

Centro de Investigación Científica y de Educación Superior de Ensenada



**RED NEURONAL PARA EL RECONOCIMIENTO DE
PATRONES Y PREDICCIÓN EN SERIES DE TIEMPO
DE COTIZACIONES DE INSTRUMENTOS
FINANCIEROS CORRELACIONADOS**

**TESIS
MAESTRIA EN CIENCIAS**

JOSE MARTIN OLGUIN ESPINOZA

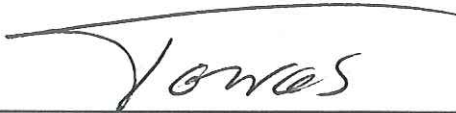
ENSENADA, B. C., JUNIO DEL 2000.

TESIS DEFENDIDA POR
José Martín Olguín Espinoza
Y APROBADA POR EL SIGUIENTE COMITÉ



Dr. Hugo Homero Hidalgo Silva

Director del Comité



M.C. Jorge Torres Rodríguez

Miembro del Comité



M.C. Jorge Enrique Preciado Velasco

Miembro del Comité



M.C. José Luis Biseño Cervantes

*Jefe del Departamento de
Ciencias de la Computación*



Dr. Federico Graef Ziehl

Director de Estudios de Posgrado

10 de julio de 2000.

CENTRO DE INVESTIGACIÓN CIENTÍFICA Y EDUCACIÓN
SUPERIOR DE ENSENADA

DIVISIÓN DE FÍSICA APLICADA

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

RED NEURONAL PARA EL RECONOCIMIENTO DE PATRONES Y
PREDICCIÓN EN SERIES DE TIEMPO DE COTIZACIONES DE
INSTRUMENTOS FINANCIEROS CORRELACIONADOS

TESIS

que para cubrir parcialmente los requisitos necesarios para obtener
el grado de MAESTRO EN CIENCIAS presenta:

JOSÉ MARTÍN OLGUÍN ESPINOZA

Ensenada, Baja California, México, junio de 2000.

RESUMEN de la tesis de José Martín Olguín Espinoza, presentada como requisito parcial para la obtención del grado de MAESTRO EN CIENCIAS en CIENCIAS DE LA COMPUTACIÓN. Ensenada, Baja California, México, junio de 2000.

RED NEURONAL PARA EL RECONOCIMIENTO DE PATRONES Y PREDICCIÓN EN
SERIES DE TIEMPO DE COTIZACIONES DE INSTRUMENTOS FINANCIEROS
CORRELACIONADOS

Resumen aprobado por:


Dr. Hugo Homero Hidalgo Silva
Director de tesis

El ser humano desde siempre ha estado preocupado por conocer los hechos futuros para poder tomar sus previsiones. La ciencia provee mecanismos que tratan de llevar a cabo esta tarea. Entre las teorías se ubican el análisis de series de tiempo, el cual consiste en identificar características que permitan clasificar y predecir el comportamiento de series de datos. En los últimos años para lograr este tipo de análisis se han venido utilizando herramientas computacionales. Dentro de estas destaca el uso de las Redes Neuronales Artificiales (RNA), las cuales se caracterizan por su poder de ajuste para sistemas no-lineales, característica que las hace especiales para el estudio de las series de tiempo.

Este trabajo de tesis aplica la teoría de las RNA para encontrar patrones y tratar de predecir el comportamiento de la correlación entre dos series de tiempo. Los datos corresponden a precios de instrumentos financieros que se cotizan en mercados de futuros. Se espera que el uso de las RNA provea formas de análisis más sencillos que los tradicionalmente utilizados y que a través del "aprendizaje" del comportamiento de las series, permita la predicción de la tendencia que presenta la correlación entre los dos instrumentos financieros.

Se presenta la teoría relacionada con las RNA y el tratamiento de las series de tiempo, las consideraciones especiales cuando se manejan series de datos financieros y los factores que influyen en la dificultad de la predicción.

Palabras clave: predicción, serie-temporal, red-neuronal

ABSTRACT of the thesis presented by **José Martín Olguín Espinoza**, as partial requirement to obtain the degree of **MASTER IN SCIENCE** with specialization in **COMPUTER SCIENCE**. Ensenada, Baja California, México, June 2000.

**NEURAL NET FOR TIME SERIES PATTERN RECOGNITION AND
PREDICTION OF CORRELATED FINANCIAL INSTRUMENTS DATA**

The human being has always been worried about knowing the future in order to foresight. Science provides ways to accomplish this task. Time series analysis is located among this theories, it consist in identifying features that allow classification and prediction of the series behavior. Over the last years, computational tools have been used in order to get this kind of analysis, artificial neural networks (ANN) stands out within this tools, their self adjusting feature for non-linear systems makes them special for time series study.

This work applies the theory of ANN for patterns finding and for trying to predict the behavior of the correlation between two time series. The data correspond to finance instruments that are cotised in the futures market. It is expected that the use of ANN provide ways of analysis simpler than the traditional ones and through the "learning" of the series behavior, allowing the trend prediction of the correlation between the two finance instruments.

In this thesis, the theory related with ANN and the time series treatment is presented, also the special considerations when finance series are used, and the factors that have influence over the prediction difficulty.

Keywords: prediction, time-series, neural-net

DEDICATORIA

A mis padres, por razones obvias.

A mi esposa Mónica y a mi hijo Héctor Manuel, porque todo lo que hago es siempre pensando en ustedes.

A mi amigo y “hermano mayor” Francisco Venegas, porque gracias a sus “amables consejos” me alentó a iniciar la aventura que concluye con este trabajo. Pancho, no perdí el tiempo...

AGRADECIMIENTOS

A mi maestro y director de tesis, Dr. Hugo Homero Hidalgo Silva, por sus atinados consejos durante el desarrollo de este trabajo.

A los miembros del comité de tesis, M.C. Jorge Torres y M.C. Jorge Preciado, por la minuciosa revisión y por haberle dedicado parte de su valioso tiempo a este trabajo.

A la Matemática Mónica Mendiola, por ayudarme a resolver mis dudas en la materia.

A mi amigo Luis Enrique Vizcarra, por sus sabios consejos en la conformación de este documento.

Al personal del Departamento de Desarrollo del Centro de Cómputo Universitario Unidad Ensenada de la UABC, por apoyarme en el soporte técnico computacional necesario para llevar a cabo este trabajo.

A la Universidad Autónoma de Baja California.

Al Consejo Nacional de Ciencia y Tecnología.

Al Centro de Investigación Científica y Educación Superior de Ensenada.

Índice General

I. INTRODUCCIÓN	1
I.1 ANTECEDENTES	2
I.1.1 <i>Series de tiempo de datos financieros.....</i>	<i>2</i>
I.1.2 <i>Redes Neuronales Artificiales aplicadas a las series de tiempo financieras.....</i>	<i>3</i>
I.2 OBJETIVOS DEL TRABAJO DE TESIS.....	5
I.2.1 <i>Objetivo General.....</i>	<i>5</i>
I.2.2 <i>Objetivos Específicos.....</i>	<i>5</i>
I.2.3 <i>Entrada y salida de la RNA.....</i>	<i>5</i>
I.2.4 <i>Alcance de la RNA.....</i>	<i>6</i>
II. MARCO TEÓRICO	7
II.1 SERIES DE TIEMPO.....	7
II.1.1 <i>Conceptos Básicos.....</i>	<i>7</i>
II.1.2 <i>Coefficiente de Correlación.....</i>	<i>9</i>
II.2 REDES NEURONALES ARTIFICIALES.....	11
II.2.1 <i>Conceptos Básicos.....</i>	<i>11</i>
II.2.2 <i>Red de Retropropagación.....</i>	<i>11</i>
II.2.3 <i>Algoritmo de aprendizaje.....</i>	<i>13</i>
II.3 REDES NEURONALES ARTIFICIALES Y LAS SERIES DE TIEMPO.....	16
II.3.1 <i>Representación de las Series de Tiempo.....</i>	<i>18</i>
II.3.2 <i>Aprendizaje de Series de Tiempo (Secuencias Temporales).....</i>	<i>18</i>
II.3.3 <i>Tipos de Arquitectura de RNA para el Aprendizaje de Series de Tiempo.....</i>	<i>19</i>
II.4 PREDICCIÓN DE SERIES DE TIEMPO CON RNA.....	23
II.4.1 <i>Dimensión inmersa.....</i>	<i>23</i>
II.4.2 <i>El espacio de estados de una serie de tiempo.....</i>	<i>24</i>
II.4.3 <i>Consideraciones en la predicción de series de tiempo financieras.....</i>	<i>25</i>
II.5 DESARROLLO DE UNA RNA.....	27
II.6 FUTUROS FINANCIEROS.....	35
II.6.1 <i>Conceptos Básicos.....</i>	<i>35</i>
II.6.2 <i>Futuros de Combustóleo y de Gasolina sin Plomo.....</i>	<i>38</i>
III. DESARROLLO	42
III.1 RECOLECCIÓN DE DATOS.....	42
III.1.1 <i>Descripción de los conjuntos de datos.....</i>	<i>42</i>
III.2 SELECCIÓN DEL CONJUNTO DE ENTRENAMIENTO Y EL CONJUNTO DE PRUEBA.....	44
III.3 ESCALAMIENTO DE LOS DATOS.....	48
III.4 ENTRENAMIENTO DE LA RED.....	49
III.4.1 <i>Primera etapa de pruebas.....</i>	<i>50</i>
III.4.2 <i>Segunda etapa de experimentos.....</i>	<i>53</i>

IV. PRESENTACIÓN Y DISCUSIÓN DE RESULTADOS.....	56
IV.1 GRÁFICAS DE PREDICCIÓN.....	56
IV.2 ANÁLISIS DEL DESEMPEÑO DE LA RNA.....	64
IV.2.1 <i>Resultados del análisis con histogramas y diagramas de dispersión.</i>	64
IV.2.2 <i>Resultados del análisis utilizando el coeficiente de Theil.</i>	73
V. CONCLUSIONES Y RECOMENDACIONES.....	75
BIBLIOGRAFÍA	78

Índice de Figuras

Figura 1. Ajuste de una función a un conjunto de datos, el objetivo es la minimización del error (E).	12
Figura 2. Representación esquemática de una RNA típica de 3 capas (entrada, oculta y salida).	13
Figura 3. Componentes de una RNA orientada a predicción en series de tiempo.	20
Figura 4. RNA con retardo temporal (TDNN). Salida en $t+1$ con una ventana de m elementos.	22
Figura 5. RNA parcialmente recurrente.	23
Figura 6. Histograma de frecuencias del tamaño del error.	32
Figura 7. Ejemplo de un diagrama de dispersión.	33
Figura 8. Comportamiento de la posición larga o de compra.	38
Figura 9. Comportamiento de la posición corta o de venta.	38
Figura 10. Cotizaciones de HOZ y HUZ de 1996.	40
Figura 11. Cotizaciones de HO y HU de 1997.	46
Figura 12. Predicción para la prueba A.	57
Figura 13. Predicción para la prueba B.	58
Figura 14. Predicción para la prueba E.	59
Figura 15. Predicción para la prueba I.	60
Figura 16. Predicción para la prueba K.	61
Figura 17. Predicción para la prueba M.	62
Figura 18. Predicción para la prueba N.	63
Figura 19. Histograma de frecuencias de los errores para la prueba A.	66
Figura 20. Diagrama de dispersión para la predicción de la prueba A.	66
Figura 21. Histograma de frecuencias de los errores para la prueba B.	67
Figura 22. Diagrama de dispersión para la predicción de la prueba B.	67
Figura 23. Histograma de frecuencias de los errores para la prueba E.	68
Figura 24. Diagrama de dispersión para la predicción de la prueba E.	68
Figura 25. Histograma de frecuencias de los errores para la prueba I.	69
Figura 26. Diagrama de dispersión para la predicción de la prueba I.	69
Figura 27. Histograma de frecuencias de los errores para la prueba K.	70
Figura 28. Diagrama de dispersión para la predicción de la prueba K.	70
Figura 29. Histograma de frecuencias de los errores para la prueba M.	71
Figura 30. Diagrama de dispersión para la predicción de la prueba M.	71
Figura 31. Histograma de frecuencias de los errores para la prueba N.	72
Figura 32. Diagrama de dispersión para la predicción de la prueba N.	72

Índice de Tablas

<i>Tabla I. Codificación de los meses utilizada por los contratos de futuros.</i>	<i>41</i>
<i>Tabla II. Ejemplos de la nomenclatura para archivos con cotizaciones de futuros HO.</i>	<i>43</i>
<i>Tabla III. Estructura de los archivos de cotizaciones de futuros HO.</i>	<i>43</i>
<i>Tabla IV. Ejemplos de la nomenclatura para archivos con cotizaciones de futuros HU.</i>	<i>44</i>
<i>Tabla V. Estructura de los archivos de cotizaciones de futuros HU.</i>	<i>44</i>
<i>Tabla VI. Matrices de correlación de las cotizaciones HO y HU para 1996 y 1997.</i>	<i>47</i>
<i>Tabla VII. Parámetros estadísticos para las series de datos de 1996 de los instrumentos HO y HU y sus diferencias.</i>	<i>48</i>
<i>Tabla VIII. Parámetros estadísticos para las series de datos de 1997 de los instrumentos HO y HU y sus diferencias.</i>	<i>49</i>
<i>Tabla IX. Concentrado de resultados de la primera etapa de pruebas con los datos de 1996.</i>	<i>51</i>
<i>Tabla X. Concentrado de resultados de la primera etapa de experimentos con los datos de 1997.</i>	<i>53</i>
<i>Tabla XI. Concentrado de resultados en la segunda etapa de pruebas con los datos de 1996.</i>	<i>54</i>
<i>Tabla XII. Concentrado de resultados en la segunda etapa de pruebas con los datos de 1997.</i>	<i>55</i>
<i>Tabla XIII. Concentrado de resultados del cálculo del coeficiente de Theil por cada prueba.</i>	<i>73</i>

RED NEURONAL PARA EL RECONOCIMIENTO DE PATRONES Y PREDICCIÓN EN SERIES DE TIEMPO DE COTIZACIONES DE INSTRUMENTOS FINANCIEROS CORRELACIONADOS

I. INTRODUCCIÓN

Las Redes Neuronales Artificiales (RNA) representan una nueva y creciente tecnología con un amplio potencial de aplicaciones en diversas áreas. Una de las áreas que ha invertido fuertes recursos en los últimos años en el campo de las RNA es el área financiera; el espectro de aplicaciones va desde la autorización de créditos hasta el manejo de estrategias de grandes inversiones. Muchas de estas aplicaciones han reportado proveer incrementos en rentabilidad de un 30 por ciento o más. [Trippi y Turban, 1996]

Una de las características básicas de las RNA que las hacen aptas para aplicarlas en el área bursátil es su capacidad de "aprendizaje"; las RNA pueden ser utilizadas para clasificación de datos, reconocimiento de patrones y para la predicción de series de tiempo. Con estas operaciones, estamos en capacidad de predecir riesgos de bancarrota de bancos o empresas, hacer diagnósticos de tensión financiera corporativa, y por supuesto la predicción de cotizaciones de instrumentos financieros [Trippi y Turban, 1996].

La información histórica de los precios de instrumentos financieros a lo largo de varios años (analizada como series de tiempo) y las técnicas de RNA, ofrecen un campo fértil de investigación que desemboca en el desarrollo de aplicaciones interesantes. Las cuales, nos permiten analizar el comportamiento de estos instrumentos financieros y predecir escenarios

de inversión propicios. Por parte de las ciencias computacionales, se pueden plantear arquitecturas de RNA que basadas en los modelos tradicionales, puedan ser modificadas para un mejor desempeño en el tratamiento de este tipo de información.

I.1 Antecedentes

I.1.1 Series de tiempo de datos financieros

Los mercados financieros mundiales se han caracterizado en los últimos años por determinar el rumbo económico de muchas naciones, en particular de aquellas en vías de desarrollo. Las fluctuaciones de estos mercados tienen un impacto cada vez mayor en la economía mundial, sobre todo en los años recientes debido al acelerado desarrollo de las tecnologías de telecomunicaciones y cómputo.

La base de los mercados financieros es la oferta y demanda de bienes como el ganado, granos o energéticos; instrumentos financieros como las acciones de las empresas; y de los instrumentos financieros derivados, de reciente aparición, como los futuros y las opciones, entre otros.

Todos estos valores se cotizan diariamente en diferentes mercados como el New York Mercantile Exchange (NYMEX), el Chicago Mercantile Exchange (CMEX), la Bolsa Mexicana de Valores (BMV) entre otros.

La principal motivación de los participantes en los mercados es generar ganancias a partir de una receta muy simple: “Comprar barato, vender caro”, y esto se da a través del análisis de

los datos de las cotizaciones de estos instrumentos. Estos datos se analizan a través de la conformación de series de tiempo, formadas con los precios de los instrumentos en los mercados, muestreados en intervalos que pueden variar, desde minutos hasta semanas o meses.

Diariamente se genera una gran cantidad de datos en todos los mercados financieros del mundo, los cuales son analizados para identificar tendencias y estar en posibilidad de predecir los movimientos a futuro, y de esta manera saber cuando hacer los movimientos de compra-venta óptimos.

I.1.2 Redes Neuronales Artificiales aplicadas a las series de tiempo financieras

Una de las áreas con mayor interés en promover el desarrollo y aplicación de las Redes Neuronales Artificiales (RNA) en series de tiempo, es la relacionada con el ámbito financiero, en la cual se han venido utilizando en diversas tareas, como la predicción del comportamiento crediticio de un consumidor, predicción de condiciones de bancarrota, reconocimiento de patrones de desorden financiero, extracción de conocimiento a partir de reportes de contabilidad, y muchas más [Trippi y Turban, 1996].

A finales de la década de los ochentas, la industria bursátil experimentó una revolución tecnológica, alentada principalmente por las nuevas posibilidades de las telecomunicaciones y la informática, la cual trajo consigo la aparición de las interrelaciones de los mercados financieros y con esto la necesidad de una perspectiva global en las transacciones [Mendelsohn, 1993]. Esta necesidad promovió la aplicación de las nuevas técnicas de

predicción basadas en el análisis de datos provenientes de múltiples fuentes, en contraste con el análisis tradicional de un solo mercado. Con este nuevo paradigma de mercados interconectados se dificulta el hacer un análisis de predicción confiable sin tomar en cuenta las interrelaciones de varios mercados al mismo tiempo.

Entre los trabajos orientados a la predicción de series de tiempo financieras tenemos la red neuronal para predecir las ganancias en acciones de IBM [White, 1988]; el sistema TOPIX, un complejo conjunto de RNAs que combinadas analizan datos de múltiples fuentes para recomendar momentos precisos para compra y venta de acciones [Kimoto, Asakawa, *et al.*, 1990]; un sistema similar, construido para determinar los movimientos de compra-venta óptimos en contratos de futuros, conformado por la aplicación de reglas de inferencia a los resultados de varias RNAs [Trippi y Turban, 1996].

Actualmente existen varios sistemas comerciales enfocados a la predicción financiera usando RNA. La red NeuralTrend de la compañía Trendy Systems es utilizada como base para dos sistemas de apoyo a transacciones:

1. Millenium, sistema enfocado en futuros.
2. Stocktrader, para acciones de empresas.

NeuralTrend hace predicciones a corto plazo utilizando cotizaciones al cierre de los mercados.

Otro sistema comercial es VantagePoint, este se enfoca en encontrar patrones ocultos y relaciones entre múltiples mercados.

Por último tenemos el resultado de investigaciones en el Instituto de Ciencias Nucleares de la UNAM, donde desarrollaron una RNA que predice el inicio y fin de las burbujas de predictibilidad en series de tiempo financieras, lapso en el cual, crea un agente artificial para explotar esta ventaja y obtener las máximas ganancias [Stephens,1996].

I.2 Objetivos del trabajo de Tesis.

I.2.1 Objetivo General.

Desarrollar una red neuronal artificial que prediga el comportamiento a corto plazo de la correlación entre dos series de tiempo correspondientes a cotizaciones de instrumentos financieros.

I.2.2 Objetivos Específicos.

- a) Explorar los tipos de RNA que han sido utilizadas para la predicción de series de tiempo.
- b) Determinar los problemas involucrados en la aplicación de RNA a la predicción de series de tiempo.

I.2.3 Entrada y salida de la RNA.

Entrada: Dos series de tiempo que corresponden a los siguientes datos:

- a) Diferencias entre las cotizaciones de los futuros del mes de Diciembre de cualquier año para combustóleo (HOZ) y gasolina (HUZ).
- b) Diferencias entre las cotizaciones de precios actuales del combustóleo y de la gasolina.

Salida: Valor correspondiente a la diferencia entre HO y HU para el día siguiente de operaciones en el mercado.

I.2.4 Alcance de la RNA.

Con la finalidad de aprovechar la ventaja que brinda el comportamiento estacional de los futuros para diciembre de combustóleo y gasolina, el modelo de RNA analizará los datos correspondientes a las cotizaciones HOZ y HUZ, así como los datos de los precios actuales de esos mismos productos energéticos (HOCash y HUCash).

II. MARCO TEÓRICO

II.1 Series de Tiempo

A través de la historia se ha visto que uno de los máximos anhelos del ser humano es el contar con el poder de la predicción, el tener la posibilidad de saber con certeza los acontecimientos que ocurrirán en el futuro, ya sea inmediato o a largo plazo. Como respuesta a esta inquietud histórica la ciencia ha desarrollado diversas teorías en múltiples áreas del conocimiento, siendo de interés especial las relacionadas con el campo de las matemáticas, en particular el análisis de las series de tiempo con propósitos de predicción.

II.1.1 Conceptos Básicos.

Una serie de tiempo se define como un conjunto de magnitudes, pertenecientes a diferentes períodos de tiempo, de cierta variable o conjunto de variables. Una serie de tiempo refleja los cambios de una variable en el tiempo, [Chou, 1990].

Una serie de tiempo también se define como una secuencia de observaciones $x_1, x_2, \dots, x_n$, usualmente ordenada en el tiempo, aunque en algunos casos la ordenación puede estar de acuerdo a otra dimensión. La característica que distingue el análisis de una serie de tiempo de otros análisis estadísticos es el reconocimiento explícito de la importancia del orden en que las observaciones fueron hechas [Anderson, 1971].

Las mediciones de los latidos del corazón de una persona enferma, la serie de datos recabados por un radiotelescopio relativos a la intensidad de la energía emitida por una

estrella, las cotizaciones del mercado de valores, etc.; todas ellas constituyen series de tiempo y la necesidad de analizar cada caso constituye un reto científico para explicar fenómenos distintos.

En términos generales existen dos tipos de análisis que pueden ser llevados a cabo sobre un conjunto de datos variables en el tiempo: la *predicción*, la cual involucra el cálculo de los valores futuros de la serie de tiempo, y la *clasificación*, la cual involucra la descripción de aspectos fundamentales de un conjunto de clases diferentes de una secuencia [Swingler, 1996]. En este trabajo de tesis el interés fundamental recae en el área de la predicción, sin embargo, muchos de los conceptos generales se aplican tanto a una tarea como a la otra.

Las técnicas utilizadas para la predicción de series de tiempo requieren un análisis de la naturaleza de los datos que componen la serie. Como resultado de este análisis se propone un modelo que represente el comportamiento de la serie de tiempo para un intervalo de tiempo dado. Para las series de tiempo, el modelo propuesto deberá ser del tipo *probabilístico*, es decir, el modelo nos proporcionará información acerca de la probabilidad de que un valor futuro esté situado dentro de dos límites específicos. A estos modelos también se les conoce como modelos *estocásticos*.

Una serie de tiempo se conoce como *estacionaria* cuando las características estadísticas son constantes en el tiempo. Cuando no existe una media natural, se le conoce como serie *no estacionaria*. Un ambiente constantemente cambiante es una de las causas de que una serie sea no estacionaria [Glass y Kaplan, 1993].

Muchas series empíricas, como el caso de las formadas por las cotizaciones de las acciones, se comportan como si no tuvieran una media fija, es decir, presentan un comportamiento no estacionario [Box y Jenkins, 1976]. Sin embargo, determinar si la serie es estacionaria o no, puede depender del intervalo de tiempo y de la longitud de los datos analizados.

Existen series estacionarias con un intervalo grande de caracterización estadística, estas series se pueden catalogar como no estacionarias si el intervalo analizado es menor al requerido para representar su comportamiento estacionario [Theiler, Linsay, *et al.*, 1993].

Si una serie de tiempo se puede caracterizar como estacionaria, la construcción del modelo de predicción contará con elementos constantes que permitirán, en teoría, extraer la tendencia subyacente de la serie, y por lo tanto, el modelo ofrecerá grados de confiabilidad más altos en relación a una serie no estacionaria.

II.1.2 Coeficiente de Correlación.

Sean X e Y dos variables aleatorias, la medida del grado de relación entre las dos variables se llama *coeficiente de correlación*, representado convencionalmente por ρ . Se define como la razón entre la covariancia entre Y y X al producto de sus respectivas desviaciones estándar [Chou, 1990]:

$$\rho = \frac{Cov(Y, X)}{\sigma_Y \sigma_X} = \frac{E[(Y - \mu_Y)(X - \mu_X)]}{\sqrt{E(Y - \mu_Y)^2} \sqrt{E(X - \mu_X)^2}} \quad (1)$$

donde $E[.]$ representa la esperanza matemática.

Cuando $Cov(Y, X) = 0$, ρ es cero, indicando que no hay relación entre las dos variables.

Cuando hay covariabilidad perfecta entre las dos variables y varían en la misma dirección, ρ es igual a 1. Análogamente si hay covariabilidad perfecta pero Y y X varían en sentidos opuestos, $\rho = -1$. Cuando existe cierto grado de covariabilidad entre las dos variables tenemos $-1 < \rho < 0$ ó $0 < \rho < 1$.

La correlación entre dos series de tiempo nos indica el grado de dependencia existente entre ellas.

En los modelos de predicción que involucran más de una serie de tiempo es importante que estas presenten un alto grado de correlación, ya que el tratamiento conjunto puede aportar más información que puede ser aprovechada por el modelo.

II.2 Redes Neuronales Artificiales.

II.2.1 Conceptos Básicos.

Las RNA surgen como una forma de imitación limitada del esquema de operación de las redes neuronales del cerebro humano, por lo menos de lo poco que se conoce sobre su funcionamiento.

Tal como en el esquema biológico, la unidad mínima de procesamiento de información en una RNA lo constituye una neurona artificial, llamada también *unidad*. La cual, recibe un vector de valores de entrada y mediante la aplicación de una *función de transferencia* a la suma ponderada de los componentes del vector, obtiene un resultado que si rebasa un límite umbral preestablecido, dispara un estímulo como salida. Una red se constituye por un gran número de unidades interconectadas entre sí en secciones denominadas *capas*, las cuales se denominan de *entrada*, *intermedias* u *ocultas* y de *salida* [Kartalopoulos, 1996].

II.2.2 Red de Retropropagación.

La RNA de retropropagación (RRP), también conocida como Perceptrón de multicapas (PMC), es actualmente el paradigma de redes neuronales de propósito general más utilizado. Las RRP deben su éxito a su capacidad para generalizar, esto es, responder a entradas nuevas de manera adecuada.

La RRP realiza el aprendizaje mediante la minimización de una función de error, utilizando descenso de gradiente, el cual se basa en intentos para ajustar una función a un

conjunto de datos empíricos, de tal forma que la solución se desvíe lo mínimo posible del valor exacto. La figura 1 ilustra el concepto de minimización de error (E) donde

$$E = \sum_{i=1}^N \left(f(x_i) - y_i \right)^2 .$$

El proceso de aprendizaje empieza con la presentación de un patrón de entrada a la RRP. El patrón de entrada (conocido como *ejemplo*) es propagado a través de toda la red, hasta que se produce un patrón de salida. La RRP utiliza la regla *delta generalizada* para determinar el error del patrón actual contribuido por cada una de las unidades de la red. Finalmente, cada unidad modifica ligeramente los pesos de sus conexiones de entrada hacia una dirección que reduzca el error, el proceso se repite para el siguiente patrón.

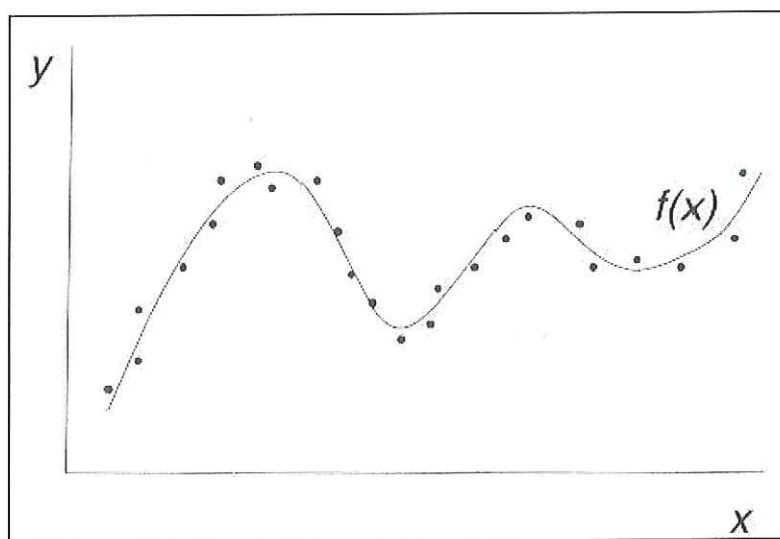


Figura 1. Ajuste de una función a un conjunto de datos, el objetivo es la minimización del error (E).

En la figura 2 está representada una RRP típica, la cual consta de una capa de entrada, una o más capas intermedias u *ocultas*, y una capa de salida. Las unidades usualmente sólo están conectadas entre capas diferentes y sólo a las inmediatamente anteriores y posteriores.

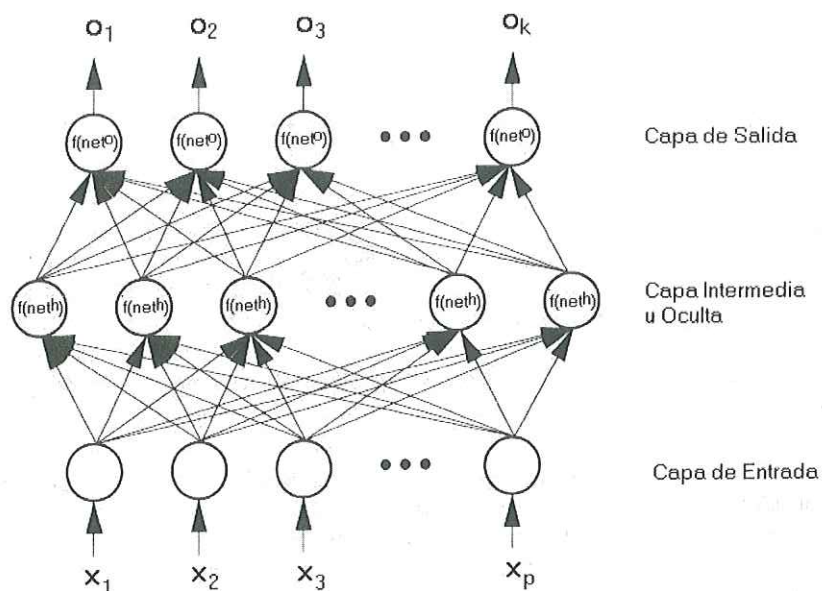


Figura 2. Representación esquemática de una RNA típica de 3 capas (entrada, oculta y salida).

II.2.3 Algoritmo de aprendizaje.

Una RNA del tipo RRP se inicia con la creación de una estructura RRP (figura 2) con valores iniciales aleatorios para los pesos. Se obtiene un conjunto de patrones de entrenamiento (*ejemplos*) que sean representativos del comportamiento que se desea aprender la RNA. Estos patrones de entrenamiento se pueden ver como un conjunto de pares ordenados de vectores $\{(x_1, y_1), (x_2, y_2), \dots, (x_p, y_p)\}$ donde cada x_i representa un vector de entrada y cada y_i representa el vector de salida asociado con el vector de entrada x_i .

La tarea del entrenamiento de la red se sujeta al siguiente algoritmo, que se deriva como resultado natural del proceso de descenso del gradiente de la superficie de error de la salida producida por la red con respecto al resultado deseado [Skapura, 1996].

1. Seleccionar el primer vector de entrenamiento ($\mathbf{x}, \mathbf{y}$).
2. Usar el vector de entrada, $\mathbf{x}$, como salida de las unidades de la capa de entrada.
3. Calcular la activación para cada unidad de la primera capa oculta utilizando:

$$net_i(t) = \sum_{j=1}^n w_{ij}(t) o_j(t) \quad (2)$$

donde:

$net_i(t)$ Señal de entrada a la i -ésima unidad de la red.

$o_j(t)$ Salida de la j -ésima unidad de la red.

$w_{ij}(t)$ Peso asociado a la conexión que va de la j -ésima unidad a la i -ésima unidad.

n Número de unidades conectadas a la entrada de la unidad i -ésima.

4. Aplicar la función de activación para cada unidad de la capa subsecuente. Para las funciones de activación se utilizará:

$f(net^h)$ para la capa oculta

$f(net^o)$ para la capa de salida

5. Repetir los pasos 3 y 4 para cada capa subsecuente de la red.
6. Calcular el error para el patrón p a través de las K unidades de la capa de salida usando la fórmula:

$$\delta_{pk}^o = (y_k - o_k) f' \left(net_k^o \right) \quad (3)$$

7. Calcular el error para las J unidades de las capas ocultas usando la fórmula recursiva:

$$\delta_{pj}^h = f' \left(net_j^h \right) \sum_{k=1}^K \delta_{pk}^o w_{kj} \quad (4)$$

8. Actualizar los pesos de las conexiones que van a las capas oculta usando la ecuación:

$$w_{ji}(t+1) = w_{ji}(t) + \eta \delta_{pj}^h x_i + \mu [w_{ji}(t) - w_{ji}(t-1)] \quad (5)$$

Donde η es la tasa de aprendizaje, un valor pequeño, entre 0 y 1, usado para limitar la magnitud del cambio permitido a cualquier conexión durante un ciclo de entrenamiento de un solo patrón. μ es el momento, un valor pequeño utilizado para evitar que la solución se quede en un mínimo local, afecta la diferencia de los valores del peso de la conexión en los pasos t y $t-1$.

9. Actualizar los pesos de las conexiones que van a la capa de salida usando la ecuación:

$$w_{kj}(t+1) = w_{kj}(t) + \eta \delta_{pk}^o f(net_j^h) + \mu [w_{kj}(t) - w_{kj}(t-1)] \quad (6)$$

10. Repetir los pasos 2 al 9 para todos los pares de vectores del conjunto de entrenamiento. A todo este ciclo se le denominará época.

11. Repetir los pasos 1 al 10 tantas épocas como sea necesario para reducir al mínimo el valor de la sumatoria al cuadrado del error (E). El cálculo de E se realiza en la capa de salida y para todos los patrones del conjunto de entrenamiento utilizando la siguiente fórmula:

$$E = \frac{1}{2} \sum_{p=1}^P \sum_{k=1}^K (y_k - o_k)^2 \quad (7)$$

II.3 Redes Neuronales Artificiales y las Series de Tiempo.

Entre las técnicas utilizadas para el análisis de las series de tiempo con fines de clasificación y predicción figuran las implementadas por las Redes Neuronales Artificiales (RNA), también llamadas redes *conexionistas* [Gershenfeld y Weigend, 1993].

En 1964 Hu aplicó la red lineal adaptiva de Widrow a la predicción del estado del tiempo. Lapedes y Farber (1987) entrenaron su red para emular la relación entre la salida y las entradas para series de tiempo caóticas generadas por computadora, y Weigend, Huberman y Rumelhart (1990, 1992) hicieron hincapié en la cuestión de encontrar redes de complejidad apropiada para predecir series de tiempo del mundo real. En todos estos casos, la información

temporal fue presentada espacialmente a la red utilizando un vector de intervalo de tiempo (Tapped Delay Line o TDNN) [Gershenfeld y Weigend, 1994].

Las RNA son usadas comúnmente en el reconocimiento de patrones, donde una colección de características es presentada a la red, y la tarea es asignar la entrada a una o más clases. Otro uso para las RNA es la regresión no lineal, donde la tarea consiste en encontrar una interpolación suave entre los puntos de entrada. En ambos casos, toda la información relevante es presentada simultáneamente a la red.

En contraste, el trabajo con series de tiempo involucra el procesamiento de patrones que evolucionan con el tiempo, esto es, la respuesta apropiada en un punto particular del tiempo depende no solamente del valor actual, sino de los datos anteriores. Se debe proveer a la RNA una forma de captar la dinámica del sistema sin que sea necesario presentarle todos los datos al mismo tiempo.

Por lo tanto, las RNA para el análisis de las series de tiempo representan un caso especial, ya que no se pueden aplicar directamente las técnicas tradicionales utilizadas en la clasificación de patrones o en el ajuste genérico de funciones. En este caso el reconocimiento de patrones temporales involucra el procesamiento de patrones que evolucionan a través del tiempo. La respuesta apropiada a un punto particular en el tiempo depende no solamente del punto actual, sino potencialmente de todos los puntos anteriores. Esto implica que la serie de datos utilizados para el entrenamiento y prueba deben ser procesados por ventanas, es decir, la entrada n a la RNA deberá ser un segmento de datos de tamaño m y la salida será el dato

$k=n+m$. Por lo tanto la RNA contará con un parámetro adicional el cual será el tamaño de la ventana.

II.3.1 Representación de las Series de Tiempo

Existen dos tipos de variables a las que se puede referir cuando se definen series de tiempo, en términos del tiempo t transcurrido desde el inicio de la serie, o en términos del número previo de valores n , inmediatamente anteriores al elemento en t . En teoría es fácil definir la serie en relación al tiempo transcurrido t , ya que la función es más sencilla. Sin embargo en la práctica, y esto se aplica especialmente a los modelos RNA, es poco aplicable usar t como entrada. Esto se debe en parte a que t puede no ser relevante y además su valor puede crecer demasiado [Swingler, 1996].

Otro concepto que apoya el uso de la representación en base a los valores previos de la serie, está dado por el *Teorema de Takens*, el cual prueba que si tomamos un vector lo suficientemente largo, constituido por valores previos de la serie, es posible reconstruir completamente la estructura subyacente de la dinámica del sistema que produjo la secuencia, [Swingler, 1996; Zhang y Hutchinson, 1993; Gershenfeld, 1993; Hertz, 1991].

II.3.2 Aprendizaje de Series de Tiempo (Secuencias Temporales).

El análisis de series de tiempo con RNA implica el aprendizaje de la estructura interna por parte de la red, este aprendizaje se puede llevar a cabo por medio de tres tareas distintas:

Reconocimiento de la secuencia, Reproducción de la secuencia y Asociación temporal [Hertz, 1991].

Reconocimiento de la Secuencia: En este caso se desea producir un patrón de salida particular cuando una secuencia de entrada específica se presenta. No es necesario reproducir la secuencia de entrada. Esta es apropiada, por ejemplo, para problemas de reconocimiento de voz, donde la salida puede indicar la palabra recién escuchada.

Reproducción de la Secuencia: En este caso la red debe ser capaz de generar el resto de sí misma cuando se le presenta una parte de la secuencia. Esta es apropiada para predecir el curso futuro de una serie de tiempo a partir de ejemplos. A este caso se le denomina *generalización de auto-asociación*.

Asociación Temporal: Es el caso más general, una secuencia particular de salida debe ser producida en respuesta a una secuencia específica de entrada. Las secuencias de entrada y de salida podrán ser totalmente diferentes. Incluye los dos casos anteriores y además el caso especial de generación pura de secuencias. A este caso se le denomina *generalización de hetero-asociación*.

II.3.3 Tipos de Arquitectura de RNA para el Aprendizaje de Series de Tiempo.

La configuración óptima de una RNA para el análisis de series de tiempo es un tema que sigue en discusión, existen diversos trabajos que sustentan el uso de muy variadas formas de configuraciones de RNA y que han aportado resultados interesantes. Sin embargo la mayoría

se aplican a casos muy específicos que impiden la generalización, en parte por la naturaleza misma de las series.

Existen dos tipos de arquitectura de RNA orientadas al análisis de las series de tiempo en las que coinciden varios autores: La red neuronal con retardo temporal o TDNN (Time Delay Neural Network) y la red neuronal parcialmente recurrente. Incluso la red de alimentación hacia adelante (feedforward) ha sido usada para este propósito [Mendelsohn, 1991].

[Mozzer, 1993] presenta una arquitectura híbrida. La predicción involucra dos componentes conceptualmente distintos. El primero consiste en construir una memoria de corto plazo que retenga aspectos de la secuencia de entrada relevantes para realizar predicciones. La segunda hace predicciones basada en esta memoria de corto plazo. Esto en el ambiente de las RNA se traduce como una red de alimentación hacia adelante (feedforward) para el predictor y una red recurrente para la representación de la memoria de corto plazo (ver figura 3).

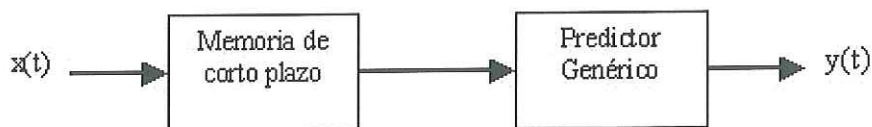


Figura 3. Componentes de una RNA orientada a predicción en series de tiempo.

II.3.3.1 RNA con retardo temporal (TDNN).

Si se convierte una secuencia temporal en un patrón espacial en la capa de entrada de la red, entonces se pueden utilizar métodos convencionales de retropropagación para aprender y reconocer la secuencia [Hertz, 1991].

El tratamiento general de la serie es el siguiente:

Sea S una serie de tiempo compuesta por los valores $s_1, s_2, \dots, s_n$. A la TDNN se le van presentando como capa de entrada los valores $s_t, s_{t-1}, s_{t-2}, \dots, s_{t-m}$ simultáneamente para cada tiempo t , donde m representa el número de pasos del retardo temporal, también conocido como tamaño de la ventana, o en otras palabras, representa la dimensión inmersa de la serie (ver II.4.1 más adelante). La representación gráfica de esta red se aprecia en la figura 4.

Este esquema se ha usado ampliamente en problemas de reconocimiento de voz por una gran cantidad de investigadores [Waibel, 1989; Sejnowski y Rosenberg, 1987; McClelland y Elman, 1986].

[Hertz, 1991] asegura que este tipo de RNA solo se aplica al reconocimiento de secuencias, existen referencias sobre su uso para la predicción (reproducción de secuencias) en [Weigend, 1994].

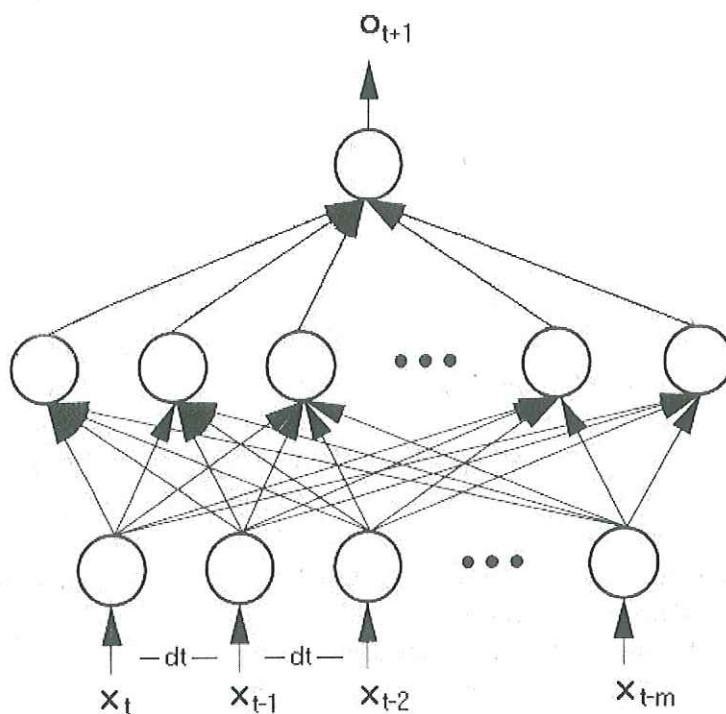


Figura 4. RNA con retardo temporal (TDNN). Salida en $t+1$ con una ventana de m elementos.

II.3.3.2 RNA Parcialmente Recurrentes.

En este tipo de arquitectura, las conexiones son principalmente de alimentación hacia adelante (feedforward), pero incluyen un conjunto de conexiones de retroalimentación (feedback) cuidadosamente seleccionados. Las unidades que almacenan los valores de retroalimentación son llamadas *unidades de contexto* (ver figura 5). La recurrencia le permite a la RNA recordar información del pasado reciente (memoria de corto plazo), pero sin complicar sustancialmente el entrenamiento. En muchos de los casos las conexiones de retroalimentación son fijas, no entrenables, de esta manera se puede fácilmente utilizar retropropagación para el entrenamiento [Hertz, 1991]. Este tipo de redes ha sido utilizado en aprendizaje de dependencias temporales en sentencias gramaticales, simulación de la máquina de Turing, aprendizaje reforzado, entre otros [Swingler, 1996].

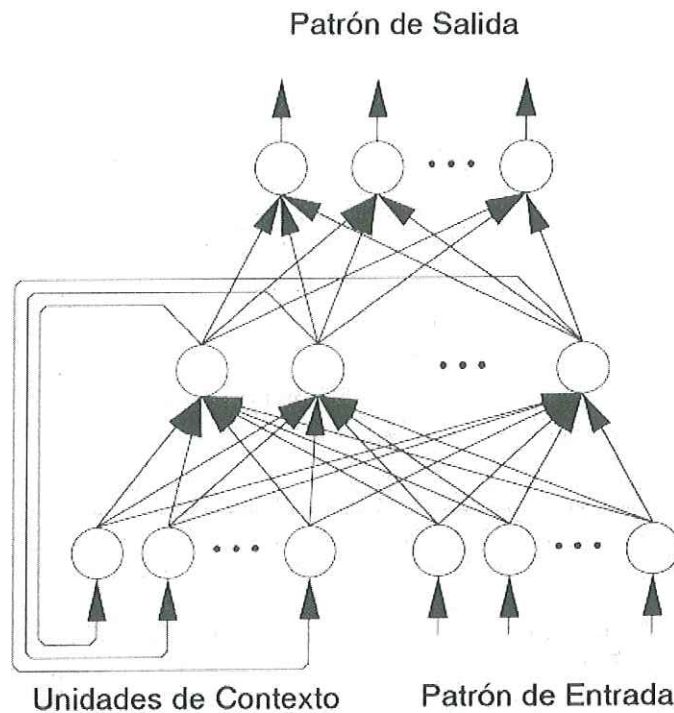


Figura 5. RNA parcialmente recurrente.

II.4 Predicción de series de tiempo con RNA

II.4.1 Dimensión inmersa.

El Teorema de Takens prueba la factibilidad de reconstruir una serie a partir de los valores previos, sin embargo no muestra cuantos valores son requeridos. El número de valores previos de la serie que determinan el siguiente es lo que se conoce como *dimensión inmersa*. No necesariamente dice que valores son importantes, simplemente cuantos. En el caso más simple, se asume que los valores recientes determinan el siguiente. Una serie de tiempo en la

cual los dos valores previos son suficientes para determinar el actual tiene una dimensión inmersa de dos [Swingler, 1996].

II.4.2 El espacio de estados de una serie de tiempo.

Como se mencionó en II.2, Takens probó la posibilidad de reconstruir la estructura de una serie a partir de un determinado número de valores previos. Por supuesto al tomar un número grande de valores anteriores de la serie se produce un número grande de grados de libertad en el sistema. Afortunadamente no es común que un sistema se mueva cubriendo completamente el espacio disponible por una larga serie de números continuos. Definir el subespacio a través del cual se mueve es un aspecto importante de la caracterización del sistema.

[Gershenfeld y Weigend, 1993] define cuatro espacios sobre los cuales una serie de tiempo puede ser considerada: El *espacio de configuración*, el *espacio de solución*, el *espacio observable* y el *espacio de estados reconstruido*.

1. El *espacio de configuración* de un sistema es todo el espacio sobre el que se puede mover sin restricción. Este espacio tiene un número de dimensiones igual al número elegido de pasos temporales hacia atrás a partir de t .
2. Más pequeño que el espacio de configuración, es el *espacio de solución*. Es el espacio en el que el sistema realmente viaja a lo largo de su desarrollo.

3. Un espacio totalmente diferente es el *espacio observable*. Es la serie de tiempo tal cual. Los observables para una serie de tiempo simple se mueven a través de un espacio unidimensional, pero usualmente son graficados contra otro espacio: *el tiempo*.
4. Es posible elaborar un *espacio de estados reconstruido* del sistema observado tomando un vector de valores previos o usando una RNA recurrente en la cual las unidades ocultas representen el espacio de la solución. De acuerdo con el teorema de Takens es posible reconstruir este espacio de estados a partir de los observables solamente.

Una RNA orientada a la predicción de series de tiempo debe ser capaz de captar la estructura del *espacio de solución* a través del *observable* y de esta forma tener la capacidad de reconstruir la secuencia. Para analizar el espacio observable, es necesario determinar la dimensión inmersa y no existe un método establecido para llevar a cabo esta tarea.

II.4.3 Consideraciones en la predicción de series de tiempo financieras.

El tratamiento de series de tiempo correspondientes a datos de instrumentos financieros es muy complejo, existen una gran diversidad de factores a considerar, a continuación se explican los más relevantes:

Hipótesis de los mercados eficientes: Existe un escepticismo justificable alrededor de la idea de que es posible hacer dinero mediante la predicción del cambio en los precios de un mercado en particular, basado solamente en su comportamiento pasado y en un número de indicadores disponibles públicamente. Este escepticismo existe debido a un número de razones, muchas de las cuales son explicadas por la hipótesis de los mercados eficientes. En su

forma más estricta, la hipótesis afirma que los mercados siguen una caminata al azar, la cual no puede ser predicha a partir de los valores previos. Cualquier oportunidad de ganancias potenciales es utilizada inmediatamente, removiendo la oportunidad casi tan rápido como fue creada. Por supuesto esta afirmación contempla la igualdad en disponibilidad de información y tecnología para todo el público.

Predicción de series univariantes y multivariantes: En algunos casos el tratamiento de series de datos financieras univariantes se resume a modelos de caminata al azar. Para estos casos el uso de RNA no significará la mejor alternativa. Sin embargo es posible identificar el espacio de solución si a la serie se le añaden variables que provean información relevante y correlacionada. Una RNA puede captar la estructura de los datos en series multivariantes, aún cuando en series univariantes no sea posible [Swingler, 1996].

Representación de los valores de salida: En el análisis de series de tiempo financieras es de vital importancia el tipo de predicción que se quiere realizar. Entre las configuraciones exploradas se encuentran las siguientes:

- Predicción del valor de la serie en el tiempo $t+1$.

Este consiste en tratar de predecir el valor exacto que la serie exhibirá en el siguiente paso. Ampliamente utilizado por ser el tipo de predicción natural.

- Predicción del movimiento del cambio.

En lugar de tratar de predecir el valor exacto, se predice si el valor de la serie en el tiempo $t+1$ será mayor, menor o igual que el del tiempo t . Se ha usado para tratar de simplificar el espacio de solución [Zhang y Hutchinson, 1993]

- Indicadores de compra-venta con umbrales.

Trata de predecir valores de movimiento de la serie y si estos sobrepasan ciertos límites, se clasifica el movimiento como ideal para comprar, vender o mantenerse [Swingler 1996].

Estos son solo algunos de los puntos que han de considerarse al realizar el análisis de series de datos financieras con fines de predicción. Son escasos los autores que presentan casos de éxito en este renglón, en el mejor de los casos solo alcanzan logros apenas por arriba del modelo de caminata al azar. Cuando se reportan casos con a alto grado de éxito, por lo general no se presentan los detalles, ya que la ventaja competitiva que implica se esfumaría tal como lo presenta la hipótesis de los mercados eficientes [Skapura, 1996].

II.5 Desarrollo de una RNA.

El desarrollo de una RNA implica las siguientes etapas: Recolección de datos, selección del conjunto de entrenamiento y el conjunto de prueba, escalamiento de los datos, entrenamiento de la red, análisis del desempeño de la red, iteración del entrenamiento y aplicación de la red. [Klimasauskas, 1991; Swingler, 1996].

1. *Recolección de datos.*

En esta etapa se especifican los datos sujetos al análisis por parte de la red. Básicamente consiste en reunir bajo un mismo formato los datos que representan el conocimiento que la red deberá obtener.

2. *Selección del conjunto de entrenamiento y el conjunto de prueba.*

Del conjunto de datos, es necesario definir cuales se utilizarán para entrenar la red y cuales serán para probar el desempeño. El criterio de selección varía de acuerdo al tipo de problema. Para el caso específico de series de tiempo, como existe una dependencia temporal entre la secuencia de valores, $x_1, x_2, \dots, x_t, x_{t+1}, \dots, x_n$, es común utilizar los primeros t valores para entrenamiento y los restantes $n-t$ como casos de prueba.

3. *Escalamiento de los datos.*

La forma en que los datos le son presentados a la RNA afecta el aprendizaje. Las RNA son capaces de procesar datos a partir de una amplia variedad de fuentes. Sin embargo, solo son capaces de procesar los datos en cierto formato, este límite está impuesto por los valores que puede producir la función de activación. Por lo tanto es necesario *escalar* los datos al mismo intervalo que maneje la función de activación. Existen diferentes métodos de escalamiento de datos, entre los más comunes tenemos:

Escalamiento lineal: Es el método más simple, consiste en ajustar los datos en un intervalo dado mediante una relación lineal. Sean $v_1, v_2, \dots, v_n$ los elementos del conjunto de

datos. Si el intervalo requerido para la función de activación está entre cero y uno, entonces el elemento escalado S_i es:

$$s_i = \frac{v_i - \min(v_{1..n})}{\max(v_{1..n}) - \min(v_{1..n})} \quad (8)$$

El escalamiento inverso lineal está dado por:

$$v_i = \min(v_{1..n}) + s(\max(v_{1..n}) - \min(v_{1..n})) \quad (9)$$

Este tipo de escalamiento se puede utilizar cuando los datos están distribuidos uniformemente, es decir, cumplen con las siguientes condiciones:

$$\max(v_{1..n}) < \bar{v} + \lambda \sigma_v \text{ y además } \min(v_{1..n}) > \bar{v} - \lambda \sigma_v \quad (10)$$

donde:

λ Determina el número de desviaciones estándar a partir de la media en la que debe caer el valor antes de considerarse un valor disparado.

$\bar{v}$ Es la media de los datos

σ Es la desviación estándar de los datos.

Escalamiento softmax: Se utiliza cuando los datos no están distribuidos uniformemente y se requiere dar peso específico a ciertos valores. El grado de escalamiento del valor es proporcional a la distancia que ese valor tenga hacia la media. Las funciones son las siguientes:

$$x = \frac{\left(v_i - \bar{v} \right)}{\lambda \frac{\sigma_v}{2 * PI}} \quad (11)$$

$$s = \frac{1}{1 + e^{-x}} \quad (12)$$

La función inversa está dada por:

$$v = \bar{v} + \frac{\lambda \sigma_v \log \sqrt{-1 + \frac{1}{1-s}}}{PI} \quad (13)$$

4. *Entrenamiento de la red.*

La selección de la arquitectura de RNA apropiada al tipo de problema es una de las tareas más difíciles de todo el proceso. La recomendación es empezar probando las más sencillas de implementar e ir aumentando el grado de complejidad conforme el aprendizaje esperado no se vaya obteniendo.

Los criterios de entrenamiento de la RNA están en función directa de la arquitectura seleccionada. En esta etapa el conjunto de datos de entrenamiento se le presentan a la red

iterativamente y de esta forma se van ajustando los pesos de las conexiones, conformando el espacio de solución. A cada iteración de todo el conjunto de datos de entrenamiento se le conoce como *época*.

En cada época de entrenamiento se monitorea la evolución del error total de entrenamiento, cuando éste error llega a un valor aceptable (ver II.2.3), se da por terminado el entrenamiento y se aplica el conjunto de prueba a la red entrenada para analizar los resultados de desempeño.

5. *Análisis del desempeño de la red.*

Una vez terminado el entrenamiento, es necesario medir el desempeño de la red, o en otras palabras, el conocimiento obtenido. Existen varios criterios a tomar en consideración para realizar esta tarea, entre los más usuales para las series de tiempo se encuentran los siguientes:

Histogramas de Error: Cuando el tamaño del error es más importante que el tipo de error (aleatorio o sistemático), la elaboración de un histograma provee un método rápido para visualizar la distribución de los errores producidos por la red. Un histograma de este tipo muestra la cuenta de la frecuencia con la que un error cae dentro de un conjunto de intervalos (ver figura 6). Una RNA bien entrenada exhibe una distribución normal en las frecuencias del error.

Gráfica de dispersión: Si los valores objetivo contra los valores producidos por la red se concentran en forma de línea recta en una gráfica de dispersión, es un indicador de un buen

desempeño de la red. El coeficiente de correlación de esta gráfica es una buena medida de la exactitud de la red (ver figura 7).

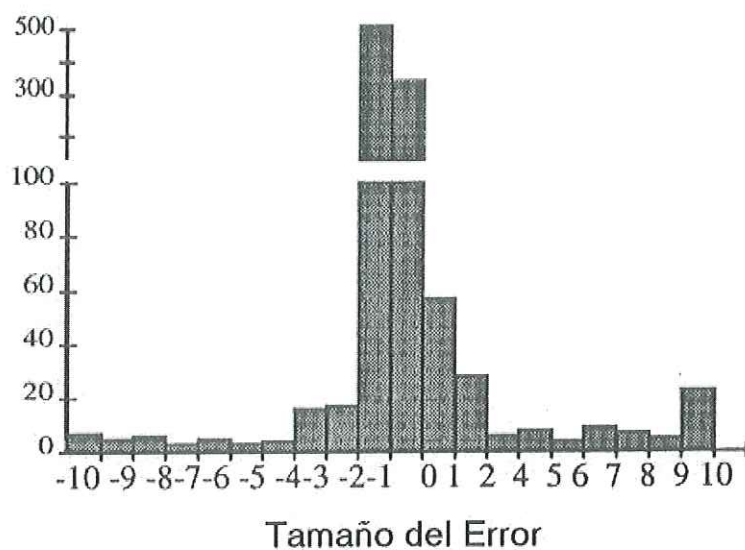


Figura 6. Histograma de frecuencias del tamaño del error.

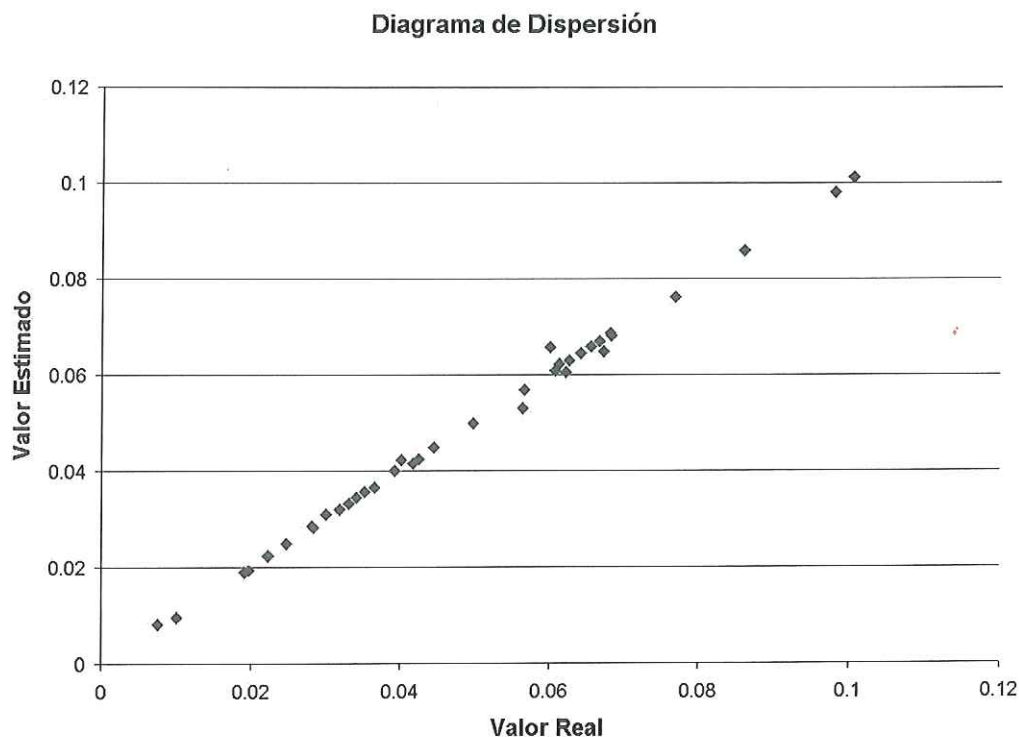


Figura 7. Ejemplo de un diagrama de dispersión.

Comparación con la caminata al azar: Una de las técnicas de análisis más utilizadas para las series de tiempo financieras es la comparación del desempeño de la red con el modelo de caminata al azar. Este modelo se basa en el hecho de que en un mercado eficiente, el precio actual es el mejor predictor del precio subsecuente, es decir $x_{t-1} = x$. La comparación de la predicción se puede realizar utilizando una variante del coeficiente de desigualdad de Theil [Swingler, 1996]:

$$T = \sqrt{\frac{\sum_{t=1}^n (y_t - x_t)^2}{\sum_{t=1}^n (x_{t-1} - x_t)^2}} \quad (14)$$

De acuerdo con este estadístico, si T es menor que uno, entonces la RNA se desempeña mejor que el modelo de caminata al azar. Un valor mayor a uno indica lo contrario y un valor igual o muy cercano a uno indica que el modelo de la RNA es equivalente a la caminata al azar [Beach, Schavey, *et al.*, 1999].

El proceso de determinar si una serie de tiempo posee un comportamiento de caminata al azar es difícil de realizar. En este caso, se utilizará el coeficiente de Theil como medida de comparación entre resultados de redes diferentes.

6. *Iteración del entrenamiento.*

El proceso de desarrollo de una RNA es iterativo, si los resultados del análisis del entrenamiento no son satisfactorios, es necesario hacer ajustes en las etapas previas al entrenamiento, ya sea un cambio en el escalamiento de los datos, modificación de la configuración de la red o cambio de arquitectura. La diversidad de parámetros que intervienen en esta técnica y la poca madurez teórica, determinan este comportamiento cíclico. En la actualidad todavía es predominante el uso de técnicas de “ensayo y error”.

7. *Aplicación de la red.*

Una vez que la red ha sido entrenada y validados los resultados, está en posibilidad de ser utilizada para conjuntos de datos distintos a los utilizados para el entrenamiento y prueba. Esta

es la etapa de producción de la RNA, es hasta este momento cuando se aprovechan las habilidades de generalización y reconstrucción del espacio de estados aprendidas durante el entrenamiento.

La RNA puede ser por sí misma toda la aplicación, pero en la mayoría de los casos está inmersa en un sistema mayor, el cual utiliza el conocimiento de la red para generar información que alimenta a otros procesos, los cuales en conjunto proveen la funcionalidad del sistema.

II.6 Futuros Financieros

II.6.1 Conceptos Básicos.

Para estar en posibilidad de analizar una serie de tiempo con fines de predicción, es necesario conocer la naturaleza de la información. Es por esto que se incluye este apartado con una descripción general del origen y el entorno que genera las series de tiempo de futuros financieros.

Dentro de la enorme gama de instrumentos financieros que se cotizan en los mercados bursátiles mundiales existen los llamados futuros. Estos son instrumentos aplicables a una gran variedad de productos que van desde los energéticos, productos agrícolas (trigo, maíz, etc.), ganadería, hasta el tipo de cambio de moneda (marco alemán, yen, etc.). Consisten básicamente en otorgar el derecho, a quien los adquiere, de comprar en la fecha pactada, un bien a un precio establecido en el contrato futuro. En términos prácticos, es un seguro contra

las fluctuaciones, tanto estacionales como las impredecibles, para la adquisición de determinados bienes, que la mayoría de las veces son materias primas [Díaz, 1996].

Los contratos están totalmente estandarizados, en el sentido de que en ellos se especifica claramente el activo en cuestión y sus características: dónde va a ser entregado, el plazo al cual se va a hacer la entrega, el monto pactado, etc.; la única variable es, pues, el precio del activo amparado.

Los futuros por sí mismos no tienen un costo, es solo el compromiso de hacerlo valer en la fecha estipulada, esto es, comprar o vender la cantidad del producto que se detalla en el contrato y al precio pactado. Sin embargo, como el contrato ampara una cantidad de producto, en términos indirectos existe un valor que variará conforme los vaivenes del costo real del producto y de muchos otros indicadores económicos. De esta manera es posible, y de hecho es común especular con este tipo de contratos. El término especulación visto como la compra y venta de futuros no por el interés del bien que asegura, sino por el hecho de aprovechar las oportunidades de diferencia entre el precio de compra y el precio de venta con el fin de generar una ganancia económica. Estos movimientos se conocen como "transacción en papel", ya que en la realidad no se lleva a cabo la entrega del bien.

Las pérdidas y ganancias que obtiene cada una de las partes participantes en el mercado, se van realizando diariamente, de acuerdo con los movimientos del valor subyacente, y por ende del precio del futuro. De acuerdo a los flujos que se generan, las operaciones con futuros resultan en un juego de suma cero, ya que lo que pierde un participante lo gana otro, esto es, la suma de las pérdidas y ganancias es igual a cero.

El patrón de ganancias y pérdidas de una *posición larga*, es decir, una posición de compra sobre un futuro, se ilustra en la figura 8. En ésta se observa que quien mantiene una posición larga acumula ganancias conforme el precio del valor subyacente sube, ya que pactó comprar el activo a un determinado precio y en el mercado dicho subyacente es cada vez más caro, con lo que su posición en el futuro se va valorando. Al vencimiento del contrato, las ganancias serán la diferencia entre el precio existente en el mercado menos el precio pactado en el futuro. Evidentemente, si el precio del valor subyacente baja en el mercado, el inversionista con una posición larga, estaría acumulando pérdidas debido a que su posición está perdiendo valor.

Para quien mantiene una *posición corta*, es decir de venta, su patrón de ganancias es el contrario al de la posición larga, es decir, conforme el precio del subyacente sube, el valor de la posición corta se reduce lo que se convierte en pérdidas para el inversionista. Por el contrario, si el precio del valor subyacente baja la posición corta se revalúa, ya que el inversionista va a vender el activo a un precio mayor que el que se observa en el mercado. Este patrón de ganancias se observa en la figura 9.

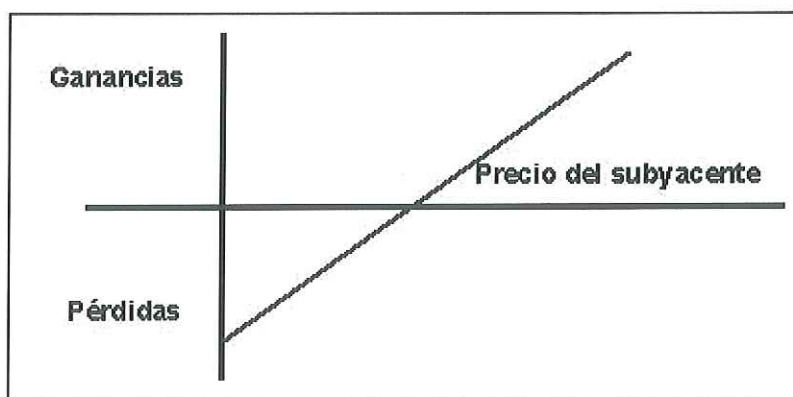


Figura 8. Comportamiento de la posición larga o de compra.

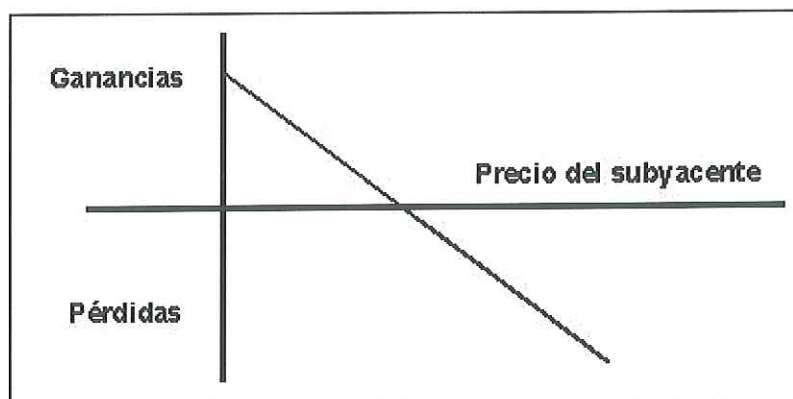


Figura 9. Comportamiento de la posición corta o de venta.

II.6.2 Futuros de Combustóleo y de Gasolina sin Plomo.

Dentro de los energéticos derivados del petróleo, existen dos en particular que llaman la atención, debido a que están muy relacionados el uno con el otro en el mercado de los Estados Unidos de América. El primero es la gasolina sin plomo (unleaded gasoline) y el segundo es el combustóleo (heating oil). Estos productos exhiben un comportamiento correlacionado estacional, ya que en el invierno crece la demanda de combustóleo y baja la demanda de

gasolina y en verano se presenta la situación contraria, baja la demanda de combustóleo y sube la demanda de gasolina. De esta manera, independientemente de la variabilidad de los precios, esta relación inversa ofrece oportunidades para generar ganancias tanto sobre el precio de los productos, como sobre el precio de los futuros.

Los futuros de combustóleo (HO) amparan la transacción de 1,000 barriles (42,000 galones americanos) de este activo real (combustible ligero No. 2 con menos de 0.2% de azufre) y los de gasolina sin plomo (HU) 1,000 barriles de gasolina cuyo contenido de plomo no exceda los 0.03 gramos por galón y un octanaje de 91 puntos o menos. Los dos están basados en la entrega en el puerto de Nueva York [Hume, 1994].

El hecho de que los contratos amparen estas cantidades de producto, hace que las operaciones que se realicen con ellos reflejen el manejo de este gran volumen.

En la figura 10 podemos apreciar el comportamiento de los HO y HU para diciembre de 1996, se observa que la diferencia entre el precio de los dos instrumentos tiende a incrementarse al final del año. Este incremento de la diferencia genera oportunidades de especulación.

Los futuros de estos dos productos se cotizan diariamente en el New York Mercantile Exchange (NYMEX) y existen técnicas de especulación que permiten, de acuerdo a la observación y seguimiento diario de los precios, determinar los momentos oportunos para efectuar transacciones que permitan generar ganancias, o en el peor de los casos, una pérdida mínima preestablecida.

II.6.2.1 Nomenclatura de los futuros HO y HU.

Los futuros de HO y HU se dividen en contratos por mes, es decir, existen futuros HO y HU para enero, febrero, etc., hasta diciembre. Se utiliza una letra para cada mes, tal como se ve en la Tabla I Por ejemplo, el contrato del futuro de combustóleo para enero se expresa como “HOF”, el de combustóleo para diciembre como “HOZ” y el contrato del futuro para gasolina de abril como “HUI”.

Cotizaciones de 1996

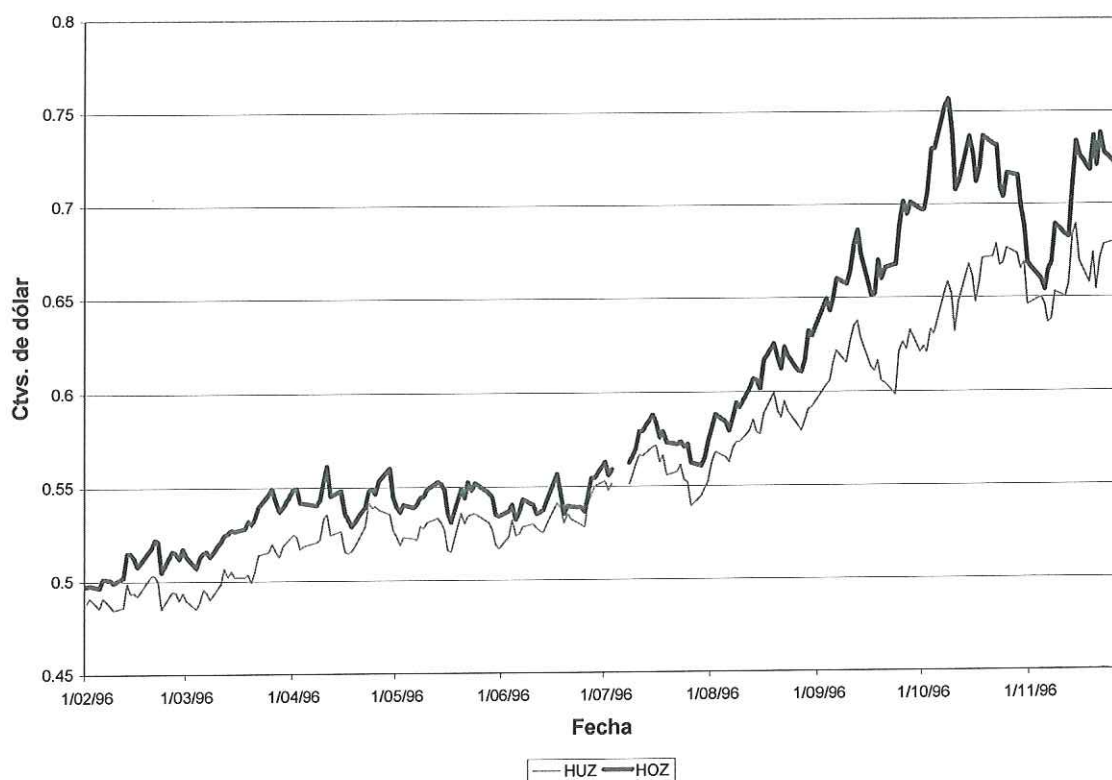


Figura 10. Cotizaciones de HOZ y HUZ de 1996.

Tabla I. Codificación de los meses utilizada por los contratos de futuros.

Clave	Mes
F	Enero
G	Febrero
H	Marzo
J	Abril
K	Mayo
M	Junio
N	Julio
Q	Agosto
U	Septiembre
V	Octubre
X	Noviembre
Z	Diciembre

III. DESARROLLO

III.1 Recolección de Datos.

Las series de datos utilizadas en este trabajo corresponden a los datos de cotizaciones diarias de futuros de combustóleo (HO) y gasolina sin plomo (HU).

Estos datos se obtuvieron de la empresa BRIDGE CRB (<http://www.bridgecb.com>) especializada en venta de datos financieros históricos.

Los conjuntos de datos adquiridos abarcan todas las cotizaciones producidas de los instrumentos hasta el año de 1997, empezando en 1979 para los HO y en 1984 para los HU. En total son 117,000 datos.

III.1.1 Descripción de los conjuntos de datos.

El conjunto HO consta de 252 archivos de datos, cada archivo corresponde a las cotizaciones diarias del futuro para un mes y año en particular, la Tabla II nos presenta un ejemplo de la nomenclatura de los archivos y su interpretación (Para la nomenclatura ver II.6.2.1).

Tabla II. Ejemplos de la nomenclatura para archivos con cotizaciones de futuros HO.

ARCHIVO	DESCRIPCION
Ho1979x	Futuro para noviembre de 1979
Ho1980m	Futuro para junio de 1980
Ho1992f	Futuro para enero de 1992
Ho1998q	Futuro para agosto de 1998
Ho1999z	Futuro para diciembre de 1999
Ho-----Y	Precios reales diarios

La estructura de cada archivo consiste en cuatro campos de información: clave del contrato (CONTRACT), fecha de la cotización (DATE), valor de apertura del mercado (OPEN), valor más alto que alcanzó durante el día (HI), valor más bajo que alcanzó durante el día (LO) y valor al cierre de operaciones (SETTLE). Ver la Tabla III.

Tabla III. Estructura de los archivos de cotizaciones de futuros HO.

CONTRACT	DATE	OPEN	HI	LO	SETTLE
HO1998Q	19970404	5347	5347	5347	5347
HO1998Q	19970407	5348	5348	5348	5348
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

El conjunto de datos HU consta de 177 archivos, ya que este instrumento se empezó a cotizar 5 años después que el HO. Presenta la misma estructura que el HO, la única diferencia es la clave del contrato. Ver Tabla IV.

Tabla IV. Ejemplos de la nomenclatura para archivos con cotizaciones de futuros HU.

ARCHIVO	DESCRIPCION
Hu1985g	Futuro para febrero de 1985
Hu1989v	Futuro para octubre de 1989
Hu1992f	Futuro para enero de 1992
Hu1998q	Futuro para agosto de 1998
Hu1999n	Futuro para diciembre de 1999
Hu----Y	Precios reales diarios

La estructura de cada uno de los archivos de datos del instrumento HU es igual a la descrita para los archivos de datos HO. Ver Tabla V.

Tabla V. Estructura de los archivos de cotizaciones de futuros HU.

CONTRACT	DATE	OPEN	HI	LO	SETTLE
HU1986U	19851127	7300	7500	7275	7500
HU1986U	19851202	7470	7470	7470	7470
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

El valor de la cotización está expresado en centésimas de centavo de dólar estadounidense, esto es, el valor 7470 equivale a 74.70 centavos o a 0.7470 dólares.

III.2 Selección del conjunto de entrenamiento y el conjunto de prueba.

De los datos anteriormente descritos, se utilizaron las cotizaciones de los futuros de diciembre (HOZ y HUZ) en los años de 1996 y 1997, también se utilizaron en algunos experimentos las cotizaciones de los precios reales de los energéticos denominados como HOCash y HUCash de los mismos años.

La razón para seleccionar estos años en particular, obedece a las características presentadas en este período. Mientras los datos correspondientes a 1996 presentan un comportamiento totalmente apegado a la variación estacional (ver II.6.2), los datos de 1997 presentan un comportamiento no esperado (ver figura 11). El análisis del desempeño bajo estas condiciones disímiles brinda mayor certidumbre en la capacidad de adaptabilidad de la RNA.

El análisis de correlación para estas series de datos nos arroja resultados interesantes, los datos de HOZ y HUZ de 1996 tal y como se esperaba muestran un alto grado de dependencia entre sí, más del 95%. Para el año de 1997 la medida de dependencia es baja, 58.5%, lo cual implica que casi el 50% del comportamiento de los datos de este año está en función de factores externos.

Del análisis anterior se puede suponer que el tratamiento combinado de estos datos (HOZ y HUZ) por parte de la RNA puede ser útil para extraer información complementaria que beneficie el proceso de aprendizaje, si bien se prevé, que este beneficio sea mayor para 1996 que para 1997 de acuerdo al valor del coeficiente de correlación. Ver la Tabla VI.

El criterio para seleccionar el conjunto de entrenamiento y prueba en ambos casos está definido por dos razones, primera, contar con un conjunto de entrenamiento con la suficiente información para que la RNA capte la dinámica subyacente de la serie, para lo cual se decidió que el entrenamiento se hiciera con al menos el 80% del total de los datos. La segunda tiene que ver con valores máximos en la serie (locales o globales), indicadores de la variación

estacional al acercarse el invierno. En teoría la diferencia en las cotizaciones de HO y HU tiende a incrementarse, siendo estos momentos propicios para la especulación.

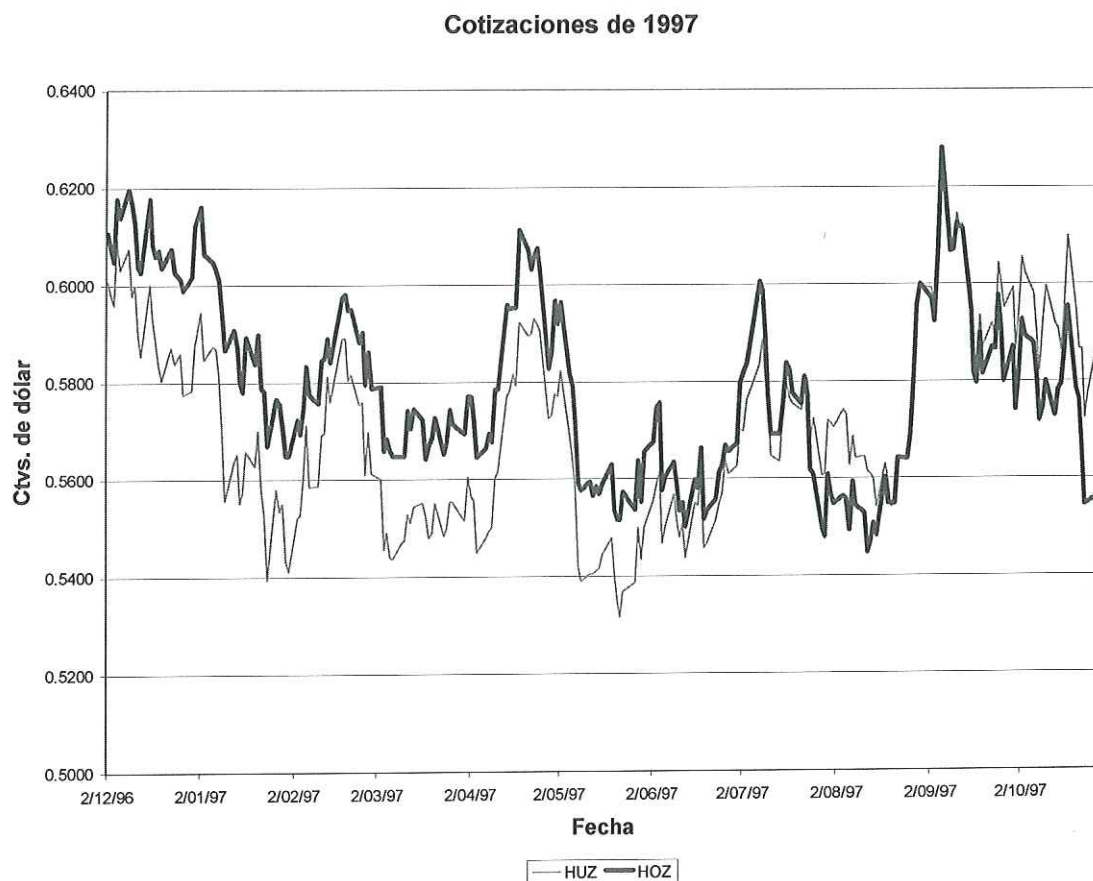


Figura 11. Cotizaciones de HO y HU de 1997.

Tabla VI. Matrices de correlación de las cotizaciones HO y HU para 1996 y 1997.

Datos de 1996				
	HUCash	HOCash	HUZ	HOZ
HUCash	1			
HOCash	0.52152819	1		
HUZ	0.53631787	0.52403515	1	
HOZ	0.49425236	0.61188758	0.97827016	1

Datos de 1996				
	HUCash	HOCash	HUZ	HOZ
HUCash	100.0%			
HOCash	27.2%	100.0%		
HUZ	28.8%	27.5%	100.0%	
HOZ	24.4%	37.4%	95.7%	100.0%

Datos de 1997				
	HUCash	HOCash	HUZ	HOZ
HUCash	1			
HOCash	0.43303973	1		
HUZ	0.36871686	0.52056793	1	
HOZ	0.44204301	0.77249984	0.76494355	1

Datos de 1997				
	HUCash	HOCash	HUZ	HOZ
HUCash	100.0%			
HOCash	18.8%	100.0%		
HUZ	13.6%	27.1%	100.0%	
HOZ	19.5%	59.7%	58.5%	100.0%

La conformación de las series finales utilizadas por la RNA están en función del tipo de prueba realizada y se explican en la sección correspondiente al entrenamiento (ver III.4).

III.3 Escalamiento de los datos.

De acuerdo a lo visto en II.5, los datos tienen que ser transformados para ajustarse al intervalo de valores manejados por la función de activación. Los datos que componen las series comprenden valores en el intervalo de -1 a 1 concentrados alrededor del cero, esto puede afectar el desempeño de la función de activación, por lo que se aplicó un escalamiento lineal a los datos para ampliar el rango sin alterar la distribución de los datos. En la Tabla VII y Tabla VIII se resumen los parámetros estadísticos de los datos.

Tabla VII. Parámetros estadísticos para las series de datos de 1996 de los instrumentos HO y HU y sus diferencias.

Datos de 1996						
Parámetro	HUCash	HOCash	HUZ	HOZ	HOC-HUC	HO-HU
Máximo	0.7464	0.799	0.6935	0.7568	0.1559	0.1012
Mínimo	0.4911	0.506	0.4845	0.4966	-0.1254	0.0041
Media	0.6243134	0.62977608	0.5655555	0.59371483	0.00546168	0.02815933
Desv. Est.	0.0534271	0.07415403	0.05939246	0.07463752	0.06493182	0.02061709

Tabla VIII. Parámetros estadísticos para las series de datos de 1997 de los instrumentos HO y HU y sus diferencias.

Datos de 1997						
Parámetro	HUCash	HOCash	HUZ	HOZ	HOC-HUC	HO-HU
Máximo	0.7639	0.7409	0.6258	0.628	0.0544	0.0312
Mínimo	0.5251	0.4955	0.5315	0.531	-0.2211	-0.0421
Media	0.61204675	0.57020779	0.57115498	0.57868139	-0.04183896	0.00752641
Desv. Est.	0.04972053	0.05331534	0.01973989	0.01885226	0.05494356	0.01325656

III.4 Entrenamiento de la red.

El punto decisivo en el entrenamiento es la definición de la arquitectura de la RNA óptima para el problema en particular, es aquí donde los autores difieren, no existe un consenso o un trabajo definitivo que apoye una arquitectura más que otra tratándose de series de tiempo (ver II.3.3). En estos casos el mejor camino a seguir es la aplicación del modelo conocido como “La navaja de Ockham”, el cual establece: “entre dos explicaciones de cualquier fenómeno natural, siempre ha de elegirse la más simple” [Trippi y Turban, 1996]. Siguiendo este modelo, el entrenamiento está basado en una serie de pruebas que empiezan con la arquitectura más simple, la red de alimentación hacia adelante con retardo temporal y algoritmo de retropropagación (TDNN), incrementando el nivel de complejidad conforme los resultados deseados no cumplan con las expectativas.

Para entrenar la RNA el primer paso es construirla, un primer esfuerzo estuvo centrado en la idea de desarrollar el software (utilizando el lenguaje C) que implementara la funcionalidad requerida. La primera parte de las pruebas se realizó utilizando este enfoque. Conforme el trabajo de investigación fue avanzando, empezó a dar un giro en relación a la importancia de

centrarse en la construcción del software o centrarse en el análisis de los datos. Como resultado de este nuevo enfoque, la segunda parte de la investigación está basada en pruebas que involucren análisis más detallados sobre el comportamiento del entrenamiento. Para facilitar este proceso se utilizó un software comercial (THINKS PRO) que proporciona la funcionalidad adecuada para acelerar el análisis exploratorio de las diferentes alternativas.

III.4.1 Primera etapa de pruebas.

En esta etapa se utilizó una arquitectura TDNN (ver II.3.3.1). Se desarrolló el software de la RNA utilizando el lenguaje C y se diseñó un formato para representar los datos de entrada tales como el conjunto de entrenamiento, el conjunto de prueba; datos de salida como el error, los valores de los pesos y los deltas.

Los parámetros comunes al entrenamiento de 1996 y 1997 fueron la ventana temporal de 5 datos, una arquitectura de 2 capas ocultas de 5 y 2 unidades respectivamente. La función sigmoide como función de activación. El criterio para seleccionar esta arquitectura fue el número de días hábiles de la semana para el tamaño de la ventana y una reducción a la mitad entre los datos de entrada y los de salida de cada capa, esto último basado en la intención de extraer sólo la información necesaria para caracterizar la serie completa y permitir su reconstrucción.

Datos de Entrada: Dos series, la primera formada por las diferencias de los precios HOCash y HUCash ($\text{HOCash} - \text{HOCash}$) y la segunda por las diferencias de los precios HOZ y HUZ ($\text{HOZ} - \text{HUZ}$). En estas pruebas se utilizaron los datos de los precios reales (Cash)

además de los precios de los futuros (HO y HU) esperando que aportaran información adicional a la RNA.

Datos de Salida: Las diferencias de los precios HOZ y HUZ (HOZ – HUZ) al día siguiente.

Los aspectos relevantes del entrenamiento fueron los siguientes:

Entrenamiento para 1996:

El parámetro tomado como criterio para determinar la mejoría de una prueba sobre otra fue el error total del entrenamiento. Aplicando esta medida se observa en la Tabla IX que la prueba I representa el mejor entrenamiento de todos.

Tabla IX. Concentrado de resultados de la primera etapa de pruebas con los datos de 1996.

Prueba	Epocas	Tamaño del conjunto de entrenamiento	Tamaño del conjunto de prueba	Tasa de aprendizaje (α)	Momento (β)	Error
A	250,000	173	38	0.3	0.3	0.014862
B	150,000	172	39	0.25	0.3	0.013087
C	50,000	172	39	0.25	0.7	0.020304
D	100,000	172	39	0.25	0.7	0.016426
E	250,000	172	39	0.25	0.7	0.010289
F	5,000	172	39	0.25	0.75	0.028250
G	50,000	172	39	0.25	0.75	0.018216
H	100,000	172	39	0.25	0.75	0.014597
I	250,000	172	39	0.25	0.75	0.010113

Las pruebas consistieron principalmente en la variación de los parámetros α y β . La heurística para llegar a los parámetros de la prueba I fue resultado de la modificación del

parámetro β y el número de épocas de entrenamiento. Se observó que conforme se aumentan estos valores, el error disminuye.

Se hicieron múltiples pruebas variando parámetros como el tamaño de la ventana, valor de la tasa de aprendizaje, número de datos para entrenamiento.

También se hicieron cambios en el enfoque general de estas pruebas, explorándose los siguientes tipos:

- Predicción al siguiente día pero teniendo como entrada el valor predicho para el día anterior.
- Predicción al día siguiente con un entrenamiento completo por cada dato a predecir.

Ninguno de estos esquemas aportó resultados significativos.

Entrenamiento para 1997:

Como los datos de 1997 exhiben un comportamiento más errático que los de 1996, la cantidad de pruebas fue menor. El menor error se obtuvo utilizando los parámetros que dieron un mejor resultado en los datos de 1996, esto es, $\alpha = 0.25$ y $\beta = 0.75$. Cabe mencionar que el valor de estos parámetros es un valor recomendado por [Swingler, 1996] y que para estas pruebas resultó adecuado.

En la Tabla X se resumen los resultados de las pruebas más significativos realizados para los datos de 1997.

Tabla X. Concentrado de resultados de la primera etapa de experimentos con los datos de 1997.

Prueba	Epocas	Tamaño del conjunto de entrenamiento	Tamaño del conjunto de prueba	Tasa de aprendizaje (α)	Momento (β)	Error
J	100,000	214	13	0.25	0.75	0.018274
K	250,000	214	13	0.25	0.75	0.013491

El análisis visual de ambos conjuntos de pruebas, tanto para los datos de 1996 como los de 1997, revela poca capacidad de predicción por parte de la RNA, sin embargo el análisis más a detalle se presenta en el próximo capítulo.

III.4.2 Segunda etapa de experimentos.

En esta etapa se retoma el análisis del comportamiento de los datos que conforman los vectores de entrada.

Debido a la poca correlación existente entre los datos correspondientes al valor real de los energéticos (Cash) y los correspondientes a los futuros (HOZ y HUZ), se optó por realizar experimentos eliminando los primeros como entrada a la RNA.

Como resultado de este nuevo planteamiento, las series quedaron compuestas de la siguiente forma:

Entrada: Los datos de las cotizaciones HOZ y HUZ solamente.

Salida: Diferencia entre las series para el día siguiente (HOZ-HUZ).

En este conjunto de experimentos se utilizó el software llamado ThinksPro, de la empresa Logical Designs Consulting, Inc.

Este producto brinda la posibilidad de que el usuario seleccione entre diferentes tipos de RNA; además provee despliegue gráfico en tiempo real conforme el entrenamiento se va desarrollando. Además permite el procesamiento por series temporales de hasta 15 datos por vector ejemplo. Estas características facilitan el seguimiento del entrenamiento en cada prueba.

Se utilizó una RNA tipo TDNN con algoritmo de retropropagación, 2 capas ocultas de 5 y 3 unidades, 2 unidades en la capa de entrada y 1 en la de salida. Como función de activación se usó la función sigmoide.

Entrenamiento para 1996:

Tomando como criterio de desempeño el valor del error, el mejor entrenamiento corresponde a la prueba M. En la Tabla XI se presenta el concentrado de resultados de las pruebas más significativos.

Tabla XI. Concentrado de resultados en la segunda etapa de pruebas con los datos de 1996.

Prueba	Epocas	Tamaño del conjunto de entrenamiento	Tamaño del conjunto de prueba	Tasa de aprendizaje (α)	Momento (β)	Error
L	25,000	170	44	0.25	0.75	0.00348
M	250,000	170	44	0.25	0.75	0.003055

Entrenamiento para 1997:

Para los datos de 1997 se utilizaron los mismos parámetros que para 1996, con excepción de la función de activación, para este caso se usó la sigmoideal bipolar, ya que los valores de las diferencias oscilan entre -1 y 1 . Una diferencia con las pruebas J y K, fue el número de elementos del conjunto de prueba, para este caso se utilizaron 52 en lugar de 13. El resultado de la prueba se aprecia en la Tabla XII.

Tabla XII. Concentrado de resultados en la segunda etapa de pruebas con los datos de 1997.

Prueba	Epocas	Tamaño del conjunto de entrenamiento	Tamaño del conjunto de prueba	Tasa de aprendizaje (α)	Momento (β)	Error
N	100,000	182	52	0.25	0.75	0.003235

El análisis preliminar de estos resultados muestra una mejoría notable del valor del error en relación a los valores obtenidos en la primera etapa. Mientras el error más bajo en la primera etapa se ubica en 0.010113 para 1996 y en 0.013491 para 1997, en la segunda etapa los valores están en 0.003055 para 1996 y 0.003235 para 1997. A primera vista estos valores hacen suponer que los datos Cash incorporan ruido al entrenamiento. En la evaluación de los modelos del siguiente capítulo, se estará en posibilidad de concluir si la suposición es válida.

IV. PRESENTACIÓN Y DISCUSIÓN DE RESULTADOS

IV.1 Gráficas de predicción.

De las pruebas presentadas en el capítulo anterior se analizan con mayor detalle aquellos que presentan características destacables en cuanto al valor del error y cuya gráfica de predicción presenta mayor aproximación a los datos reales, esto es, aquellas con el menor valor del error. Basados en este criterio fueron seleccionadas las pruebas A, B, E, I, K, M y N.

El análisis más simple y directo que se hace a los modelos de predicción es el análisis visual de la gráfica de predicción que resulta de la aplicación del modelo propuesto. En este tipo de análisis se aprecian los detalles más obvios de ajuste a los datos reales.

En el análisis de la gráfica de predicción se considera un elemento de juicio importante el hecho de que la tendencia general de la gráfica coincida con la tendencia de la gráfica de los datos reales, es decir, que los cambios de inflexión concuerden lo más posible. Esta concordancia independiente de la exactitud nominal con los datos reales, nos indica que la RNA está aprendiendo la relación subyacente del movimiento de alzas y bajas de las cotizaciones, solo que en una escala diferente.

Es importante hacer notar que los cambios de inflexión desfasados en el tiempo, implican una predicción similar a la caminata al azar, donde la mejor aproximación es asumir que el valor de mañana será igual al de hoy.

Prueba A
Gráfica de predicción

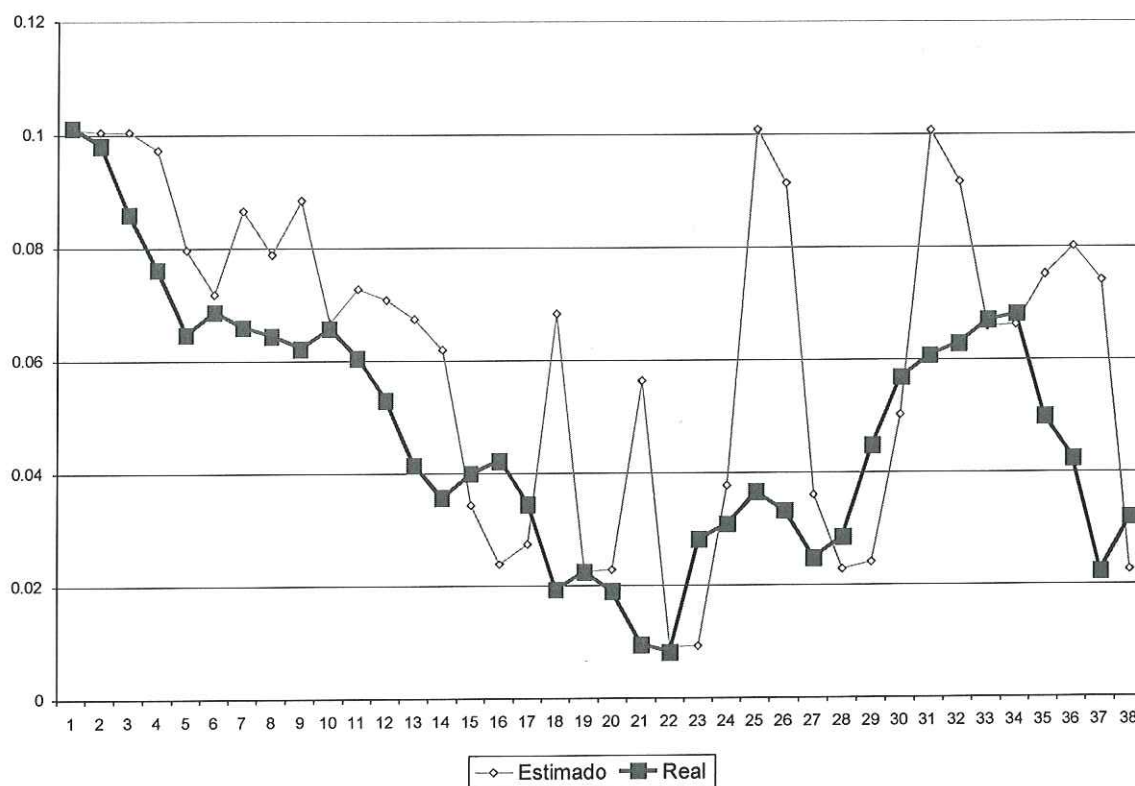


Figura 12. Predicción para la prueba A.

En la prueba A sólo los primeros valores de predicción siguen un patrón similar a los reales (figura 12), pero se aprecia un desfase de un valor de la serie en los cambios de inflexión, esto implica que el resultado de la RNA predice que el valor del día siguiente será similar al del día anterior.

Prueba B Gráfica de predicción

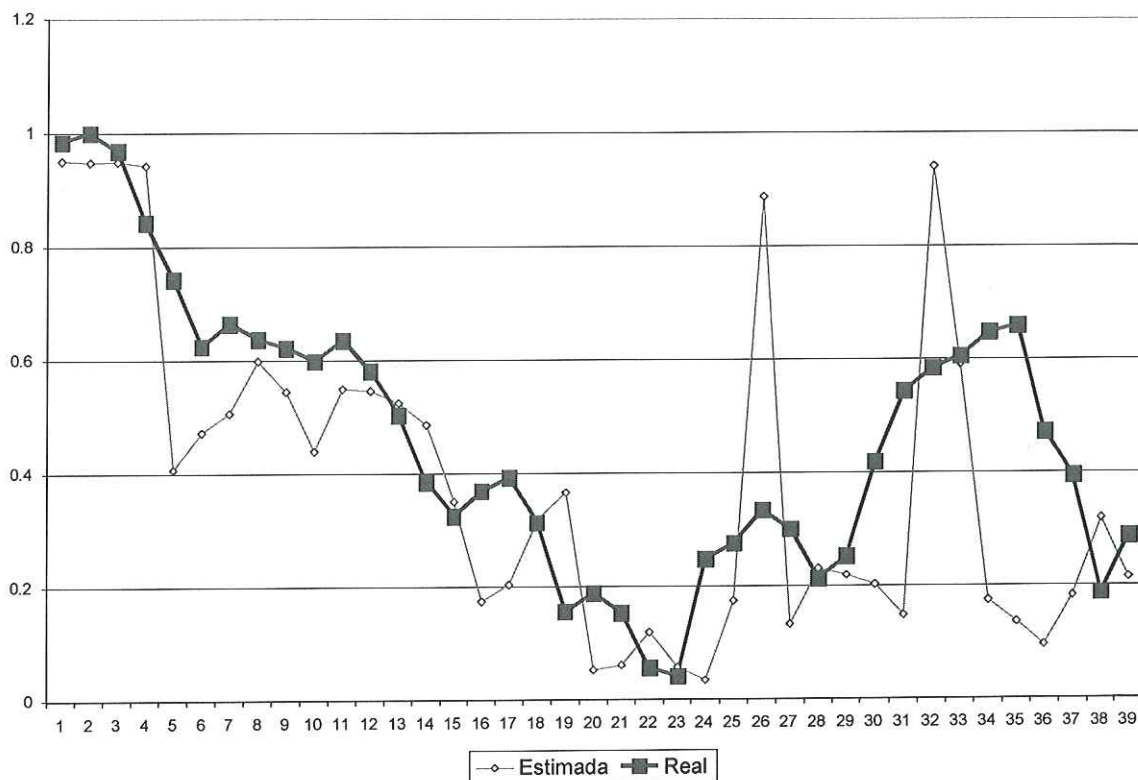


Figura 13. Predicción para la prueba B.

Este mismo comportamiento lo apreciamos en la prueba B (figura 13), con la diferencia que la mejor aproximación se observa alrededor del quinto valor de la serie.

En términos generales las pruebas A y B no se consideran exitosas, ya que la tendencia general de los valores es poco coincidente con los datos reales.

Prueba E Gráfica de predicción

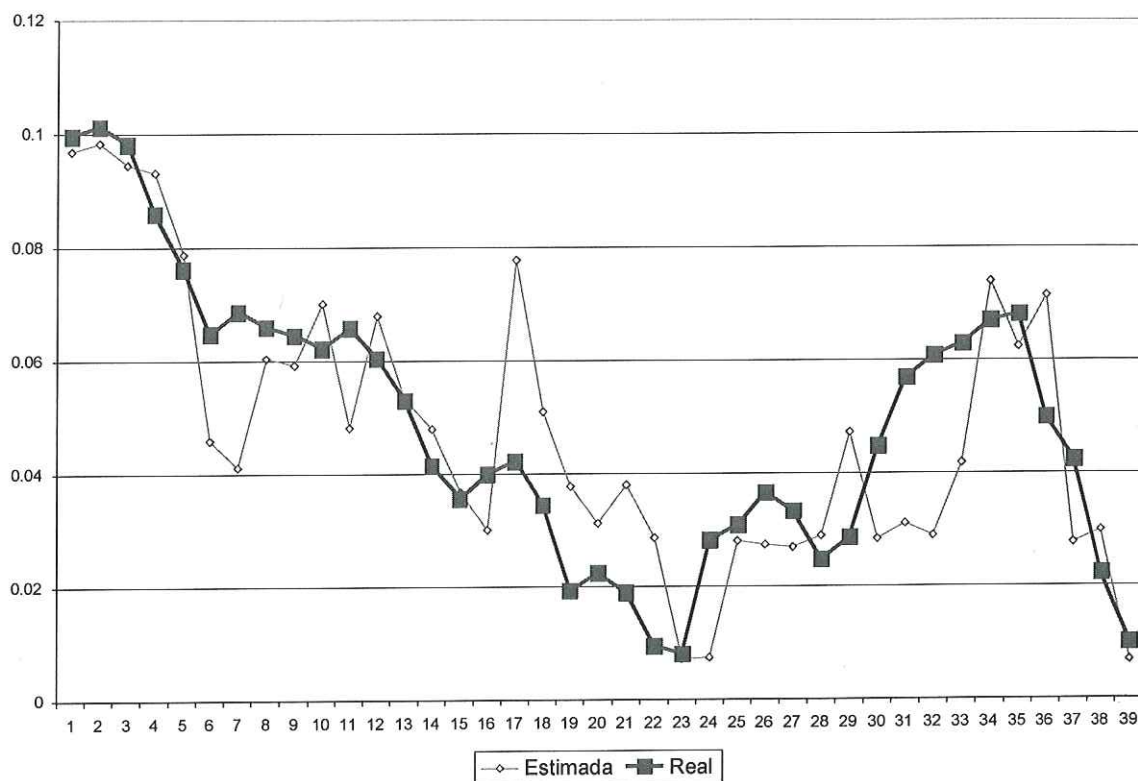


Figura 14. Predicción para la prueba E.

La prueba E (figura 14) muestra una muy buena aproximación para los primeros seis valores, se observa una tendencia similar a la serie real, aunque las características del resto de la gráfica son similares a las pruebas presentadas anteriormente.

Los primeros tres valores muestran un comportamiento igual al de la serie real, esto la califica como una de las mejores aproximaciones bajo la perspectiva de el análisis visual.

Prueba I Gráfica de predicción

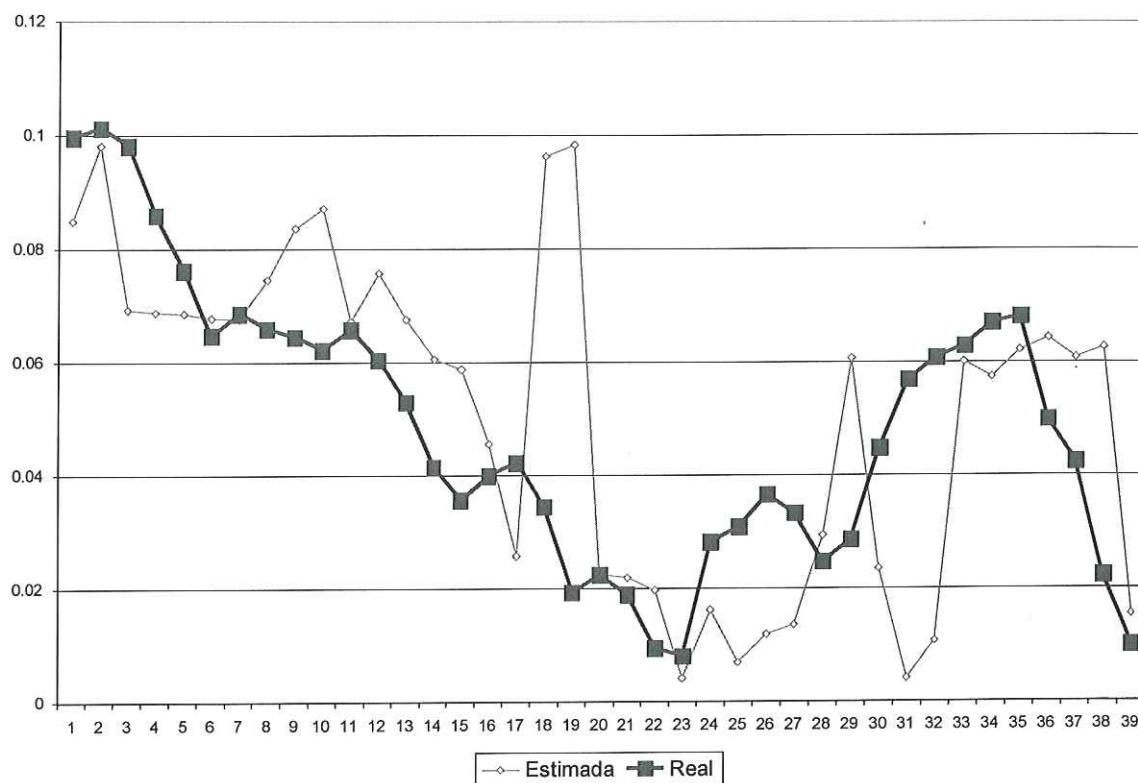


Figura 15. Predicción para la prueba I.

La prueba I (figura 15) muestra una comportamiento similar a la prueba E para los tres primeros datos, pero con valores más disparados. El resto de los valores se comportan de manera similar a los obtenidos en las pruebas A y B.

Prueba K Gráfica de predicción

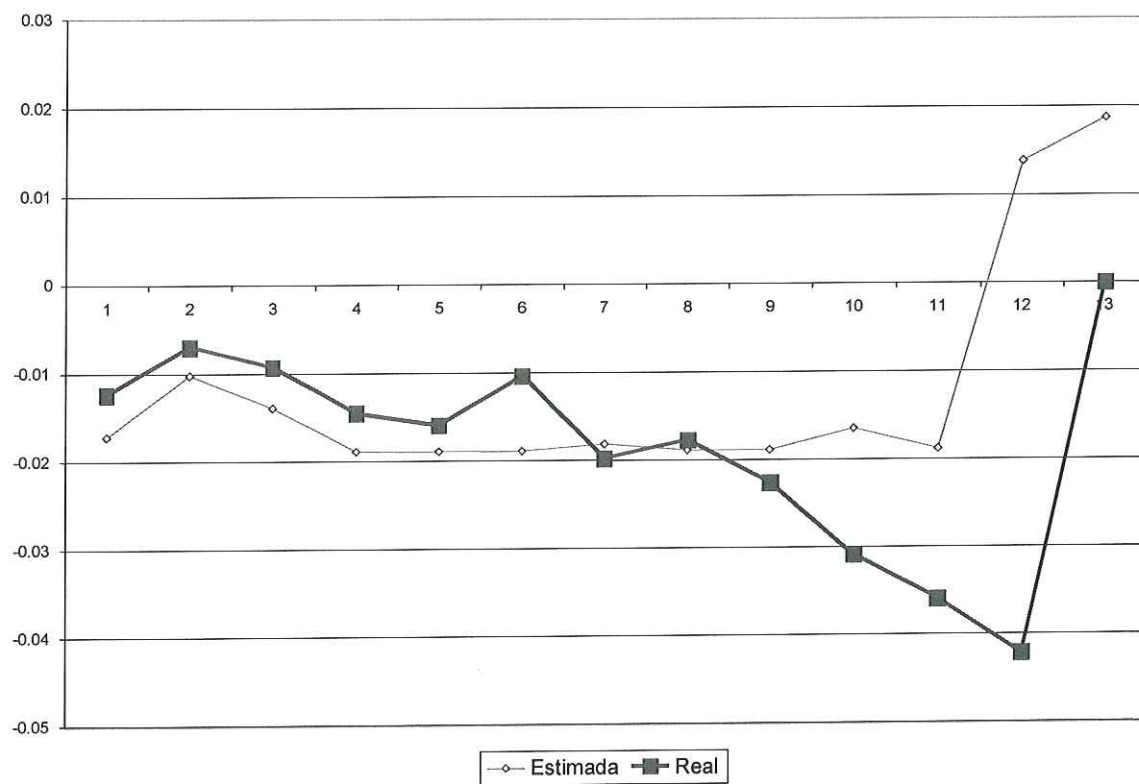


Figura 16. Predicción para la prueba K.

En la gráfica de la prueba K (figura 16) destaca la aproximación de los primeros cuatro datos, los cuales presentan el mismo comportamiento que los reales. Esta característica la hace la mejor prueba para los datos de 1997.

Prueba M
Gráfica de predicción

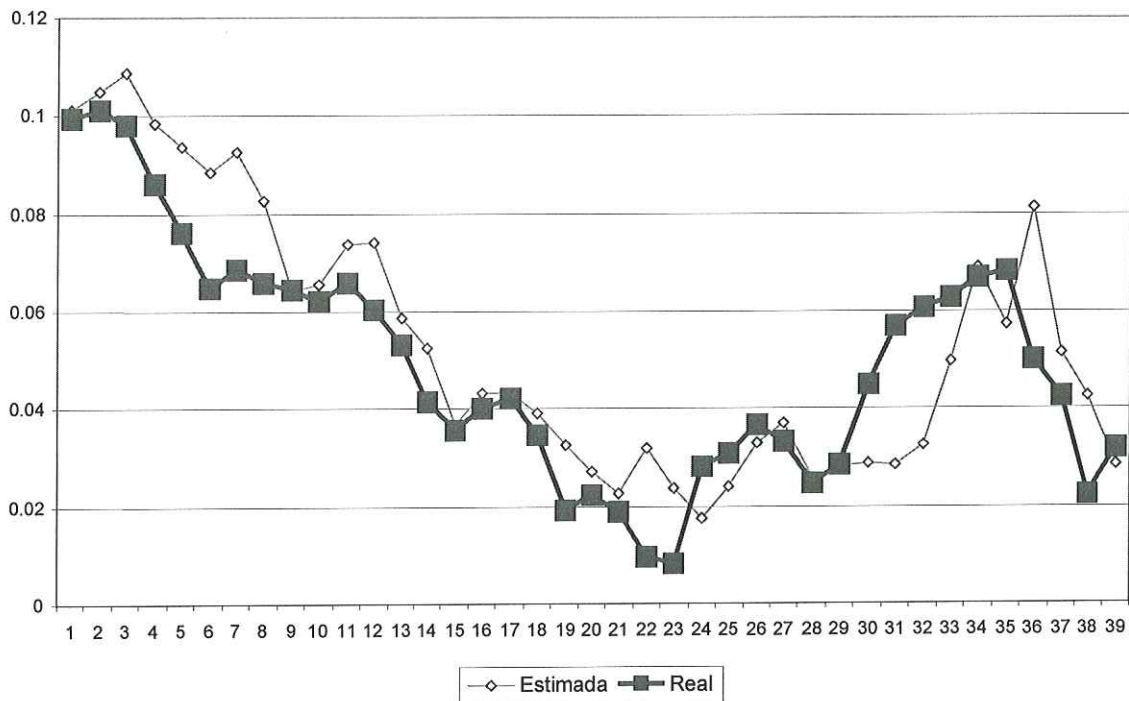


Figura 17. Predicción para la prueba M.

La prueba M (figura 17) muestra una mayor aproximación nominal de los datos, y además una mayor número de coincidencias en los cambios de inflexión. Estas características le asignan una alta calificación al modelo propuesto para esta prueba.

Prueba N Gráfica de predicción

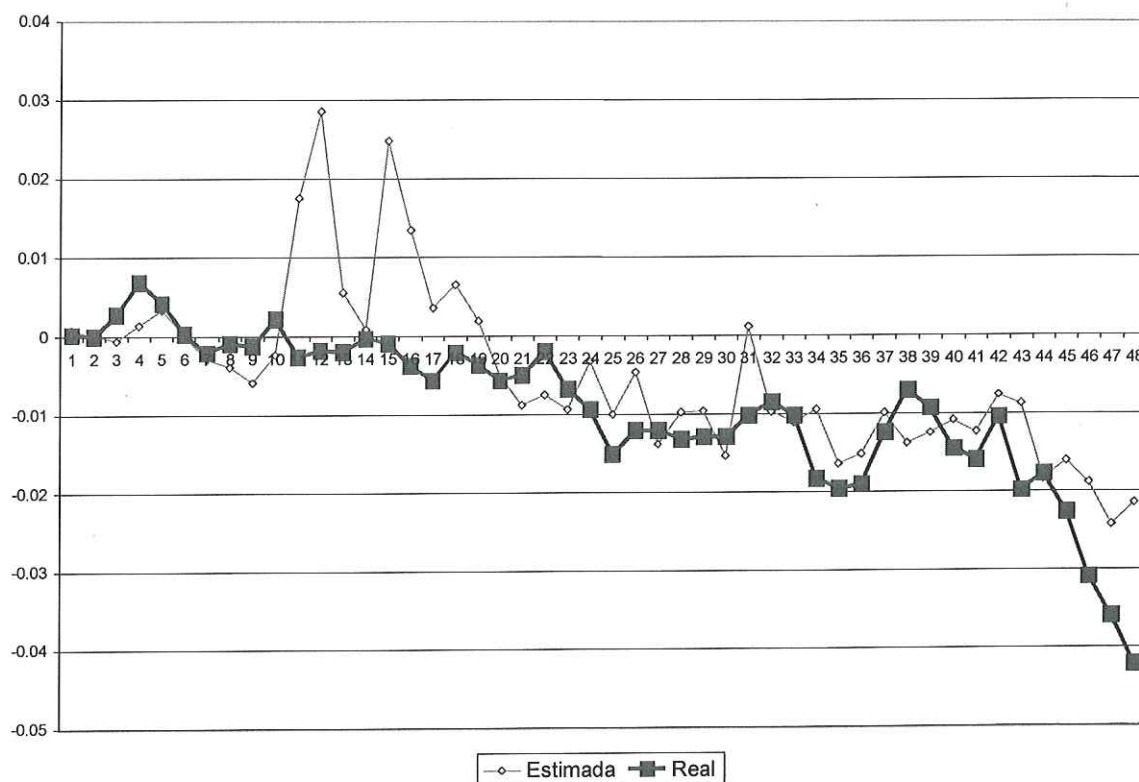


Figura 18. Predicción para la prueba N.

Los datos arrojados por la prueba I (figura 15) son muy similares a los datos reales, sin embargo en la gráfica se aprecia un desfase en los cambios de inflexión, y esta característica, como se discutió al inicio de la sección, hace suponer que el modelo no es bueno.

En resumen, la revisión visual de las gráficas de predicción permite ubicar a las pruebas E (figura 14) y M (figura 17) como los mejores entrenamientos para los datos de 1996 y a la prueba K (figura 16) para 1997.

En particular la prueba M ofrece una buena predicción en los cambios de dirección de las cotizaciones, esto es, si el valor se incrementa, decrementa o permanece igual al de la cotización del día anterior.

IV.2 Análisis del desempeño de la RNA.

Los puntos a considerar en esta etapa son los que marca la metodología propuesta (ver II.5): elaboración de histogramas de error, gráficas de dispersión y la comparación con el modelo de caminata al azar. Estos tres análisis se aplican a las pruebas seleccionadas.

IV.2.1 Resultados del análisis con histogramas y diagramas de dispersión.

En los histogramas de frecuencias de los errores para la prueba A (ver figura 19), la prueba B (ver figura 21), la prueba E (ver figura 23) y la prueba I (ver figura 25), se observan que las frecuencias del tamaño del error de entrenamiento pueden ser aproximadas a una distribución normal, presentando un promedio del tamaño del error de entrenamiento cercano a cero. Debido a esto, se podría decir que la RNA está bien entrenada. Esto es un indicativo de un grado de confiabilidad aceptable para los entrenamientos de acuerdo a lo presentado en II.5.

En caso contrario, en los histogramas de frecuencias de los errores para la prueba K (ver figura 27), la prueba M (ver figura 29) y la prueba N (ver figura 31), las frecuencias del tamaño del error de entrenamiento no podrían ser aproximadas a una distribución normal. Esto significa que el entrenamiento de la red no fue exitoso.

Los diagramas de dispersión de la prueba A (ver figura 20), la prueba B (ver figura 22), la prueba E (ver figura 24) y la prueba I (ver figura 26), muestran mejores resultados para el año de 1996. Esto se debe, a que los valores reales y los valores estimados por la red se concentran en forma de línea recta, lo cual indica un buen desempeño de la RNA.

Cabe hacer notar, que en el diagrama de dispersión de la prueba M (ver figura 30) para el año de 1996, y en los diagramas de dispersión de la prueba K (ver figura 28) y la prueba N (ver figura 32) para el año de 1997, los valores reales y los valores estimados por la RNA no se concentran en forma de una línea recta. Esto implica que a la RNA se le dificultó más el aprendizaje sobre los datos del año de 1997.

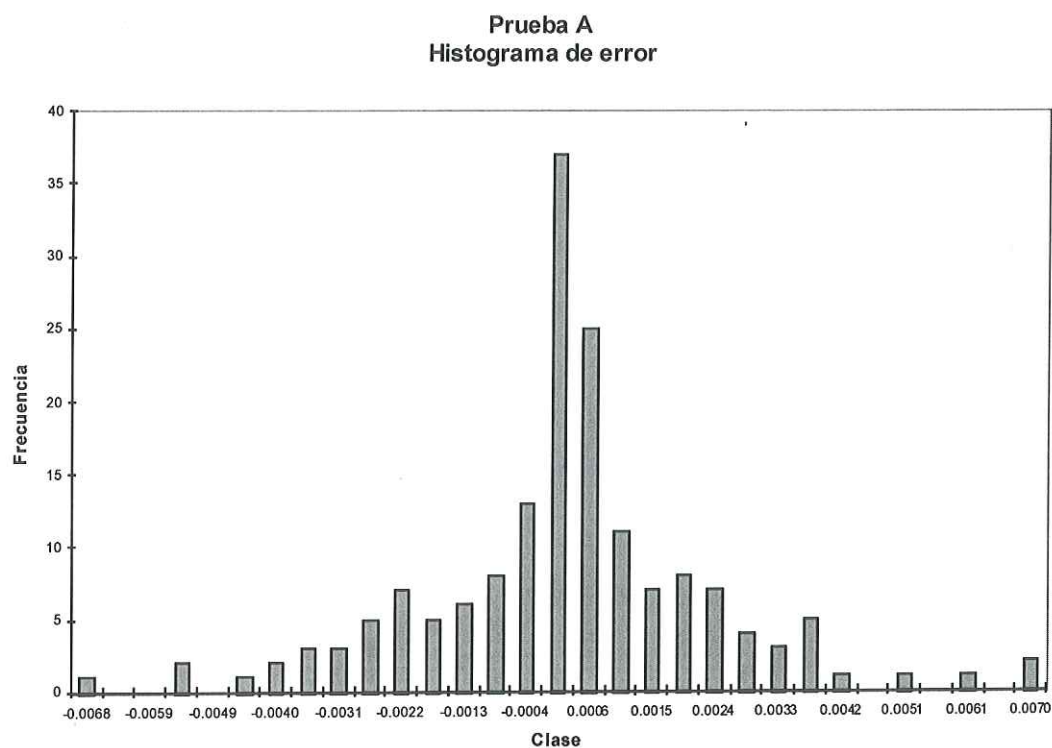


Figura 19. Histograma de frecuencias de los errores para la prueba A.

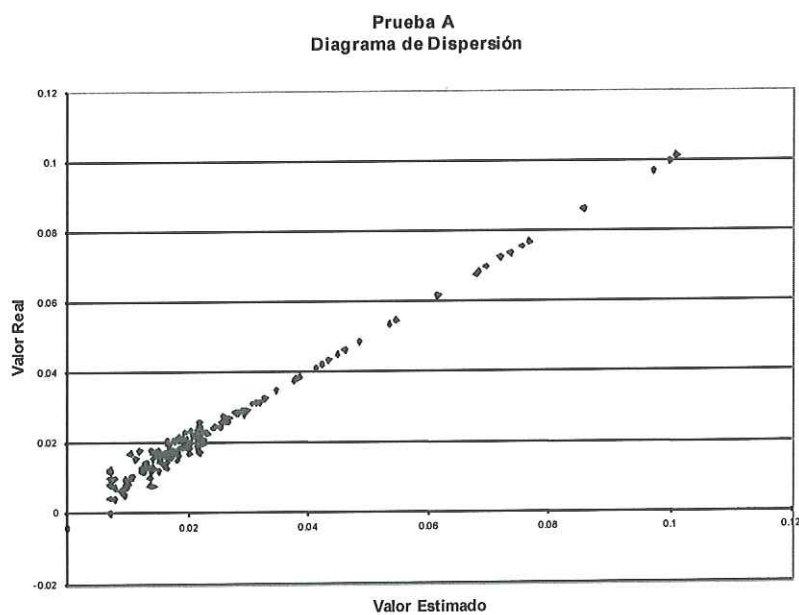


Figura 20. Diagrama de dispersión para la predicción de la prueba A.

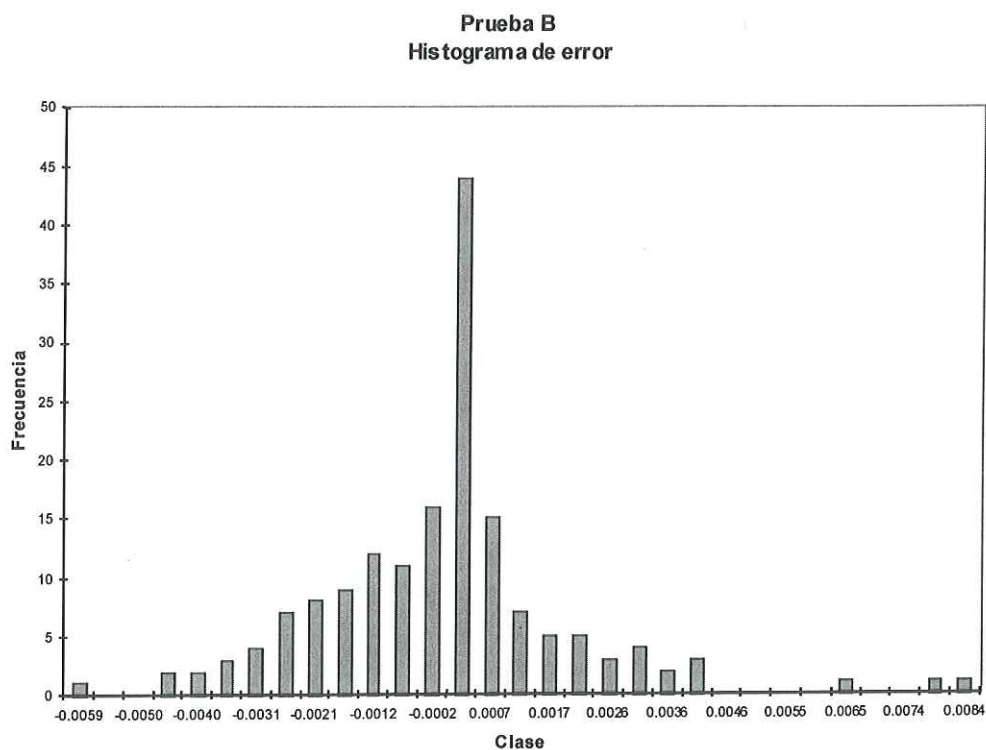


Figura 21. Histograma de frecuencias de los errores para la prueba B.

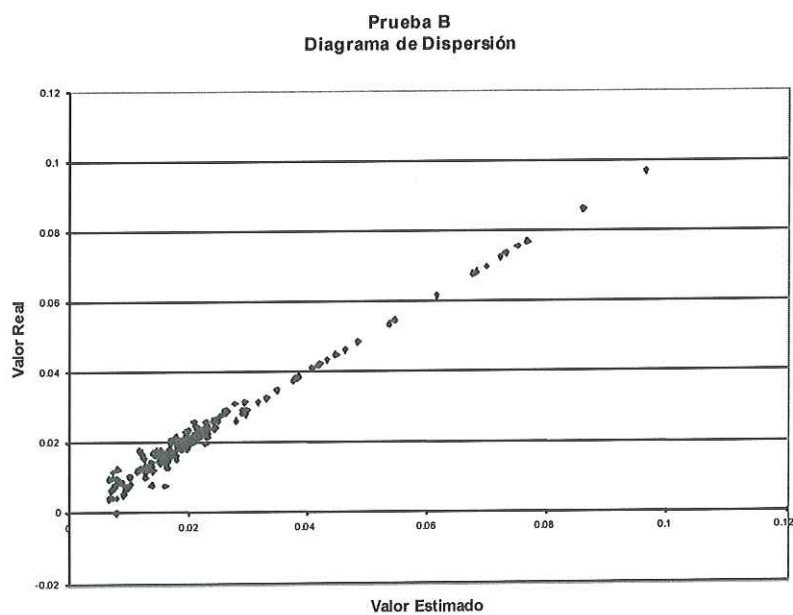


Figura 22. Diagrama de dispersión para la predicción de la prueba B.

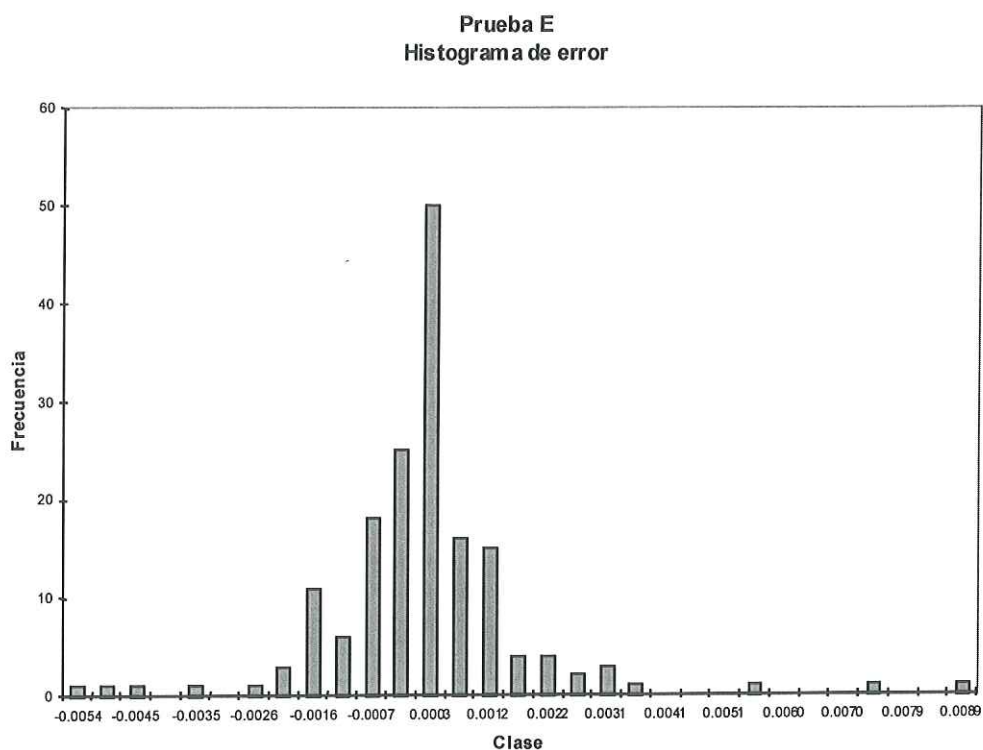


Figura 23. Histograma de frecuencias de los errores para la prueba E.

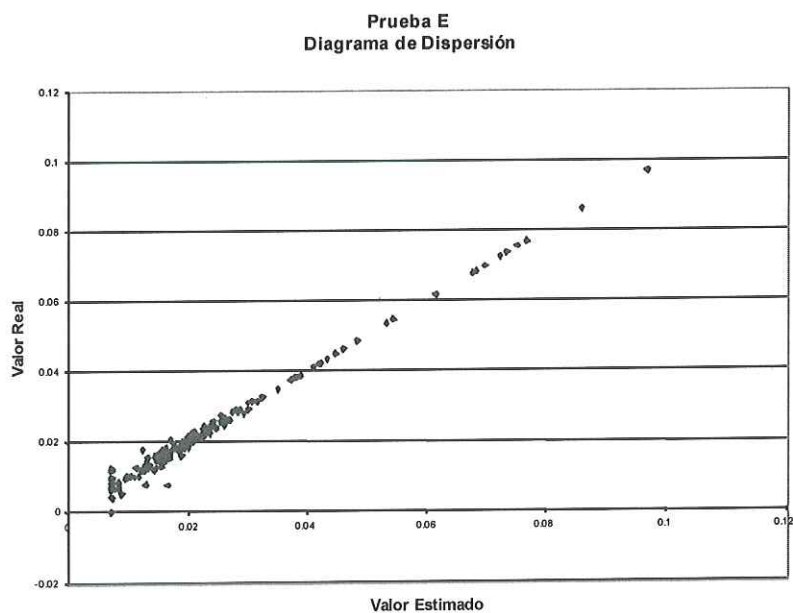


Figura 24. Diagrama de dispersión para la predicción de la prueba E.

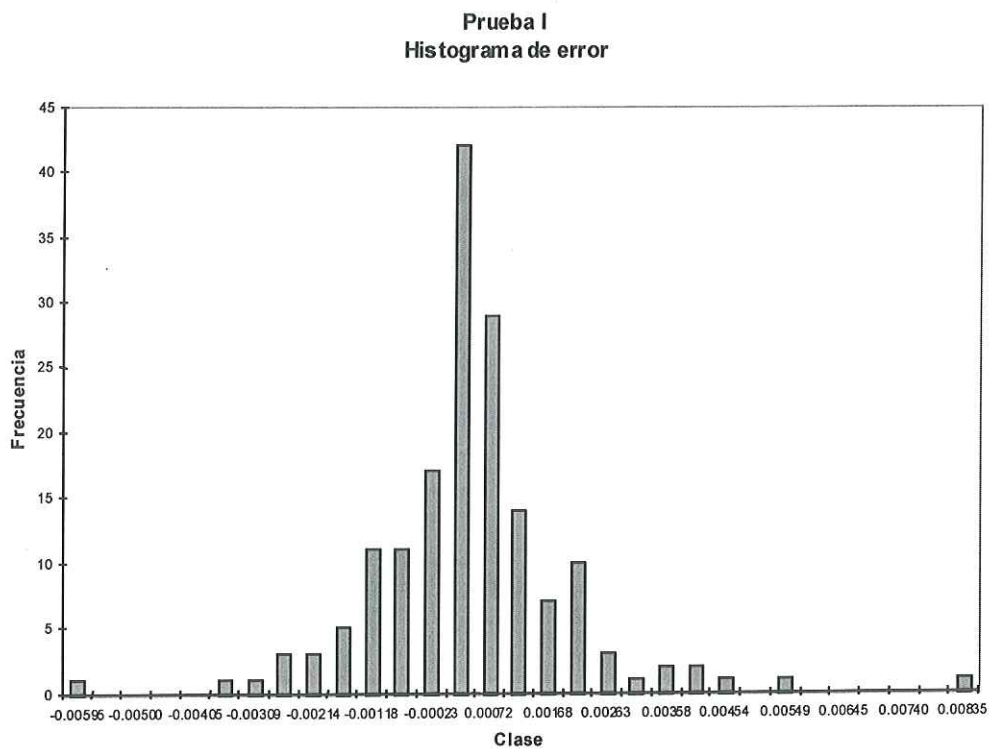


Figura 25. Histograma de frecuencias de los errores para la prueba I.

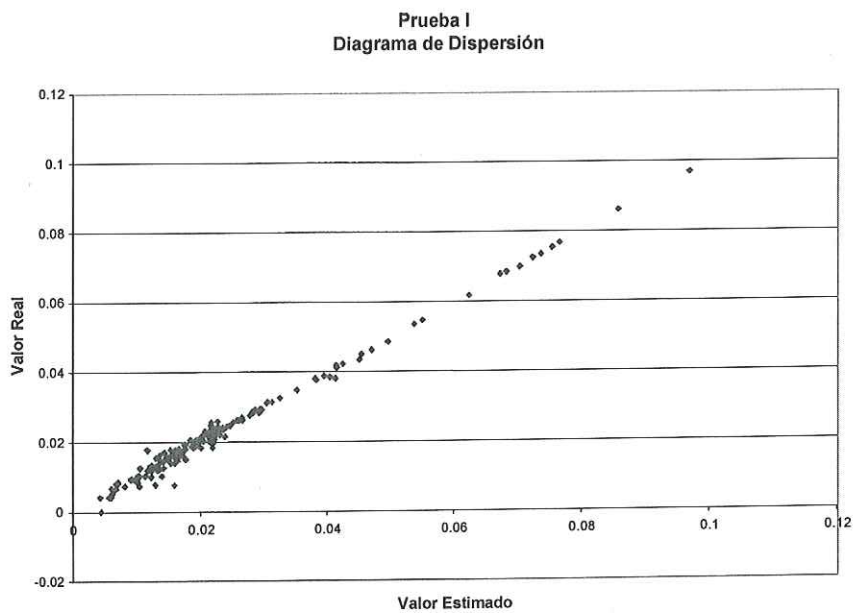


Figura 26. Diagrama de dispersión para la predicción de la prueba I.

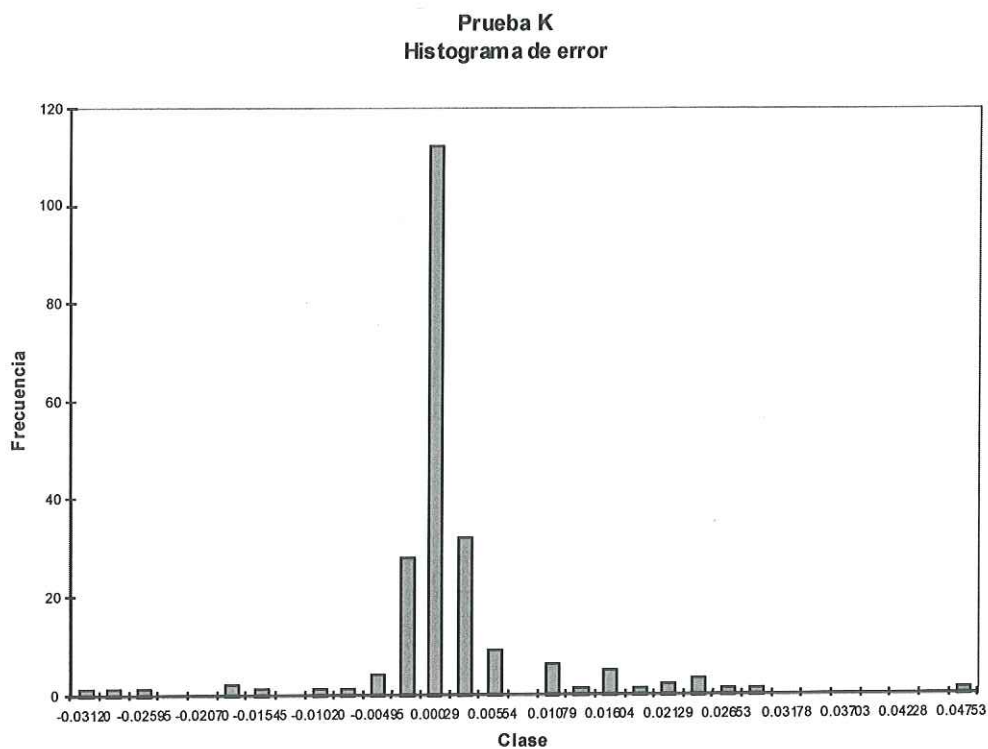


Figura 27. Histograma de frecuencias de los errores para la prueba K.

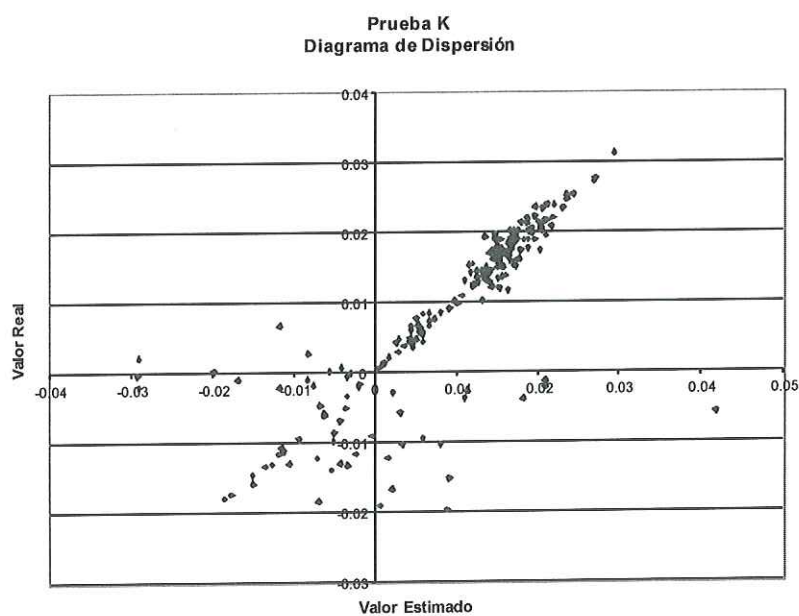


Figura 28. Diagrama de dispersión para la predicción de la prueba K.

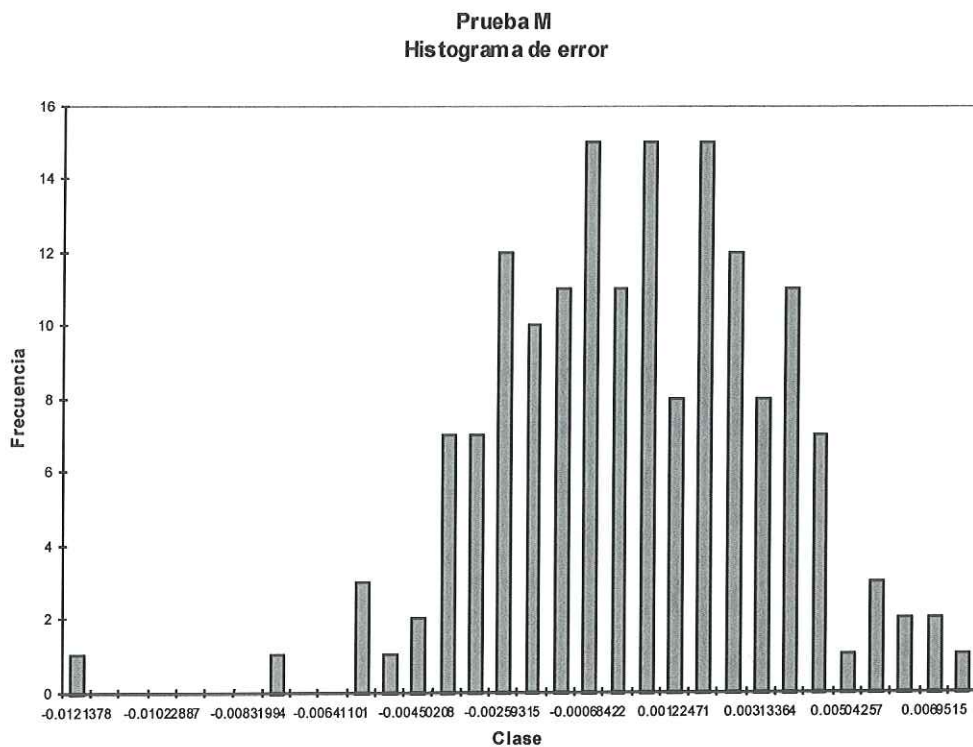


Figura 29. Histograma de frecuencias de los errores para la prueba M.

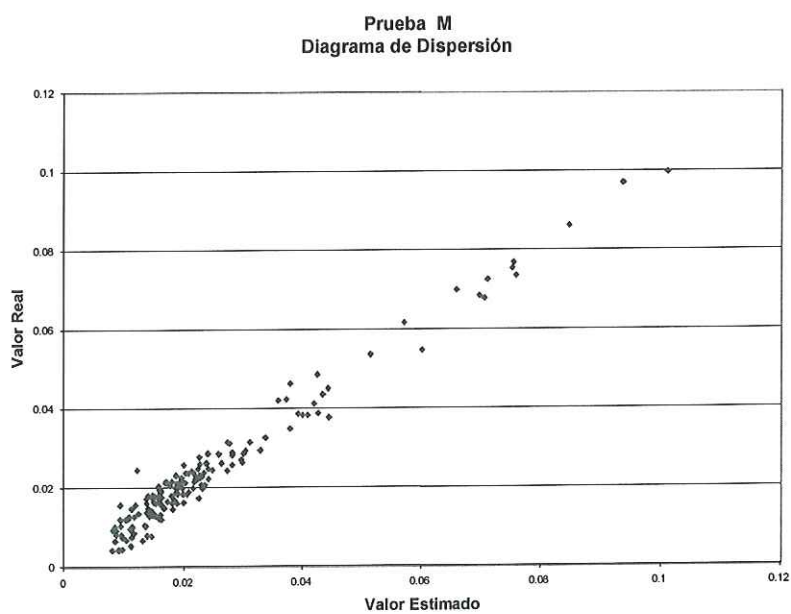


Figura 30. Diagrama de dispersión para la predicción de la prueba M.

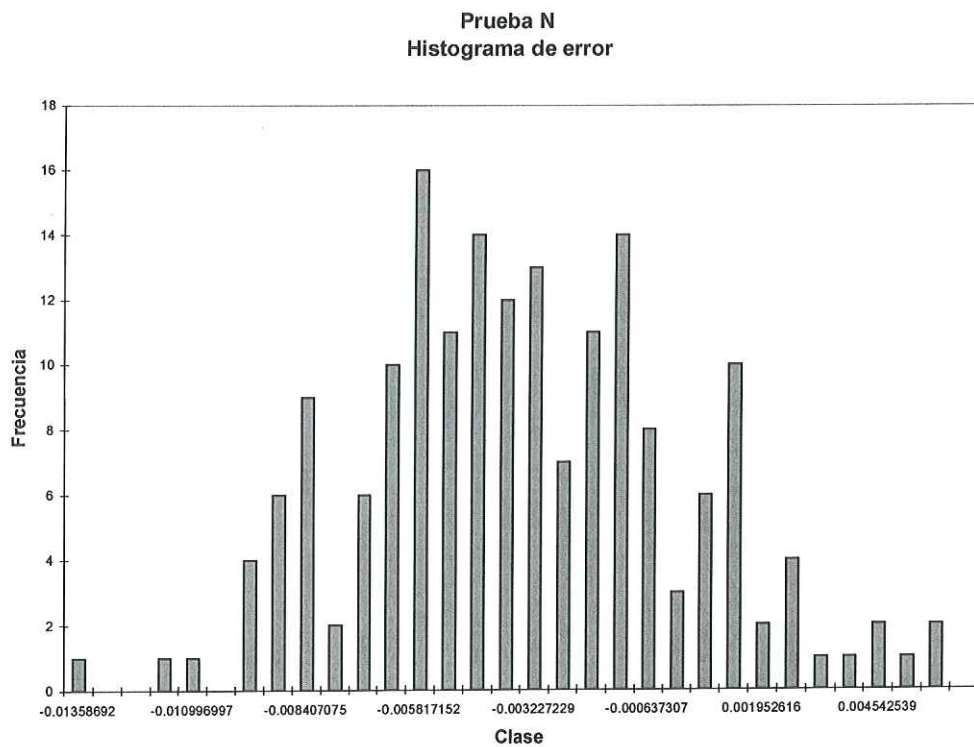


Figura 31. Histograma de frecuencias de los errores para la prueba N.

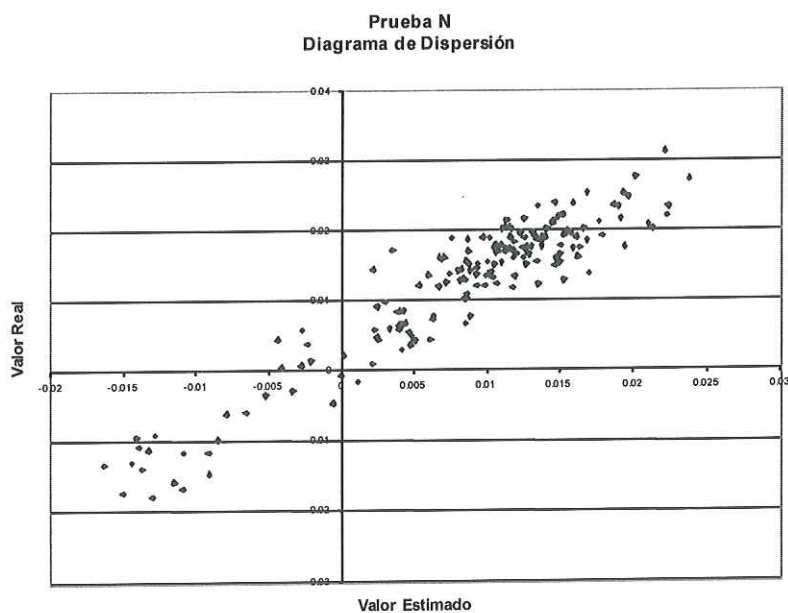


Figura 32. Diagrama de dispersión para la predicción de la prueba N.

IV.2.2 Resultados del análisis utilizando el coeficiente de Theil.

La sección anterior se enfocó en el análisis visual de los datos generados durante el entrenamiento. En esta sección se lleva a cabo un análisis utilizando el cálculo del coeficiente de Theil, para determinar el comportamiento del modelo desarrollado con la RNA, en comparación con el modelo de caminata al azar (Random Walk Model) tal como se expuso en la sección II.5. Mediante éste análisis se podrán descubrir aspectos difícilmente apreciables por la simple observación gráfica.

Los coeficientes de Theil obtenidos para cada una de las pruebas aquí realizadas se muestran en la Tabla XIII.

Tabla XIII. Concentrado de resultados del cálculo del coeficiente de Theil por cada prueba.

Prueba	Coeficiente de Theil
A	2.88766193
B	2.36630682
E	1.6884839
I	2.880034
K	1.400391
M	1.55537267
N	2.4479022

Con respecto a estos resultados se observa que, de acuerdo con este estadístico, ninguna de las pruebas es mejor que el modelo de caminata al azar. Todos los resultados se situaron por arriba del valor unitario. Esto implica, que el modelo de RNA con los parámetros seleccionados no se desempeña mejor que el utilizar solo el valor del día anterior para predecir el del día siguiente.

Las pruebas que muestran un mejor desempeño son la E, la K y la M (ver Tabla XIII). Estos resultados rechazan la suposición de que incluir los precios reales de los energéticos (Cash) aportan ruido al modelo, ya que dos de las pruebas con mejor desempeño son precisamente aquellas que incluyen estos datos en las series de entrada. Por lo tanto, los datos Cash sí aportan información adicional que la RNA utiliza para mejorar su aprendizaje.

En resumen, de acuerdo a los diferentes tipos de análisis aplicados a las pruebas, se está en posibilidad de asegurar que de las pruebas aquí documentadas, los mejores resultados se obtuvieron con las siguientes:

- *Prueba E para 1996.*
- *Prueba K para 1997.*

V. CONCLUSIONES Y RECOMENDACIONES

El análisis de series de datos es un área muy extensa y complicada, la cual ha contribuido a la exploración de nuevas herramientas para su tratamiento, tal es el caso de las redes neuronales artificiales. En este trabajo se exploraron alternativas de implementación y configuración de redes neuronales artificiales para la predicción de series de tiempo.

Las pruebas aquí documentadas se basaron en una arquitectura tipo TDNN con dos capas ocultas de 5 y 2 unidades respectivamente y una ventana de retardo temporal de 5 datos. Los valores para la tasa de aprendizaje (α), el momento (β) y el número de épocas se variaron para cada prueba. Los resultados obtenidos de acuerdo al coeficiente de Theil, indican que los modelos desarrollados no exhiben un comportamiento mejor al que se hubiera obtenido con un modelo de caminata al azar.

El objetivo general de este trabajo, fue crear un sistema capaz de predecir las fluctuaciones de futuros financieros, el cual no se alcanzó con el grado de confiabilidad deseada. Este resultado puede tener su origen en las siguientes afirmaciones:

- a) Los parámetros de entrenamiento de la RNA no son los adecuados.
- b) La serie de datos es no estacionaria, su comportamiento depende de múltiples factores, los cuales no están totalmente contenidos en la propia serie, lo cual implica que la RNA es incapaz de extraer la información suficiente.

Si bien los entrenamientos de las prueba E y K resultaron ser los mejores, de acuerdo a los análisis aplicados, no existe un índice de confiabilidad que avale la utilización de los valores de la configuración de la RNA para aplicarlo a datos fuera del intervalo utilizado en la prueba.

Sin embargo, el camino recorrido a lo largo de este trabajo de investigación, arroja conclusiones sobre el proceso mismo:

- a) La parte fundamental del desarrollo de la RNA se basa en la adecuada selección de la arquitectura misma. No existe una metodología probada que se aplique a la construcción de una RNA para el manejo de series de tiempo. La mayor parte del trabajo se tiene que hacer con base en el ensayo y error.
- b) El análisis previo de la información que recibirá de entrada la RNA, es crucial para la selección de la arquitectura y los parámetros de la misma. El tratamiento preliminar de los datos deberá estar apoyado con herramientas estadísticas que nos presenten información relevante acerca del comportamiento de los datos, de tal manera que apoye la selección de los parámetros adecuados de la RNA.

En general se concluye que el trabajo con RNA para series de tiempo es difícil por la gran cantidad de parámetros no determinísticos que intervienen en el desarrollo del modelo. Además, si las series presentan características que las clasifiquen como no estacionarias, el aprendizaje de la RNA se complica.

Adicionalmente, tal como se expuso en la sección II.4.3, la hipótesis de los mercados eficientes ofrece una explicación a la poca capacidad de predicción de cualquier método aplicado a las series de datos financieros.

Se recomienda como continuación de este trabajo, la exploración de los modelos de RNA con arquitectura recurrente, los cuales no fueron probados en esta investigación. También un análisis matemático de los datos para comprobar si la serie es no estacionaria. La construcción de un modelo de promedios móviles autorregresivos integrados (ARIMA) para compararlo con el desempeño de la RNA.

Por último, la construcción de un paquete de software que permita la creación de modelos de RNA de una manera sencilla y con capacidad de importación y exportación de datos en diferentes formatos para una fácil integración con paquetes de graficado y análisis estadístico.

BIBLIOGRAFÍA

- Anderson, T. 1971. "The statistical analysis of time series". John Wiley & Sons.
- Beach, W., Schavey, A., Isidro, I. 1999. "How Reliable are IMF Economic Forecasts?". The Heritage Foundation. Washington, D.C.
<http://www.heritage.org>
- Box, G.E., Jenkins, G.M. 1976. "Time series analysis: forecasting and control".
- Chou, Y. 1990. "Análisis Estadístico". McGraw-Hill, Segunda edición.
- Díaz Tinoco, J. 1996. "Futuros y opciones financieras: una introducción". Limusa-BMV
- Gershenfeld, N., Weigend, A. 1993. "The future of time series: Learning and understanding". En: Weigend, A., Gershenfeld, N. Editores. 1994. "Time Series Prediction: Forecasting the future and understanding the past". Addison-Wesley, Reading MA. 1-70 p.
- Glass, L., Kaplan, D. 1993. "Complex dynamics in physiology and medicine". En: Weigend, A., Gershenfeld, N. Editores. 1994. "Time Series Prediction: Forecasting the future and understanding the past". Addison-Wesley, Reading MA. 513-528 p.
- Hertz, J., Krogh, A., Palmer, R. 1991. "Introduction to the theory of neural computation". Addison-Wesley. Primera edición. Redwood City. 327 pp.
- Hume. 1994. "The Petro Parlay" serie The Superinvestor Files. Hume & Associates
- Hutchinson, J. 1994. "A Radial Basis Function Approach to Financial Time Series Analysis". Tesis de Doctorado. Department of Electrical Engineering and Computer Science. Massachusetts Institute of Technology.
- Kartalopoulos, S. 1996. "Understanding Neural Networks and Fuzzy Logic: basic concepts and applications". IEEE Press, ISBN 0-7803-1128-0.
- Kimoto, T., Asakawa, K., Yoida, M. y Takeoka, M. 1990. "Stock market prediction system with modular neural networks". En: Trippi, R., Turban, E. editores. 1996. "Neural Networks in Finance and Investing: using artificial intelligence to improve real-world performance". IRWIN - Professional Publishing, Primera edición. Chicago. 497-510 p.

- Klimasauskas, C. 1991. "Applying Neural Networks". En: : Trippi, R., Turban, E. editores. 1996. "Neural Networks in Finance and Investing: using artificial intelligence to improve real-world performance". IRWIN - Professional Publishing, Primera edición. Chicago.45-69 p.
- McClelland, J.L., Elman, J.L. 1986. "Interactive Processes in Speech Perception: The TRACE Model". En: Parallel Distributed Processing, vol. 2, cap.15
- Mendelsohn, L. 1991. "The Basics of Developing a Neural Trading Systems". En: Technical Analysis of Stocks&Commodities Jun. 1991
- Mendelsohn, L. 1993. "Neural Network Development for Financial Forecasting". En: Technical Analysis of Stocks&Commodities Sep. 1993.
- Mozer, M. 1993. "Neural Net Architectures for Temporal Sequence Processing". En: Weigend, A., Gershenfeld, N. Editores. 1994. "Time Series Prediction: Forecasting the future and understanding the past". Addison-Wesley, Reading MA. 243-264 p.
- Sejnowski, T.J., Rosenberg, C.R. 1986. "Parallel Networks that Learn to Pronounce English Text". En: Complex Systems 1, 1986.
- Skapura, D. 1996. "Building Neural Networks". Addison-Wesley, ACM Press. Primera edición. New York. 286 pp. ISBN 0-201-53921-7.
- Stephens, C. R. 1996. "Mercados Financieros Adaptivos". Instituto de Ciencias Nucleares, UNAM <http://www.nuclecu.unam.mx/nncp>
- Swingler, K. 1996. "Applying Neural Networks: A Practical Guide". Academic Press Inc., Primera edición. San Diego, CA. 303 pp. ISBN 0-12-679170-8
- Thelie, J., Linsay, P., Rubin, D. 1993. "Detecting Nonlinearity in Data with Long Coherence Times". En: Weigend, A., Gershenfeld, N. Editores. 1994. "Time Series Prediction: Forecasting the future and understanding the past". Addison-Wesley, Reading MA. 429-455 p.
- Trippi, R., Turban, E. editores. 1996. "Neural Networks in Finance and Investing: using artificial intelligence to improve real-world performance". IRWIN - Professional Publishing, Primera edición. Chicago. 821 pp. ISBN 1-55738-919-5
- Weigend, A., Gershenfeld, N. Editores. 1994. "Time Series Prediction: Forecasting the future and understanding the past". Addison-Wesley, Reading MA. 641 pp.

- Waibel, A. 1989. "Modular Construction of Time Delay Neural Networks for Speech Recognition". En: Neural Computation, Vol. 1, No. 1. 1989.
- White, H. 1988. "Economic prediction using Neural Networks: The case of IBM daily stocks returns". En: Trippi, R., Turban, E. editores. 1996. "Neural Networks in Finance and Investing: using artificial intelligence to improve real-world performance". IRWIN - Professional Publishing, Primera edición. Chicago. 469-481 p.
- Zhang, X., Hutchinson, J. 1993. "Simple Architectures on Fast Machines: Practical Issues in Nonlinear Time Series Prediction". En: Weigend, A., Gershenfeld, N. Editores. 1994. "Time Series Prediction: Forecasting the future and understanding the past". Addison-Wesley, Reading MA. 219-241 p.

