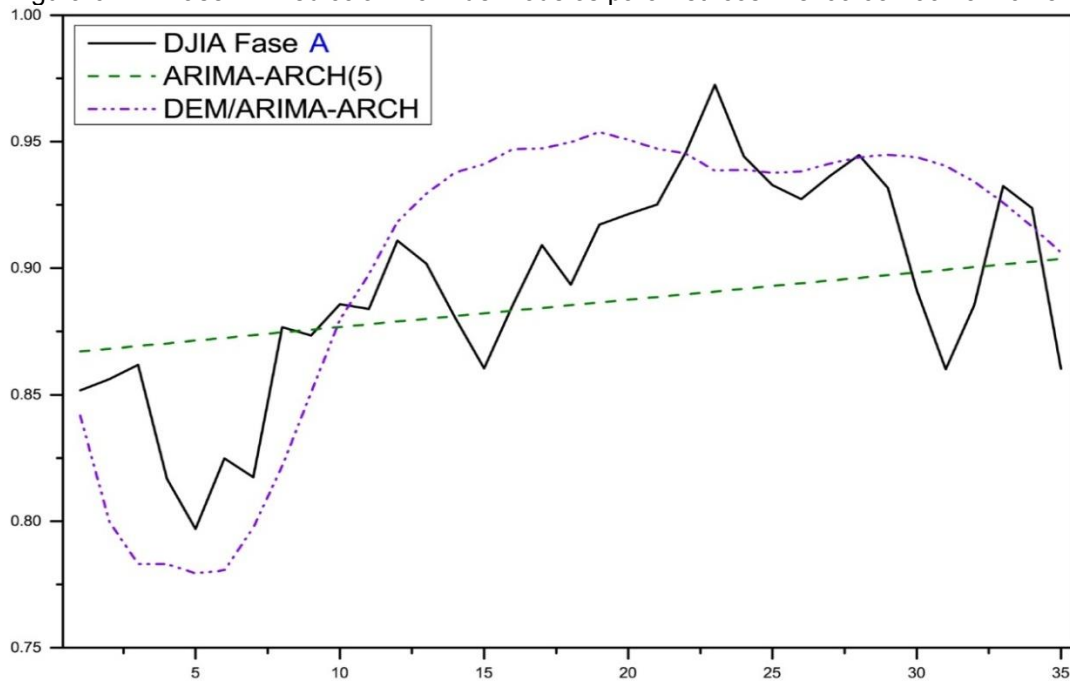
Figura 6. 13: Fase A. Predicción de los modelos no paramétricos del DJIA. Del 09/05/1997 al 10/23/1997.



Fuente: Elaboración propia.

La siguiente figura que se muestra en esta etapa, es la figura 6.13 que ilustra el comportamiento de los modelos paramétricos y la evolución del DJIA.

Figura 6. 14: Fase A. Predicción DJIA de modelos paramétricos. Del 09/05/1997 al 10/23/1997.



Fuente: Elaboración propia.

La tabla 6.10 muestra el desempeño de los diferentes modelos utilizados en esta etapa mediante los diferentes criterios de error. La tabla 6.10 indica que para los 5 y 15 periodos la fórmula que mejor se ajusta los datos es DEM/FS-EPNet/RNA. Para los 35 días el que tiene un error menor es el modelo que por medio del promedio simple combina los modelos DEM/ARIMA-ARCH y DEM/FS-EPNet/RNA, denominado como Combinación de Modelos Predictivos, mostrado en la figura 6.14.

El modelo que en las medidas de desempeño se comporta de manera menos adecuada es el modelo DEM/ARIMA-ARCH, sin embargo como se aprecia en la figura 6.13, dicho modelo se ajusta a los cambios de dirección, en contraste el modelo ARIMA-ARCH(5) bajo los criterios de desempeño no es el peor modelo. Sin embargo, cabe destacar que mediante una inspección visual de la figura 6.13 esta indica todo lo contrario, ya que es evidente el pobre desempeño del modelo paramétrico ARIMA-ARCH(5). Donde es posible apreciar como la predicción del modelo ARIMA-ARCH(5) se comporta como una línea recta que no captura el proceso subyacente, ni se ajusta a los cambios en la dirección de la serie del DJIA.

Tabla 6. 10: Error: DJIA Fase A, para 5, 15 y 35 días.

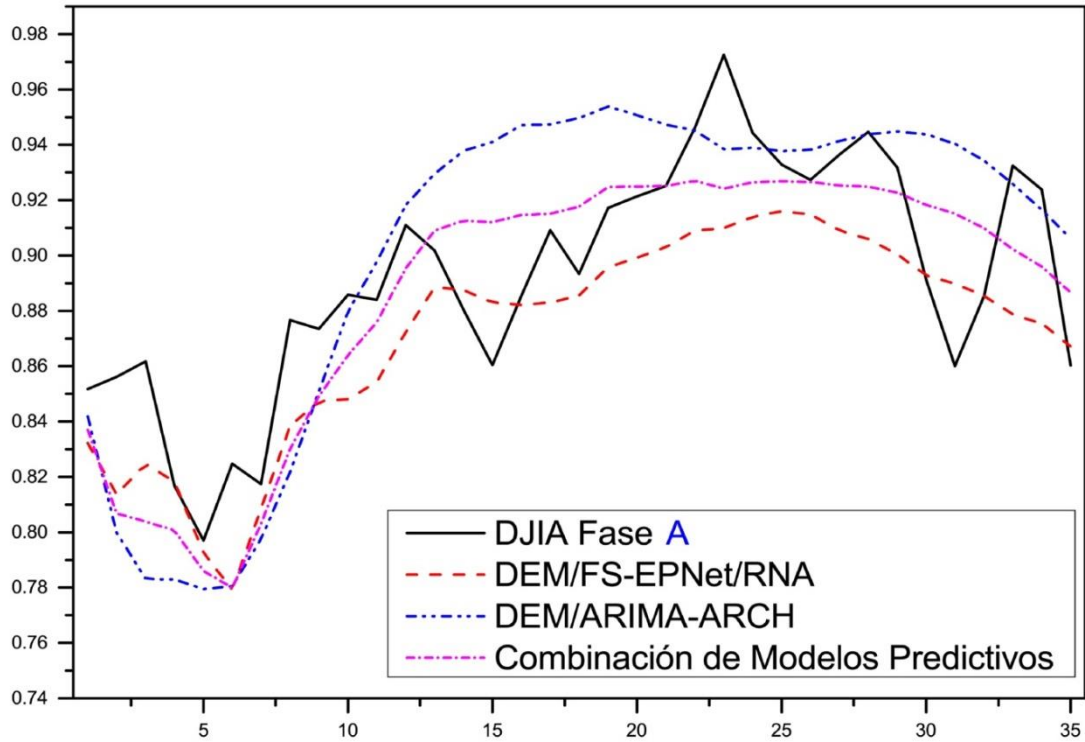
<b>5 Periodos de Predicción</b>					
CRITERIO	DEM/ARIMA-ARCH	ARIMA-ARCH(5)	FS-EPNet/RNA	DEM/FS-EPNet/RNA	Combinación de Modelos Predictivos
MSE	0.00218	0.00177	0.00093	<b>0.00072</b>	0.00127
RMSE	0.04666	0.04206	0.03048	<b>0.02675</b>	0.03570
NRMSE	0.76019	0.68522	0.49658	<b>0.43585</b>	0.58168
<b>15 Periodos de Predicción</b>					
CRITERIO	DEM/ARIMA-ARCH	ARIMA-ARCH(5)	FS-EPNet/RNA	DEM/FS-EPNet/RNA	Combinación de Modelos Predictivos
MSE	0.00184	0.00109	0.00085	<b>0.00083</b>	0.00106
RMSE	0.04291	0.03304	0.02909	<b>0.02874</b>	0.03256
NRMSE	0.93746	0.72179	0.63562	<b>0.62791</b>	0.71129
<b>35 Periodos de Predicción</b>					
CRITERIO	DEM/ARIMA-ARCH	ARIMA-ARCH(5)	FS-EPNet/RNA	DEM/FS-EPNet/RNA	Combinación de Modelos Predictivos
MSE	0.00154	0.00132	0.00136	0.00088	<b>0.00079</b>
RMSE	0.03930	0.03631	0.03687	0.02958	<b>0.02819</b>
NRMSE	0.95008	0.87791	0.89143	0.71515	<b>0.68155</b>

Fuente: Elaboración propia

Finalmente, en la figura 6.14 se exhibe el comportamiento de los modelos DEM/FS-

EPNet/RNA y DEM/ARIMA-ARCH, así como también la combinación de ambos (Combinación de Modelos Predictivos) bajo un esquema de promedio simple.

Figura 6. 15: Fase A. Combinación de pronósticos del DJIA.



Fuente: Elaboración propia.

### 6.5.3.2 Fase experimental B

La tabla 6.11, muestra los parámetros utilizados para cada descomposición por los modelos paramétricos y no paramétricos.

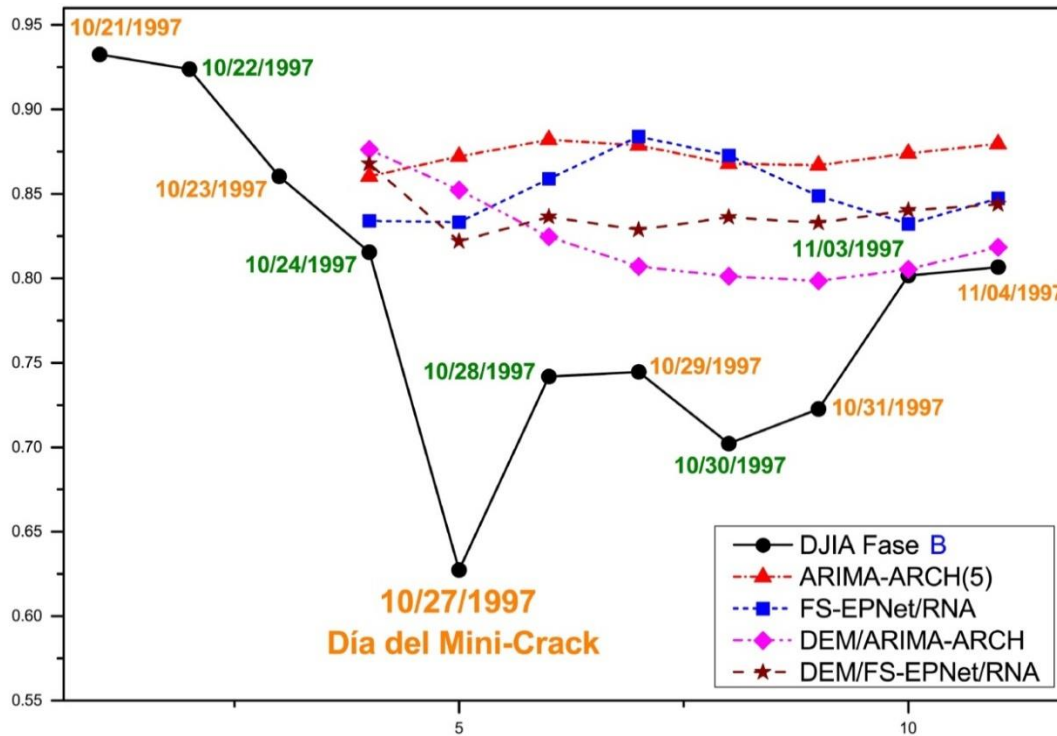
Tabla 6. 11: Parámetros: modelos Paramétrico y No-Paramétrico. Etapa B, DJIA.

Modelo No-Paramétrico	Parámetros del Modelo	Modelo Paramétrico	Parámetros del Modelo
<b>FMI<sub>1</sub></b>			
FS-EPNet/RNA	Entradas:13; Ocultas:6; Conexiones:46; Retardos:7;	ARIMA-GARCH(1,1)	p=AR(1)AR(5),d=0, q=MA(1)MA(3)
<b>FMI<sub>2</sub></b>			
FS-EPNet/RNA	Entradas:8; Ocultas:5; Conexiones:45; Retardos:6;	ARIMA-ARCH(4)	p=AR(1)AR(2)AR(4), d=0,q=MA(2)MA(5)
<b>FMI<sub>3</sub></b>			
FS-EPNet/RNA	Entradas:6; Ocultas:7; Conexiones:46; Retardos:2;	ARIMA-GARCH(1,1)	p=AR(2)AR(4)AR(6), d=0,q=MA(1)MA(3)MA(4)
<b>FMI<sub>4</sub></b>			
FS-EPNet/RNA	Entradas:2; Ocultas:14; Conexiones:43; Retardos:1;	ARIMA-GARCH(1,1)	p=AR(1)AR(2)AR(4)AR(7), d=0,q=MA(1)MA(3)MA(4)MA(5)
<b>FMI<sub>5</sub></b>			
FS-EPNet/RNA	Entradas:4; Ocultas:10; Conexiones:31; Retardos:5;	SARIMA	(p=AR(1)AR(2)AR(5)AR(12), d=0,q=MA(3)MA(4)MA(5)) x(P=0,D=1,Q=SMA(1))
<b>FMI<sub>6</sub></b>			
FS-EPNet/RNA	Entradas:3; Ocultas:11; Conexiones:35; Retardos:1;	SARIMA	(p=AR(1)AR(3)AR(4)AR(8),d=0, q=MA(3)MA(12)) x(P=SAR(1),D=1,Q=SMA(1))
<b>FMI<sub>7</sub></b>			
FS-EPNet/RNA	Entradas:3; Ocultas:17; Conexiones:74; Retardos:4;	SARIMA	p=AR(1)AR(12)AR(27),d=0, q=MA(4)MA(27))x(P=SAR(1) SAR(2),D=1,Q=SMA(1))
<b>FMI<sub>8</sub></b>			
FS-EPNet/RNA	Entradas:3; Ocultas:4; Conexiones:9; Retardos:1;	SARIMA	(p=AR(1)AR(2)AR(3)AR(4), d=0,q=MA(7)) x(P=0,D=1,Q=SMA(1)SMA(2))
<b>FMI<sub>9</sub></b>			
FS-EPNet/RNA	Entradas:3; Ocultas:9; Conexiones:35; Retardos:5;	ARIMA	p=AR(1)AR(4),d=1, q=MA(1)MA(2)
<b>FMI<sub>10</sub></b>			
FS-EPNet/RNA	Entradas:3; Ocultas:7; Conexiones:27; Retardos:5;	ARIMA	p=AR(1)AR(2)AR(5)AR(8), d=1,q=MA(1)MA(2)MA(3)
<b>FMI<sub>11</sub></b>			
FS-EPNet/RNA	Entradas:2; Ocultas:18; Conexiones:72; Retardos:5;	ARIMA	p=AR(1)AR(2)AR(3)AR(4), d=1,q=MA(1)MA(2)MA(3)
<b>Residuo</b>			
FS-EPNet/RNA	Entradas:2; Ocultas:10; Conexiones:32; Retardos:6;	ARIMA	p=AR(2)AR(3)AR(6)AR(7),d=5, q=MA(1)MA(2)MA(6)MA(8)

Fuente: Elaboración propia.

La figura 6.15 ilustra el comportamiento del DJIA en el periodo que transcurre del 10/21/1997 al 11/04/1997 un total de 11 observaciones, así como también el comportamiento de los diferentes modelos predictivos por un periodo de 8 días del 10/24/1997 al 11/04/1997.

Figura 6. 16: Fase B. Predicción 8 días, DJIA. Del 10/24/1997 al 11/04/1997.



Fuente: Elaboración propia.

La tabla 6.12 indica el desempeño de los modelos no paramétricos y paramétricos para 8 puntos de pronóstico. El modelo DEM/ARIMA-ARCH es la fórmula que se ajusta mejor a los 8 periodos de pronóstico y por el contrario el modelo paramétrico ARIMA-ARCH(5) es el que obtuvo el desempeño más pobre.

Tabla 6. 12: Errores: DJIA Fase B, a 8 días del 20/10/2008 al 31/10/2008.

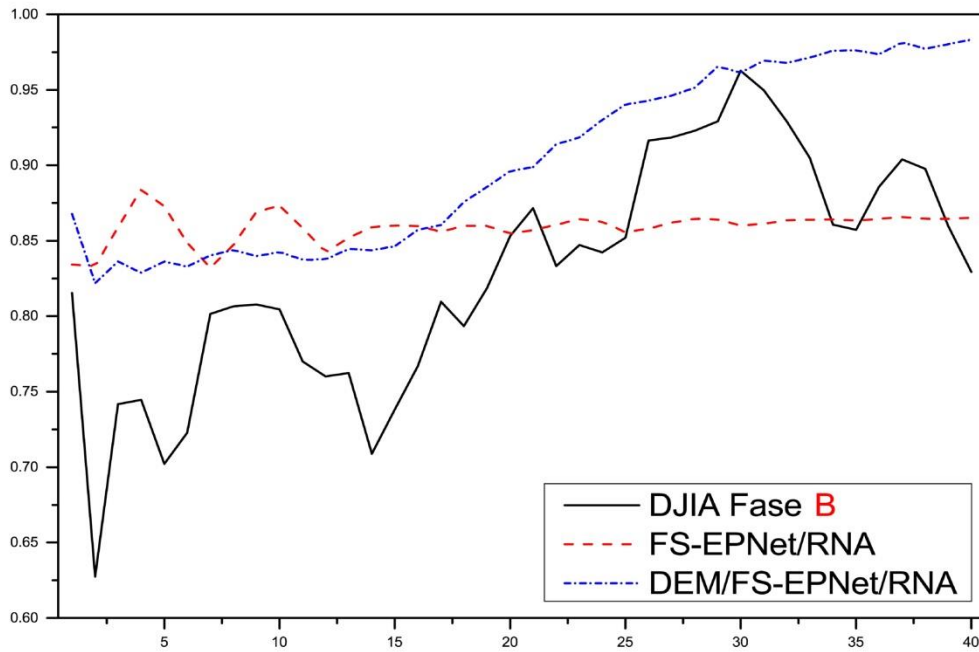
CRITERIO	ARIMA-ARCH(5)	FS-EPNet/RNA	DEM/ARIMA-ARCH	DEM/FS-EPNet/RNA
MSE	0.01666	0.01320	<b>0.00825</b>	0.00920
RMSE	0.12908	0.11488	<b>0.09081</b>	0.09593
NRMSE	1.37993	1.32004	<b>1.06741</b>	1.12423

Fuente: Elaboración propia.

La gráfica del comportamiento de las predicciones de los modelos no paramétricos, así como la evolución del DJIA en el periodo que transcurre del 10/24/1997 al

12/19/1997 un total de 40 días, se presenta en la figura 6.16.

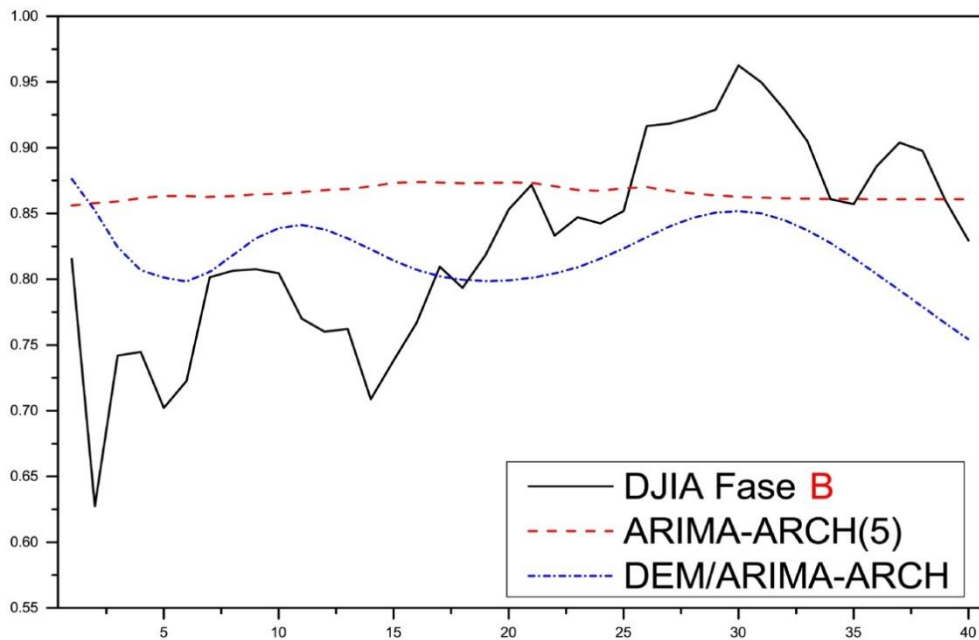
Figura 6. 17: Fase B. Predicción de modelos no paramétricos a 40 días del DJIA.



Fuente:Elaboración propia.

Los pronósticos de los modelos paramétricos y el comportamiento del DJIA son ilustrados en la figura 6.17, por el periodo de observación que abarca del 10/24/1997 al 12/19/1997 para contabilizar un total de 40 días.

Figura 6. 18: Fase B. Predicción de modelos paramétricos a 40 días del DJIA.



Fuente:Elaboración propia.

### 6.5.3.3 Fase C. Combinación de pronósticos.

Para realizar la combinación ponderada, se adquirió el peso de cada modelo tomando en cuenta el desempeño de los modelos que utilizaron la herramienta DEM en la fase A. Se utilizó la misma estrategia y ecuación 6.1 que en la experimentación del IPC. La fórmula para determinar la ponderación de pesos se presenta en la tabla 6.13.

Tabla 6. 13: Formulación de la Combinación Ponderada del DJIA.

EGNRMSE = Error Global NRMSE	
MPNRMSE = Modelo Paramétrico–NRMSE	
MNP NRMSE = Modelo No-Paramétrico–NRMSE	
EGNRMSE = MPNRMSE + MNP NRMSE	
Para determinar el peso se tomó como tasa de valoración el criterio NRMSE a 11 días de la fase A.	
MPNRMSE = DEM/ARIMA-ARCH = 0.95008	
MNP NRMSE = DEM/FS-EPNet/RNA = 0.71515	
Dando como resultado el error global:	
EGNRMSE = 0.95008 + 0.71515 = 1.66523	
<b>PMP = Peso Modelo Paramétrico</b>	
El peso del Modelo Paramétrico se determinó:	
$PMP = 1 - \frac{0.95008}{1.66523} \qquad \qquad \qquad PMP = 0.42946$	
<b>PMNP = Peso Modelo No-Paramétrico</b>	
El peso del Modelo No-Paramétrico se determinó:	
$PMNP = 1 - \frac{0.71515}{1.66523} \qquad \qquad \qquad PMNP = 0.57054$	
Y <sub>t</sub> = predicción del modelo paramétrico	X <sub>t</sub> = predicción del modelo no paramétrico
CP = Combinación Ponderada	CP = PMP(Y <sub>t</sub> ) + PMNP(X <sub>t</sub> )
CP = PMP( <b>Predicciones-DEM/ARIMA-ARCH</b> ) + PMNP( <b>Predicciones-DEM/FS-EPNet/RNA</b> )	

Fuente: Elaboración propia.

De este modo cada una de las predicciones provenientes de la formula

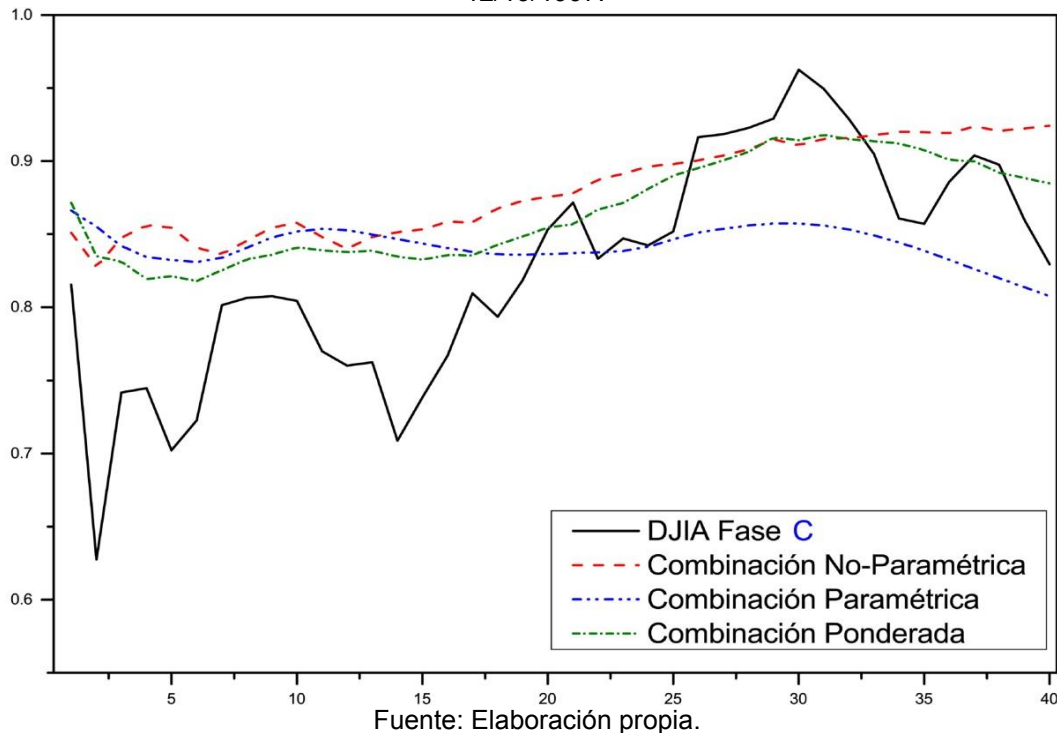
DEM/ARIMA-ARCH se multiplicó por 0.42946, por el otro lado, los pronósticos emanados del modelo DEM/FS-EPNet/RNA fueron multiplicados por 0.57054. Cabe recalcar que la suma de ambos coeficientes ( $PMP + PMNP = 1$ ) dan como resultado la unidad. Así como la unión de ambas predicciones multiplicadas por su respectivo coeficiente da como resultado la Combinación Ponderada (CP). Además con el propósito de comparar otro tipo de combinaciones se propone incluir combinaciones de pesos igualitarios (PS), que a continuación se muestran:

Combinación Paramétrica = **ARIMA-ARCH(5) + DEM/ARIMA-ARCH**

Combinación No Paramétrica = **FS-EPNet/RNA + DEM/FS-EPNet/RNA**

La figura 6.18 es una representación gráfica de la actuación de las diferentes combinaciones de predicciones propuestas en esta fase, además de la serie original del DJIA en el mismo periodo de tiempo de la Fase B.

Figura 6. 19: Fase C. Combinación de pronósticos para el DJIA, 40 días. Del 10/24/1997 al 12/19/1997.



La tabla 6.13 reafirma lo que la figura 6.18 ilustra, que la combinación ponderada es la predicción que se ajusta de manera más adecuada para todos los periodos de predicción mostrados en la tabla 6.13 en esta fase experimental.



Tabla 6. 14: Combinación de pronósticos, Etapa C. Medición de errores: predicción del DJIA para 5, 10, 20 y 40 días del 10/24/1997 al 12/19/1997.

5 Periodos de Predicción				20 Periodos de Predicción			
CRITERIO	Combinación No-Paramétrica	Combinación Paramétrica	Combinación Ponderada	CRITERIO	Combinación No-Paramétrica	Combinación Paramétrica	Combinación Ponderada
MSE	0.01764	0.01788	<b>0.0148</b>	MSE	0.00915	0.00846	<b>0.00683</b>
RMSE	0.13283	0.13372	<b>0.12164</b>	RMSE	0.09565	0.092	<b>0.08263</b>
NRMSE	1.11575	1.12328	<b>1.02177</b>	NRMSE	1.21033	1.16411	<b>1.04559</b>
10 Periodos de Predicción				40 Periodos de Predicción			
CRITERIO	Combinación No-Paramétrica	Combinación Paramétrica	Combinación Ponderada	CRITERIO	Combinación No-Paramétrica	Combinación Paramétrica	Combinación Ponderada
MSE	0.01099	0.01072	<b>0.00865</b>	MSE	0.00551	0.00589	<b>0.00389</b>
RMSE	0.10484	0.10354	<b>0.09299</b>	RMSE	0.07423	0.07672	<b>0.06237</b>
NRMSE	1.14282	1.12861	<b>1.01357</b>	NRMSE	0.9821	1.01507	<b>0.82518</b>

Fuente: Elaboración propia.

## Capítulo 7. Conclusiones y trabajos futuros

### 7.1 Conclusiones

Las Series de Tiempo provenientes del mercado bursátil, poseen rasgos que en su conjunto las distinguen de otras series, ingredientes como el exceso de curtosis, la frecuente manifestación de la heteroscedasticidad, que en otras palabras, son los frecuentes y sistemáticos cambios en la varianza de la serie, lo que a su vez se traduce en un nivel de precios irregular muy similar al paseo aleatorio, además de presentar un coeficiente de autocorrelación muy bajo, incluso Hawawini y Keim (1995) concluyen que la correlación serial que presentan las series sobre los precios del mercado accionario son estadísticamente insignificantes, así como también el proceso subyacente generador de datos es cambiante a través del tiempo, esto se ve reflejado en el nivel de ruido y en el agrupamiento de la volatilidad, esto a su vez demuestra como el mercado accionario oscila entre periodos de tensa calma y momentos de turbulencia. Esta inestabilidad se traduce en alta volatilidad que se refleja en los desequilibrios del mercado de valores, donde eventos como la burbuja especulativa suelen suceder y en algún momento esta tiene que estallar, dando como resultado el crack bursátil. Por ello y en base a la evidencia histórica, es posible determinar que las crisis financieras son acontecimientos recurrentes y que en los últimos años tienden a ser más frecuentes, y los cuales no tienden a desaparecer. Puesto que una crisis financiera puede despojar a una nación de años de desarrollo económico, es tarea necesaria que políticos y analistas, sean capaces de manejar la crisis con el fin de minimizar las pérdidas que estos eventos podrían acarrear.

A pesar de que la mayor parte de la literatura académica y de las creencias habituales sobre la dificultad y en algunos casos la imposibilidad de poder predecir una crisis financiera, en base a evidencia empírica es posible que en algunos casos se pueda predecir. De igual forma con pruebas apropiadas y determinantes es factible rechazar en algunos casos la renombrada hipótesis de los mercados eficientes. En base a estos antecedentes las preguntas que trata de responder esta investigación, es por un lado, demostrar la ineficiencia del mercado bursátil en un

entorno altamente turbulento, por el otro la posibilidad de que los modelos propuestos en esta investigación sean capaces de anticipar en cierta medida la caída del mercado bursátil. Aunque existe una vasta literatura donde se analiza y se pronostica con diferentes técnicas el mercado de valores, son muy escasos los estudios que abordan de manera efectiva el pronóstico de las caídas bursátiles con técnicas cuantitativas y con variables endógenas. La mayoría de estos estudios analizan el entorno de los mercados financieros para detectar señales de alarma, donde son las variables exógenas las causantes de este fenómeno. En esta investigación después de evaluar los resultados se puede afirmar que en estos casos el mercado de valores tiene memoria, además es evidente que el mercado bursátil tiene caídas y subidas en el nivel de precios provocadas por motivos endógenos, y que posiblemente las variables exógenas cumplen la función de ser un detonante de las vertiginosas caídas bursátiles. Además es posible afirmar que a través de la comprobación experimental, las dos series analizadas mediante su descomposición y en el intervalo de tiempo en el cual fueron muestreadas, estas no siguen un paseo aleatorio, por lo tanto en estos casos son predecibles en un corto plazo.

En base al rendimiento mostrado por los diferentes modelos, no es posible determinar la superioridad de un modelo, y como era de esperarse, es evidente que el modelo paramétrico maneja de manera más adecuada las series con características lineales (ver Apéndice), por el otro lado, las redes Neuronales Artificiales son mejores capturando el proceso no lineal de las series, además de poseer una memoria más larga a diferencia del modelo paramétrico que pierde su fuerza o precisión predictiva en un horizonte de más corto tiempo. Sin embargo, la mejor predicción la podemos encontrar con una combinación de las predicciones de ambas técnicas, ya que estos beneficios implican considerar ambos enfoques en lugar de considerar un sólo modelo. Aunque se puede estimar que la llave de la efectividad de los modelos aquí propuestos se encuentra en la descomposición de las señales. De igual forma, se puede afirmar que el uso de la herramienta DEM puede ayudar a emerger los patrones que subyacen ocultos en las serie financieras que por lo regular se encuentran saturadas de ruido intrínseco, además el proceso

de cernido facilita la estimación de los parámetros de los modelos paramétricos y disminuye la complejidad en la búsqueda de patrones para los modelos no paramétricos. Finalmente, la investigación demuestra la valía que produce la combinación de las diferentes herramientas en el manejo de la inherente volatilidad del mercado de valores, y que los analistas, políticos e investigadores, no deben permitirse estar ausentes del empleo de estas técnicas para analizar y predecir el mercado de valores.

## **7.2 Recomendaciones para trabajos futuros**

De la presente investigación, se pueden derivar futuras tareas como:

- Utilizar variables explicativas, que enriquezcan el modelo, y coadyuven para mejorar la precisión de la predicción.
- Tratar de mejorar el desempeño del modelo paramétrico utilizando algoritmos evolutivos en la elección de los parámetros.
- Utilizar lógica difusa para mejorar el sistema neuronal incluyendo variables cualitativas, de igual forma es factible emplear otro algoritmo evolutivo para el diseño de la topología neuronal.
- Utilizar otras técnicas predictivas, tanto paramétricas como no paramétricas, para verificar su rendimiento.
- Utilizar herramientas más sofisticadas de combinación de datos, en la búsqueda de un perfeccionamiento en la combinación de las predicciones.

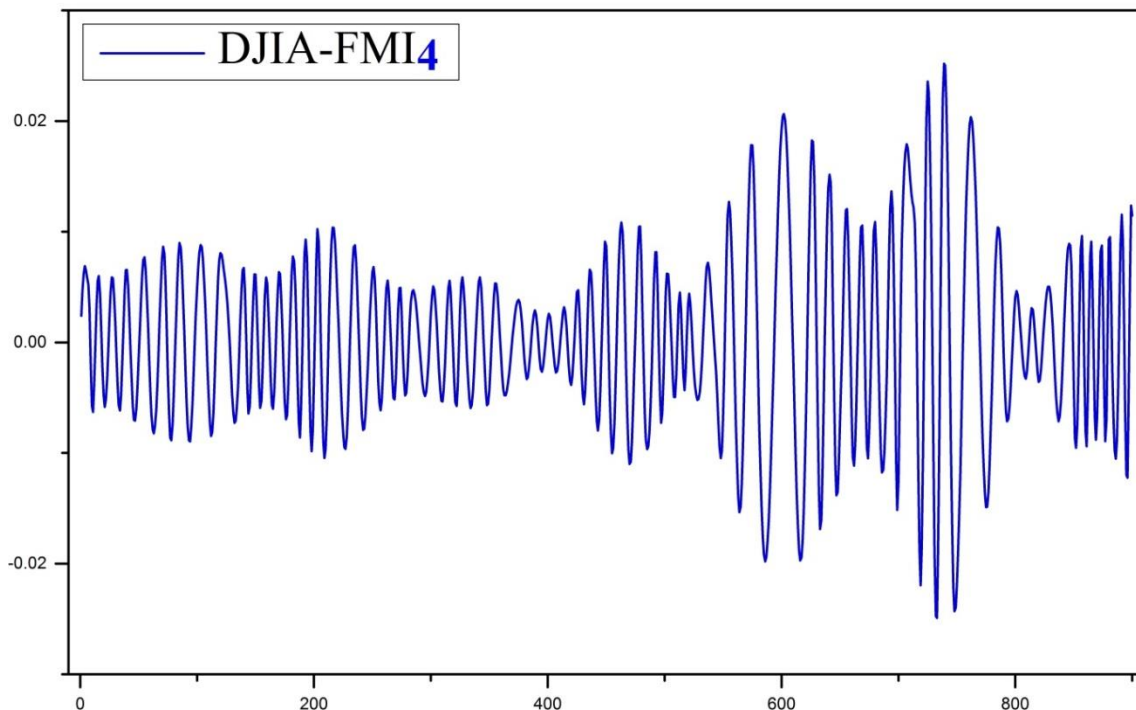


## Apéndice

### Ejemplo metodología ARIMA-ARCH.

Con el fin de ilustrar los pasos más comunes a seguir en la edificación de un modelo ARIMA-ARCH. A continuación se desarrolla un ejemplo, donde se analiza la  $FMI_4$  que resulto de la descomposición de la serie del DJIA de la fase A. Para estimar los parámetros del modelo se partió de una muestra de 1100 datos que abarcan del 10/20/1993 al 09/04/1997 y la predicción abarco del 05/09/1997 al 23/10/1997 un total de 35 datos. La estrategia en la construcción de los modelos paramétricos se encuentra basada en 4 autores: Fabris (2009), Pérez Ramírez (2008), Meza Carvajalino (2007) y Brooks (2008). Se exponen a continuación los pasos más importantes y típicos para realizar la edificación de un modelo paramétrico ARIMA-ARCH bajo la metodología Box-Jenkins. El paso inicial es graficar la ST que es el objeto de análisis. El propósito es detectar si no existen valores atípicos, ya que estos pudieran afectar el análisis, algunas posibles causas a estas fallas son por ejemplo una inserción deficiente de datos en la muestra. La ST de la  $FMI_4$  se puede visualizar en la figura A.1.

Figura A. 1: Serie de Tiempo DJIA-FMI<sub>4</sub>.

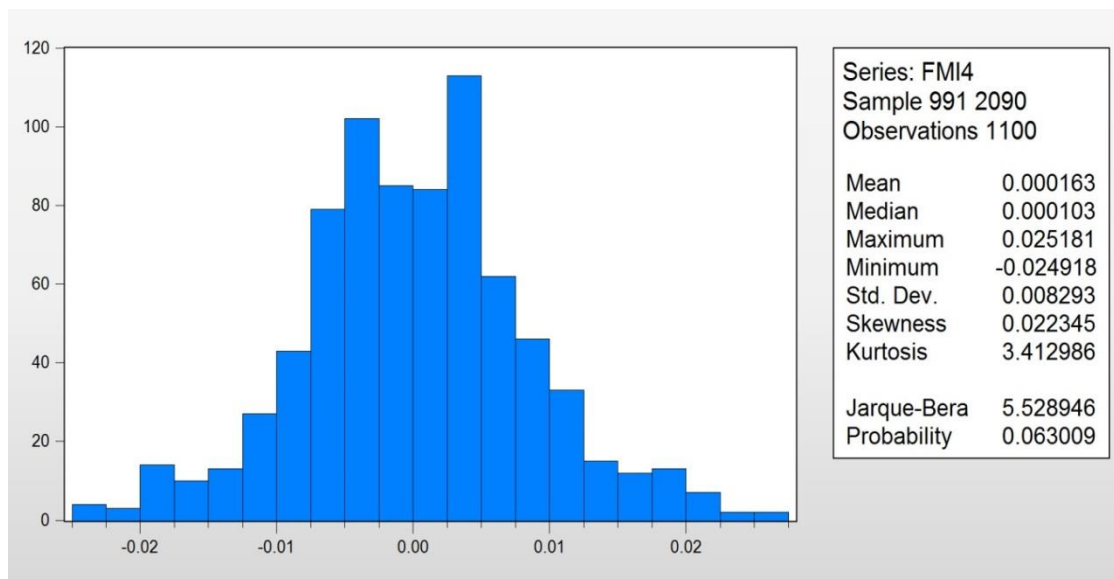


Fuente: Elaboración propia.

Otro punto que se pretende acometer en esta etapa, es que mediante un análisis visual de la ST se puede ir descubriendo los rasgos que caracterizan al fenómeno en cuestión, es decir, detectar los patrones típicos de una ST como son la tendencia, estacionalidad y el comportamiento cíclico, donde en cierto grado se pudiera empezar a discernir si una ST es estacionaria o no lo es, la cualidad estacionaria en una ST financieras es infrecuente. Sin embargo, las FMI4 provenientes de la descomposición de la serie original poseen características diferentes a la de la ST de la cual provienen.

Posteriormente se visualiza el histograma y los principales estadísticos descriptivos de la ST como es: media, mediana, valor máximo, valor mínimo, desviación estándar, coeficiente de asimetría (Skewness) y coeficiente de curtosis (Kurtosis), ver figura A.2.

Figura A. 2: Histograma del DJIA-FMI<sub>4</sub>

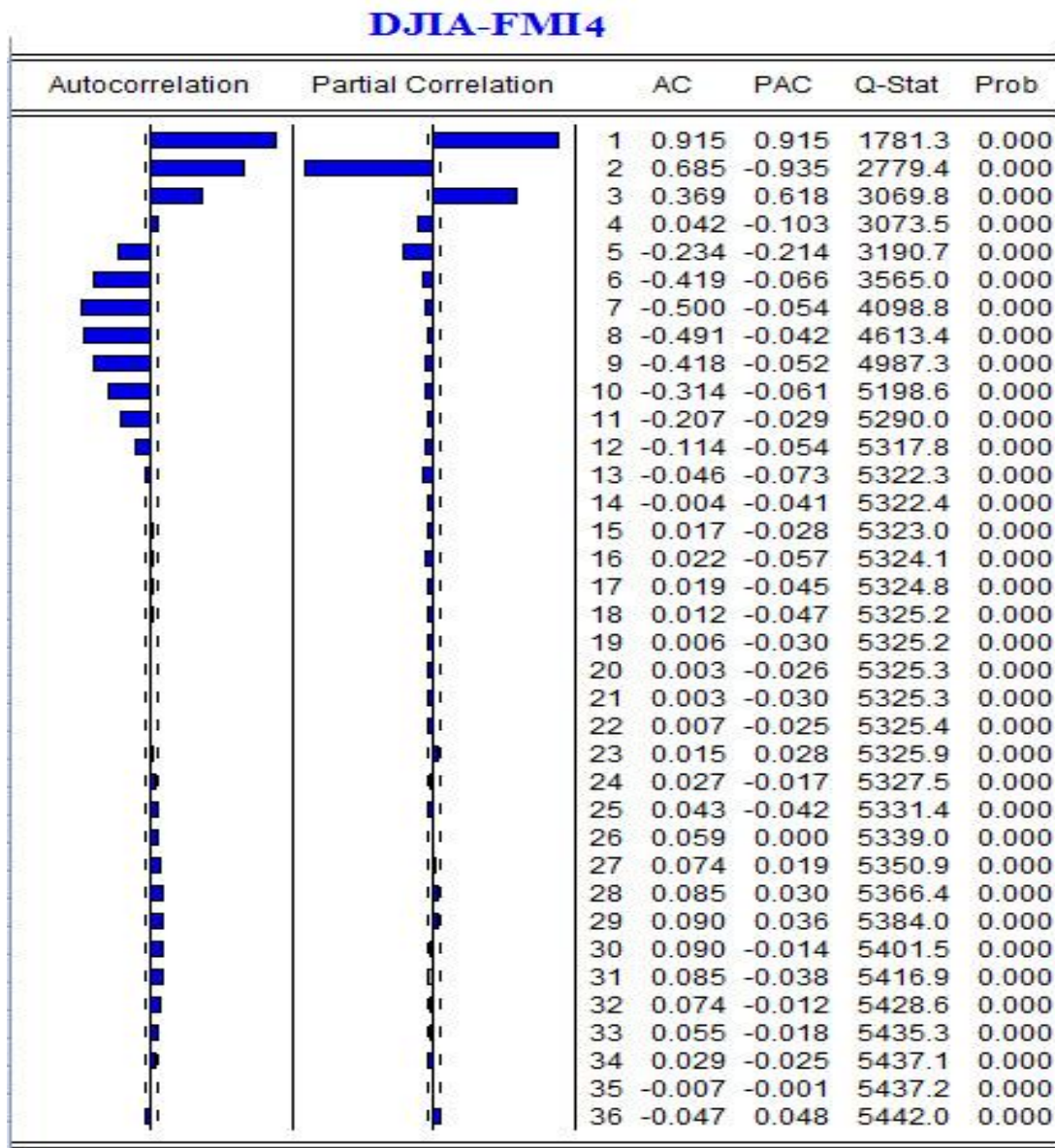


Fuente: Elaboración propia.

Una de las características que se puede apreciar en el histograma de la figura A.2 y que se confirma mediante el coeficiente de curtosis o kurtosis es que la serie DJIA-FMI<sub>4</sub> posee rasgos cercanos a las de una distribución leptocúrtica, que indica que la mayoría de los datos se encuentran aglutinados en torno a la media, dicha cualidad se aprecia visualmente en la forma de la curva como una distribución

puntiaguda o escarpada. El valor del sesgo o skewness es cercano a cero, esto significa que la distribución de DJIA-FMI<sub>4</sub> es simétrica y muy similar a la de la distribución normal. El valor (“probability”) de la prueba Jarque Bera es mayor al 5%, esto indica no rechazar la hipótesis nula de normalidad en los datos. Por lo tanto, se puede determinar que los datos se adecuan a los de una distribución normal. A continuación, se busca estimar si una ST es estacionaria o no estacionaria. La prueba del correlograma de la ST del IPC se muestra en la figura A.3.

Figura A. 3: Correlograma del DJIA- FMI4.



Fuente: Elaboración propia.



Por medio de una inspección visual de la figura A.3 se aprecia como la función de Autocorrelación (FAC) no desciende suavemente (síntoma de no estacionariedad en los datos), sino todo lo contrario, ya que la teoría indica que cuando no hay presencia de raíces unitarias el descenso es exponencial. Es decir, esta prueba indica que la serie es estacionaria. Debido a que estas pruebas son de carácter descriptivo y que en ocasiones no resulta evidente determinar si una ST es estacionaria o no lo es, se utiliza la prueba de raíz unitaria. La versión más popular de pruebas de raíz unitaria es la prueba de Dickey-Fuller Aumentado (ADF), donde en caso de aceptar la prueba, la hipótesis nula indica que la ST tiene raíz unitaria y por lo tanto es no estacionaria. La tabla A.1 muestra el valor de  $p=0.0000$  y el valor estadístico de ADF es  $t=-21.07105$ , este resultado indica rechazar la hipótesis nula, puesto que el valor del estadístico  $t$  es mayor en términos absolutos que los valores críticos al 1%, 5% y 10%. Es decir la ST es estacionaria. Por lo tanto no hay que diferenciar la serie. Por el contrario, a la serie original del DJIA hubo que diferenciarla para trabajar con ella.

Tabla A. 1: Prueba de raíz unitaria para la serie de DJIA-FMI4.

Prueba Estadística ADF	-21.07105	Valor Critico 1 %	-2.566053
		Valor Critico 5 %	-1.940973
		Valor Critico 10 %	-1.616599
Prob(Prueba Estadística ADF) = 0.0000			

Fuente: Elaboración propia.

Después se estiman los parámetros del modelo. Por medio del correlograma de la ST de DJIA-FMI<sub>4</sub> y basándose en la función de Autocorrelación (FAC) y la función de Autocorrelación Parcial (FACP) se encuentran los parámetros para estimar el modelo. En esta ocasión se estiman dos modelos, pero se pueden estimar un número indeterminado de modelos. Para determinar cuál es el mejor modelo se emplean los criterios de Criterio de Información de Akaike (AIC), Criterio de información Bayesiano (BIC) o de Schwarz, y Criterio de información de Hannan-Quinn (HQC), es posible seleccionar uno de los criterios o utilizarlos en conjunto, el mejor modelo de acuerdo a estos métodos es aquel que tenga el valor menor:

Tabla A. 2: Criterios de selección: AIC, BIC y HQC.

<b>CRITERIO</b>	Modelo A	Modelo B
Criterio de Akaike	-12.86135	-12.84516
Criterio de Schwarz	-12.82181	-12.80561
Criterio de Hannan-Quinn	-12.84632	-12.83012

Fuente: Elaboración propia.

A)  $FMI_4$   $p=AR(1)AR(2)AR(3)AR(7), d=0, q=MA(1)MA(3)MA(4)MA(5)$

B)  $FMI_4$   $p=AR(1)AR(2)AR(3)AR(7), d=0, q=MA(1)MA(3)MA(4)MA(28)$

En la tabla A.2 se muestran los diferentes valores para los criterios, donde el modelo A es el que en conjunto tiene un valor menor, por lo tanto este será el seleccionado. Una fórmula para diagnosticar si el modelo escogido se encuentra especificado correctamente es el correlograma de los residuos del modelo, este se conoce como el estadístico Q de Ljung-Box.

Figura A. 4: Estadístico Q de Ljung-Box para DJI-FMI<sub>4</sub>.

Q-statistic probabilities adjusted for 8 ARMA term(s)						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.003	-0.003	0.0064	
		2	-0.003	-0.003	0.0155	
		3	-0.004	-0.004	0.0318	
		4	-0.000	-0.000	0.0318	
		5	0.032	0.032	1.0785	
		6	0.027	0.027	1.8060	
		7	-0.011	-0.010	1.9179	
		8	0.025	0.026	2.5573	
		9	-0.028	-0.027	3.3255	0.190
		10	-0.008	-0.009	3.3860	0.336
		11	-0.020	-0.021	3.7734	0.438
		12	-0.020	-0.021	4.1945	0.522
		13	0.001	-0.000	4.1958	0.650
		14	-0.021	-0.021	4.6266	0.705
		15	0.017	0.019	4.9102	0.767
		16	-0.043	-0.043	6.7798	0.660
		17	-0.032	-0.028	7.7953	0.649
		18	-0.009	-0.009	7.8791	0.724
		19	0.002	0.003	7.8832	0.794
		20	0.011	0.010	7.9992	0.844
		21	0.057	0.057	11.315	0.661
		22	-0.020	-0.015	11.737	0.699
		23	-0.044	-0.046	13.710	0.620
		24	-0.029	-0.027	14.539	0.629
		25	0.045	0.043	16.647	0.548
		26	-0.012	-0.019	16.794	0.604
		27	0.020	0.017	17.197	0.640
		28	0.033	0.035	18.296	0.630
		29	0.053	0.055	21.207	0.508
		30	-0.013	-0.014	21.385	0.558
		31	-0.048	-0.048	23.774	0.475
		32	0.069	0.071	28.670	0.278
		33	0.043	0.036	30.604	0.243
		34	0.027	0.022	31.348	0.257
		35	-0.007	-0.010	31.401	0.300
		36	-0.043	-0.040	33.283	0.266

Fuente: Elaboración propia.

La figura A.4 muestra el correlograma de los residuos del modelo A propuesto del FMI<sub>4</sub>, donde todas las probabilidades de las observaciones tiene un p-valor superior al 0.05 (séptima columna encabezado “Prob”), esto demuestra que los residuales del modelo son ruido blanco y por lo tanto los residuales no se encuentran autocorrelacionados, y por lo tanto el modelo estimado se encuentra correctamente especificado. Seguidamente se trata de encontrar efecto ARCH en el modelo. Para diagnosticar la existencia de volatilidad condicional heteroscedástica en el modelo ajustado del IPC se efectúa el correlograma de los residuos al cuadrado.

Figura A. 5: Correlograma de los residuales al cuadrado DJIA–FMI<sub>4</sub>.

Q-statistic probabilities adjusted for 8 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.247	0.247	60.812	
		2	0.516	0.484	325.63	
		3	0.232	0.065	379.38	
		4	0.219	-0.091	427.10	
		5	0.178	0.027	458.60	
		6	0.218	0.160	505.88	
		7	0.157	0.037	530.53	
		8	0.147	-0.061	552.27	
		9	0.122	-0.008	567.10	0.000
		10	0.122	0.065	582.01	0.000
		11	0.102	0.026	592.53	0.000
		12	0.100	-0.021	602.61	0.000
		13	0.105	0.020	613.71	0.000
		14	0.129	0.088	630.39	0.000
		15	0.203	0.165	671.94	0.000
		16	0.267	0.182	743.81	0.000
		17	0.175	-0.075	774.77	0.000
		18	0.282	0.042	855.14	0.000
		19	0.120	-0.012	869.70	0.000
		20	0.167	-0.032	897.85	0.000
		21	0.147	0.025	919.66	0.000
		22	0.140	0.010	939.49	0.000
		23	0.111	-0.034	952.01	0.000
		24	0.115	-0.014	965.42	0.000
		25	0.062	-0.023	969.40	0.000
		26	0.134	0.094	987.71	0.000
		27	0.069	0.017	992.58	0.000
		28	0.129	0.014	1009.6	0.000
		29	0.092	0.025	1018.3	0.000
		30	0.035	-0.109	1019.5	0.000
		31	0.039	-0.110	1021.1	0.000
		32	0.059	0.031	1024.6	0.000
		33	0.035	-0.000	1025.9	0.000
		34	0.061	-0.053	1029.7	0.000
		35	0.052	0.016	1032.5	0.000
		36	0.007	-0.056	1032.5	0.000

Fuente: Elaboración propia.

En la figura A.5 se advierte de la posible existencia de heteroscedasticidad, ya que

se percibe que los valores de autocorrelación de los residuos al cuadrado tienen un valor de  $p$  menor a 0.05, evidencia inequívoca que los residuales son significativos. Se realiza la prueba ARCH-LM a los residuales al cuadrado del modelo, con el fin de verificar la existencia de heteroscedasticidad condicional en la varianza. Donde la hipótesis nula indica la no existencia de efecto ARCH. El resultado se aprecia en la tabla A.3.

$H_0$ : no se encuentra efecto ARCH.

$H_a$ : se encuentra efecto ARCH.

Tabla A. 3: Prueba ARCH a los residuos del modelo ajustado de DJIA-FMI4.

Prueba de Heteroscedasticidad: ARCH

F-Estadística	198.4584	Probabilidad	0.0000
Obs.*R cuadrado	283.8571	Probabilidad	0.0000

Fuente: Elaboración propia.

En base al valor de  $p=0.0$  ligado al estadístico F, esta indica refutar la hipótesis nula de la no existencia de efecto ARCH, como tal esto lleva a inferir la existencia de heteroscedasticidad y que la varianza no es constante, por lo tanto hay que modelar el efecto ARCH. En la tabla A.4 se muestran los valores de los criterios para diferentes modelos que se estimaron. El modelo que en base a los criterios de selección se debe escoger es el ARCH(5), ya que es el que se ajusta mejor a la asimetría de la varianza condicional del modelo, y por ende es el que se encuentra más capacitado para el modelado del proceso heteroscedástico.

Tabla A. 4: Resultados de los criterios AIC y BIC para los modelos: ARCH(5), GARCH(1,1), para la serie DJIA-FMI4.

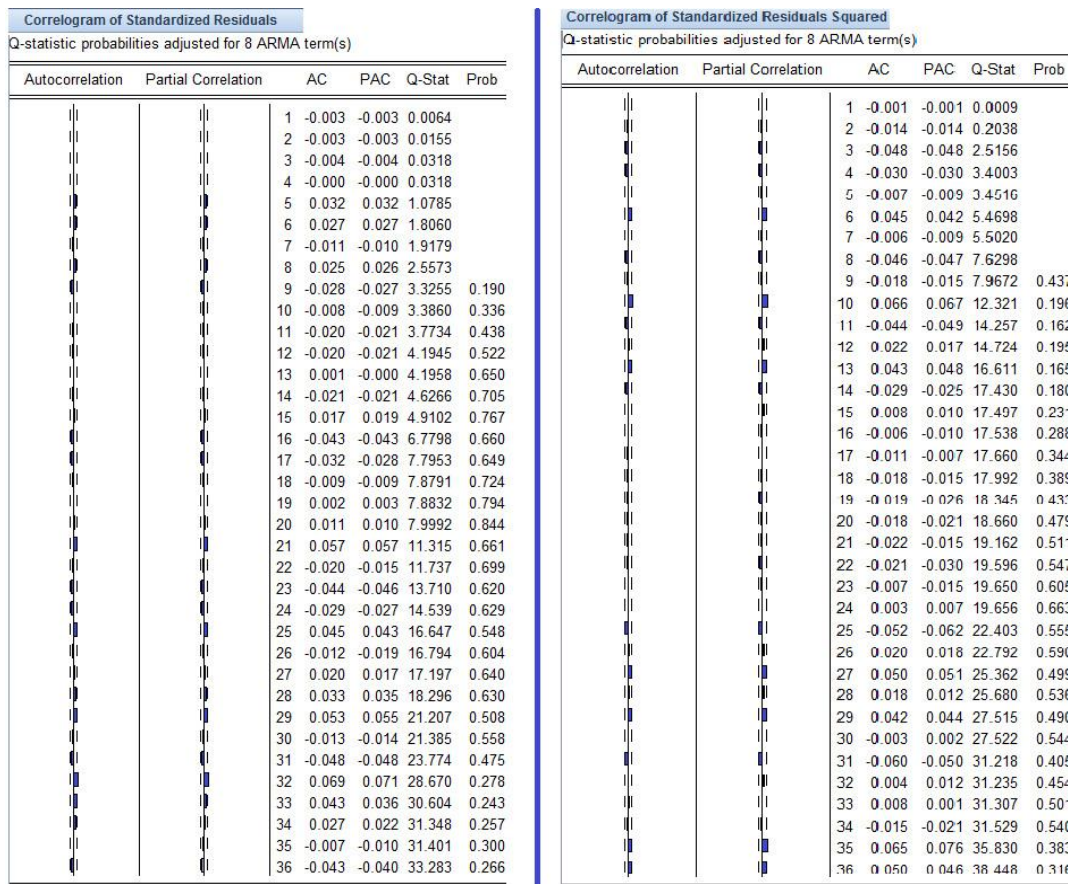
Modelo	Ecuación	Criterio de Akaike	Criterio de Schwarz
ARCH(5)	FMI <sub>4</sub> AR(1) AR(5) AR(19) MA(1) MA(36)	-13.94626	-13.90836
GARCH(1,1)	FMI <sub>4</sub> AR(1) AR(2) AR(5) AR(6) MA(1) MA(3) MA(4) MA(5)	-13.84458	-13.81520

Fuente: Elaboración propia

El último paso antes de realizar el pronóstico es verificar si el modelo seleccionado

se encuentra especificado correctamente a través del correlograma residual al cuadrado y del correlograma muestral de los residuos, así como de la prueba ARCH que indique si aún hay presencia de heteroscedasticidad. A continuación, se verifican los correlogramas: a) Correlograma de los residuales (estadístico Q de Ljung-Box) y b) Correlograma de los residuales al cuadrado. Ambos se muestran en la figura A.6, donde se evidencia que en ambos correlogramas los valores de cada observación de p son mayores a 0.05 (séptima columna a la derecha respectivamente), esto significa que todos los coeficientes no son significativos y se concluye que ambos residuales del modelo ARCH(5) son ruido blanco.

Figura A.6: Correlograma de los residuales y correlograma al cuadrado de los residuales de DJIA-FMI<sub>4</sub>.



Fuente: Elaboración propia.

La última prueba para validar la no existencia de heteroscedasticidad en el modelo elegido es la prueba ARCH. Ver tabla A.5.

Tabla A. 5: Prueba ARCH a los residuales del modelo ajustado ARCH(5).



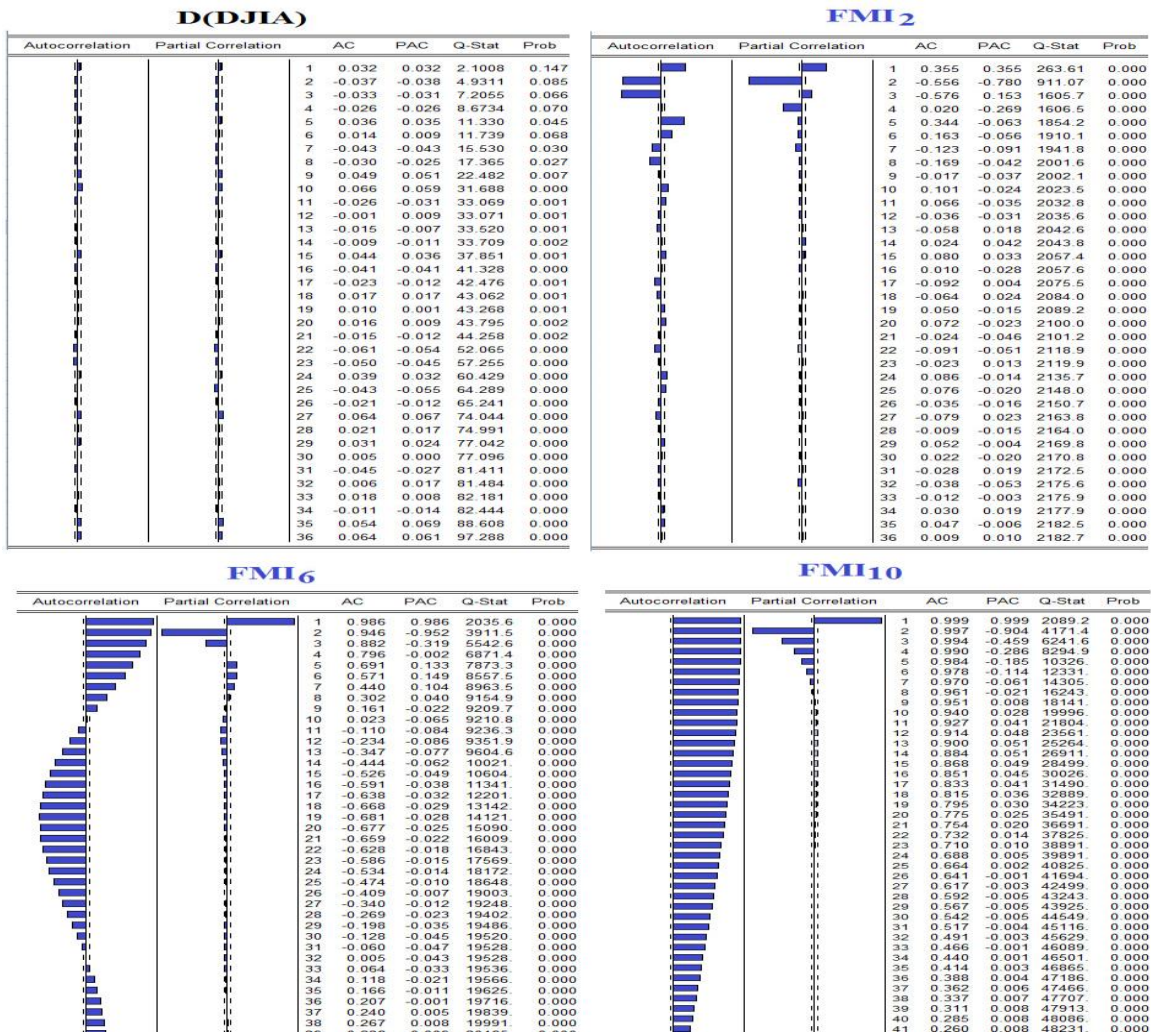
Prueba de Heteroscedasticidad: ARCH

F-Estadística	0.06810	Probabilidad	0.4094
Obs.*R cuadrado	0.06819	Probabilidad	0.4089

Fuente: Elaboración propia.

Una vez que se ha llegado hasta este punto, se está en una posición donde ya es posible realizar la predicción del modelo seleccionado. Como anteriormente se había mencionado para estimar los parámetros de los modelos se utiliza el correlograma. En la figura A.7 se muestran varios correlogramas: en primer lugar el correlograma de la serie del DJIA con una diferencia, y a continuación el de sus descomposiciones FMI<sub>2</sub>, FMI<sub>6</sub> y FMI<sub>10</sub>.

Figura A. 7: Correlograma de DJIA, DJIA-FMI<sub>2</sub>, DJIA-FMI<sub>6</sub> y DJIA-FMI<sub>10</sub>.

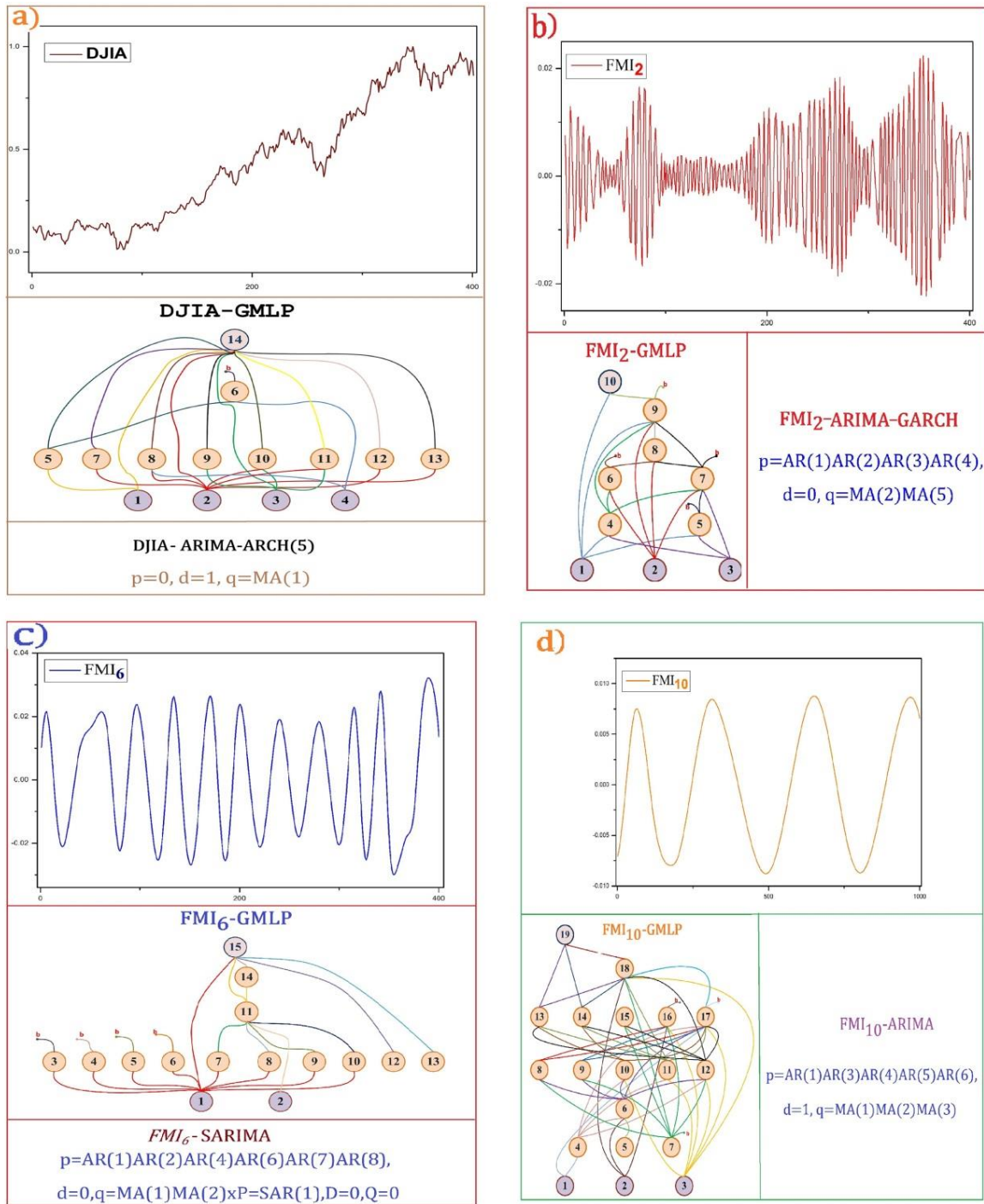


Fuente: Elaboración propia.

Se puede apreciar en la figura A.7, por medio del correlograma la posible dificultad

para encontrar los parámetros que ajusten a la serie original del DJIA con una diferencia, ya que se puede apreciar que la serie se asemeja al ruido blanco.

Figura A. 8: RNAs diseñadas por FS-EPNet y los parámetros del modelo ARIMA-ARCH para resolver las series del DJIA, DJIA-FMI<sub>2</sub>, DJIA-FMI<sub>6</sub> y DJIA-FMI<sub>10</sub>.



Fuente: Elaboración propia.

Por último, en la figura A.8 se ilustran cuatro series de las cuales solo se muestran

400 datos, a diferencia de la serie  $FMI_{10}$  que se muestran 1000 datos, ya que así es posible apreciar mejor las características de estas.

La series mostradas son las del DJIA (A.8a), así como las  $FMI_2$  (A.8b),  $FMI_6$  (A.8c), y  $FMI_{10}$  (A.8d) que resultaron de su descomposición, además se muestran las topologías de las diferentes RNAs que diseño el algoritmo FS-EPNet para resolver cada una de estas series, y de igual forma se muestran los parámetros estimados por el modelo paramétrico para las mismas series de tiempo. Cabe recalcar la dificultad para estimar los parámetros del modelo paramétrico para la serie del DJIA (A.8a), y que puede llevar a concluir que el proceso de cernido realizado mediante la herramienta DEM, les facilitaría a los modelos ARIMA-ARCH la estimación de sus parámetros, ya que eliminaría gran parte del ruido intrínseco de la serie. Por otro lado, se puede apreciar que para el modelo no paramétrico la  $FMI_{10}$  (A.8d), es la red topológicamente con mayores dimensiones, lo cual podría deberse a que la red aumento en tamaño para evitar el sobreajuste a los datos, ya que es evidente que esta serie es la menos compleja de las cuatro. Además para el modelo paramétrico la  $FMI_{10}$  (A.8d), dadas las características de linealidad podría resultar la serie con un menor grado de dificultad para ajustarse de manera adecuada de entre estas cuatro series presentadas en la figura A.8.





## Bibliografía

- Abraham, A., & Nath, B. (2000). *Hybrid Intelligent Systems Design -- A Review of a Decade of Research*. Auburn: Machine Intelligence Research Labs.
- Adams, A. J. (1978). *Modeling and forecasting seasonal time series of product sales across territories: a comparative study*. Master's thesis, University of Iowa.
- Aiolfi, M., Capistrán, C., y Timmermann, A. G. (2010). *Forecast combinations*. No. 2010-04, BANXICO.
- Aiolfi, M. y Timmermann, A. (2006). *Persistence in forecasting performance and conditional combination strategies*. *Journal of Econometrics*: 135 (1-2), 31-53.
- Andersson, Magnus & Palm, Johan. (2009). *Forecasting the Stock Market - A Neural Network Approach*. Tesis de Maestría en matemáticas, Mälardalen University.
- Angeline, P., Saunders, G., y Pollack, J. (1994). *An evolutionary algorithm that constructs recurrent neural networks*. *Neural Networks, IEEE Transactions*, 5(1):54-65.
- Aparicio, M., Unda, E., Cuanalo, R. (2014). *Análisis fractal del electroencefalograma exponente de hurst como un indicador de complejidad*. Estudiantes de Medicina en Movimiento.
- Ardalani-Farsa, M. y Zolfaghari, S. (2010). *Chaotic time series prediction with residual analysis method using hybrid elman-narx neural networks*. *Neurocomputing*, 73(13- 15): 2540-2553.
- Ashlock, D. (2010). *Evolutionary Computation for Modeling and Optimization*. Springer Publishing, 1st ed.
- Balacco, H. (2012). *Mercados Financieros: Análisis Estadístico y Evidencia Empírica: No Linealidades, Dinámica Caótica y Predicción utilizando Redes Neuronales*. EAE.
- Bassett, D. y Gazzaniga, M. (2011). *Understanding complexity in the human brain*. *Trends in Cognitive Sciences*, 15(5):200-209.
- Bates, J. M. y Granger, C. W. J. (Dec., 1969). *The combination of forecasts*. *Operational Research Society*, 20:451-468.
- Belew, R. K., Mcinerney, J., y Schraudolph, N. N. (1990). *Evolving networks: Using the genetic algorithm with connectionist learning*. *In Proc. Second Conference on Artificial Life (1991)*, 511-547. Addison-Wesley.
- Bell, T. B. (1997). *Neural nets or the logit model? a comparison of each model's ability to predict commercial bank failures*. *International Journal of Intelligent Systems in Accounting, Finance & Management*, 6:249-264.
- Bessler, D. A. y Chamberlain, P. J. (1987). *On bayesian composite forecasting*. *Omega*, 15(1):43-48.
- Blázquez, J., García-Berrocal, A., Balbás, M., y Montalvo, C. (noviembre 2010). *Aplicación de las transformadas de hilbert a la dinámica de una pala de aerogenerador*. *XVIII Congreso Nacional de Ingeniería Mecánica CNIM 2010*.
- Bollerslev, T. (1986). *Generalized autoregressive conditional heteroskedasticity*. *Journal of Econometrics*, vol. 31, issue 3, pages 307-327.

- Box, G. y Jenkins, G. (1976). Time series analysis: forecasting and control. *Holden-Day series in time series analysis and digital processing*. Holden-Day.
- Brooks, C. (2008). *Introductory Econometrics for Finance*. Cambridge University Press.
- Bullinaria, J. A. (2005). Evolving neural networks: Is it really worth the effort? *In Proceedings of the European Symposium on ANN*, p. 267–272. Evere, Belgium.
- Bullinaria, J. A. (2006). Understanding the advantages of modularity in neural systems. *In Proceedings of the Twenty-eighth Annual Conference of the Cognitive Science Society*, 119–124, Mahwah, NJ: Lawrence Erlbaum Associates.
- Cao, Y., Li, Y., y Liao, X. (2005). Applying modified fuzzy neural network to customer classification of e-business. *Internet and Network Economics: First International Workshop*, Proceedings WINE 2005, pages 356–365.
- Carles, P. G. (1996). Economía dinámica caótica: una aplicación al mercado de capitales español. Tesis Doctoral, Universidad Complutense de Madrid.
- Castellano, G., Fanelli, A., y Pelillo, M. (1997). *An iterative pruning algorithm for feedforward neural networks*. Neural Networks, IEEE Transactions on, 8(3):519–531.
- Castillo, P., Castellano, J., Merelo, J., y Prieto, A. (2001). Diseño de redes neuronales artificiales mediante algoritmos evolutivos. *Revista Iberoamericana de Inteligencia Artificial*, 5:2– 32.
- Castillo Valdivieso, P. A., García Arenas, M., García Castellano, F. J., Merelo Guervós, J. J., Prieto, A., Rivas, V. M., y Romero, G. (2006). Lamarckian evolution and the baldwin effect in evolutionary neural networks. *eprint arXiv:cs/0603004*.
- Chao, J., Shen, F., and Zhao, J. (2011). Forecasting exchange rate with deep belief networks. *In International Symposium on Neural Networks*, pages 1259–1266.
- Chen, P.C., Lo, C.Y., y Chang, H.T. (2009). An empirical study of the artificial neural network for currency exchange rate time series prediction. In Wang, H., Shen, Y., Huang, T., y Zeng, Z., editors, *The Sixth International Symposium on Neural Networks (ISNN 2009)*, vol.56 of *Advances in Intelligent and Soft Computing*, pages 543–549. Springer Berlin.
- Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4):559–583.
- Cruz Torres, I. (2007). *Pronósticos en el mercado de derivados utilizando redes neuronales y modelos arima: una aplicación al cete de 91 días en el mexder*. Tesis Maestría, UNAM.
- Edgerton, H. y Kolbe, L. (1936). *The method of minimum variation for the combination of criteria*. Psychometrika, 1(3):183–187.
- Elton, E. J., Gruber, M. J., Brown, S. J. y Goetzmann, W. N. (2014). *Modern portfolio theory and investment analysis*. New York: Wiley.
- Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987–1007.

- Ertay, T. y cekyay, B. (2005). *Integrated clustering modeling with backpropagation neural network for efficient customer relationship management*. D., Chen, G., E. Kerre, E., y Wets, G., editors, *Intelligent Data Mining, volume 5 of Studies in Computational Intelligence*, 355-373. Springer Berlin Heidelberg.
- Fabris, J. (2009). *Econometría Financiera: Modelos y Pronósticos utilizando QMS EViews®* Omicron.
- Fama, E. (1970). *Efficient capital markets: A review of theory and empirical work*. *Journal of Finance*, 25:383-417.
- Fama, E. F. (1965a). *The behaviour of stock market prices*. *Journal of Business*, 34–105.
- Fama, E. F. (1965b). *Random walks in stock-market prices*. *Financial Analysts Journal*, 21:55-59.
- Floreano, D., Dürr, P., y Mattiussi, C. (2008). *Neuroevolution: from architectures to learning*. *Evolutionary Intelligence*, 1(1):47–62.
- Flórez López, R. (2007). *Análisis de los determinantes del riesgo de crédito en presencia de carteras de bajo incumplimiento. Una nueva propuesta de aplicación*. *Revista Europea de Dirección y Economía de la Empresa*.
- Fogel, L. J., Owens, A., & Walsh, M. (1966). *Artificial Intelligence Through Simulated Evolution*. N.Y. Wiley.
- García Salgado, O. and Morales Castro, A. (2012). *Las redes neuronales artificiales como una herramienta de análisis en la determinación de las empresas que permanecen listadas o deslistadas dentro de la bolsa mexicana de valores*. CIENCIA@UAQ.
- Garratt, A., Lee, K., Pesaran, M. H., & Shin, Y. (2003). *Forecast uncertainties in macroeconomic modeling: An application to the UK economy*. *Journal of the American Statistical Association*, 98(464):829–838.
- Gómez Ramos, E. L., Venegas Martínez, F., y Allier Campuzano, H. (2011). *Análisis comparativo entre modelos garch y redes neuronales en el pronóstico de los índices bursátiles ipc y dow jones*. *Eseconomía*.
- Gonzales Martínez, R. and Luna Acevedo, B. (2011). *Predicciones del WTI para el año 2011: Combinaciones de los pronósticos de un modelo autorregresivo de redes neuronales artificiales y de un modelo arima-garch*. 4to Encuentro de economistas de Bolivia.
- Granger, C. W. J. y Ramanathan, R. (1984). *Improved methods of combining forecasts*. *Journal of Forecasting*, 3(2):197–204.
- Guzmán Plata, M. d. I. P. (2008). *Fundamentos, causalidad y predicción en los mercados cambiario y bursátil mexicano*. Tesis de Maestría, IPN.
- Guzmán Plata, M. d. I. P., Leyva López, S., y Cárdenas Almagro, A. (2007). *El futuro del índice de precios y cotizaciones de la bolsa mexicana de valores*. *Análisis Económico*.
- Hall, M. A. (1999). *Correlation-based feature selection for machine learning*. Technical report.
- Hamid, S. A. y Iqbal, Z. (2004). *Using neural networks for forecasting volatility of S&P500 index futures prices*. *Journal of Business Research*, 57(10):1116–1125.

- Hancock, P. (1992). *Genetic algorithms and permutation problems: a comparison of recombination operators for neural net structure specification*. In *Combinations of Genetic Algorithms and Neural Networks, 1992.*, COGANN-92. International Workshop on (108-122). IEEE.
- Hanke, J. y Reitsch, A. (1986). *Business forecasting*. Allyn and Bacon.
- Hawawini, G. y Keim, D. B. (1995). *On the predictability of common stock returns: World-wide evidence*. *Handbooks in Operations Research and Management Science*, 9:497–544.
- Haykin, S. (1999). *Neural Networks. A comprehensive Foundation*. Prentice Hall, 2nd edition.
- Herrera, J. B. (2008). *El caos y las redes neuronales en la economía y finanzas interpretando el desorden perfecto: Modelado y predicción*. Pensamiento Crítico.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI:University of Michigan Press, Cambridge, MA, USA.
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N. C., Tung, C. C., y Liu, H. H. (1998). *The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis*. In *Proceedings of the Royal Society of London, volume 454 of Series A: Mathematical, Physical and Engineering Sciences*, 903–995.
- Huang, N. E. y Wu, Z. (2008). A review on hilbert-huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, 46(2).
- Hyndman, R. J. y Kostenko, A. V. (2007). *Minimum sample size requirements for seasonal forecasting models*. *Foresight: The International Journal of Applied Forecasting*, (6):12–15.
- Jacobsson, E. (2009). *How to predict crashes in financial markets with the Log-Periodic Power Law*. Master's thesis, Stockholm University.
- Ludlow, J. (2006). *Volatilidad del IPC, Nasdaq y S&P500: un modelo Garch multivariado*. *Análisis Económico N.48*.
- Koehn, P. (1994). *Combining genetic algorithms and neural networks: The encoding problem*. Tesis de Maetría, The University of Tennessee, Knoxville.
- Krishnamurti, T. N., Kishtawal, C. M., LaRow, T. E., Bachiochi, D. R., Zhang, Z., Williford, C. E., Gadgil, S., y Surendran, S. (1999). *Improved weather and seasonal climate forecasts from multimodel superensemble*. *Science*, 1548-1550.
- Landassuri-Moreno, V. y Bullinaria, J. A. (2009a). *Feature selection in evolved artificial neural networks using the evolutionary algorithm EPNNet*. *Proceedings of the 2009 UK Workshop on Computational Intelligence, UKCI '2009, Nottingham, UK: University of Nottingham*.
- Landassuri-Moreno, V. y Bullinaria, J. A. (2009b). *Neural network ensembles for time series forecasting*. In *Proceedings of the 11th Annual conference on Genetic and evolutionary computation, GECCO '09, 1235–1242, New York, USA. ACM*.
- Landassuri-Moreno, V., Bustillo-Hernández, C., y Carbajal-Hernández, J. (2012). *Single-step-ahead and Multi-step-ahead Prediction with Evolutionary Artificial Neural Networks*, volume 60 of *Research in Computer Science*, chapter *Advances in intelligent and Information Technologies*, pages 3–12.

Landassuri-Moreno, V., Marcial-Romero, J., Montes-Venegas, A., y Ramos, M. (2011). *Chaotic time series prediction with feature selection evolution*. In *Electronics, Robotics and Automotive Mechanics Conference (CERMA), 2011 IEEE*, 71–76.

Landassuri-Moreno, V. M. (2012). *Evolution of modular neural networks*. Tesis Doctoral, University of Birmingham.

Landassuri-Moreno, V. M. y Bullinaria, J. A. (2011). *Biasing the evolution of modular neural networks*. In Smith, A. E., editor, *Proceedings of the 2011 IEEE Congress on Evolutionary Computation, 1952-1959*, New Orleans, USA. IEEE Computational Intelligence Society, IEEE Press.

Lawrence, S., Giles, C. L., and Tsoi, A. C. (1996). What size neural network gives optimal generalization? convergence properties of backpropagation. Technical report.

León-Anaya, L., Landassuri-Moreno, V., Bustillo-Hernández, C., y Orozco, H. (2013). *Predicción de la demanda de energía eléctrica comparando métodos clásicos contra redes neuronales artificiales evolucionadas*. In RVP-AI/2013, IEEE Sección México, Acapulco Gro., Memoria Técnica.

Leondes, C. T. (2002). *Intelligent Systems: Technology and Applications, Six Volume Set*. Taylor & Francis.

Leung, M. T., Daouk, H., y Chen, A. (2000). *Forecasting stock indices: a comparison of classification and level estimation models*. International journal of forecasting.

Lorenzo Valdéz, A. (2005). No linealidad en los mercados accionarios latinoamericanos. *Análisis Económico*.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., y Winkler, R. (1982). *The accuracy of extrapolation (time series) methods: Results of a forecasting competition*. Journal of Forecasting.

Makridakis, S., Wheelwright, S. y Hyndman, R. (1998). *Forecasting: methods and applications*. 3rd ed. New York: John Wiley & Sons.

Malkiel, B. G. (1973). *A Random Walk Down Wall Street*. Norton, New York.

Man, K. F., Tang, K. S., y Kwong, S. (1999). *Genetic Algorithms: Concepts and Designs*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2nd edition.

Mandelbrot, B. B. (2010). *Fractals and Scaling In Finance: Discontinuity, Concentration, Risk*. Springer Publishing Company, Incorporated, 1st edition.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Matamoras, O. M. (2004). *Modelos Mecanicos de la Dinamica Fractal del Mercado Petrolero*. Tesis Doctoral, Escuela Superior de Ingeniería Mecánica y Eléctrica. Instituto Politecnico Nacional.

Melin, P. y Castillo, O. (2005). Hybrid Intelligent Systems for Pattern Recognition Using Soft Computing: An Evolutionary Approach for Neural Networks and Fuzzy Systems. *Studies in Fuzziness and Soft Computing*. Springer.

Mendoza-Velázquez, A. y Gómez-Gil, P. (2011). Neural networks, ordered probit models and multiple

discriminants. Evaluating risk rating forecasts of local governments in México. *Technical report*, Centro de Investigación e Inteligencia Económica (CIIE), UPAEP.

Meza Carvajalino, C. A. (2007). *Econometría Fundamental*. Universidad de La Salle.

Michalewicz, Z. (1996). *Heuristic methods for evolutionary computation techniques*. Journal of Heuristics, 1(2):177-206.

Mizuta, S., Sato, T., Lao, D., Ikeda, M., y Shimizu, T. (2001). *Structure design of neural networks using genetic algorithms*. Complex Systems, 161–175.

Monroy Olivares, C. (1997). *Teoría del caos*. Tecnologías Emergentes de Cómputo. Alfaomega.

Cervantes, M. T., Carmona, E. G., y Villalba, S. I. (2014). *Aplicación de los modelos CAPM y D-CAMP para la medición del riesgo sistemático de las emisoras que han conformado el Índice de Precios y Cotizaciones en periodos de crisis*. Investigación Multidisciplinaria.

Morales-Matamoros, O., Contreras-Troya, T. I., Mota-Hernández, C. I., y Trueba-Ríos, B. (2009). *Fractal analysis of epilepsy*. In Proceedings of the 53rd Annual Meeting of the International Society for the Systems Sciences.

Moreno, V. M. L. (2006). Predicción de series de tiempo con descomposición empírica en modos, algoritmo genético y redes neuronales artificiales. Tesis de Maestría, Centro de Investigación en Computación. Instituto Politécnico Nacional.

Müller, K.-R., Finke, M., Murata, N., Schulten, K., y Amari, S. (1996). *A numerical study on learning curves in stochastic multi-layer feed-forward networks*. Neural Computation, 1085-1106

Ortiz-Rodríguez, J. M., Martínez-Blanco, M. d. R., Cervantes Viramontes, J. M., y Vega-Carrillo, H. R. (2013). Robust design of artificial neural networks methodology in neutron spectrometry. , Artificial Neural Networks-Architectures and Applications, Kenji Suzuki (Ed.).

Peters, E. E. (July/August 1989). *Fractal structure in the capital markets*. Financial Analysts Journal, Volume 45 Issue 4.

Plagiannakos, V., Magoulas, G., Androulakis, G., y Vrahatis, M. (1999). *Global search methods for neural network training*. In Proceedings of the 3rd IEEE-IMACS World Multiconference on Circuits, Systems, Communications and Computers, vol 1, 3651–3656.

Radcliffe, N. J. (1993). *Genetic set recombination and its application to neural network topology optimisation*. Neural Computing and Applications, 1(1):67–90.

Ramírez, F. (2008). *Modelos ARIMA-ARCH: algunas aplicaciones a las series de tiempo financieras*. Universidad de Medellín, Sello Editorial.

Ravisankar, P. y Ravi, V. (2009). *Failure prediction of banks using threshold accepting trained kernel principal component neural network*. In World Congress on Nature & Biologically Inspired Computing, 7-12.

Reed, R. y Marks, R. (1999). *Neural Smothing: Supervised Learning in Feedforward Artificial Neural Networks*. A Bradford book. MIT Press.

- Renders, J.M. and Flasse, S. P. (1996). *Hybrid methods using genetic algorithms for global optimization*. IEEE Trans. System Man Cybernet, 243–258.
- Riessen, G. A., Williams, G. J., y Yao, X. (1997). *Pepnet: parallel evolutionary programming for constructing artificial neural networks*. In Proceedings of the 6th International Conference, Evolutionary Programming, 35–45. Springer-Verlag.
- Riley, M., Thompson, C. P., y Jenkins, K. W. (2010). *Improving the performance of cascade correlation neural networks on multimodal functions*. In Proceedings of the World Congress on Engineering.
- Rilling, G., Flandrin, P., y Gonçalves, P. (2003). *On empirical mode decomposition and its algorithms*. In Proceedings of the 6th IEEE/EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP '03), Grado, Italy.
- Rodríguez-González, A., Guldrís-Iglesias, F., Colomo-Palacios, R., Gomez-Berbis, J., Jimenez-Domingo, E., Alor-Hernandez, G., Posada-Gomez, R., y Cortes-Robles, G. (2010). Improving trading systems using the RSI financial indicator and neural networks. *In Proceedings of the 11th International Conference on Knowledge Management and Acquisition for Smart Systems and Services, PKAW'10, 27–37*, Berlin, Heidelberg. Springer-Verlag.
- Rodríguez López, M., Piñeiro Sánchez, C. y de Llano Monelos, P. (2014). *Determinación del riesgo de fracaso financiero mediante la utilización de modelos paramétricos, de inteligencia artificial, y de información de auditoría*. Estudios de economía, 41:187–217.
- Sammut, C. y Webb, G. (2011). *Encyclopedia of Machine Learning*. Springer Publishing Company, Inc., 1st ed.
- Samuelson, P. A. (1965). *Proof that Properly Anticipated Prices Fluctuate Randomly*. Industrial Management Review, 6:41–49.
- Saravanan, N. y Fogel, D. (1995). Evolving neural control systems. *IEEE Expert*, 10(3):23–27.
- Schaffer, J., Whitley, D., and Eshelman, L. (1992). *Combinations of genetic algorithms and neural networks: a survey of the state of the art*. In Whitley, D. and Schaffer, J., editors, International Workshop on Combinations of Genetic Algorithms and Neural Networks (COGANN-92), 1–37, Piscataway, New Jersey. IEEE Press.
- Schwefel, H. P. (1981). *Numerical Optimization of Computer Models*. John Wiley & Sons, Inc., New York, USA.
- Schwefel, H.P. (1993). *Evolution and Optimum Seeking: The Sixth Generation*. John Wiley & Sons, Inc., New York, USA.
- Seiler, M. J. y Rom, W. (1997). *A historical analysis of market efficiency: Do historical returns follow a random walk?* Journal of Financial and Strategic Decisions.
- Sekhar Behera, S., & Chattopadhyay, S. (2012). *A Comparative Study of Back Propagation and Simulated Annealing Algorithms for Neural Net Classifier Optimization*. Procedia Engineering, Volume 38, 448-455.



Sexton, R., Dorsey, R., y Johnson, J. (1999). Beyond backpropagation: Using simulated Annealing for training neural networks. *Journal of Organizational and End User Computing*, 11(3):3-10.

Sierra, G. (2012). *Relación de procesos de persistencia estimados a través de los coeficientes Hurst y volatilidad implícita en el mercado mexicano*. ECORFAN Optimización-Estocástica- Recursiva-Coherente-Sistémica y sus variantes.

Slutzky, E. (1937). *The summation of random causes as the source of cyclic processes*. *Econometrica*, Vol. 5, No. 2.

Srinivasulu, S. y Jain, A. (2006). A comparative analysis of training methods for artificial neural network rainfall-runoff models. *Applied Soft Computing*, 6(3):295–306.

Surkan, A. y Singleton, J. (1990). *Neural networks for bond rating improved by multiple hidden layers*. In *Neural Networks, 1990, IJCNN International Joint Conference on*, 157–162, vol.2.

Tetko, I. V. y Villa, A. (1997). An enhancement of generalization ability in cascade correlation algorithm by avoidance of overfitting/overtraining problem. *Neural Process. Lett.*, 6(1-2):43–50.

Timmermann, A. (2006). Forecast Combinations. *Handbook of Economic Forecasting, Chapter 04, vol. 1, pp 135-196*. Elsevier, 1 edition.

Trippi, R. y Turban, E. (1996). *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real-world Performance*. Irwin Professional Pub.

Urrutia, J. L. (1995). Tests of random walk and market efficiency for Latin American emerging equity markets. *Journal of Financial Research*, 18(3).

Valdez-Cepeda, R. D. (2000). *Complejidad de los precios nacionales de productos hortícolas e índices económicos mensuales*. *Revista Chapingo Serie Horticultura*, VI: 203–211.

Valle, M. V., Díaz Padilla, G., Villalobos, J. (2011). *Aleatoriedad de una serie de precipitación en el estado de Veracruz México*. *Revista Mexicana de Ciencias Agrícolas* Núm. 1 p. 41-55.

Van Horne, J. y Wachowicz, J. (1998). *Fundamentals of Financial Management*. Prentice Hall.

van Wezel, M., Kok, J., y Sere, K. (1996). Determining the number of dimensions underlying customer-choices with a competitive neural network. In *Neural Networks, 1996. IEEE International Conference on*, volume 1, 484–489 vol.1.

Velasco Robles, G. (2008). *Medición de la relación riesgo-rendimiento y desempeño de los fondos de inversión en México mediante modelos Garch in mean y Egarch in mean*. Tesis de Maestría, UDLA.

Wheelock, D. C. y Wilson, P. W. (2000). Why do banks disappear? the determinants of U.S. bank failures and acquisitions. *Review of Economics and Statistics*, 82:127–138.

White, H. (1988). Economic prediction using neural networks: the case of IBM daily stock returns. In *International Symposium on Neural Networks*.

Whitley, L. D., Gordon, V. S., y Mathias, K. E. (1994). Lamarckian evolution, the baldwin effect and function optimization. In *Proceedings of the International Conference on Evolutionary*

*Computation. The Third Conference on Parallel Problem Solving from Nature: Parallel Problem Solving from Nature*, PPSN III, 6–15, London, UK. Springer-Verlag.

Granger, C.W.J. y Newbold, C. G. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, vol. 2, no. 2 111-20.

Yao, X. (1999). Evolving artificial neural networks. *Proceedings of the IEEE*, 87(9):1423-1447.

Yao, X. y Liu, Y. (1997a). *EPNet for chaotic time-series prediction*. In SEAL'96: Selected papers from the First Asia-Pacific Conference on Simulated Evolution and Learning, 146-156, London, UK. Springer-Verlag.

Yao, X. y Liu, Y. (1997b). A new evolutionary system for evolving artificial neural networks. *IEEE Transactions on Neural Networks*, 8(3):694–713.

Yule, G. U. (1927). *On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers*. Philosophical Transactions of the Royal Society.

Yule, G. U. (1926). *Why do we sometimes get nonsense-correlations between time-series? a study in sampling and the nature of time-series*. Journal of the Royal Statistical Society, Vol. 89, No. 1.

Yule, G. U. (July 1921). *On the time-correlation problem, with especial reference to the variate-difference correlation method*. Journal of the Royal Statistical Society, Vol. 84, No. 4.

Zhang, G. (2003). *Time series forecasting using a hybrid ARIMA and neural network model*. Neurocomputing, 50(0):159–175.

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. P., Patuwo, B. E., y Hu, M. Y. (2001). *A simulation study of artificial neural networks for nonlinear time-series forecasting*. Computers & Operations Research.

Zorin, A. y Borisov, A. (2002). *Modelling Riga stock exchange index using neural networks*. In Proceedings of the International Conference "Traditions and Innovations in Sustainable Development of Society. Information Technologies", 312–320.

# Anexo



Mariana Rodríguez

para mí

Estimado Luis Manuel León,

Muchas gracias por su envío. Confirmando que he recibido su artículo, mismo que será enviado al Consejo Editorial quien determinará si procede a dictamen.

Sin más por el momento reciba un cordial saludo,

M. Lucía Rodríguez  
Coordinadora Editorial Interina REMEF  
+5215551557080

Leer

El El sáb, 10 de junio de 2017 a la(s) 00:58, Luis leon <luisleonanaya@gmail.com> escribió:  
A quien corresponda

Por medio del presente correo electrónico, hago del conocimiento que tengo interés de que tomen en consideración el artículo denominado como:

**"DESCOMPOSICIÓN Y PREDICCIÓN DEL ÍNDICE IPC UTILIZANDO LA DESCOMPOSICIÓN EMPÍRICA EN MODOS, REDES NEURONALES"**

El artículo, resumen, currículo y foto son adjuntados en el presente mensaje. Esperando contar con una opinión favorable.

Sin más por el momento y agradeciendo de antemano su valioso tiempo, me despido de manera respetuosa y quedo a sus órdenes para cualquier duda o aclaración.

Atentamente

Luis Manuel León Anaya

Dr. Gerardo Dubcovsky

Editor Principal de la REMEF, Revista Mexicana de Economía y Finanzas, Nueva Época

EL INSTITUTO DE MÁS ALTO NIVEL  
DE LAS FINANZAS EN MÉXICO

www.imef.org.mx



SISTEMA DE CLASIFICACIÓN DE REVISTAS (<http://www.revistascyconacyt.mx/index.php/inicio>) | CONVOCATORIA 2017  
(<http://www.revistascyconacyt.mx/index.php/Convocatoria>) | RESULTADOS (<http://www.revistascyconacyt.mx/convocatoria/resultados.pdf>) | MANUAL DEL SISTEMA

Sistema de Clasificación de Revistas Mexicanas de Ciencia y Tecnología

Buscar revista (Por Título, ISSN, E-ISSN, Institución Editora o Editores)



Inicio (/) / V. Ciencias Sociales / Revista Mexicana de Economía y Finanzas REMEF

## Revista Mexicana de Economía y Finanzas REMEF

ISSN: 1665-5346

Descargar PDF ([http://www.revistascyconacyt.mx/index.php/print\\_pdf](http://www.revistascyconacyt.mx/index.php/print_pdf))

Dimensión	Puntaje Dimensión	Criterio	Puntaje Criterio
Política y gestión editorial 4.44 % / 15.00 %	8 pts / 27 pts	1.1 Disponibilidad de título paralelo y abstracts en inglés	0 pts
		1.2 Tipo de revisión por pares	4 pts
		1.3 Aplicación de criterios éticos	0 pts
		1.4 Internacionalidad de los editores y del comité científico	1 pts
		1.5 Internacionalidad de los autores	0 pts
		1.6 Porcentaje de autores de la propia institución	3 pts
		1.7 Número de artículos de producción citables publicados	0 pts
Calidad del contenido 3.75 % / 25.00 %	3 pts / 20 pts	2.1 Claridad de los abstracts	2 pts
		2.2 Calidad y conformidad con el alcance (scope) de la revista	1 pts
		2.3 Legibilidad de los artículos	0 pts
Nivel de citación 8.33 % / 25.00 %	4 pts / 12 pts	3.1 Citación de los artículos de la revista en Scopus	2 pts
		3.2 Citación de los artículos de la revista en Google Scholar	2 pts
		3.3 Citación de los editores de la revista en Scopus	0 pts
Cumplimiento de la frecuencia de publicación 6.25 % / 10.00 %	5 pts / 8 pts	4.1 Publicación al inicio del periodo programado	4 pts
		4.2 Tiempo entre aceptación, revisión y aprobación	1 pts
		4.3 Uso de plataforma de edición en línea	0 pts
Accesibilidad 5.38 % / 10.00 %	7 pts / 13 pts	5.1 Contenido disponible on-line	3 pts
		5.2 Disponibilidad en inglés del home page de la revista	2 pts
		5.3 Calidad del home page de la revista	2 pts
		5.4 Edición de artículos en formato XML	0 pts
Visibilidad internacional 11.25 % / 15.00 %	15 pts / 20 pts	6.1 Indización en Scielo	6 pts
		6.2 Indización en Redalyc	6 pts
		6.3 H-Index Google Scholar	0 pts
		6.4 Puntaje según características editoriales de Latindex Catálogo	3 pts
		6.5 La revista está incluida en el Directory of Open Access Journals (DOAJ)	0 pts
<b>La revista obtuvo 39.40 pts.</b>			

## TRABAJOS DERIVADOS DE LA TESIS

### ARTICULOS

2013; **PREDICCIÓN DE LA DEMANDA DE ENERGÍA ELÉCTRICA COMPARANDO MÉTODOS CLÁSICOS CONTRA REDES NEURONALES ARTIFICIALES EVOLUCIONADAS** ISBN:978-607-95630-4-2, Luis León Anaya, Víctor M. Landassuri Moreno, C. Bustillo–Hernández, Héctor, Rafael Orozco Aguirre, **RVP-AI/2013, IEEE Sección México, Memorias de congresos.**

### POR PUBLICAR

2017; **DESCOMPOSICIÓN Y PREDICCIÓN DEL ÍNDICE IPC UTILIZANDO LA DESCOMPOSICIÓN EMPÍRICA EN MODOS, REDES NEURONALES ARTIFICIALES EVOLUTIVAS Y ARIMA-ARCH**, Luis Manuel León Anaya, Víctor Manuel Landassuri Moreno, Héctor Rafael Orozco Aguirre, Maricela Quintana López. Revista Mexicana de Economía y Finanzas

### PARTICIPACION EN CONGRESOS

- 2012; **PRONÓSTICO DE SERIES DE TIEMPO ECONÓMICO-CAÓTICAS APLICANDO PROGRAMACIÓN EVOLUTIVA DE REDES NEURONALES ARTIFICIALES**, Nacional, **Coloquio de Investigación de la Maestría en Ciencias de la Computación 2012-A**, Luis Manuel León Anaya, Víctor M. Landassuri Moreno, **México.**
- 2012; **COMPARACIÓN DE MODELOS DE PREDICCIÓN PARA DATOS MACROECONÓMICOS**, Nacional, **LV Congreso Nacional de Física**, Luis León Anaya, Víctor M. Landassuri Moreno, C. Bustillo-Hernández, **México.**
- 2013; **PREDICCIÓN DE LA DEMANDA DE ENERGIA ELECTRICA COMPARANDO METODOS CLASICOS CONTRA REDES NEURONALES ARTIFICIALES EVOLUCIONADAS**, Nacional, **RVP-AI/2013**, Luis Manuel León Anaya, **México.**
- 2013; **COMPARACIÓN ENTRE MÉTODOS DE PREDICCIÓN PARA LA DEMANDA DE ENERGÍA ELÉCTRICA EN CORTO PLAZO**, Nacional, **5ta semana Nacional de Sistemas y Computación**, Luis Manuel León Anaya, Víctor M. Landassuri Moreno, **México.**
- 2013; **COMPARACIÓN DE MODELOS ESTADÍSTICOS Y NO PARAMÉTRICOS PARA PREDICCIÓN DE CONSUMO ELÉCTRICO**, Nacional, **LVI Congreso Nacional de Física**, Luis Manuel León Anaya, Víctor M. Landassuri Moreno, Héctor Rafael Orozco Aguirre, C. Bustillo-Hernández, **México.**