



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FISICAS Y MATEMATICAS
DEPARTAMENTO DE INGENIERIA INDUSTRIAL**

**PROYECCIÓN DEL PRECIO DEL COBRE:
¿HERRAMIENTAS DE INTELIGENCIA COMPUTACIONAL O
SERIES DE TIEMPO?.**

**En busca de pronósticos ajustados
para el precio del cobre en el corto y mediano plazo.**

**TESIS PARA OPTAR AL GRADO DE MAGISTER EN GESTION DE
OPERACIONES**

CRISTIÁN ISAAC FOIX CASTILLO

**PROFESOR GUÍA:
SR. RICHARD WEBER HAAS.**

**MIEMBROS DE LA COMISIÓN:
SR. LUIS ABURTO LAFOURCADE.
SR. FABIÁN MEDEL GARCÍA.
SR. ENRIQUE SILVA RAMOS.**

**SANTIAGO DE CHILE
DICIEMBRE 2006**

**Resumen de Tesis para optar al Grado de Magíster en
Gestión de Operaciones
Alumno: Cristian Foix Castillo
Prof. Guía: Richard Weber H.
Fecha: 25/04/2007**

**PROYECCIÓN DEL PRECIO DEL COBRE: ¿HERRAMIENTAS DE INTELIGENCIA
COMPUTACIONAL O SERIES DE TIEMPO?.
En busca de pronósticos ajustados para el precio del cobre en el corto y mediano plazo.**

Esta tesis tiene como objetivo la entrega de evidencias respecto a la potencia de las redes neuronales como herramienta para el pronóstico del precio anual del cobre. Con este fin se evaluó el desempeño predictivo, fuera de muestra, de diferentes redes neuronales de tipo multilayer perceptron, construidas sobre la base del precio rezagado del cobre y variables derivadas del mismo, tales como la última variación del precio y la desviación estándar de los últimos periodos. Los resultados conseguidos se contrastaron con los generados mediante la aplicación de los más tradicionales y exitosos modelos de series de tiempo (ARIMA, caminata aleatoria y promedio móvil).

Adicionalmente, se construyeron modelos híbridos combinando modelos ARIMA y redes neuronales. En este caso, los errores de pronóstico fuera de muestra de los modelos de series de tiempo sirvieron para entrenar redes destinadas a la proyección de residuos. Con los residuos así pronosticados, se corrigieron las proyecciones iniciales de los modelos de series de tiempo.

En términos generales, la metodología de trabajo aplicada en este estudio comprendió las siguientes tareas: selección de datos (precio anual del cobre refinado de la Bolsa de Metales de Londres entre los años 1913 y 2006); aplicación de transformaciones en los datos (escalamiento y transformación logarítmica); aplicación de redes neuronales, modelos de series de tiempo y modelos híbridos, junto con la programación de rutinas computacionales para la realización de pronósticos con un alcance de hasta 6 años; evaluación de resultados e identificación de las características de los modelos más exitosos.

La calidad de los pronósticos generados se midió comparando el promedio y la desviación estándar de los errores porcentuales absolutos en el periodo comprendido entre los años 1977 y 2006, ventana de tiempo que abarca tanto fases ascendentes como descendentes del precio.

Los resultados revelaron un mejor desempeño de los modelos de pronóstico basados en redes neuronales, con una reducción relativa del promedio del error porcentual absoluto (MAPE) de 30% respecto de los mejores modelos ARIMA y de hasta un 49% respecto a la caminata aleatoria, en pronósticos a más de dos años.

De este modo, se estableció que el modelamiento no lineal a través de redes neuronales es capaz de aprovechar mejor la información contenida en los precios históricos considerados en el estudio.

Por otra parte, en pronósticos a 1 año, el menor error alcanzado con la aplicación de un modelo híbrido (reducción relativa del MAPE de 1,5% respecto a redes neuronales, de 5,6% respecto a modelos ARIMA y de 13% respecto al camino aleatorio) evidenció que, mediante el trabajo conjunto con modelos de series de tiempo y redes neuronales, es posible conseguir pronósticos de mayor precisión a los generados con cada técnica por separado.

Finalmente, la comparación de los pronósticos publicados por un prestigioso analista del mercado del cobre versus los resultados conseguidos con los mejores modelos seleccionados reveló la mayor precisión de estos últimos, con reducciones relativas del MAPE de 33% y 43% en proyecciones a 1 y 2 años respectivamente.

El impacto que tiene el precio del cobre en la actividad minera (evaluación de proyectos, planificación de la producción, evaluación y control de los resultados de gestión y de negocios) y en la economía de nuestro país (alta participación del cobre en las exportaciones y en los ingresos fiscales), conjuntamente con los resultados observados en este estudio, hacen atractivo continuar con la investigación en torno al uso de herramientas de Inteligencia Computacional en el pronóstico del precio del cobre. Trabajos futuros podrían enfocarse en otras estrategias de aprendizaje, arquitecturas alternativas, incorporación de otras variables explicativas (stocks en semanas de consumos, índices de actividad económica, etc.), aplicación de heurísticas para el diseño de redes, evaluación de diferentes modelos híbridos y la aplicación de máquinas de soporte vectorial.

AGRADECIMIENTOS

El autor agradece las contribuciones y correcciones del profesor guía, Sr. Richard Weber H. y de los profesores miembros de la comisión, Sr. Luis Aburto L., Sr. Fabián Medel G. y Sr. Enrique Silva R.

También se agradece a la Srta. Julie Lagos, por todas sus gestiones y el apoyo entregado durante el curso del Magíster en Gestión de Operaciones.

TABLA DE CONTENIDOS

I. INTRODUCCIÓN.	4
1.1. Motivación del estudio.	4
1.2. Objetivos.	4
1.2.1. Objetivo general.	4
1.2.2. Objetivos específicos.	4
1.3. Metodología.	5
1.4. Alcances.	7
II. LA IMPORTANCIA DEL PRECIO DEL COBRE Y UNA MIRADA A SU EVOLUCIÓN HISTÓRICA.	8
2.1. El impacto del precio del cobre en la actividad minera y en la economía nacional.	8
2.2. ¿Es pronosticable el precio del cobre?.	9
III. PRONÓSTICOS: ALGUNAS DEFINICIONES Y HERRAMIENTAS.	13
3.1. Distinciones relevantes.	13
3.1.1. Pronósticos dentro y fuera de muestra.	14
3.1.2. Pronósticos estáticos y dinámicos.	14
3.1.3. Conjuntos de entrenamiento y evaluación.	14
3.1.4. Origen fijo versus origen móvil de los pronósticos (fixed origin versus rolling origin).	15
3.1.5. Recalibración de los modelos.	16
3.1.6. Conjunto entrenamiento de tamaño creciente versus conjunto de entrenamiento de tamaño constante.	17
3.1.7. Alcance de los pronósticos y toma de decisiones.	17
3.2. Evaluación del desempeño predictivo: medición del error.	18
3.3. Herramientas de Pronóstico: Modelos Econométricos.	20
3.4. Modelos de Series de Tiempo.	21
3.5. Inteligencia Computacional, herramientas y aplicaciones.	26

3.6. Redes Neuronales.	28
3.6.1. Caracterización y funcionamiento de una neurona.	29
3.6.2. Características y funcionamiento de las Redes Neuronales.	32
3.7. Sistemas Inteligentes Híbridos.	40
IV. GENERACIÓN DE PRONÓSTICOS Y EVALUACIÓN DE RESULTADOS.	41
4.1. Datos históricos y diseño general del ejercicio de pronóstico.	41
4.2. Modelamiento, pronóstico y resultados.	43
4.2.1. Modelos de Series de Tiempo.	43
4.2.2. Redes Neuronales.	51
4.2.3. Comparación de resultados.	61
4.2.4. Sistemas Inteligentes Híbridos.	70
4.2.5. Comparación con proyecciones de expertos.	74
V. CONCLUSIONES Y TRABAJOS FUTUROS.	77
5.1. Conclusiones.	77
5.2. Trabajos Futuros.	79
VI. BIBLIOGRAFÍA.	81
ANEXO A: Test de Dickey-Fuller y correlograma del precio anual del cobre.	85
ANEXO B: Descripción del algoritmo backpropagation.	87
ANEXO C: Descripción del método de Levenberg-Marquardt.	89
ANEXO D: Programa computacional para la generación y evaluación de proyecciones fuera de muestra con Modelos de Series de Tiempo.	90
ANEXO E: Programa computacional para la generación y evaluación de proyecciones fuera de muestra con Redes Neuronales.	94
ANEXO F: Redes Neuronales: MAPE en los conjuntos de entrenamiento y testeo.	100
ANEXO G: Redes Neuronales: Desvío Estándar del Error Porcentual Absoluto antes y después de la inclusión de variables adicionales.	101

CAPÍTULO I

INTRODUCCIÓN

1.1. MOTIVACIÓN DEL ESTUDIO.

Durante el transcurso del año 2006, el precio del cobre ha llegado a rozar los 4 dólares por libra, con lo que su promedio al mes de junio alcanzó los 275 centavos por libra, situándose en un nivel que no se veía desde los años setenta. Este hecho, sumado al impacto que tiene el precio del cobre en la actividad minera y en la economía del país, reaviva el interés por contar con buenas herramientas para su pronóstico.

Con el propósito de explorar alternativas a los métodos econométricos de uso más común, en esta tesis se evalúa la exactitud, en el pronóstico del precio anual del cobre, de numerosos modelos basados en Redes Neuronales (multilayer perceptron) buscando entregar indicios respecto a la potencia de los mismos.

1.2. OBJETIVOS.

1.2.1. OBJETIVO GENERAL.

- Entregar evidencias respecto a la potencia de las Redes Neuronales como herramienta para el pronóstico del precio anual del cobre.

1.2.2. OBJETIVOS ESPECIFICOS.

- Evaluar el desempeño predictivo de diferentes Redes Neuronales de tipo multilayer perceptron, construidas y alimentadas sólo con el precio rezagado del cobre y variables derivadas de éste.
- Contrastar dicho desempeño con el de los más tradicionales y exitosos Modelos de

Series de Tiempo, estableciendo el tipo de modelamiento que es capaz de aprovechar mejor la información contenida en los precios históricos considerados en el estudio.

1.3. METODOLOGÍA.

En este apartado, se hará una descripción breve del trabajo realizado, utilizando como marco el proceso de extracción de conocimiento a partir de datos, conocido como KDD (Knowledge Discovery in Databases).

El problema básico que enfrenta el proceso KDD es el de transformar datos de bajo nivel (que típicamente son muy numerosos para ser asimilados con facilidad) en información que pueda ser más compacta (por ejemplo, un reporte breve), o más abstracta (por ejemplo, una aproximación descriptiva o un modelo del proceso generador de datos) o más útil (por ejemplo, un modelo predictivo para la estimación de valores futuros) [11].

El proceso KDD parte por una comprensión básica del dominio en que se da el fenómeno bajo estudio y la identificación del objetivo deseado, para luego continuar con 5 etapas principales cuya aplicación podría ser recursiva [11]:

1. Selección de los datos a partir de los cuales se extraerá conocimiento.

Los datos utilizados en esta tesis corresponden al precio anual del cobre refinado de la Bolsa de Metales de Londres durante el periodo comprendido entre los años 1913 y 2006.

2. Limpieza y preprocesamiento.

De ser necesario, en esta etapa se remueven los datos erróneos y se define el procedimiento frente a datos faltantes. Estas tareas no se requirieron en los ejercicios desarrollados.

3. Aplicación de transformaciones que faciliten la manipulación, interpretación o modelamiento de los datos.

Los precios históricos fueron escalados para expresarlos en moneda constante y luego se les aplicó la función logaritmo natural. Adicionalmente, se crearon las variables “última variación del precio” y “desviación estándar de los últimos periodos”.

4. Data mining: aplicación de herramientas analíticas en la búsqueda de patrones e información relevante.

Se construyeron rutinas computacionales para la generación de pronósticos fuera de muestra en un horizonte de 1 a 6 años. Se aplicaron Modelos de Series de Tiempo y Redes Neuronales (multilayer perceptron) con distintas frecuencias de actualización y, en el caso de las Redes Neuronales, probando diversas combinaciones de parámetros de diseño (número de neuronas en la capa de entrada, número de neuronas en la capa oculta, fracción de datos muestrales en los conjunto de entrenamiento y validación). Junto a lo anterior, también se construyó un Modelo Híbrido, aplicando conjuntamente Modelos de Series de Tiempo y Redes Neuronales.

5. Interpretación y evaluación.

Se evaluó el desempeño predictivo de los modelos construidos, comparando sus resultados en todo el periodo de evaluación (1977-2006) y en circunstancias particulares como cambios bruscos de precios. Los mejores modelos fueron seleccionados de acuerdo al menor MAPE, registrándose también el desvío estándar de los errores porcentuales absolutos y el RMSE. Junto con esto, se destacaron las características particulares de los mejores modelos lineales y no lineales estimados.

1.4. ALCANCES.

Como se señala en sus objetivos, este trabajo busca aportar evidencias respecto a la potencia de las Redes Neuronales como herramienta para el pronóstico del precio del cobre.

Para tal efecto, utilizando los mismos datos históricos del precio anual del cobre, se construyeron Redes Neuronales y Modelos de Series de Tiempo con los que se generaron pronósticos fuera de muestra en un horizonte de 1 a 6 años, para luego evaluar sus correspondientes errores de proyección. De este modo, se estableció el tipo de modelamiento que es capaz de aprovechar mejor la información contenida en los precios históricos considerados en el estudio.

Sin embargo, a diferencia de los Modelos de Series de Tiempo utilizados como referente, las Redes Neuronales pueden ser alimentadas por variables distintas al precio del cobre, por lo que la comparación realizada podría subestimar el potencial de las redes. Los resultados presentados deben ser considerados como un antecedente para futuras investigaciones que evalúen la capacidad de pronóstico de redes construidas sobre la base del precio histórico y otros datos adicionales, como por ejemplo: stocks de cobre refinado, índices de actividad económica, posiciones netas de inversionistas no comerciales, índices accionarios, etc.

Fuera del alcance de esta tesis está la proposición de metodologías para el diseño y selección de redes, tema en el que se trabaja arduamente hoy en día, buscando heurísticas que permitan capturar el máximo del potencial de generalización de las redes [7]. Es justamente dicha capacidad la que este trabajo desea insinuar.

CAPÍTULO II

LA IMPORTANCIA DEL PRECIO DEL COBRE Y UNA MIRADA A SU EVOLUCIÓN HISTÓRICA

2.1. EL IMPACTO DEL PRECIO DEL COBRE EN LA ACTIVIDAD MINERA Y EN LA ECONOMÍA NACIONAL.

La minería del cobre se ve directamente afectada por los vaivenes del precio:

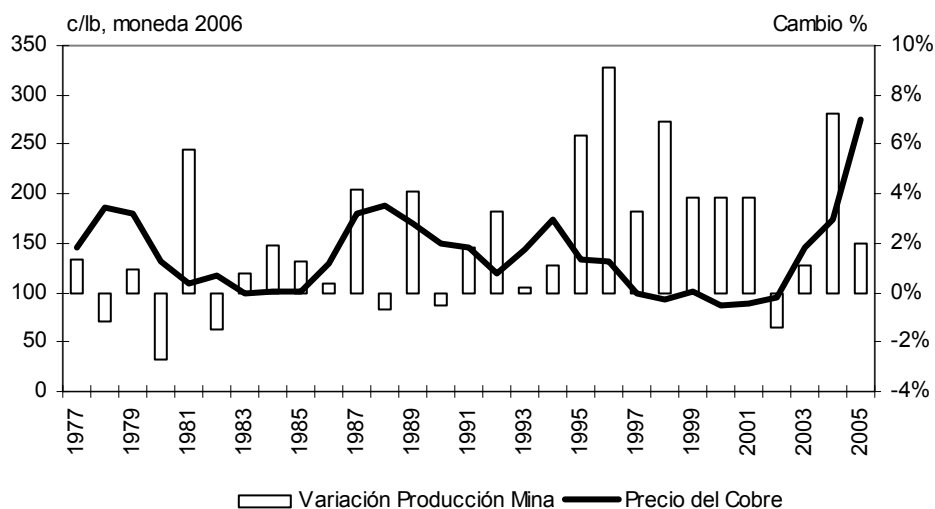
- En los periodos de altos precios, aumentan las utilidades de las faenas en operación y se reactivan las exploraciones, nuevos proyectos, expansiones y reaperturas.
- En los periodos de bajos precios, disminuyen las utilidades de las operaciones existentes, disminuyen las inversiones en exploraciones, se postergan proyectos y se cierran faenas de altos costos.

De este modo las fluctuaciones del precio del cobre se traducirán en variaciones de la producción minera, aunque con un desfase dado por la capacidad de reacción de la industria (Gráfico N°1).

En Chile, estas variaciones de la actividad minera repercuten en las inversiones, en el tipo de cambio vía exportaciones, y en el presupuesto fiscal vía tributación, transmitiéndose al resto de la economía gracias al efecto de estas variables sobre la demanda agregada, la inflación, los precios relativos y el tipo de cambio real [23][28]. Así, en los últimos 20 años, mientras el precio registró un mínimo anual de 87,4 c/lb y un máximo anual de 188,5 c/lb (valores en moneda 2006), la minería del cobre representó entre el 34% y el 50% de las exportaciones y los aportes de la Corporación del Cobre (Codelco) llegaron a significar entre el 2% y el 25% de los ingresos fiscales [5].

Por lo anterior, contar con proyecciones precisas del precio anual del cobre tiene atractivo no sólo para los productores del metal, quienes resultarían favorecidos por el menor error en un parámetro clave del proceso de planificación minera y evaluación de proyectos, sino que también para el gobierno de la nación. Por ejemplo, la conveniencia de ajustar el gasto fiscal o de recurrir a endeudamiento en un contexto de bajos precios del cobre dependerá de la duración pronosticada del periodo de malos precios. Si la fase depresiva del precio se proyecta como prolongada, lo conveniente será la reducción del gasto fiscal, mientras que si se trata de un fenómeno de corta duración, la mejor estrategia podría ser el endeudamiento [9].

Gráfico N°1: Precio del Cobre y Variación Porcentual de la Producción Minera Mundial



Nota: Precio del Cobre entre los años 1977 y 2006, c/lb moneda 2006. 2006, promedio a junio.
Fuente: Cochilco y World Bureau of Metal Statistics.

2.2. ¿ES PRONOSTICABLE EL PRECIO DEL COBRE?.

Una variable será pronosticable en la medida que su comportamiento histórico nos revele patrones o relaciones que nos permitan estimar su evolución futura. De este modo, el éxito de los pronósticos basados en modelos depende de [4]:

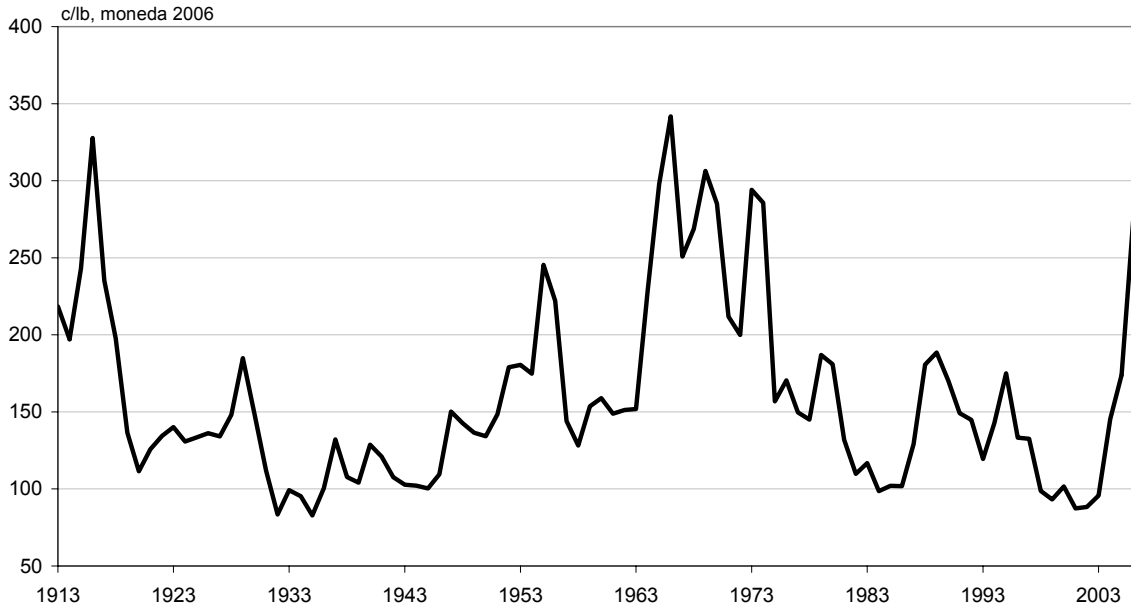
- La existencia de regularidades a ser capturadas.
- El que las regularidades sean informativas acerca del futuro.
- La adecuada captura de estas regularidades por parte del modelo construido.

Con esta definición como referente podemos examinar el caso del cobre. La evolución del precio del cobre desde 1913 en adelante está marcada por su alta volatilidad (Gráfico N°2). Las causas que provocan estas variaciones son diversas y de distinta naturaleza [22]:

- Fuertes cambios en la industria derivados ya sea de shocks en el precio de los factores productivos, grandes innovaciones tecnológicas o del desarrollo de nuevos mercados, cuyos efectos pueden extenderse por decenas de años. A estos trastornos se les denomina shocks con efecto permanente.
- Desajustes transitorios entre la oferta y la demanda de cobre provocados por los ciclos de la actividad económica mundial y la falta de flexibilidad de los productores para alterar sus tasas de operación, cerrar/reabrir faenas o adelantar el desarrollo de proyectos.
- Fallas de equipos, accidentes, huelgas, terremotos, conflictos políticos y, desde los años noventa, cada vez con mayor fuerza, operaciones financieras, tanto de cobertura como de inversión o especulación, que podrían afectar el precio por días, semanas o meses.

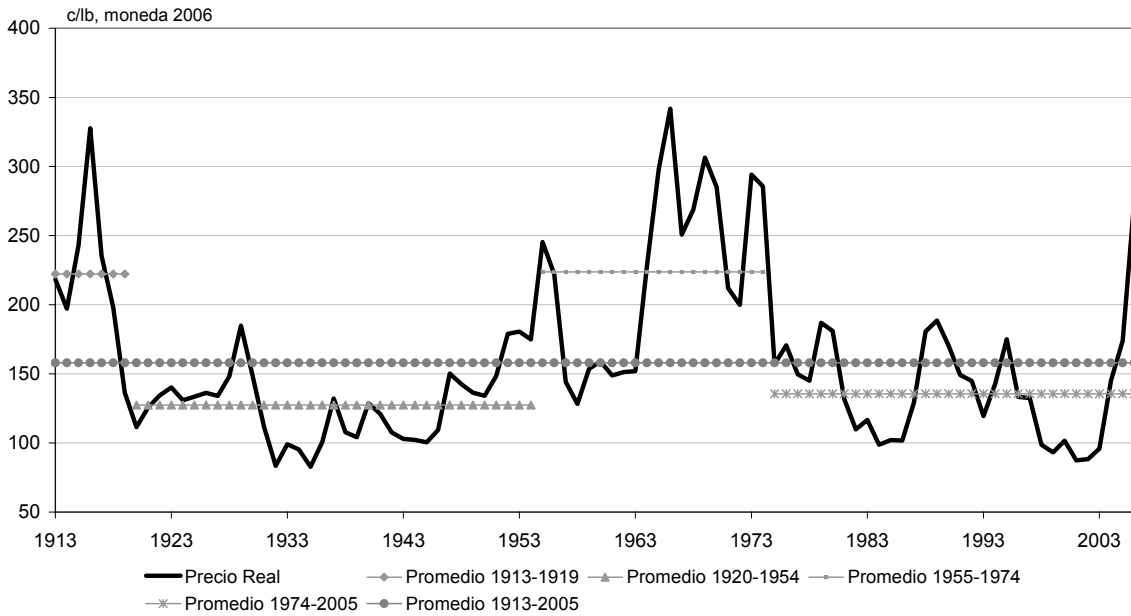
La suma de todos estos eventos determinará el precio del cobre en cada momento. La serie histórica de precios refleja esta descripción: el precio del cobre se mueve por decenas de años, sujeto a shocks temporales, en torno a un nivel definido por distintos shocks permanentes (Gráfico N°3).

Gráfico N°2: Precio del Cobre entre 1913 y 2006 (promedio a junio), c/lb moneda 2006.



Fuente: Precio del Cobre: CRU y Cochilco. Deflactor del Precio: U.S. Department of Labor, Bureau of Labor Statistics.

Gráfico N°3: Precio Promedio del Cobre a través del Tiempo, c/lb moneda 2006.



Fuente: Precio del Cobre: CRU y Cochilco. Deflactor del Precio: U.S. Department of Labor, Bureau of Labor Statistics.

La dificultad en el pronóstico del precio tiene que ver con lo complejo que resulta adelantarse a los shocks mencionados. Por otra parte, el potencial de pronóstico del precio del cobre se basa en el decaimiento esperado de los shocks temporales recientemente producidos [25]. La predictibilidad del decaimiento de los shocks temporales obedece al comportamiento de la producción de cobre que, siguiendo a los precios, siempre terminará ajustándose a la demanda en el mediano y largo plazo. A modo de ejemplo, desde fines del año 2003, la demanda mundial de cobre ha superado con creces a la oferta, lo que ha elevado considerablemente los precios. La oferta ha reaccionado lentamente a esta mayor demanda, esperándose que recién supere al consumo el año 2007 o 2008.

Distinta podría ser la situación de más corto plazo, donde la mayor participación de los agentes financieros en las bolsas de metales, realizando operaciones que involucran cobre en función de la rentabilidad de otros activos, dificulta la proyección de la evolución del precio en los próximos minutos, semanas o meses [22].

En síntesis, estando el precio del cobre en el mediano y largo plazo ligado fuertemente al mercado físico, y evidenciando éste un patrón de comportamiento, las regularidades resultantes podrían ser captadas por modelos de pronóstico.

CAPÍTULO III

PRONÓSTICOS: ALGUNAS DEFINICIONES Y HERRAMIENTAS

3.1. DISTINCIONES RELEVANTES.

En este trabajo, junto al desarrollo de nuevos modelos de pronóstico, se busca evaluar su desempeño predictivo. Ambas tareas requieren el conocimiento de ciertos conceptos que van más allá de la herramienta de pronóstico particular que se desee utilizar. La descripción que se realiza a continuación está basada en [31].

3.1.1. PRONÓSTICOS DENTRO Y FUERA DE MUESTRA.

Al hablar de pronósticos, se distingue entre proyecciones dentro y fuera de muestra. En las primeras, las proyecciones realizadas se refieren a los mismos datos que se emplearon para la construcción y calibración del modelo (la muestra), mientras que en las segundas las proyecciones se refieren a datos ajenos a dicha muestra.

En la búsqueda de metodologías que generen pronósticos precisos de los valores futuros de una variable, sólo son relevantes las proyecciones fuera de muestra por las siguientes razones:

- Las proyecciones fuera de muestra replican el funcionamiento de la herramienta de pronósticos en la práctica, por lo que la evaluación de su desempeño predictivo será un referente válido para los futuros errores de pronóstico.
- Los modelos de pronóstico se construyen minimizando los errores dentro de muestra por lo que los errores de pronósticos intramuestrales sobrestiman el potencial predictivo de las herramientas.

- Un modelo con buen desempeño intramuestral podría tener un muy mal desempeño en proyecciones fuera de muestra. Esto se debe a un sobreajuste (overfitting) o memorización de los datos muestrales, con lo que el modelo resultante será incapaz de responder de buena manera a nuevos valores.

Por lo anterior, a lo largo de este documento, cuando se haga mención a pronósticos, siempre se estará haciendo referencia a pronósticos fuera de muestra.

3.1.2. PRONÓSTICOS ESTÁTICOS Y DINÁMICOS.

Los pronósticos estáticos son aquellos que están basados en la última información efectiva disponible, por lo que están limitados a las proyecciones a un periodo hacia adelante. En este trabajo se emplearán pronósticos dinámicos caracterizados por utilizar el último pronóstico disponible como dato para el siguiente pronóstico, permitiendo la realización de proyecciones a dos y más periodos hacia adelante.

3.1.3. CONJUNTOS DE ENTRENAMIENTO Y EVALUACIÓN.

Existen dos formas de evaluar la precisión de los pronósticos fuera de muestra:

- Esperar hasta que se cuente con los valores reales para los periodos pronosticados. Por ejemplo: si, durante el año 2006, se pronostica el precio del año siguiente, esperar hasta conocer el valor efectivo del año 2007.
- Evaluar la precisión sobre la base de un conjunto de datos que previamente se separó de la muestra disponible y que no participó de la construcción del modelo. Al conjunto de datos empleados para la construcción del modelo se le denomina conjunto de entrenamiento, mientras que el resto de los datos conforma el conjunto de evaluación.

Por las limitaciones prácticas que tiene la estrategia de espera, en esta tesis se aplicó la división de los datos muestrales en los conjuntos de entrenamiento y evaluación (estrategia más comúnmente empleada).

La división de los datos muestrales es una decisión trascendental en la generación de pronósticos ya que determina la cantidad de datos para la construcción del modelo y la cantidad de pronósticos fuera de muestra que se podrán evaluar.

La definición del tamaño y composición de los conjuntos de entrenamiento y evaluación deberá considerar factores tales como:

- Tamaño total de la muestra: en muestras pequeñas, grandes conjuntos de evaluación podrían comprometer la calidad del modelo construido, si es que el conjunto de entrenamiento no consigue un tamaño que lo haga representativo [19].
- Tipo de metodología de pronóstico a emplear: distintas metodologías demandan conjuntos de entrenamiento más o menos numerosos.
- Representatividad: los componentes del conjunto de entrenamiento deben ser diversos para asegurar que el modelo pueda captar los diversos patrones de comportamiento de la serie bajo estudio (por ejemplo: precios en fases depresivas, precios en fases expansivas) [19].

Respecto a los tamaños relativos de ambos conjuntos no existe un consenso y se opera en función de la experiencia.

3.1.4. ORIGEN FIJO VERSUS ORIGEN MÓVIL DE LOS PRONÓSTICOS (FIXED ORIGIN VERSUS ROLLING ORIGIN).

Dados los conjuntos de entrenamiento y evaluación, se define como el origen de los pronósticos al índice T correspondiente al último dato del conjunto de entrenamiento y se define como N al tamaño del conjunto de evaluación.

Los pronósticos de origen fijo, predicen la variable de interés a partir del dato T , esto es para los periodos $T+1$, $T+2$, ..., $T+N$. De este modo, para un origen fijo sólo se calcularán N pronósticos y sólo un pronóstico para cada alcance (un pronóstico a un periodo, un pronóstico a dos periodos, etc.), lo que es insuficiente para evaluar el desempeño de una metodología.

Por el contrario, en los pronósticos de origen móvil, se actualiza sucesivamente el origen de los pronósticos, lo que incrementa el número de proyecciones para cada alcance. Así en la situación recién descrita, una vez que se proyectaron los valores a partir de T, se calculan los pronósticos a partir de T+1 (T+2, T+3, ..., T+N), a partir de T+2 (T+3, T+4, ..., T+N) y así sucesivamente. El total de pronósticos calculados será: $N \times (N + 1) / 2$

En cuanto al número de pronósticos por alcance, la relación se describe en la Tabla N°1.

Tabla N°1: Alcance de Pronósticos y Número de Evaluaciones Empleando Origen Móvil.

Alcance del pronóstico	Número de evaluaciones
a 1 periodo	N
a 2 periodos	N-1
a H periodos	N-H+1
a N-1 periodos	2
a N periodos	1

Esta última relación entre el alcance de los pronósticos y el número de evaluaciones, nos permite dimensionar el tamaño absoluto del conjunto de evaluación. Sea H el máximo alcance de los pronósticos que se desea evaluar y sea M el número mínimo de evaluaciones que se desea realizar a dicho alcance, el tamaño del conjunto de evaluación estará dado por: $N \geq M + H - 1$.

Por otra parte, el uso de origen móvil disminuye la influencia de un determinado origen en los resultados (por ejemplo, fase depresiva de un ciclo económico).

Las ventajas del origen móvil por sobre el origen fijo hacen que el origen móvil sea la técnica preferida en evaluaciones fuera de muestra, lo que fundamenta su aplicación en este trabajo.

3.1.5. RECALIBRACIÓN DE LOS MODELOS.

El empleo de origen móvil plantea la posibilidad de reestimar el modelo de pronóstico en cada actualización. Este procedimiento es el más usado ya que disminuye la influencia del conjunto de entrenamiento original, aunque esto signifique un aumento de los cálculos necesarios. Por lo anterior, en esta tesis, los modelos serán reestimados periódicamente.

3.1.6. CONJUNTO ENTRENAMIENTO DE TAMAÑO CRECIENTE VERSUS CONJUNTO DE ENTRENAMIENTO DE TAMAÑO CONSTANTE.

Al utilizar la técnica de origen móvil, cada nueva evaluación significa la adición de un nuevo dato al conjunto de entrenamiento, por lo que se puede optar entre la realización de proyecciones sobre la base de un conjunto de entrenamiento de tamaño creciente o de tamaño constante (fixed size rolling window). El uso de un conjunto de entrenamiento de tamaño constante implicaría que al agregar un nuevo dato, se descarte la observación más antigua (pruning), lo que parece recomendable si la trayectoria de precios a través del tiempo sigue un patrón notoriamente distinto al del pasado, situación que parece no aplicar al caso del cobre [25].

3.1.7. ALCANCE DE LOS PRONÓSTICOS Y TOMA DE DECISIONES.

Todo pronóstico tiene asociado un alcance, pudiendo ser éste de corto, mediano o largo plazo. Los horizontes de tiempo correspondientes a dichos alcances dependerán de la industria bajo estudio. En cuanto al atractivo de uno u otro pronóstico, éste estará sujeto al tipo de decisión que se desea tomar o de acción en desarrollo.

A modo de ejemplo, en la industria del cobre, alcances convencionales y decisiones comunes en el mercado y la industria son:

Tabla N°2: Alcance de Pronósticos y Toma de Decisiones.

Tipo de Pronóstico	Alcance	Decisiones
Cortísimo Plazo	Minutos, horas	Operaciones especulativas
Corto Plazo	Días, semanas, meses, un año	Operaciones especulativas, de cobertura y de gestión comercial
Mediano Plazo	Uno a seis años	Evaluación y control de los resultados de la gestión y de los negocios de una empresa. Planificación de la producción y evaluación de proyectos ¹
Largo Plazo	6 a 50 años	Planificación de la producción y evaluación de proyectos ¹

¹ El desarrollo de un proyecto minero, desde su etapa exploratoria hasta el fin de su construcción puede tomar, en promedio, unos 5 a 7 años. La planificación minera de un yacimiento abarca toda la vida económica del mismo, pudiendo extenderse ésta por más de 50 años.

Relacionado con el alcance de un pronóstico está su nivel de incertidumbre. A mayor alcance del pronóstico, mayor es el nivel de incertidumbre que se debe enfrentar. Esta consideración no debe olvidarse al momento de tomar decisiones basadas en datos proyectados [13][21][26].

En el presente trabajo, el foco estará puesto en las proyecciones de corto y mediano plazo, en un horizonte de 1 a 6 años por sus potenciales aplicaciones en la industria del cobre y en las finanzas públicas (prepuesto fiscal y Fondo de Compensación del Cobre o similar). No se evaluarán pronósticos de largo plazo, dado que los Modelos ARIMA, utilizados como referente, no tienen un buen desempeño en proyecciones de largo alcance [9].

3.2. EVALUACIÓN DEL DESEMPEÑO PREDICTIVO: MEDICIÓN DEL ERROR.

Para la evaluación del desempeño predictivo se emplean diferentes indicadores que cuantifican qué tan cerca está la variable pronosticada de su serie de datos correspondiente [26]. Una de las medidas más utilizadas es el Promedio del Error Porcentual Absoluto (MAPE) [14][26]:

$$MAPE = \frac{1}{T} \left(\sum_{t=1}^T APE_t \right) = \frac{1}{T} \left(\sum_{t=1}^T \frac{|Y_t^s - Y_t^a|}{Y_t^a} \right) \times 100$$

donde,

APE : error porcentual absoluto.

Y_t^s : valor pronosticado de Y_t .

Y_t^a : valor real de Y_t .

T : número de periodos.

El MAPE mide el valor medio del error absoluto en términos porcentuales al valor real de la variable.

Para evaluar la dispersión de los errores se puede calcular el Desvío Estándar del Error porcentual absoluto (APE).

$$Desvío\ Estándar\ APE = \sqrt{\frac{1}{T} \sum_{t=1}^T (APE_t - MAPE)^2}$$

Otra medida del error de pronóstico comúnmente empleada es la Raíz Cuadrática Media del Error (RMSE):

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (Y_t^s - Y_t^a)^2}$$

donde,

Y_t^s : valor pronosticado de Y_t

Y_t^a : valor real de Y_t

T: número de periodos

El RMSE mide la dispersión de la variable simulada en el curso del tiempo, penalizando fuertemente los errores grandes al elevarlos al cuadrado. Esta característica hace que el RMSE se recomiende cuando el costo de cometer un error es aproximadamente proporcional al cuadrado de dicho error [24].

No siempre el modelo que genere pronósticos con un menor MAPE generará los pronósticos con el menor RMSE y viceversa, por lo que en la selección de los mejores modelos de pronóstico se hace necesario establecer la medida de error a utilizar para la elaboración del ranking de desempeño.

Dado que una mala estimación del precio futuro del cobre se traduce en una pérdida de ingresos proporcional al tamaño del error, el MAPE, y no el RMSE, parece ser la medida de desempeño más adecuada. A esto se suma la ventaja práctica del MAPE de no requerir ser acompañado por la media para dimensionar la magnitud del error. Luego, la medida de error que se empleará para identificar los modelos de mejor desempeño será el MAPE.

Sin perjuicio de lo anterior, para los mejores modelos identificados se presentarán tanto el MAPE, el desvío estándar del APE y el RMSE de sus pronósticos.

3.3. HERRAMIENTAS DE PRONÓSTICO: MODELOS ECONOMÉTRICOS.

En el campo de la econometría, se acostumbra distinguir dos áreas de modelamiento [21][26]:

- Modelos Estructurales. En estos modelos se representa la variable de interés (variable dependiente) en función de sus relaciones económicas con un conjunto de variables explicativas, a través de una o más ecuaciones.
- Modelos de Series de Tiempo. En los Modelos de Series de Tiempo, a diferencia del modelamiento estructural, no se busca explicar la variable de interés, sino que sólo pronosticarla de acuerdo al comportamiento pasado de la misma variable.

Ambos tipos de modelos se aplican al pronóstico del precio del cobre. Sin embargo, en la generación de proyecciones de corto y mediano plazo, existe evidencia de un mejor desempeño de los Modelos de Series de Tiempo [9]. Llama la atención el que los Modelos de Series de Tiempo sean más precisos en sus proyecciones que los modelos econométricos más complejos, algunos de los cuales tienen múltiples ecuaciones y decenas de variables [3][18][34][35]. Entre las razones que explican este fenómeno está la dificultad asociada a la selección de las variables explicativas de un modelo estructural y la dificultad que conlleva el pronóstico de las mismas, problema que podría ser aún más difícil que el pronóstico de la variable de interés.

Siendo el foco de este estudio la generación de proyecciones de corto y mediano plazo, y siendo la metodología benchmark de pronósticos los Modelos de Series de Tiempo [9], se hace interesante contrastar el desempeño de estos métodos con otras metodologías de aplicación más reciente en la industria, como son las herramientas propias del campo de la Inteligencia Computacional (por ejemplo: Redes Neuronales).

3.4. MODELOS DE SERIES DE TIEMPO

Una serie de tiempo es un conjunto ordenado de observaciones de los valores que toma una variable en diferentes momentos [13].

Los Modelos de Series de Tiempo tienen un enfoque netamente predictivo y en ellos los pronósticos se elaborarán sólo con base al comportamiento pasado de la variable de interés.

Podemos distinguir dos tipos de Modelos de Series de Tiempo [26]:

- Modelos deterministas: se trata de métodos de extrapolación sencillos en los que no se hace referencia a las fuentes o naturaleza de la aleatoriedad subyacente en la serie. Su simplicidad relativa generalmente va acompañada de menor precisión.

Ejemplo de modelos deterministas son los modelos de promedio móvil en los que se calcula el pronóstico de la variable a partir de un promedio de los “n” valores inmediatamente anteriores.

$$\text{Promedio móvil (n=12): } y_{t+1} = \frac{1}{12}(y_t + y_{t-1} + \dots + y_{t-11})$$

- Modelos estocásticos: se basan en la descripción simplificada del proceso aleatorio subyacente en la serie. En términos sencillos, se asume que la serie observada $y_1, y_2, y_3, \dots, y_T$ se extrae de un grupo de variables aleatorias con una cierta distribución conjunta difícil de determinar, por lo que se construyen modelos aproximados que sean útiles para la generación de pronósticos.

La serie y_T podrá ser estacionaria o no estacionaria:

- *Serie no estacionaria*: es aquella cuyas características de media, varianza y covarianza cambian a través del tiempo lo que dificulta su modelamiento. Sin embargo, en muchas ocasiones, si dicha serie es diferenciada una o más veces la serie resultante será estacionaria (procesos no estacionarios homogéneos).

- *Serie estacionaria*: es aquella cuya media y varianza no cambian a través del tiempo y cuya covarianza sólo es función del rezago. Gracias a estas características podremos modelar el proceso subyacente a través de una ecuación con coeficientes fijos estimados a partir de los datos pasados.

Media: $E(y_t) = E(y_{t+m})$ para todo t, m.

Varianza: $\sigma(y_t) = \sigma(y_{t+m})$ para todo t, m.

Covarianza: $Cov(y_t, y_{t+k}) = Cov(y_{t+m}, y_{t+m+k})$ para todo t, m.

A continuación se describirán los modelos estocásticos de series de tiempo de uso más común:

Modelamiento de series no estacionarias.

- Caminata aleatoria: $y_t = \delta + y_{t-1} + \varepsilon_t$

donde, δ es una constante (drift) y cada perturbación ε_t (error) es una variable aleatoria con distribución normal con media cero, varianza constante y covarianza cero (el proceso $\varepsilon_1, \varepsilon_2, \dots$ se denomina ruido blanco).

En el modelo de caminata aleatoria más simple (sin drift), el pronóstico para y_t es su valor más reciente. La inclusión de un drift intenta reproducir una tendencia existente en la variable de interés.

En un modelo de caminata aleatoria, la varianza de y_t aumenta a través del tiempo, lo que es propio de un proceso no estacionario.

Cuando una serie se comporta como una caminata aleatoria se dice que ésta presenta raíz unitaria.

Modelamiento de series estacionarias².

- Modelos de Media Móvil, MA(q): $y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q}$

En los modelos de media móvil, el proceso se representa como una suma ponderada de errores actuales y anteriores. El número de rezagos del error considerados (q) determina el orden del modelo de media móvil.

- Modelos Autorregresivos, AR(p): $y_t = \delta + \phi_1y_{t-1} + \phi_2y_{t-2} + \dots + \phi_py_{t-p} + \varepsilon_t$

En los modelos autorregresivos, el proceso se representa como una suma ponderada de observaciones pasadas de la variable. El número de rezagos (p) determina el orden del modelo autorregresivo.

- Modelos Mixtos Autorregresivos – Media Móvil, ARMA(p,q):

$$y_t = \delta + \phi_1y_{t-1} + \phi_2y_{t-2} + \dots + \phi_py_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q}$$

En estos modelos, el proceso se representa en función de observaciones pasadas de la variable y de los valores actuales y rezagados del error. El número de rezagos de la variable de interés (p) y el número de rezagos del error (q) determinan el orden del modelo mixto.

² Cada perturbación ε_t se asume generada por un proceso de ruido blanco.

- Modelos Autorregresivos Integrados de Promedio Móvil, ARIMA(p,d,q):

Muchas series de tiempo no son estacionarias, por ejemplo el Producto Nacional Bruto o la Producción Industrial [8]. Un tipo especial de series no estacionarias, son las no estacionarias homogéneas que se caracterizan porque, al ser diferenciadas una o más veces, se vuelven estacionarias.

La serie y_t será no estacionaria homogénea de orden d si $w_t = \Delta^d y_t$ es estacionaria, donde, a modo de ejemplo, las diferencias de primer y segundo orden se definen como:

$$\Delta y_t = y_t - y_{t-1}$$

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$$

Si después de haber diferenciado la serie y_t se consigue una serie estacionaria w_t y dicha serie obedece a un proceso ARMA(p,q), se dice que y_t responde a un proceso ARIMA(p,d,q):

$$\Delta^d y_t = \delta + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

Para la correcta identificación del Modelo ARIMA representativo de una serie se hace necesario:

- Determinar el grado de homogeneidad u orden de integración de la serie. Para determinar el orden de integración se utilizan herramientas como el correlograma y tests de raíz unitaria. Cabe mencionar que los tests de raíz unitaria, como el test de Dickey-Fuller Aumentado (ADF), tienen baja potencia presentando, ante situaciones de difícil discriminación, un sesgo al no rechazo de la hipótesis nula de presencia de raíz unitaria [8][21].
- Determinar el orden de las partes de promedio móvil y autorregresivas del modelo. El examen de las funciones de autocorrelación total y parcial ayuda en esta tarea, aunque habitualmente la selección correcta no será clara, por lo que se recomienda probar distintas formulaciones guiándose por el conocimiento que se tenga del fenómeno analizado [21].

- Evaluar los distintos modelos construidos. Se descartan las estructuras que arrojen coeficientes no significativos y/o que fueron mal evaluadas de acuerdo a indicadores como el Criterio de Información de Schwarz (SIC). Un buen modelo tendrá un buen ajuste (coeficiente de determinación cercano a la unidad) y arrojará residuos que se comportarán como ruido blanco.
- Modelos con varianza cambiante: ARCH y GARCH.

En los Modelos de Series de Tiempo, podemos distinguir dos tipos de pronósticos, los condicionales, que están condicionados a la información disponible hasta el momento, y los no condicionales. Por ejemplo, en el modelo AR(1): $y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t$, el pronóstico condicional de y_{t+1} es $E(y_{t+1}) = \delta + \phi_1 y_t$. Este valor contrasta con el pronóstico no condicional que es, simplemente, la media de largo plazo de la serie, $\delta/(1 - \phi_1)$. Se puede demostrar que la varianza del error de pronóstico condicional es menor a la varianza del error de pronóstico no condicional, por lo que los Modelos ARIMA presentados se utilizarán para la generación de pronósticos condicionales [8].

En muchas aplicaciones (modelamiento de la inflación, tasas de interés y rendimientos de acciones), la varianza condicional de una variable cambia a través, dependiendo de la magnitud de los errores del pasado. Se observa un “agrupamiento” de errores, esto es, periodos de alta volatilidad (y grandes errores) seguidos de periodos de baja volatilidad (y errores menores) [26]. Este fenómeno también parece producirse en el precio histórico del cobre (Gráfico N°2).

La adecuada representación de las variables que muestran este tipo de heterocedasticidad requiere la definición de un modelo para su varianza condicional. Una vez hecho esto se procede a la estimación simultánea de los modelos de la media y de la varianza condicional [8].

- Modelos de Heterocedasticidad Condicional Autorregresiva, ARCH(p): se basan en la existencia de una relación entre la varianza del error y los rezagos del error al cuadrado. La cantidad de rezagos utilizada determina el orden del proceso ARCH:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_p \varepsilon_{t-p}^2$$

- Modelos de Heterocedasticidad Condicional Autorregresiva Generalizado, GARCH(p,q): suponen una relación entre la varianza del error, los rezagos del error al cuadrado y los rezagos de la varianza. La cantidad de rezagos utilizada determina el orden del proceso GARCH:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 + \lambda_1 \sigma_{t-1}^2 + \dots + \lambda_q \sigma_{t-q}^2$$

Para el adecuado modelamiento de la varianza del error se deberá recurrir a herramientas como el correlograma de los residuos al cuadrado y tests para detectar la presencia de heterocedasticidad condicional autorregresiva.

3.5. INTELIGENCIA COMPUTACIONAL, HERRAMIENTAS Y APLICACIONES.

La Inteligencia Computacional es una disciplina que tiene sus orígenes en la Inteligencia Artificial. La Inteligencia Artificial nace formalmente en la conferencia de Dartmouth de 1956, siendo definida por Marvin Minsky como: “la ciencia de construir máquinas para que hagan cosas que, si las hicieran los humanos, requerirían inteligencia” [17]. El desafío de esta naciente ciencia fue la traducción de los problemas a resolver a símbolos, los cuales serían manipulados por un computador usando reglas.

La Inteligencia Artificial tuvo un gran desarrollo hasta principios de los años ochenta, resolviendo en forma adecuada diversos problemas y viendo el surgimiento de los primeros sistemas expertos y programas de juegos (por ejemplo: ajedrez). Sin embargo, las altas expectativas generadas y el fracaso en la resolución de problemas triviales para los seres humanos, como la traducción de lenguajes naturales y el reconocimiento de patrones,

generaron fuertes críticas que llevaron cierto descrédito a la disciplina.

Frente a este panorama, en la segunda mitad de los años 80s aparecen nuevos métodos de análisis y conceptos tales como: manejo de incertidumbre, complejidad, autoorganización, aprendizaje, representación fuzzy, procesamiento neural, etc. con lo que se configura una nueva Inteligencia Artificial que adopta los nombres de Inteligencia Computacional o Soft-Computing.

La Inteligencia Computacional es una colección de paradigmas computacionales con inspiración biológica y lingüística dedicada al diseño, desarrollo y aplicación de metodologías de procesamiento tales como [16]:

- Redes Neuronales.
- Sistemas Difusos (fuzzy).
- Computación Evolutiva.
- Sistemas Swarm (inteligencia distribuida).

Las áreas de aplicación de las herramientas de Inteligencia Computacional son diversas abarcando desde la medicina hasta la manufactura y desde los negocios hasta las defensa. Ejemplos de esto son:

- Medicina:
 - Diagnóstico a partir de datos analíticos (electrocardiogramas, encefalogramas).
 - Modelamiento del sistema nervioso.
- Manufactura:
 - Robots automatizados y sistemas de control (visión artificial y sensores).
 - Inspección de calidad.

- Negocios:
 - Pronóstico de precios.
 - Valoración del riesgo crediticio.
 - Segmentación.
- Defensa:
 - Clasificación de señales de radar.

En el presente documento, se explorará el uso de Redes Neuronales y Sistemas Inteligentes Híbridos, comparando su desempeño con los Modelos de Series de Tiempo.

3.6. REDES NEURONALES.

Una Red Neuronal puede ser descrita como un modelo de regresión no lineal cuya estructura se inspira en el funcionamiento del sistema nervioso. En términos generales, una red consiste en un gran número de unidades simples de proceso, denominadas neuronas, que actúan en paralelo y están conectadas mediante vínculos ponderados [15][27].

Cada neurona recibe inputs desde otras neuronas y genera un resultado que depende sólo de la información localmente disponible, ya sea almacenada internamente o plasmada en los ponderadores de las conexiones. El output generado por la neurona servirá de input para otras neuronas.

Mediante la adecuada modificación de los ponderadores de la red, en un proceso denominado aprendizaje, la red mejorará su desempeño en el desarrollo de la tarea para la cual fue construida.

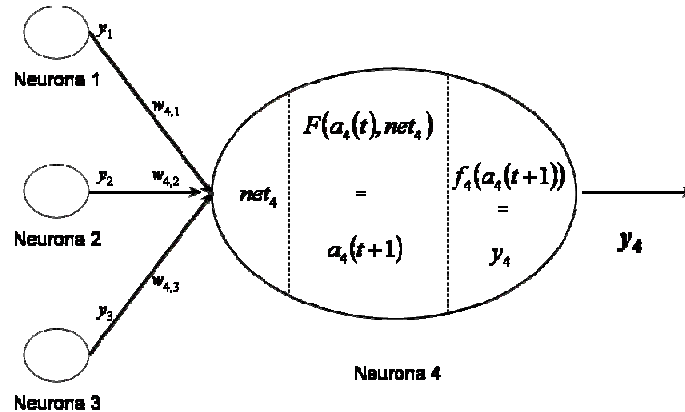
Las Redes Neuronales tienen el potencial de implementar funciones complejas. Se puede demostrar que una Red Neuronal suficientemente grande, con una estructura y ponderadores adecuados, es capaz de aproximar cualquier función con el nivel de precisión

que se desee [27].

3.6.1. CARACTERIZACIÓN Y FUNCIONAMIENTO DE UNA NEURONA.

La figura N°1 representa una Red Neuronal sencilla, destacando una neurona particular.

Figura N°1: Inputs, funcionamiento y output de una neurona.



Fuente: Hilera, J. y Martínez, V. Redes Neuronales Artificiales. Fundamentos, Modelos y Aplicaciones. [15]

Para describir una neurona artificial necesitamos definir los siguientes términos [15]:

- **Regla de propagación:** procedimiento a seguir para combinar los valores de entrada a una neurona (inputs) con los ponderadores de las conexiones que llegan a esa neurona.

Por ejemplo:

$$net_j = \sum_i^N w_{j,i} y_i$$

donde,

$w_{j,i}$: es el ponderador de la conexión entre la neurona i y la neurona j ,

y_i : es el valor de la "salida" de la neurona i ,

net_j : es el valor combinado de las entradas a la neurona j .

- **Estado de activación, $a(t)$:** valor continuo o discreto que señala la situación de la neurona, por ejemplo: reposo ó excitado. La señal que envía cada una de las neuronas a sus vecinas puede depender de su propio estado de activación.

- Función de activación, $F(a(t),net)$: relación entre el estado de activación de una neurona en el periodo $t+1$, $a(t+1)$, su estado de activación anterior, $a(t)$, y sus inputs. Normalmente, el estado de activación anterior no se toma en cuenta y el nuevo estado de activación será simplemente igual a la combinación de inputs y ponderadores (net).
- Función de salida o transferencia, $f(a(t))$: relación que transforma el estado de activación en una señal de salida. Las funciones de transferencia más comunes son: escalón, lineal (identidad), mixta, sigmoïdal y tangente hiperbólica (\tanh).

Las próximas figuras ejemplifican estas funciones.

Figura N°2: Función Escalón.

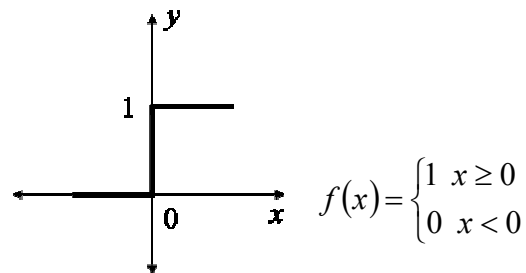


Figura N°3: Función Lineal (Identidad).

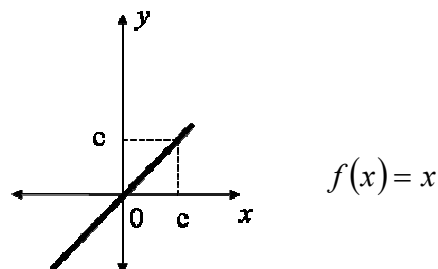


Figura N°4: Función Mixta.

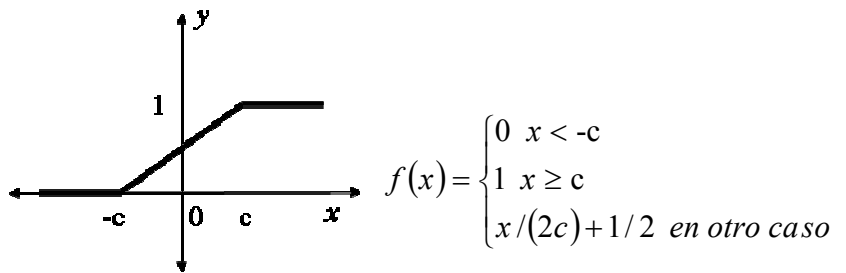


Figura N°5: Función Sigmoidal.

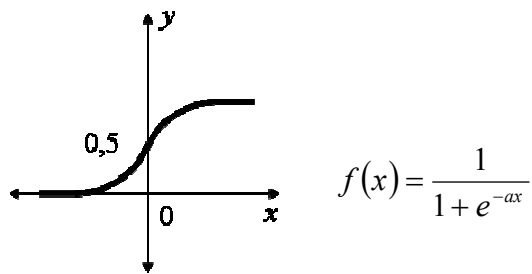
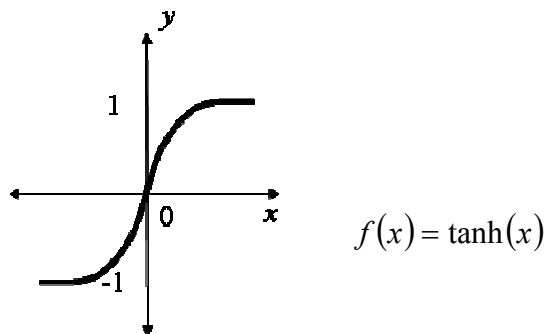


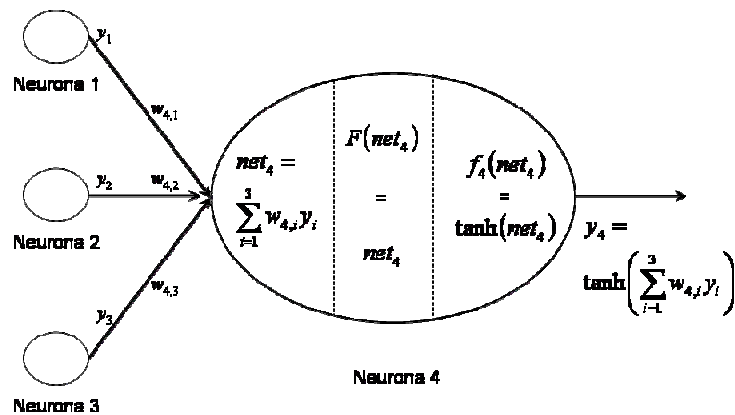
Figura N°6: Función Tanh.



En la figura N°7 se ejemplifica el funcionamiento de una neurona con características particulares:

- Regla de propagación: suma ponderada de los inputs.
- Función de activación: función identidad.
- Función de salida o transferencia: tangente hiperbólica.

Figura N°7: Inputs, Funcionamiento y Output de una Neurona.



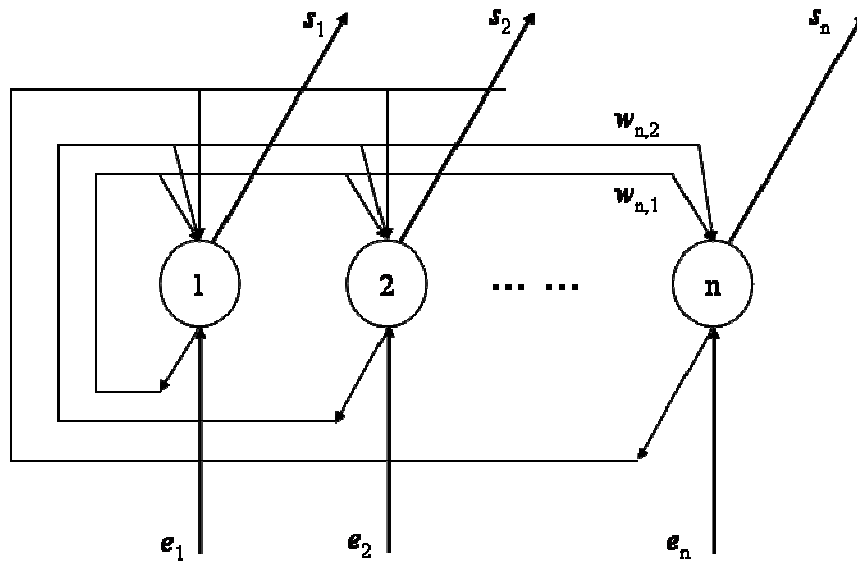
Fuente: Hilera, J. y Martínez, V. Redes Neuronales Artificiales. Fundamentos, Modelos y Aplicaciones. [15]

3.6.2. CARACTERÍSTICAS Y FUNCIONAMIENTO DE LAS REDES NEURONALES.

Una Red Neuronal corresponde a un conjunto organizado de neuronas. Para caracterizar una Red Neuronal se debe prestar atención a dos aspectos básicos: topología y mecanismo de aprendizaje [15]:

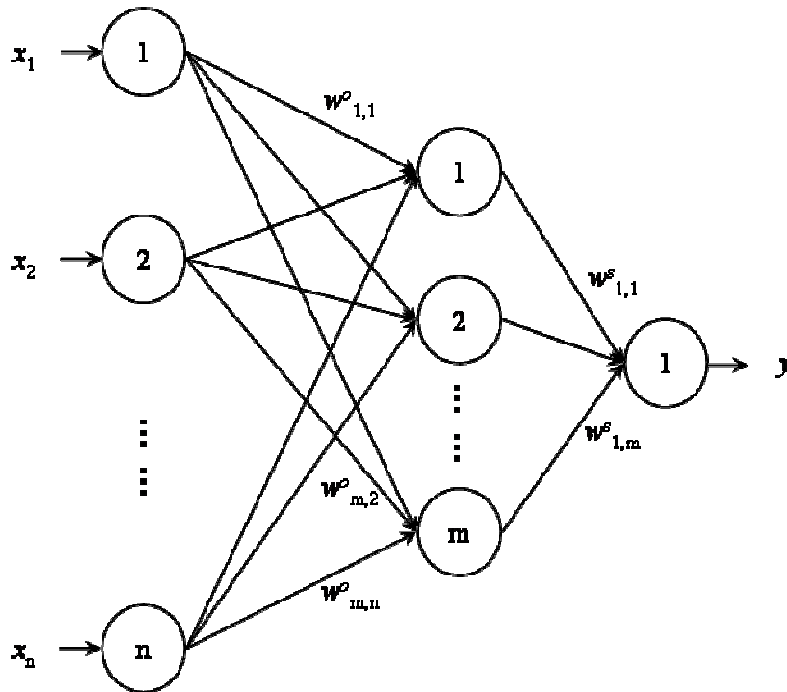
- Topología o arquitectura: corresponde a la disposición de las neuronas en la red formando capas o agrupaciones. Rasgos topológicos de una red son:
 - Número de capas: donde se distingue entre redes monocapa que tienen sólo una capa de neuronas y redes multicapa. Ejemplo de red monocapa es la Red de Hopfield (figura N°8) y ejemplo de red multicapa es la Red Multilayer Perceptron (MLP) (figura N°9).
 - Número de neuronas por capa.
 - Tipo de conexiones:
 - Conexiones hacia adelante: son las que conectan una neurona de una capa con otra neurona de la capa siguiente o posterior (figura N°9).
 - Conexiones hacia atrás: son las que vinculan una neurona de una capa con otra neurona de la capa previa o anterior.
 - Conexiones laterales: conexiones entre neuronas de una misma capa (figura N°8).
 - Conexiones autorecurrentes: donde la salida de una neurona está conectada a su propia entrada (figura N°8).
 - Grado de interconexión:
 - Red total o parcialmente conectada.
 - Red con o sin conexiones entre capas no contiguas (shortcuts).

Figura N°8: Red de Hopfield.



Fuente: Hilera, J. y Martínez, V. Redes Neuronales Artificiales. Fundamentos, Modelos y Aplicaciones. [15]

Figura N°9: Red Multilayer Perceptron.



Fuente: Hilera, J. y Martínez, V. Redes Neuronales Artificiales. Fundamentos, Modelos y Aplicaciones. [15]

A cada una de las neuronas de la red (capa oculta y capa de salida) se agrega una entrada de valor unitario cuyo ponderador (sesgo) influirá en el término constante de la regresión.

Para ejemplificar la descripción de la arquitectura de una red se hará uso de la figura N°9:

- Red multicapa.
- Número de neuronas en la capa de entrada: n .
- Número de neuronas en la capa oculta: m .
- Número de neuronas en la capa de salida: 1.
- Tipo de conexiones: conexiones hacia delante.
- Grado de interconexión: red totalmente conectada y sin conexiones entre capas no vecinas.

Continuando con el ejemplo, se pueden especificar las características de las neuronas dicha red:

- Regla de propagación y función de activación: suma ponderada de los inputs.
- Función de transferencia de las neuronas de la capa oculta: tangente hiperbólica.
- Función de transferencia de las neuronas de la capa de salida: identidad.

La red así construida, al recibir el vector de entrada (x_1, x_2, \dots, x_n) , generará la salida:

$$y = w^{sc} + \sum_{j=1}^m w_{1,j}^s \tanh\left(w_j^{oc} + \sum_{i=1}^n w_{j,i}^o x_i\right)$$

donde,

w_j^{oc} : ponderador de la conexión con entrada de valor unitario y la neurona j de la capa oculta,

w^{sc} : ponderador de la conexión con entrada de valor unitario y la neurona de la capa de salida,

$w_{j,i}^o$: ponderador de la conexión entre la neurona i de la capa de entrada y la neurona j de la capa oculta,

$w_{1,j}^s$ ponderador de la conexión entre la neurona j de la capa oculta y la neurona 1 de la capa de salida.

Si, en vez de usar la función \tanh como función de transferencia de las neuronas ocultas, se hubiera empleado la función identidad estaríamos frente a una regresión lineal.

Es importante tener en cuenta que el número ponderadores de las conexiones de una Red Neuronal es típicamente mayor al número de parámetros de un modelo tradicional de series de tiempo (coeficientes de las variables) [10]. Por ejemplo, para una red con n_i neuronas en la capa de entrada, n_h neuronas en la capa oculta y una neurona en la capa de salida, el número de ponderadores es $p=(n_i+2)*n_h + 1$, lo que con $n_i=10$ y $n_h=2$, da un total de 25 parámetros. Este gran número de parámetros se traduce en un importante riesgo de sobreajuste de la red a los datos del conjunto de entrenamiento (overfitting) y repercute en una mayor carga computacional y una mayor duración del proceso de aprendizaje.

- Mecanismo de aprendizaje: En una Red Neuronal, el conocimiento se encuentra representado en los ponderadores de las conexiones entre neuronas y el aprendizaje de la red obedecerá a su modificación de acuerdo con la información que se le entregue a la red y a la aplicación de cierto algoritmo.

Existen distintos mecanismos de aprendizaje. A continuación se ofrecen dos clasificaciones de los mismos:

- De acuerdo a la presencia de supervisión:
 - Aprendizaje supervisado: se trata de un entrenamiento controlado, donde se compara la salida de la red con la salida deseada, procediéndose a la modificación de los ponderadores si es que existen diferencias significativas.
 - Aprendizaje por corrección de error: se ajustan los ponderadores de las conexiones en función de las diferencias entre los valores deseados y los obtenidos en la salida de la red.
 - Aprendizaje por refuerzo: a diferencia del mecanismo anterior, no se cuenta con una referencia exacta respecto de la salida deseada, sino que solamente con una señal de refuerzo que indica éxito o fracaso.
 - Aprendizaje estocástico: se realizan cambios aleatorios en los valores de los ponderadores de las conexiones, evaluando su efecto a partir del resultado deseado y de distribuciones de probabilidad.
 - Aprendizaje no supervisado: la red no recibe ninguna información respecto a si la salida generada es o no correcta.
 - Aprendizaje hebbiano: se ajustan los pesos de acuerdo con la correlación de los valores de las salidas de las neuronas conectadas.
 - Aprendizaje competitivo o cooperativo: se ajustan los pesos de tal modo que cuando la red reciba cierta información de entrada sólo algunas neuronas se activen.
- De acuerdo a la dinámica de actualización de los pesos:
 - Aprendizaje off line: en él se distingue la fase de aprendizaje o entrenamiento, en que se modifican los ponderadores haciendo uso de datos de entrenamiento, de la fase de operación o funcionamiento, donde los ponderadores quedan fijos.

- Aprendizaje on line: los ponderadores varían dinámicamente cada vez que se presenta nueva información al sistema.

El tiempo de aprendizaje de una red no se conoce a priori y, en vista del riesgo de memorización de los datos del conjunto de entrenamiento después de un entrenamiento suficientemente largo (overfitting), se hace necesario establecer condiciones para el término del aprendizaje.

En el aprendizaje por corrección de error, las condiciones de término del aprendizaje se clasifican en tres categorías:

- Límite admisible para el error, donde se establecen metas para indicadores críticos, como por ejemplo: máximo error cuadrático de entrenamiento de la red, máximo error cuadrático medio para el conjunto de datos de entrenamiento.
- Número máximo de épocas, donde una época corresponde a la presentación de todos los ejemplos del conjunto de entrenamiento.
- Desempeño de la red en un conjunto de testeo. Parte de los ejemplos disponibles se marginan del proceso de actualización de ponderadores de la red para conformar un conjunto de testeo. Durante el proceso de entrenamiento, a intervalos regulares, se evalúa el desempeño de la red en el conjunto de testeo para verificar el cumplimiento de la meta en algún indicador crítico, como por ejemplo máximo error cuadrático medio. Una vez alcanzado el valor meta se detiene el entrenamiento (early-stopping).

La selección de una cierta arquitectura y mecanismo de aprendizaje guarda relación con la aplicación que se esté desarrollando. Por ejemplo, para reconocimiento de imágenes se acostumbra utilizar Redes de Hopfield con aprendizaje hebbiano, mientras que para la generación de pronósticos se acostumbra utilizar Redes Multilayer Perceptron con aprendizaje por corrección de error. En este último caso, uno de los algoritmos más empleados es el backpropagation, donde los ponderadores de la red se actualizan siguiendo la dirección negativa del gradiente del error (método de descenso del gradiente)[19] (para detalles ver Anexo B).

El diseño de una red para resolver un problema con éxito puede ser una tarea muy compleja y larga dado la gran cantidad de decisiones de diseño que se deben tomar y la gran cantidad de parámetros que se deben definir. La enumeración completa de todas las alternativas no es práctica por requerir de un elevado número de evaluaciones, motivándose la aparición de variadas heurísticas y reglas basadas en la experiencia. Sin embargo, ninguna heurística ha mostrado la capacidad de entregar modelos con buen desempeño predictivo en cualquier conjunto de datos [7]. Temas como la determinación automática del número de capas o neuronas ocultas están actualmente bajo investigación, lo que hace que, en la práctica, el método más común para el diseño de Redes Neuronales sea el de “prueba y error” [19], cuya duración podría ser prolongada dado que no se debe descartar una red mientras ésta no haya completado su aprendizaje.

3.7. SISTEMAS INTELIGENTES HÍBRIDOS.

Hoy en día, la comunidad científica reconoce que los grandes desafíos de la Inteligencia Artificial no podrán ser resueltos exitosamente con aproximaciones metodológicas puras. Esto ha llevado al surgimiento de los Sistemas Inteligentes Híbridos que son el resultado de la asociación de metodologías de procesamiento que, sinérgicamente, proveen la base para la conceptualización y diseño de nuevos sistemas inteligentes [12].

Durante los últimos años los Sistemas Híbridos han alcanzado cierta madurez en muchos campos, lo que se revela en importantes desarrollos teóricos y exitosas aplicaciones en el mundo real en las áreas de: modelamiento, control, optimización, reconocimiento de patrones y pronósticos. Sin embargo, aún quedan desafíos por delante en estos tópicos y en otros como el de procesamiento semántico.

Los intereses actuales de investigación están enfocados en la integración de diferentes paradigmas computacionales: lógica difusa, computación neuronal, computación evolutiva, computación probabilística, aprendizajes de máquina, entre otros.

Algunos ejemplos de Modelos Híbridos son:

- Optimización de Redes Neuronales Artificiales usando técnicas de optimización global.
- Sistemas Híbridos involucrando Máquinas de Soporte Vectorial, Rough Sets, Redes Bayesianas, Razonamiento Probabilístico, etc..
- Computación Híbrida usando Redes Neuronales, Sistemas Difusos y Algoritmos Evolutivos.
- Híbridos de Computación Soft y Técnicas de Aprendizaje Estadístico.
- Técnicas de Optimización Híbrida (Algoritmos Evolutivos, Simulated Annealing, Tabu Search).

En el desarrollo de pronósticos de series de tiempo, existen antecedentes de buenos resultados mediante el empleo combinado de Modelos de Series de Tiempo (ARIMA) y Redes Neuronales [1], por lo que también se explorará su uso en el pronóstico de precio del cobre.

CAPÍTULO IV

GENERACIÓN DE PRONÓSTICOS Y EVALUACIÓN DE RESULTADOS

4.1. DATOS HISTÓRICOS Y DISEÑO GENERAL DEL EJERCICIO DE PRONÓSTICO.

Las proyecciones del precio del cobre refinado que se realizarán en este estudio sólo se basarán en los valores rezagados de la misma variable por las siguientes razones:

- Se desea establecer qué modelo es capaz de aprovechar mejor la información contenida en los precios históricos.
- La generación de proyecciones del precio empleando otras variables explicativas tiene el potencial problema de requerir el pronóstico de dichas variables, lo que complica su utilización práctica.

Los datos de precios serán sometidos a dos transformaciones para la construcción de modelos y la realización de pronósticos:

- Escalamiento de valores para expresar el precio en moneda constante: los precios históricos del cobre expresados en moneda corriente (moneda de cada periodo) reflejan dos efectos: la evolución del precio del cobre y el cambio del valor de la moneda. Dado que se desea construir modelos para el pronóstico del precio del cobre, se debe eliminar el efecto de la inflación mediante la aplicación de un deflactor que permitan expresar el precio en moneda de un periodo de referencia.

- Aplicación de logaritmo natural a los datos en moneda constante: con frecuencia, en los modelos econométricos se utiliza la transformación logarítmica de los datos para reducir su dispersión y atacar problemas de heterocedasticidad [20][21].

Los datos utilizados en esta tesis corresponden al precio anual del cobre refinado de la Bolsa de Metales de Londres durante el periodo comprendido entre los años 1913 y 2006³ (2006, promedio a junio). La elección de este periodo se fundamenta a continuación:

- Si bien existen datos para el precio cobre en los Estados Unidos durante el siglo XIX, dichos valores corresponden a un precio de productores, no a un precio de mercado y obedecen a un tipo de minería muy distinta a la gran minería que comenzó a desarrollarse a partir de las primeras décadas del siglo XX (flotación, minería a rajo abierto, mecanización). Por este motivo se excluyeron los datos previos a 1910.
- El deflactor empleado por la Comisión Chilena del Cobre (COCHILCO) para expresar el precio del metal rojo en moneda constante es el Índice de Precios de Productores de los Estados Unidos (PPI All Commodities, Not Seasonally Adjusted) que registra valores desde 1913 en adelante [32].
- En el apartado “¿Es pronosticable el precio del cobre?” se describió al precio como una variable que presenta reversión a la media, en particular, a distintas medias a través del tiempo. Este comportamiento de mediano/largo plazo se observa en la ventana seleccionada, donde el logaritmo natural del precio no presenta tendencia y los test practicados rechazan la presencia de raíz unitaria.

Los conjuntos iniciales de entrenamiento y de evaluación que se utilizaron en la construcción de modelos y en la generación de pronósticos son los siguientes:

Tabla N°3: Conjuntos de Entrenamiento y Evaluación.

Conjunto de Entrenamiento		Conjunto de Evaluación	
Datos	Número de datos	Datos	Número de datos
1913 - 1976	64	1977 - 2006	30

³ Precio del Cobre: 1913-1934, de acuerdo a informe de CRU International Ltd; 1935-2006, Cochilco [5].

Estos conjuntos se denominan como iniciales ya que, con la aplicación de recalibración, los conjuntos de entrenamiento y evaluación se redefinirán periódicamente.

Se optó por un periodo de evaluación de 30 años, similar al empleado en otros trabajos [9][25] que permitió contar con un número de pronósticos de entre 30 (proyecciones a un año) y 25 (proyecciones a 6 años), abarcando tanto periodos con precios ascendentes como descendentes.

Se trabajó con pronósticos de origen móvil, incorporando la nueva información al modelamiento gracias al uso de un conjunto de entrenamiento de tamaño creciente y la aplicación de recalibración. No se eliminaron las observaciones más antiguas del conjunto de entrenamiento dado que, a partir del comportamiento histórico, se consideró que podían ser de utilidad [25].

4.2. MODELAMIENTO, PRONÓSTICO Y RESULTADOS.

En esta sección se detallarán los modelos generados con las distintas metodologías de pronóstico empleadas y sus resultados.

4.2.1. MODELOS DE SERIES DE TIEMPO.

El enfoque de Box - Jenkins es una de las metodologías de uso más amplio para el modelamiento estocástico de series de tiempo. Es popular debido a su generalidad, ya que puede manejar cualquier serie, estacionaria o no estacionaria, y por haber sido implementado en numerosos programas computacionales [21].

Los pasos básicos de la metodología de Box-Jenkins son [21]:

1. Verificar la estacionariedad de la serie. Si ésta no es estacionaria, diferenciarla hasta alcanzar estacionariedad.
2. Identificar un modelo tentativo.
3. Estimar el modelo.
4. Verificar el diagnóstico (si este no es adecuado, volver al paso 2).

5. Usar el modelo para pronosticar.

Una serie puede ser evidentemente no estacionaria, presentando, por ejemplo, una tendencia definida. En el caso del precio del cobre, la respuesta dependerá de la ventana de tiempo bajo estudio [33]. En el periodo comprendido entre los años 1913 y 2006 (promedio a junio), el precio del cobre no presenta una tendencia estadísticamente significativa y el test de Dickey-Fuller Aumentado (ADF) aplicado a la serie en logaritmos permite rechazar la hipótesis nula de presencia de raíz unitaria (ver Anexo A). Esto es consistente con la lógica económica del mercado físico del cobre y con los resultados conseguidos en otros estudios [9][26][33].

Sin embargo, si se examina una serie más corta, el precio puede no evidenciar reversión a la media o los tests podrían no ser suficientemente potentes para rechazar la hipótesis nula de presencia de raíz unitaria. El periodo comprendido entre 1913 y 1976, donde el test ADF no rechaza la hipótesis nula, es un ejemplo de esta última situación. Por lo anterior, también se estimaron modelos a partir de datos en primera diferencia, dado que en dicha serie el test ADF rechaza la presencia de raíz unitaria en todas las submuestras de interés.

La función de autocorrelación parcial de la serie en niveles revela la presencia de autocorrelación de primer orden, por lo que el modelo AR(1) es uno de los modelos tentativos para la generación de pronósticos (ver Anexo A). A pesar de esto, el análisis de los residuos reveló la existencia de autocorrelación, motivando la búsqueda de alternativas. Evaluaciones preliminares que permitieron descartar modelos de orden superior, y los resultados conseguidos en otros estudios [9][25], llevaron a la construcción de un set de modelos tentativos cuyos coeficientes son significativos para los datos en niveles en el periodo de interés⁴:

⁴ Para los modelos de datos en primeras diferencias sólo se obtienen coeficientes significativos con los modelos AR2, MA1 y MA2.

Tabla N°4: Modelos ARIMA Utilizados en el Ejercicio de Proyección.

AR1	AR1MA2
AR2	AR2MA1
AR3	AR3MA1
AR4	MA1MA2
MA1	AR3MA1MA2
MA2	AR1AR2AR3
MA3	AR1AR2AR3MA1MA2MA3
MA4	MA1MA2MA3MA4
AR1MA1	MA1MA2MA3MA4MA5

Adicionalmente, y sólo con fines comparativos, se modeló el precio del cobre como un camino aleatorio con y sin drift.

Por último, en la línea de los modelos deterministas, se modeló el precio del cobre como un promedio móvil de 3, 6 y 9 años. Es interesante incluir el promedio móvil, dado que, hasta hace algunos años, fue el método utilizado para determinar el precio de referencia del Fondo de Compensación del Cobre (promedio móvil de 6 periodos) [2][9].

Para la construcción de modelos estocásticos se utilizó el software EViews 4.1 (Quantitative Micro Software). Las rutinas programadas en EViews para la generación y evaluación de proyecciones fuera de muestra sólo necesitaron un par de segundos para ejecutar los cálculos correspondientes a cada modelo, y la evaluación del set completo de modelos tentativos fue cosa de pocos minutos (Anexo D). Para el resto de los modelos se empleó la planilla de cálculo Microsoft Excel 2002 (Microsoft Corporation).

Las evaluaciones realizadas mostraron que los mejores resultados se conseguían recalibrando el modelo cada vez que un nuevo dato histórico se incorporaba al conjunto de entrenamiento.

La Tabla N°5 muestra el Promedio del Error Porcentual Absoluto (MAPE) conseguido en las proyecciones que abarcan el periodo 1977 – 2006 (2006, promedio a Junio). Se incluyen en dicha tabla modelos con volatilidad variable (GARCH) contruidos sobre la base de los modelos de varianza constante con mejor desempeño en proyecciones a 1 y 2 años (AR1 y AR2MA1).

Junto al valor medio del error de pronóstico, la dispersión de los errores también es relevante en la selección de un modelo de predictivo. Los valores de desviación estándar del Error Porcentual Absoluto (APE) y de la Raíz Cuadrática Media del Error (RMSE) se entregan en las Tablas N°6 y N°7.

Tabla N°5: Promedio del Error Porcentual Absoluto (MAPE) %.

	a 1 año	a 2 años	a 3 años	a 4 años	a 5 años	a 6 años
AR1	15,036	23,596	28,579	30,926	33,244	34,592
AR2	23,123	24,102	30,427	31,133	33,468	34,000
AR2_dif*	16,952	25,802	28,001	31,616	35,702	38,100
AR3	27,929	27,639	28,468	32,270	33,379	33,939
AR4	30,138	30,182	29,745	30,237	33,091	32,983
MA1	18,751	30,604	31,605	32,369	33,212	33,817
MA1_dif*	16,005	24,700	28,675	32,694	37,405	40,822
MA2	24,100	24,932	31,718	32,480	33,291	33,897
MA2_dif*	15,780	24,925	28,407	32,223	36,149	38,518
MA3	28,522	28,665	29,569	32,440	33,319	33,909
MA4	34,354	34,692	34,335	35,022	33,272	33,783
AR1MA1	14,822	23,978	28,576	30,746	32,549	33,520
AR1MA2	15,756	25,212	28,025	30,806	34,043	35,268
AR2MA1	14,078	23,738	28,320	30,545	32,999	34,354
AR3MA1	19,265	27,724	28,659	30,474	32,250	32,852
MA1MA2	18,132	27,790	31,027	31,774	32,576	33,230
AR3MA1MA2	16,233	24,378	28,116	29,589	31,235	32,323
AR1AR2AR3	16,026	25,290	28,107	30,237	33,148	34,686
MA1MA2MA3MA4	18,581	25,995	31,198	32,652	32,083	32,800
MA1MA2MA3MA4MA5	18,651	27,479	29,197	30,092	31,827	32,574
AR1AR2AR3MA1MA2MA3	17,491	28,301	33,233	35,323	37,962	36,116
AR2MA1.GARCH(0,2)	14,347	23,059	27,020	28,867	31,083	31,067
AR2MA1.GARCH(1,4)	14,891	24,713	28,866	31,763	33,402	34,273
AR1.GARCH(0,2)	14,878	22,979	27,269	29,348	31,859	33,117
Caminata Aleatoria	15,264	23,823	29,201	33,295	37,371	40,135
Caminata Aleatoria c/drift	15,287	23,603	28,575	32,869	36,558	38,832
Promedio Móvil 3 años	21,518	26,101	29,493	33,909	38,047	39,693
Promedio Móvil 6 años	27,248	30,232	32,483	35,991	39,361	41,072
Promedio Móvil 9 años	32,035	34,100	35,806	38,853	41,758	42,670

*: Corresponden a modelos construidos sobre la base de datos en primera diferencia.
Se destacan los mejores resultados conseguidos.

Tabla N°6: Desvío Estándar del Error Porcentual Absoluto %.

	a 1 año	a 2 años	a 3 años	a 4 años	a 5 años	a 6 años
AR1	11,423	13,870	16,739	19,800	19,837	20,447
AR2	13,597	13,450	19,189	19,436	20,675	21,055
AR2_dif*	13,500	15,936	19,527	20,824	19,619	21,016
AR3	16,652	16,681	16,673	20,384	20,223	20,539
AR4	18,331	18,733	18,936	19,131	20,601	21,389
MA1	14,014	22,050	22,391	22,887	23,148	23,610
MA1_dif*	10,995	15,332	22,797	26,120	25,413	25,293
MA2	15,112	15,227	22,407	22,901	23,268	23,666
MA2_dif*	13,422	15,975	18,332	20,136	19,950	22,057
MA3	18,762	19,344	19,341	22,964	23,238	23,602
MA4	22,075	22,815	23,094	23,396	23,183	23,594
AR1MA1	12,288	13,659	16,803	20,219	20,879	22,190
AR1MA2	12,623	14,157	17,586	19,649	18,695	20,326
AR2MA1	11,292	13,625	15,926	20,584	20,394	20,608
AR3MA1	12,630	16,850	16,973	20,964	21,186	21,600
MA1MA2	14,944	18,349	21,959	22,436	22,783	23,213
AR3MA1MA2	12,864	13,875	16,476	19,589	19,957	20,620
AR1AR2AR3	13,291	14,577	17,051	19,389	18,860	20,091
MA1MA2MA3MA4	13,742	17,211	19,948	23,144	22,483	22,993
MA1MA2MA3MA4MA5	14,763	17,406	18,961	21,684	22,241	23,064
AR1AR2AR3MA1MA2MA3	14,369	15,968	19,082	23,907	23,095	21,704
AR2MA1.GARCH(0,2)	11,441	13,545	15,152	18,450	17,961	18,854
AR2MA1.GARCH(1,4)	11,391	13,236	16,033	19,512	20,058	20,430
AR1.GARCH(0,2)	11,088	13,462	16,647	18,825	19,383	20,836
Caminata Aleatoria	11,613	16,155	21,090	24,054	22,893	23,296
Caminata Aleatoria c/drift	11,530	16,302	21,567	23,800	22,606	23,232
Promedio Móvil 3 años	14,755	17,302	20,443	21,407	19,858	21,630
Promedio Móvil 6 años	16,202	17,538	18,625	19,229	20,684	24,725
Promedio Móvil 9 años	19,429	20,438	21,166	23,037	25,969	28,920

*: Corresponden a modelos construidos sobre la base de datos en primera diferencia.
Se destacan los mejores resultados conseguidos.

Sumando y restando un desvío estándar al Promedio del Error Porcentual Absoluto (MAPE) se puede definir un intervalo donde se esperaría encontrar buena parte o la mayoría de los errores de pronóstico. Los gráficos N°4 y N°5 muestran dichos intervalos para los distintos modelos en pronósticos a 1 y 6 años.

Gráfico N°4: Pronósticos a 1 Año, Modelos de Series de Tiempo.
MAPE Más Menos un Desvío Estándar del Error Porcentual Absoluto %.

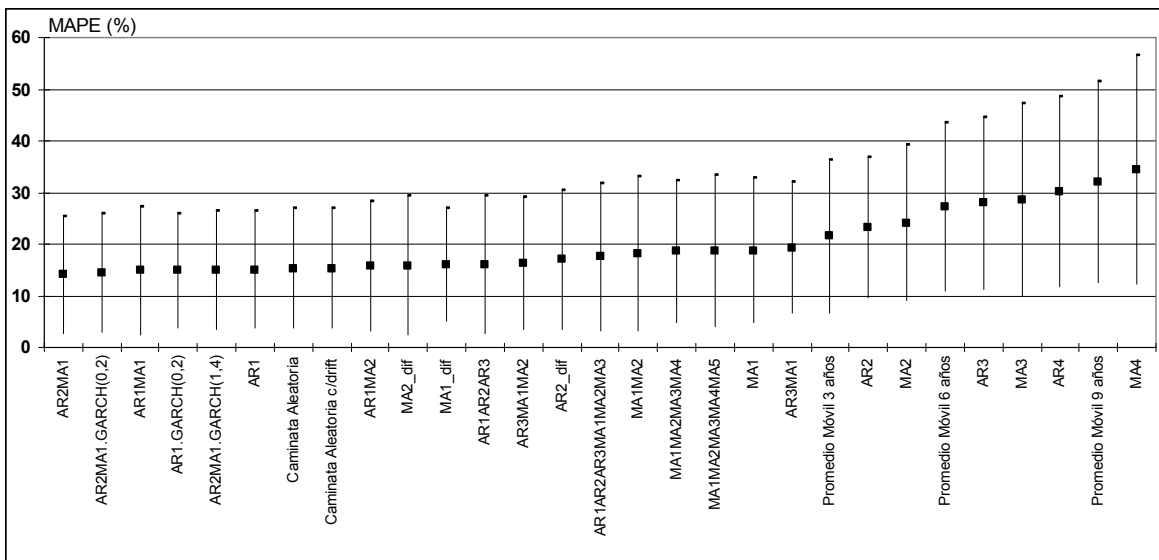


Gráfico N°5: Pronósticos a 6 Años, Modelos de Series de Tiempo..
MAPE Más Menos un Desvío Estándar del Error Porcentual Absoluto %.

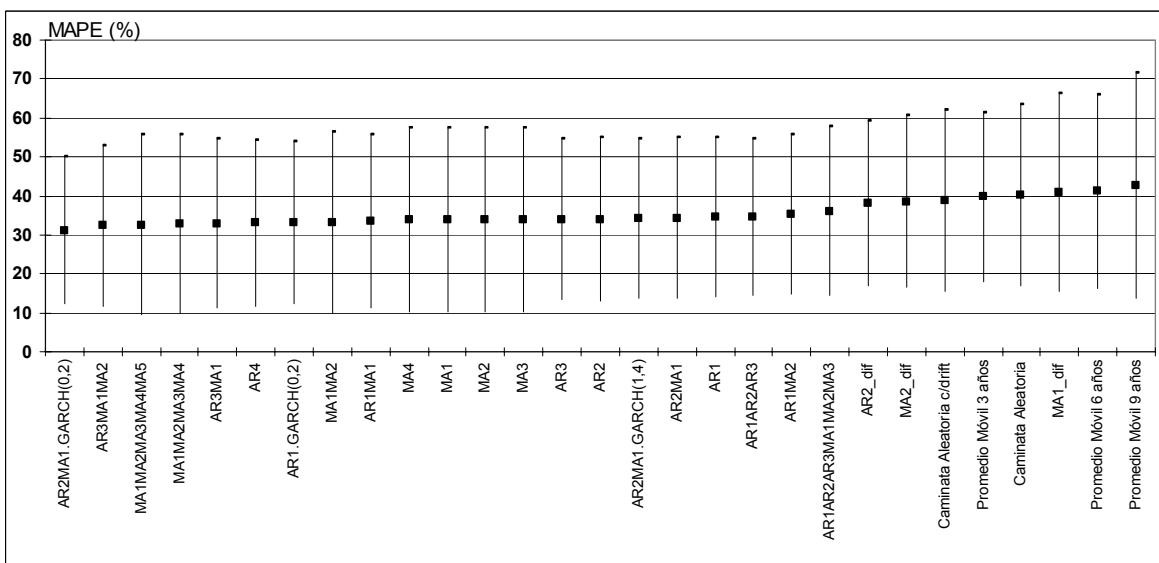


Tabla N°7: Raíz Cuadrática Media del Error (RMSE).

	a 1 año	a 2 años	a 3 años	a 4 años	a 5 años	a 6 años
AR1	30,608	42,310	49,488	51,385	51,615	50,957
AR2	40,857	41,918	49,509	50,029	49,145	49,634
AR2_dif*	35,647	49,297	55,304	58,515	61,186	61,589
AR3	48,210	47,771	48,730	48,582	49,318	49,824
AR4	50,073	50,230	49,626	49,946	48,770	48,753
MA1	33,074	44,672	45,707	46,344	47,025	47,652
MA1_dif*	30,671	45,429	57,235	62,190	65,243	65,729
MA2	37,453	<u>38,312</u>	45,817	46,456	47,115	47,723
MA2_dif*	34,704	47,960	54,088	57,475	59,963	60,708
MA3	46,792	46,954	47,840	46,453	47,176	47,775
MA4	54,375	54,948	54,286	54,973	47,129	47,621
AR1MA1	<u>30,009</u>	40,138	46,759	48,176	48,756	48,479
AR1MA2	33,196	44,931	50,180	52,235	53,318	53,028
AR2MA1	30,193	42,089	48,996	50,967	51,384	51,384
AR3MA1	35,939	45,625	46,814	47,571	47,340	47,903
MA1MA2	33,364	41,719	<u>45,284</u>	<u>45,860</u>	46,589	47,194
AR3MA1MA2	31,491	41,594	47,012	47,661	47,952	47,675
AR1AR2AR3	33,034	44,334	49,223	50,960	52,201	51,561
MA1MA2MA3MA4	33,358	41,477	48,265	48,167	<u>46,161</u>	46,850
MA1MA2MA3MA4MA5	35,158	46,211	47,608	47,182	47,426	<u>46,826</u>
AR1AR2AR3MA1MA2MA3	34,536	46,932	53,411	55,984	57,162	52,556
AR2MA1.GARCH(0,2)	30,534	41,858	47,539	48,881	49,211	47,528
AR2MA1.GARCH(1,4)	30,568	42,653	48,633	50,494	50,494	50,716
AR1.GARCH(0,2)	30,564	41,655	48,323	49,841	50,196	50,204
Caminata Aleatoria	30,981	45,454	55,873	60,156	62,128	62,466
Caminata Aleatoria c/drift	31,103	45,816	56,471	60,830	62,803	63,182
Promedio Móvil 3 años	41,942	49,516	55,433	58,449	60,364	60,850
Promedio Móvil 6 años	48,783	53,052	55,685	57,944	59,775	60,311
Promedio Móvil 9 años	53,694	55,882	56,958	59,481	62,125	62,448

*: Corresponden a modelos construidos sobre la base de datos en primera diferencia.
Se destacan los mejores resultados conseguidos.

En términos del MAPE, el mejor resultado conseguido en proyecciones a un año lo entrega el modelo AR2MA1 con un error de 14,078%, cifra que, en términos relativos, es un 7,8% inferior al Camino Aleatorio Sin Drift. A medida que el pronóstico es a más años plazo, el error se incrementa importantemente, llegando a 30,067% en el mejor modelo, AR2MA1.GARCH(0,2), con una reducción de 22,6% respecto al Camino Aleatorio.

La revisión del Desvío Estándar del Error Porcentual Absoluto revela que, en general, los modelos con menores MAPE generan errores de proyección con baja dispersión. Es destacable que en el caso de los pronósticos a 3 y más años, el modelo con menor MAPE también genera errores con menor dispersión (AR2MA1.GARCH(0,2)).

En cuanto al RMSE, el mejor resultado en proyecciones a un año se consigue con el modelo AR1MA1 con un error de 30,009 que significa una reducción de 3,1% respecto al Camino Aleatorio Sin Drift. El RMSE aumenta en los pronósticos a más años plazo llegando a 46,826 en el mejor modelo, MA1MA2MA3MA4MA5, con una reducción de 25% respecto al Camino Aleatorio.

Se observa que, con la excepción de las proyecciones a 4 y 5 años evaluadas con RMSE, los mejores resultados se consiguen con formulaciones de orden igual o menor a 2.

4.2.2. REDES NEURONALES.

Como ya se ha señalado, el diseño de una Red Neuronal es una tarea compleja por la gran cantidad de decisiones que involucra. Por los buenos resultados conseguidos en numerosas aplicaciones [19], en este estudio se emplearon redes MLP con una sola capa oculta, una sola neurona en la capa de salida y conexiones hacia delante.

Cada una de las neuronas utilizadas tiene las siguientes características:

- Regla de propagación y función de activación: suma ponderada de inputs.
- Función de transferencia de las neuronas de la capa oculta: tangente hiperbólica.
- Función de transferencia de la neurona de la capa de salida: identidad.

El mecanismo de aprendizaje aplicado fue off-line y por corrección de error, siendo el método de aprendizaje elegido el de Levenberg-Marquardt (ver Anexo C). Dicha elección se fundamentó en la mayor velocidad de convergencia de este último método en comparación con el método de descenso del gradiente (ver Anexo B). Adicionalmente, evaluaciones preliminares mostraron la obtención de un menor error de pronóstico. Para poner fin al aprendizaje se utilizó early-stopping, estrategia con la que se hizo frente al peligro de sobreajuste.

Para el resto de los parámetros de diseño se siguió una estrategia de evaluación de distintos de valores, mediante el uso de grillas [7].

La aplicación de early-stopping hizo necesario definir un conjunto de testeo, seleccionándose para tal efecto una fracción variable de los últimos ejemplos disponibles dentro de la muestra: 10%, 20% o 30% [20], que no participó, directamente, en el entrenamiento de las Redes Neuronales.

Respecto al número de neuronas en la capa de entrada, se construyeron y entrenaron redes con 3, 6 o 9 neuronas en la capa de entrada, alimentadas con los correspondientes rezagos del precio, mientras que para la capa oculta se utilizaron entre 1 y 6 neuronas. Las combinaciones evaluadas fueron determinadas por el número de conexiones asociadas y el número de ejemplos del conjunto de entrenamiento. Si bien sólo existen teoremas que relacionan el número de conexiones de la red con el número de ejemplos de entrenamiento para ciertos tipos de redes (multicapas, con entradas y salidas binarias y función de activación escalón) [29], al menos debe respetarse que el número de ejemplos supere, con alguna holgura, el número de coeficientes a estimar (ponderadores) para así evitar el overfitting. Las combinaciones presentadas en la tabla N°8 cumplen con este requisito, ya que, en promedio, el número de ejemplos de entrenamiento quintuplica el número de coeficientes [20] (el número de ejemplos de entrenamiento depende del tamaño de la muestra, de la fracción de datos muestrales destinada al conjunto de testeo y del número de neuronas de la capa de entrada, moviéndose en un rango de entre 55 y 90 ejemplos).

Tabla N°8: Neuronas en la Capa de Entrada, en la Capa Oculta y Número de Conexiones.

Neuronas en la capa de entrada	Neuronas en la capa oculta	Número de conexiones
3	1	6
3	2	11
3	3	16
3	4	21
3	5	26
3	6	31
6	1	9
6	2	17
6	3	25
6	4	33
9	1	12
9	2	23
9	3	34

Nota: El número de conexiones considera las conexiones entre capas (incluyendo la capa de salida) y las conexiones correspondientes a las entradas de valor unitario.

Dado que los ponderadores iniciales de las conexiones determinan el punto de origen del proceso de aprendizaje (minimización de error del modelo), se recomienda repetir el entrenamiento de cada arquitectura unas 15 veces, redefiniendo aleatoriamente dichos pesos [7][30], obteniéndose así 15 redes distintas para una misma arquitectura. Si a esto agregamos que, como ya fue mencionado, cada arquitectura se entrenó utilizando conjuntos de testeo definidos de 3 maneras distintas, el total de redes construidas para una misma combinación de neuronas en la capa de entrada y neuronas en la capa oculta llega a 45 y el total de redes a entrenadas y evaluadas llega a 585.

Al igual que los Modelos de Series de Tiempo, las Redes Neuronales se recalibraron para incorporar la información adicional disponible a medida que se realizaban las evaluaciones. Sin embargo, a diferencia del caso anterior, no se realizaron 30 recalibraciones para cada red debido a que, en evaluaciones preliminares, se observó que un mayor número de recalibraciones no siempre se traducían en un mejor pronóstico. Por este motivo, cada red se recalibró 3, 6 y 10 veces.

Para entrenar y evaluar las distintas redes, se empleó el lenguaje de programación MATLAB y su toolbox de Redes Neuronales (MATLAB 6.1, The MathWorks Inc.) (Anexo E). Si bien el entrenamiento y evaluación de cada red tomó pocos segundos, el gran número de redes estimadas requirió de tiempos de cálculo de hasta 6 horas (Pentium IV, 3.0 GHz, 1 GB RAM).

Las siguientes tablas y gráficos muestran el desempeño predictivo de los mejores modelos de pronóstico para los distintos alcances y frecuencias de recalibración considerados. Los modelos más exitosos fueron seleccionados de acuerdo al menor MAPE en el conjunto de evaluación⁵. La notación empleada para identificar las redes es $\text{Red}(a_1; a_2; a_3; a_4; a_5)$, donde:

a_1 : Número de neuronas de la capa de entrada.

a_2 : Número de neuronas de la capa oculta.

a_3 : Razón entre el número de ejemplos efectivamente utilizados en el entrenamiento y el número de ejemplos empleados en el testeo (early-stopping).

a_4 : Número correlativo de la inicialización de la red.

a_5 : Número de recalibraciones efectuadas.

⁵ Para valores de MAPE en los conjuntos de entrenamiento y testeo ver Anexo F.

Tabla N°9: Promedio del Error Porcentual Absoluto (MAPE) %.

	a 1 año	a 2 años	a 3 años	a 4 años	a 5 años	a 6 años
Red(3; 5; 80/20; 14; 3)	13,481	21,233	25,069	30,137	35,656	41,214
Red(6; 1; 70/30; 3; 3)	15,409	19,483	22,631	23,272	26,111	28,238
Red(6; 2; 70/30; 10; 3)	16,025	21,630	23,068	21,989	24,360	28,525
Red(6; 2; 80/20; 9; 3)	16,151	21,240	23,442	23,317	22,653	23,487
Red(6; 3; 70/30; 13; 3)	14,917	19,811	22,299	24,035	25,987	28,394
Red(9; 1; 80/20; 8; 3)	17,848	27,426	29,541	28,887	26,050	22,838
Red(6; 2; 80/20; 9; 6)	14,785	18,211	22,194	25,234	27,015	29,482
Red(6; 3; 70/30; 9; 6)	25,292	30,488	33,508	28,311	26,308	22,668
Red(6; 3; 80/20; 2; 6)	14,184	19,247	25,107	28,651	32,708	37,120
Red(6; 3; 80/20; 15; 6)	16,300	22,854	26,543	24,930	22,641	23,611
Red(6; 4; 70/30; 12; 6)	15,431	19,912	20,467	25,601	31,221	34,805
Red(9; 2; 70/30; 15; 6)	17,499	21,330	21,872	22,604	26,371	27,548
Red(6; 2; 70/30; 4; 10)	16,490	22,152	20,255	25,847	29,443	32,472
Red(6; 2; 70/30; 7; 10)	17,180	20,795	22,696	21,801	24,449	27,099
Red(6; 2; 70/30; 10; 10)	24,992	28,753	32,711	30,112	27,001	23,257
Red(6; 2; 70/30; 13; 10)	16,923	21,432	25,693	24,149	23,822	23,471
Red(6; 3; 70/30; 8; 10)	14,391	16,130	22,522	29,002	35,847	45,193

Se destacan los mejores resultados conseguidos.

Tabla N°10: Desvío Estándar del Error Porcentual Absoluto %.

	a 1 año	a 2 años	a 3 años	a 4 años	a 5 años	a 6 años
Red(3; 5; 80/20; 14; 3)	11,337	14,181	18,906	22,863	22,433	28,413
Red(6; 1; 70/30; 3; 3)	12,003	13,452	<u>13,220</u>	16,851	16,397	17,789
Red(6; 2; 70/30; 10; 3)	12,641	16,042	14,643	16,661	17,481	19,235
Red(6; 2; 80/20; 9; 3)	12,459	15,022	13,951	<u>14,087</u>	<u>16,272</u>	15,599
Red(6; 3; 70/30; 13; 3)	11,760	12,779	13,676	16,417	16,754	18,012
Red(9; 1; 80/20; 8; 3)	<u>10,423</u>	27,748	25,556	25,099	18,180	<u>10,166</u>
Red(6; 2; 80/20; 9; 6)	10,819	<u>12,049</u>	13,952	16,922	16,864	18,533
Red(6; 3; 70/30; 9; 6)	17,702	16,619	15,599	17,132	17,117	17,279
Red(6; 3; 80/20; 2; 6)	12,528	14,555	14,961	19,238	22,134	23,516
Red(6; 3; 80/20; 15; 6)	11,413	13,748	15,249	19,986	19,091	20,713
Red(6; 4; 70/30; 12; 6)	12,941	13,341	16,730	21,304	22,480	24,469
Red(9; 2; 70/30; 15; 6)	11,875	16,433	16,561	15,963	16,519	14,980
Red(6; 2; 70/30; 4; 10)	11,731	13,690	15,571	17,601	18,224	22,949
Red(6; 2; 70/30; 7; 10)	13,270	14,479	17,199	19,515	17,515	20,444
Red(6; 2; 70/30; 10; 10)	21,823	20,130	23,421	20,875	19,431	19,433
Red(6; 2; 70/30; 13; 10)	12,078	17,021	21,445	17,000	16,568	17,738
Red(6; 3; 70/30; 8; 10)	11,835	12,383	19,905	25,287	21,867	29,124

Se destacan los mejores resultados conseguidos.

Los gráficos N°6 y N°7 muestran los intervalos de MAPE más menos un Desvío Estándar del Error porcentual absoluto para los pronósticos a 1 y 6 años.

Gráfico N°6: Pronósticos a 1 año, Redes Neuronales.
MAPE Más Menos un Desvío Estándar del Error Porcentual Absoluto %.

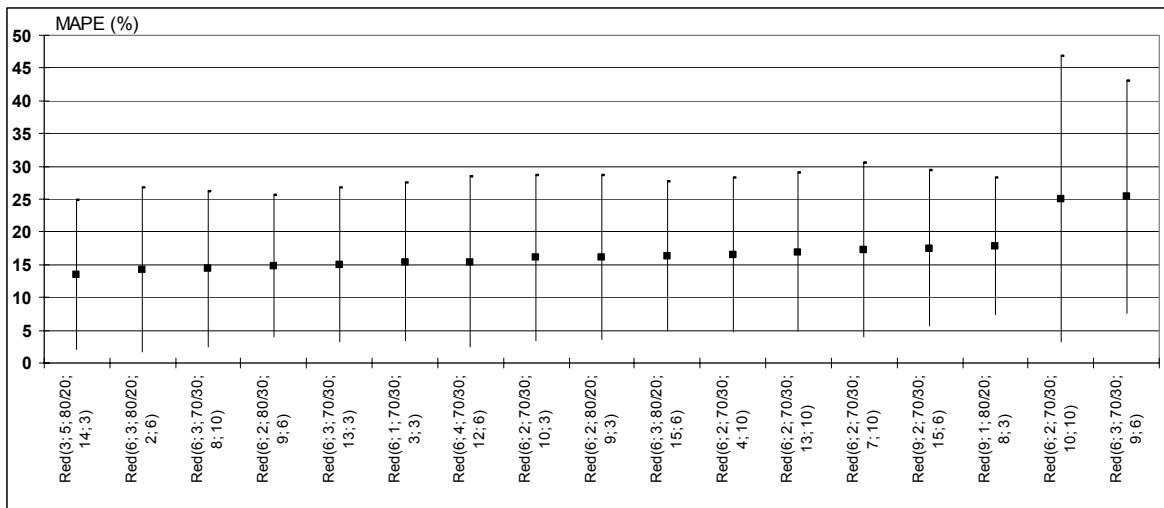


Gráfico N°7: Pronósticos a 6 años, Redes Neuronales.
MAPE Más Menos un Desvío Estándar del Error Porcentual Absoluto %.

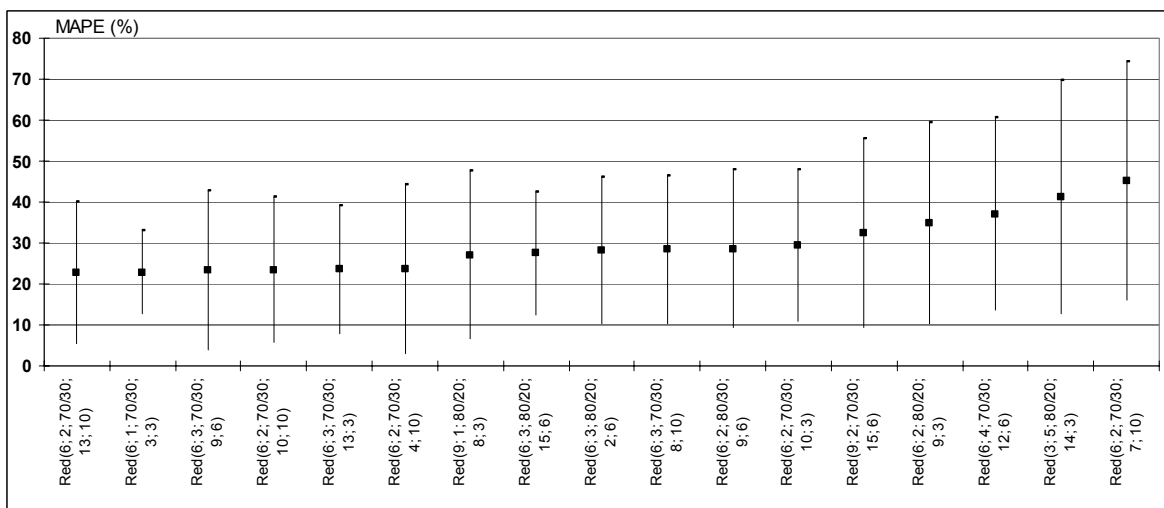


Tabla N°11: Raíz Cuadrática Media del Error (RMSE).

	a 1 año	a 2 años	a 3 años	a 4 años	a 5 años	a 6 años
Red(3; 5; 80/20; 14; 3)	<u>29,752</u>	40,586	48,053	52,435	56,350	61,353
Red(6; 1; 70/30; 3; 3)	30,714	35,941	41,327	<u>40,959</u>	41,518	44,367
Red(6; 2; 70/30; 10; 3)	32,534	40,451	45,574	43,421	44,373	47,105
Red(6; 2; 80/20; 9; 3)	34,634	43,974	48,345	46,002	44,122	44,518
Red(6; 3; 70/30; 13; 3)	31,271	37,645	43,490	45,220	45,348	46,786
Red(9; 1; 80/20; 8; 3)	30,151	45,642	46,741	45,457	40,857	<u>36,687</u>
Red(6; 2; 80/20; 9; 6)	30,982	38,287	42,824	46,285	48,741	50,968
Red(6; 3; 70/30; 9; 6)	41,070	50,151	53,826	49,268	45,163	41,424
Red(6; 3; 80/20; 2; 6)	31,499	38,794	46,651	50,066	52,864	54,775
Red(6; 3; 80/20; 15; 6)	33,641	45,735	51,634	50,618	47,480	46,259
Red(6; 4; 70/30; 12; 6)	32,848	39,301	42,184	47,791	53,690	59,090
Red(9; 2; 70/30; 15; 6)	35,868	46,239	46,828	45,039	48,861	51,436
Red(6; 2; 70/30; 4; 10)	30,573	39,784	<u>41,185</u>	47,454	48,533	53,422
Red(6; 2; 70/30; 7; 10)	34,290	41,048	44,471	41,430	<u>40,805</u>	42,937
Red(6; 2; 70/30; 10; 10)	42,914	46,357	52,481	46,963	43,249	40,948
Red(6; 2; 70/30; 13; 10)	33,547	42,635	50,722	45,864	44,214	43,810
Red(6; 3; 70/30; 8; 10)	31,608	<u>32,012</u>	44,646	51,611	57,222	67,222

Se destacan los mejores resultados conseguidos.

En las proyecciones a un año, evaluando ya sea con el MAPE o con el RMSE, los mejores resultados, entre los modelos seleccionados, se consiguieron con la misma red, Red(3; 5; 80/20; 14; 3), con sólo tres neuronas en la capa de entrada. El MAPE de los pronósticos de dicha red, 13,481%, fue un 11,7% inferior al de la Caminata Aleatoria, mientras que su RMSE, 29,752, fue un 4% menor al del mismo referente.

Al igual que con los Modelos de Series de Tiempo, a medida que el pronóstico es a más años plazo, el error aumenta de forma importante. Así en pronósticos a 6 años plazo, la red con menor MAPE, Red(6; 3; 70/30; 9; 6), registró un error de 22,668% y la red con menor RMSE, Red(9; 1; 80/20; 8; 3), registró un error de 36,687, con reducciones de 43,5% y 41,3% respecto al Camino Aleatorio.

En las proyecciones a 2 y más años, los menores errores se consiguieron con redes dotadas de 6 neuronas en la capa de entrada y 2 o 3 neuronas en la capa oculta. Así, con un modelamiento no lineal, los últimos 6 rezagos del precio aportan información útil para la generación de pronósticos, lo que contrasta con lo observado en los Modelos ARIMA, donde modelos de ese orden fueron descartados por tener coeficientes no significativos.

En este punto, resulta interesante probar cómo cambian los resultados si se alimentan las redes con “variables adicionales” derivadas de los precios rezagados. Para tal efecto, se crearon las variables “última variación del precio” y “desviación estándar del precio en los últimos periodos” y se modelaron nuevas redes que incorporaban, además de los rezagos del precio, una de estas dos variables.

Definición de variables adicionales:

- Última variación del precio (uvp): diferencia entre los dos últimos rezagos del precio en logaritmo natural.
- Desviación estándar del precio en los últimos periodos (dep): desviación estándar de los precios rezagados que alimentan la capa de entrada.

En las Tablas N°12 y N°13, se resumen los resultados de los mejores modelos, seleccionados en términos de MAPE, antes y después de la inclusión de las variables adicionales⁶.

Tabla N°12: Promedio del Error Porcentual Absoluto (MAPE) %.

	a 1 año	a 2 años	a 3 años	a 4 años	a 5 años	a 6 años
Red(3; 5; 80/20; 14; 3)	13,481	21,233	25,069	30,137	35,656	41,214
Red(6; 3; 70/30; 9; 6)	25,292	30,488	33,508	28,311	26,308	22,668
Red(6; 3; 80/20; 15; 6)	16,300	22,854	26,543	24,930	22,641	23,611
Red(6; 2; 70/30; 4; 10)	16,490	22,152	20,255	25,847	29,443	32,472
Red(6; 2; 70/30; 7; 10)	17,180	20,795	22,696	21,801	24,449	27,099
Red(6; 3; 70/30; 8; 10)	14,391	16,130	22,522	29,002	35,847	45,193
Red(3; 1; 80/20; 13; 3)(dep)	17,824	22,897	22,694	22,443	21,538	22,709
Red(3; 2; 80/20; 4; 3)(dep)	16,076	16,373	18,657	20,993	22,827	26,759
Red(3; 6; 80/20; 11; 3)(dep)	16,677	29,171	24,934	20,627	27,685	24,692
Red(6; 3; 70/30; 3; 6)(uvp)	13,928	20,566	26,109	30,678	31,561	35,679
Red(6; 3; 70/30; 10; 6)(uvp)	17,709	21,181	24,722	25,929	23,608	20,329

Se destacan los mejores resultados conseguidos.

Tabla N°13: Raíz Cuadrática Media del Error (RMSE).

	a 1 año	a 2 años	a 3 años	a 4 años	a 5 años	a 6 años
Red(3; 5; 80/20; 14; 3)	29,752	40,586	48,053	52,435	56,350	61,353
Red(6; 3; 70/30; 9; 6)	41,070	50,151	53,826	49,268	45,163	41,424
Red(6; 3; 80/20; 15; 6)	33,641	45,735	51,634	50,618	47,480	46,259
Red(6; 2; 70/30; 4; 10)	30,573	39,784	41,185	47,454	48,533	53,422
Red(6; 2; 70/30; 7; 10)	34,290	41,048	44,471	41,430	40,805	42,937
Red(6; 3; 70/30; 8; 10)	31,608	32,012	44,646	51,611	57,222	67,222
Red(3; 1; 80/20; 13; 3)(dep)	37,424	55,371	56,049	55,222	54,362	55,551
Red(3; 2; 80/20; 4; 3)(dep)	35,245	45,458	48,339	50,130	51,379	54,315
Red(3; 6; 80/20; 11; 3)(dep)	34,851	68,056	45,009	48,341	70,488	41,858
Red(6; 3; 70/30; 3; 6)(uvp)	29,137	40,264	51,491	55,410	55,710	59,397
Red(6; 3; 70/30; 10; 6)(uvp)	35,624	38,851	47,874	47,111	42,852	39,060

Se destacan los mejores resultados conseguidos.

⁶ Para resultados de Desvío Estándar del Error Porcentual Absoluto conseguidos con la inclusión de variables derivadas, ver Anexo G.

La inclusión de las variables derivadas del precio permite reducir entre un 5% y un 10% el MAPE de los pronósticos a 3 y más años. También se aprecia que, en términos de MAPE, la mayoría de los mejores modelos sólo utilizan tres rezagos del precio.

Por otra parte, en términos de RMSE, sólo se consigue una mejora de 2% en los pronósticos a 1 año y de 5,7% en pronósticos a 6 años.

4.2.3. COMPARACIÓN DE RESULTADOS.

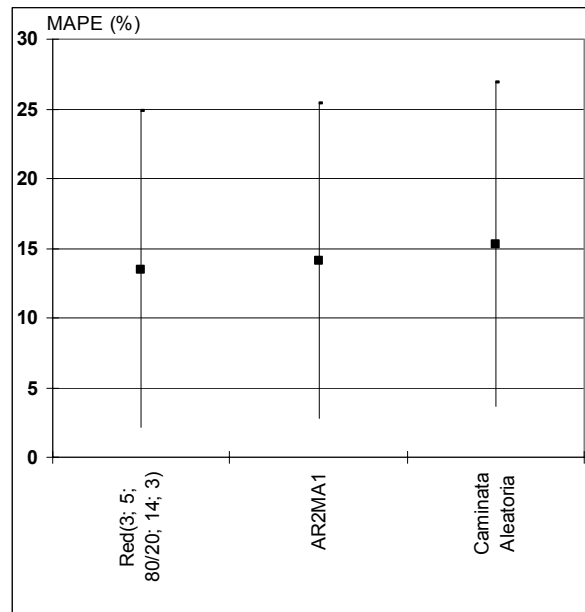
En esta sección se comparan los errores de pronóstico, a distintos plazos, arrojados por los mejores Modelos de Series de Tiempo, las mejores Redes Neuronales y la Caminata Aleatoria. En todos los alcances, las Redes Neuronales entregaron mejores pronósticos (menor MAPE) que los Modelos de Series de Tiempo, y estos últimos, a su vez, superaron a la Caminata Aleatoria. La mayor exactitud que registran las Redes Neuronales respecto a los Modelos de Series de Tiempo es marginal en pronósticos a 1 año, con una reducción del MAPE de 4%, pero se hace más significativa en pronósticos a 2 y más años, con una reducción de 30%. En cuanto a la dispersión del error, las redes que generan pronósticos de menor MAPE muestran una dispersión en sus errores similar a la de los mejores Modelos de Series de Tiempo.

En la medición del error aplicando RMSE, nuevamente los mejores modelos de Redes Neuronales tuvieron un mejor desempeño que los Modelos ARIMA y la Caminata Aleatoria, con reducciones del RMSE de entre 1% y 23% y de entre 4% y 37% respectivamente.

Tabla N°14: Comparación de Resultados. MAPE (%) en Pronósticos a 1 Año.

a 1 año	
Red(3; 5; 80/20; 14; 3)	13,481%
AR2MA1	14,078%
Caminata Aleatoria	15,264%
% Reducción Error	
Redes v/s ARIMA	-4,2%
Redes v/s Caminata Aleatoria	-11,7%

Gráfico N°8: Pronósticos a 1 Año. Redes Neuronales, ARIMA y Caminata Aleatoria. MAPE Más Menos un Desvío Estándar del Error Porcentual Absoluto %.



El mejor resultado conseguido por la Red Neuronal en el pronóstico fuera de muestra a 1 año, es una evidencia de su capacidad de generalización. Otro indicador de esta capacidad se obtiene a partir de la revisión de los errores conseguidos en los conjuntos de entrenamiento y testeo. La Tabla N°15 ofrece los valores de MAPE de esta Red Neuronal para los conjuntos de entrenamiento, testeo y evaluación.

Tabla N°15: MAPE (%) de la Red(3; 5; 80/20; 14; 3)
en los Conjuntos de Entrenamiento, Testeo y Evaluación.

Entrenamiento	Testeo	Evaluación
12,225	17,642	13,481

El mayor error conseguido en el conjunto de testeo es una señal de alerta respecto a las limitaciones de la capacidad de generalización de la Red(3; 5; 80/20; 14; 3), revelando que el desempeño conseguido en el conjunto de evaluación puede desmejorarse en otros periodos de evaluación. Una mayor capacidad de generalización se podría conseguir en futuros experimentos probando con una estrategia de selección aleatoria para la construcción de los conjuntos de entrenamiento y testeo [20].

En este punto vale la pena hacer mención a las heurísticas para el diseño y selección de redes. Entre las 585 redes construidas, la aplicación de la estrategia más común de selección indicaría la simple elección de la red con menor error en el conjunto de testeo. Esto conduciría a un modelo con un muy pobre desempeño en el conjunto de evaluación (Tabla N°16), delatando las fuertes inconsistencias que pueden presentarse entre el rendimiento en ambos conjuntos, y motivando la investigación de metodologías de diseño y selección más sofisticadas que permitan acceder a redes con la mayor capacidad de generalización y, con ello, a resultados robustos [7].

Tabla N°16: MAPE (%) de la Red(6; 3; 90/10; 1; 3)
en los Conjuntos de Entrenamiento, Testeo y Evaluación.

Entrenamiento	Testeo	Evaluación
18,480	10,285	25,710

A continuación se presenta la comparación de los errores de pronóstico a más años plazo.

Tabla N°17: Comparación de Resultados. MAPE (%) en Pronósticos a 2 Años.

a 2 años	
Red(6; 3; 70/30; 8; 10)	16,130%
AR1.GARCH(0,2)	22,979%
Caminata Aleatoria	23,823%
% Reducción Error	
Redes v/s ARIMA	-29,8%
Redes v/s Caminata Aleatoria	-32,3%

Gráfico N°9: Pronósticos a 2 Años. Redes Neuronales, ARIMA y Caminata Aleatoria. MAPE Más Menos un Desvío Estándar del Error Porcentual Absoluto %.

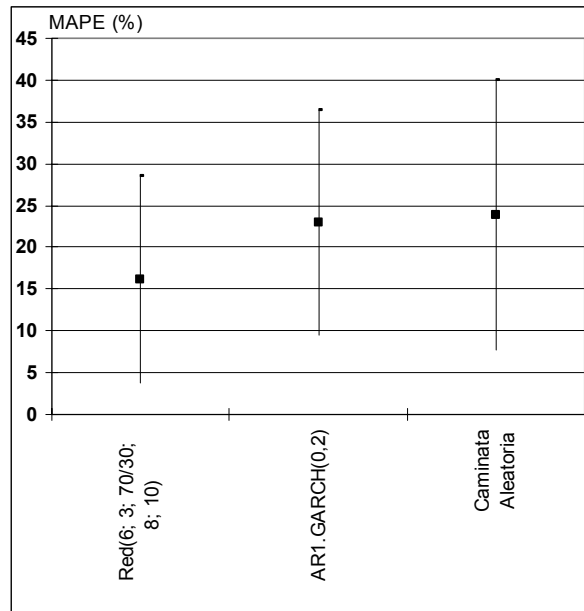


Tabla N°18: Comparación de Resultados. MAPE (%) en Pronósticos a 3 Años.

a 3 años	
Red(3; 2; 80/20; 4; 3)(dep)	18,657%
AR2MA1.GARCH(0,2)	27,020%
Caminata Aleatoria	29,201%
% Reducción Error	
Redes v/s ARIMA	-30,9%
Redes v/s Caminata Aleatoria	-36,1%

Gráfico N°10: Pronósticos a 3 Años. Redes Neuronales, ARIMA y Caminata Aleatoria. MAPE Más Menos un Desvío Estándar del Error Porcentual Absoluto %.

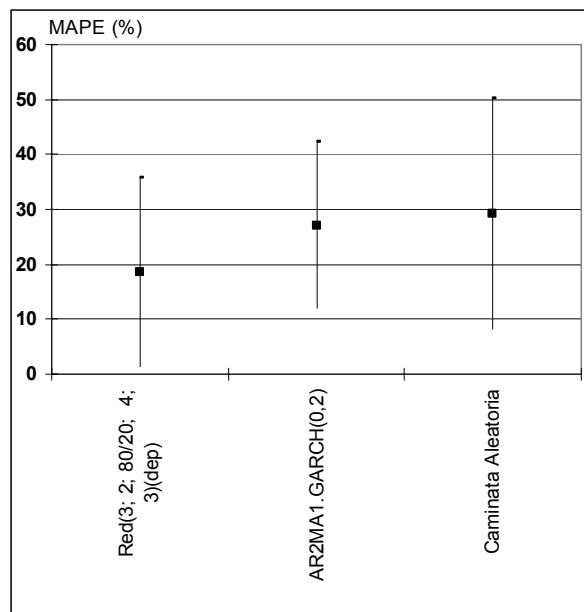


Tabla N°19: Comparación de Resultados. MAPE (%) en Pronósticos a 4 Años.

a 4 años	
Red(3; 6; 80/20; 11; 3)(dep)	20,627%
AR2MA1.GARCH(0,2)	28,867%
Caminata Aleatoria	33,295%
% Reducción Error	
Redes v/s ARIMA	-28,5%
Redes v/s Caminata Aleatoria	-38,0%

Gráfico N°11: Pronósticos a 4 Años. Redes Neuronales, ARIMA y Caminata Aleatoria. MAPE Más Menos un Desvío Estándar del Error Porcentual Absoluto %.

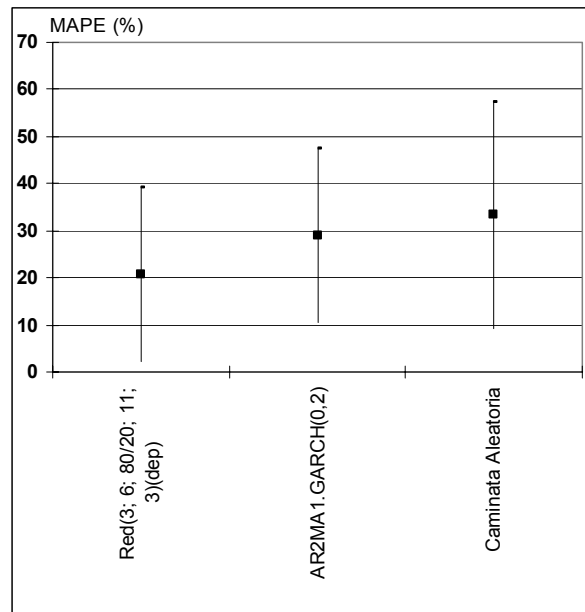


Tabla N°20: Comparación de Resultados. MAPE (%) en Pronósticos a 5 Años.

a 5 años	
Red(3; 1; 80/20; 13; 3)(dep)	21,538%
AR2MA1.GARCH(0,2)	31,083%
Caminata Aleatoria	37,371%
% Reducción Error	
Redes v/s ARIMA	-30,7%
Redes v/s Caminata Aleatoria	-42,4%

Gráfico N°12: Pronósticos a 5 Años. Redes Neuronales, ARIMA y Caminata Aleatoria. MAPE Más Menos un Desvío Estándar del Error Porcentual Absoluto %.

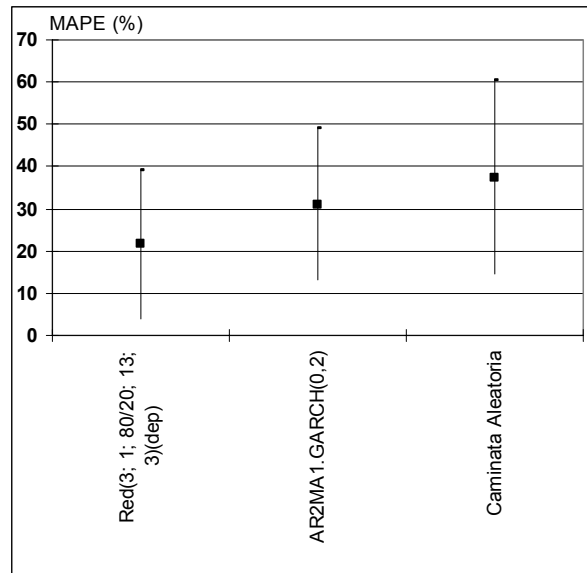


Tabla N°21: Comparación de Resultados. MAPE (%) en Pronósticos a 6 Años.

a 6 años	
Red(6; 3; 70/30; 9; 6)(uvp)	20,329%
AR2MA1.GARCH(0,2)	31,067%
Caminata Aleatoria	40,135%
% Reducción Error	
Redes v/s ARIMA	-34,6%
Redes v/s Caminata Aleatoria	-49,3%

Gráfico N°13: Pronósticos a 6 Años. Redes Neuronales, ARIMA y Caminata Aleatoria. MAPE Más Menos un Desvío Estándar del Error Porcentual Absoluto %.

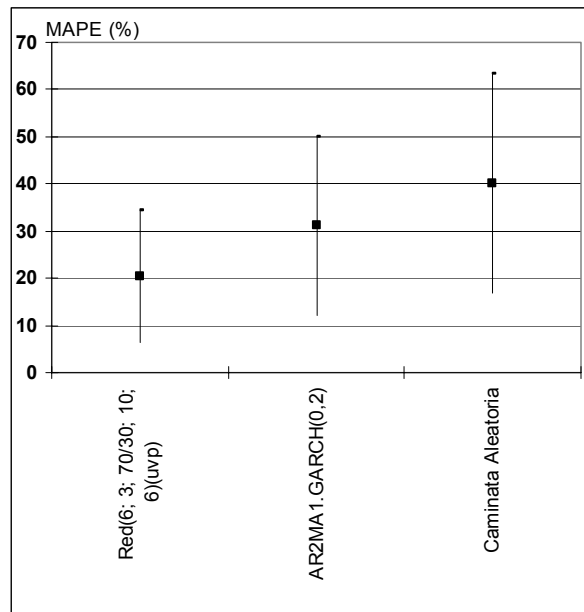


Tabla N°22: Comparación de Resultados. Disminución del MAPE Conseguido con las Mejores Redes Neuronales Seleccionadas.

Reducción del MAPE respecto a:	a 1 año	a 2 años	a 3 años	a 4 años	a 5 años	a 6 años
ARIMA	-4,2%	-29,8%	-30,9%	-28,5%	-30,7%	-34,6%
Caminata Aleatoria	-11,7%	-32,3%	-36,1%	-38,0%	-42,4%	-49,3%

Tabla N°23: Comparación de Resultados. Disminución del RMSE Conseguido con las Mejores Redes Neuronales Seleccionadas.

Reducción del RMSE respecto a:	a 1 año	a 2 años	a 3 años	a 4 años	a 5 años	a 6 años
ARIMA	-1%	-23%	-6%	-4%	-13%	-18%
Caminata Aleatoria	-4%	-30%	-20%	-22%	-31%	-37%

En las Tabla N°22 y N°23, se observa que la ventaja conseguida por las mejores Redes Neuronales por sobre los Modelos ARIMA al evaluar mediante MAPE, se hace menos significativa al emplear una medida como el RMSE que penaliza los errores de mayor tamaño. Sin embargo, se debe recordar que los mejores modelos de Redes Neuronales fueron seleccionados por su menor MAPE y no por su menor RMSE.

4.2.4. SISTEMAS INTELIGENTES HÍBRIDOS.

Los buenos resultados obtenidos mediante la aplicación conjunta de Modelos de Series de Tiempo y Redes Neuronales en la generación de pronósticos [1], motivan su utilización en este estudio.

Se procede de la siguiente manera:

- Con Modelos de Series de Tiempo:
 - Se realizan proyecciones fuera de muestra.
 - Se calculan los errores de proyección fuera de muestra.
- Con Redes Neuronales:
 - A partir de los errores calculados con Series de Tiempo, se entrena una Red Neuronal.
 - Se pronostican los errores de proyección fuera de muestra.
- Se corrigen las proyecciones fuera de muestra vía Modelos de Series de Tiempo con los errores pronosticados mediante Redes Neuronales.

Esta metodología se utilizará sólo en las proyecciones a un año por dos razones:

- Las proyecciones a un año son las que tienen menor error, siendo especialmente interesante evaluar la reducción del error que se consigue con un Modelo Híbrido.
- La aplicación de un Modelo Híbrido en pronósticos a más de un periodo requeriría el entrenamiento de redes con errores a 2 y más periodos, lo que aumenta de manera considerable la cantidad de cálculos a realizar y el tiempo de entrenamiento.

Para la aplicación del Modelo Híbrido se trabajó con el modelo AR2MA1, por ser éste el de mejor desempeño en pronósticos a un año.

En cuanto a las redes, se aplicó la misma estrategia de búsqueda en torno a una grilla descrita en la sección de Redes Neuronales. Sin embargo, dado que se debe trabajar con residuos de proyecciones fuera de muestra y que para construir el primer modelo AR2MA1 se utilizan los datos históricos del periodo 1913 a 1942, sólo se cuenta con 64 datos de residuos para entrenar y evaluar redes. Debido a esto se trabajó con una grilla más pequeña cuidando mantener una adecuada razón entre el número de datos disponibles y el número de conexiones.

Tabla N°24: Neuronas en la Capa de Entrada, en la Capa Oculta y Número de Conexiones.

Neuronas en la capa de entrada	Neuronas en la capa oculta	Número de conexiones
3	1	6
3	2	11
3	3	16
3	4	21
6	1	9
6	2	17
6	3	25

Nota: El número de conexiones considera las conexiones entre capas (incluyendo la capa de salida) y las conexiones correspondientes a las entradas de valor unitario.

A diferencia de lo ocurrido con las redes empleadas para proyectar el precio, en la proyección de los residuos, los mejores resultados no se obtuvieron actualizando los ponderadores con el método de Levenberg-Marquardt, sino que con el método de descenso del gradiente. Esto podría indicar que la función de error para este problema es menos suave, por lo que el Método de Levenberg-Marquardt no converge al mínimo global. La mejor red resultante fue Red(3; 2; 80/20; 9; 3).

Las siguientes tablas muestran los precios históricos, las proyecciones generadas con el mejor Modelo de Series de Tiempo (AR2MA1) y las proyecciones generadas con el Método Híbrido.

Tabla N°25: Precio, Pronósticos y Error Porcentual Absoluto (APE).

	Precio Histórico	Pronóstico AR2MA1	APE	Método Híbrido	APE
1977	149,7	181,4	21,208%	157,5	5,216%
1978	145,1	137,6	5,153%	141,0	2,846%
1979	187	158,2	15,418%	144,2	22,862%
1980	181	172,1	4,924%	148,9	17,752%
1981	132	184,0	39,429%	152,4	15,458%
1982	109,9	128,0	16,470%	116,8	6,318%
1983	116,7	116,4	0,232%	104,4	10,569%
1984	98,7	115,9	17,406%	112,0	13,491%
1985	102,1	108,1	5,894%	102,4	0,273%
1986	101,9	106,1	4,163%	92,0	9,747%
1987	129,2	110,1	14,755%	111,0	14,120%
1988	180,8	134,2	25,752%	136,1	24,734%
1989	188,5	180,3	4,342%	177,7	5,718%
1990	170,3	175,8	3,258%	160,2	5,923%
1991	149,2	170,0	13,930%	153,6	2,925%
1992	144,8	145,4	0,401%	128,8	11,035%
1993	119,5	141,4	18,320%	127,7	6,846%
1994	142,8	118,4	17,069%	99,5	30,341%
1995	175	152,3	12,956%	153,0	12,549%
1996	133,3	162,8	22,117%	154,2	15,657%
1997	132,5	136,5	2,982%	135,0	1,923%
1998	98,7	134,2	35,987%	134,3	36,032%
1999	93,2	101,6	9,045%	99,3	6,515%
2000	101,6	103,6	1,956%	106,0	4,369%
2001	87,4	104,8	19,954%	104,7	19,739%
2002	88,3	94,6	7,154%	93,4	5,832%
2003	95,8	96,6	0,887%	97,3	1,522%
2004	145,3	103,7	28,645%	100,2	31,039%
2005	173,9	153,5	11,717%	148,6	14,547%
2006	275,3	162,9	40,814%	158,0	42,624%
		MAPE 1977-2006	14,078%	MAPE 1977-2006	13,284%
		Desvío Estándar APE	11,292%	Desvío Estándar APE	10,657%

Tabla N°26: Comparación de Resultados. MAPE (%) en Pronósticos a 1 Año.

	a 1 año
Método Híbrido	13,284%
Red(3; 5; 80/20; 14; 3)	13,481%
AR2MA1	14,078%
Caminata Aleatoria	15,264%
% Reducción Error	
Método Híbrido v/s Red	-1,5%
Método Híbrido v/s ARIMA	-5,6%
Método Híbrido v/s Caminata Aleatoria	-13,0%

En el conjunto de 30 pronósticos evaluados, el Modelo Híbrido entregó mejores resultados que el más exitoso Modelo ARIMA en 16 oportunidades, logrando una disminución del MAPE de 5,6%.

Otro dato interesante al momento de elegir una herramienta de pronóstico es la capacidad de respuesta frente a grandes cambios de precios. Entre 1977 y 2006, en 14 oportunidades el cambio del precio respecto al año anterior superó el 15%, entregando el Modelo Híbrido los mejores pronósticos en 8 ocasiones.

Si la ventana de pronósticos se divide en 3 subperiodos de igual duración, se aprecia que el MAPE del Modelo Híbrido es menor al MAPE del Modelo ARIMA durante los primeros 20 años. En el tercer subperiodo, la comparación favorece al Modelo ARIMA, gracias a los menores errores conseguidos en los últimos 3 años.

A pesar de estos buenos resultados, la comparación del Modelo Híbrido con la mejor Red Neuronal muestra sólo una mejora marginal del MAPE y un deterioro, también marginal, del RMSE.

Tabla N°27: MAPE (%) por subperiodo.

	Pronóstico AR2MA1	Método Híbrido
1977-1986	13,030%	10,453%
1987-1996	13,290%	12,985%
1997-2006	15,914%	16,414%

4.2.5. COMPARACIÓN CON PROYECCION DE EXPERTOS.

Es interesante comparar los resultados conseguidos con las proyecciones publicadas en el pasado por analistas del mercado. Lamentablemente, no fue posible encontrar pronósticos de expertos para el periodo completo entre los años 1977 y 2006, por lo que la comparación se restringió a los últimos 20 años. Este es un elemento que debe ser considerado en la comparación, dado que los mejores modelos de pronóstico identificados se seleccionaron por la calidad de su desempeño en el periodo de 30 años mencionado y no en la ventana de tiempo 1987 - 2006, por lo que el contraste puede subestimar la precisión de los Modelos Híbridos y de las Redes Neuronales.

La Tabla N°28 muestra los errores de pronóstico conseguidos con los mejores modelos identificados y los errores de proyección del consultor Brook Hunt & Associates, una de las empresas más reputadas de la industria.

Tabla N°28: Precio, Pronósticos y Error Porcentual Absoluto (APE).

	Precio Histórico	Pronósticos a 1 año		Pronósticos a 2 años	
		APE Método Híbrido	Proyección Analista	APE Red	Proyección Analista
1987	129,2	14,120%	21,441%		
1988	180,8	24,734%	33,015%	37,414%	39,714%
1989	188,5	5,718%	14,628%	11,768%	36,490%
1990	170,3	5,923%	22,744%	3,698%	34,645%
1991	149,2	2,925%	27,533%	9,405%	38,021%
1992	144,8	11,035%	35,748%	7,800%	43,536%
1993	119,5	6,846%	9,712%	18,976%	20,265%
1994	142,8	30,341%	34,611%	4,327%	38,038%
1995	175	12,549%	18,167%	20,231%	42,942%
1996	133,3	15,657%	20,315%	10,137%	12,411%
1997	132,5	1,923%	7,967%	20,919%	1,721%
1998	98,7	36,032%	6,704%	11,381%	13,373%
1999	93,2	6,515%	4,735%	45,606%	6,472%
2000	101,6	4,369%	1,572%	15,780%	14,939%
2001	87,4	19,739%	39,731%	0,340%	24,361%
2002	88,3	5,832%	4,746%	17,570%	55,704%
2003	95,8	1,522%	6,483%	0,673%	15,193%
2004	145,3	31,039%	21,208%	21,782%	35,871%
2005	173,9	14,547%	23,920%	29,818%	31,506%
2006	275,3	42,624%	43,568%	36,949%	63,820%
MAPE					
1987-2006		13,284%	19,928%	17,083%	29,910%
Desvío Estándar APE					
		10,657%	12,463%	12,527%	15,991%

Fuente: Brook Hunt & Associates. Datos corresponden a proyecciones realizadas en Diciembre, salvo para los años 1990, 1991 y 1994, donde las proyecciones fueron realizadas en Septiembre.

En el conjunto de 20 pronósticos a 1 año, el Modelo Híbrido entregó mejores resultados que el analista en 15 oportunidades, consiguiendo disminuir el MAPE en 33% y disminuir el desvío estándar del APE en 14%.

En proyecciones a 2 años (19 evaluaciones), el Modelo Híbrido entregó mejores resultados que el analista en 16 oportunidades, consiguiendo disminuir el MAPE en 43% y el desvío estándar del APE en 22%.

En cuanto al pronóstico en situaciones con grandes cambios de precios, entre 1987 y 2006, en 10 oportunidades el cambio del precio respecto al año anterior superó el 15%, entregando el Modelo Híbrido los mejores pronósticos, en términos de MAPE, en 8 ocasiones.

Finalmente, en términos de RMSE, los Modelos de Inteligencia Computacional entregaron menores errores con reducciones de 16% en proyecciones a 1 año y de 40% en proyecciones a 2 años.

Así los mejores modelos de pronóstico identificados generaron proyecciones que, en promedio y en desviación estándar del error, son más precisas que las proyecciones publicadas por Brook Hunt & Associates.

CAPÍTULO V

CONCLUSIONES Y TRABAJOS FUTUROS

5.1. CONCLUSIONES.

- En las evaluaciones realizadas, los modelos no lineales de pronóstico, basados en Redes Neuronales multilayer perceptron, superaron la exactitud de los Modelos Lineales de Series de Tiempo más comúnmente usados. De esta manera, las Redes Neuronales demostraron ser capaces de aprovechar mejor la información contenida en los precios históricos considerados en el estudio.
- El modelo con mejores pronósticos a un año, para el periodo 1977 – 2006, fue un Modelo Híbrido que, utilizando conjuntamente modelos AR2MA1 y Redes Neuronales entrenadas con residuos, promedió un error porcentual absoluto (MAPE) de 13,284%. Esta cifra significa una reducción del error, en términos relativos, de 5,6% respecto al mejor Modelo ARIMA, y de 13% respecto a la Camina Aleatoria.
- Dicho Modelo Híbrido también mostró un mejor desempeño que ARIMA en situaciones con grandes cambios de precios (variaciones de precios superiores a 15% de un año a otro), generando proyecciones con menor MAPE en 8 de 14 eventos ocurridos entre 1977 y 2006.
- La ventaja de los modelos basados en Redes Neuronales por sobre los Modelos ARIMA se acrecienta en pronósticos a 2 y más periodos, llegando a una reducción del MAPE del orden de 30% en términos relativos.
- Aún cuando los mejores modelos basados en Redes Neuronales fueron seleccionados por su menor MAPE y no por su menor RMSE, estos también consiguieron una

reducción del RMSE respecto a los mejores Modelos de Series de Tiempo, aunque de menor tamaño (entre 1% y 23%).

- Así como la estimación con modelos GARCH permitió mejorar los resultados en los Modelos ARIMA, la inclusión de variables derivadas del precio, en especial su desviación estándar, también mejoró las proyecciones de las Redes Neuronales.
- En cuanto a la arquitectura de las Redes Neuronales más exitosas, éstas presentaron 3 o 6 neuronas alimentadas por rezagos del precio en la capa de entrada, y, en proyecciones a 3 y más años, una neurona adicional alimentada con la desviación estándar del precio o su última variación. En la capa oculta, las Redes Neuronales más exitosas mostraron un diverso número de neuronas (1, 2, 3, 5, 6).
- Se observa que, con un modelamiento no lineal, los últimos 6 rezagos del precio aportan información útil para la generación de pronósticos, hecho que no se aprecia en los Modelos ARIMA, donde modelos de ese orden fueron descartados por tener coeficientes no significativos.
- La comparación de los pronósticos publicados por expertos en el mercado del cobre versus los resultados conseguidos con Modelos de Inteligencia Computacional reveló la mayor precisión de estos últimos con reducciones del MAPE de 33% y 43% en proyecciones a 1 y 2 años respectivamente.
- El mejor Modelo Híbrido también mostró un mejor desempeño que las proyecciones de expertos en situaciones con grandes cambios de precios (variaciones de precios superiores a 15% de un año a otro), generando proyecciones con menor MAPE en 8 de 10 eventos ocurridos entre 1987 y 2006.
- La mayor exactitud de los pronósticos conseguidos utilizando Redes Neuronales va acompañada por una mayor complejidad en el diseño. En esta tesis se optó por sensibilizar sólo algunos parámetros del diseño de una red y, aún así, el número de cálculos y el tiempo requerido fue muy superior al de los Modelos de Series de Tiempo (horas versus minutos).

- Si bien las Redes Neuronales superaron a los Modelos de Series de Tiempo, los resultados conseguidos con la aplicación de Modelos Híbridos muestran que, mediante el trabajo conjunto con Modelos de Series de Tiempo y Redes Neuronales, es posible conseguir pronósticos más precisos para el precio del cobre. Así la pregunta planteada en el título esta tesis, en vez de tener una respuesta excluyente, tiene una respuesta integradora que alienta nuevos estudios en el área.
- Las evidencias entregadas avalan a las Redes Neuronales como una herramienta atractiva para el pronóstico del precio del cobre, animando el desarrollo de futuros estudios que aporten más antecedentes sobre su potencial y motivando la investigación de metodologías de diseño que posibiliten el aprovechamiento de dicha capacidad de generalización.

5.2. TRABAJOS FUTUROS.

- Continuando con la evaluación del potencial de las Redes Neuronales como herramienta de pronóstico para el precio del cobre, podría evaluarse el uso de otras estrategias de aprendizaje: otros algoritmos, aplicación de momentum y pruning. Así mismo, también resulta interesante probar el efecto de una definición aleatoria de los conjuntos de entrenamiento y testeo en la capacidad de generalización.
- A nivel de la arquitectura de las redes, se podría explorar el uso de n neuronas en la capa de salida, siendo n el número de periodos hacia adelante que se desea pronosticar.
- En el área de los Modelos Híbridos, sería un aporte contar con una evaluación de su desempeño en pronósticos a 2 y más periodos.
- Otro trabajo interesante para la entrega de evidencias sobre el atractivo de las Redes Neuronales consistiría en el entrenamiento de redes incluyendo variables adicionales al precio y sus derivaciones, como por ejemplo: stocks en semanas de consumo, indicadores de actividad económica y expectativas.

- En la línea del diseño de Redes Neuronales, futuros estudios deberían explorar algunas de las heurísticas propuestas, evaluándolas y proponiendo variantes que aseguren la construcción de redes con elevada capacidad de generalización.
- Finalmente, los resultados conseguidos en este trabajo pueden ser un referente para la evaluación del potencial de otras herramientas de pronóstico, como las Máquinas de Soporte Vectorial (Support Vector Regresión).

CAPÍTULO VI

BIBLIOGRAFÍA

1. ABURTO, L. y WEBER, R. 2005. Improved supply chain management based on hybrid demand forecasts.[en línea] Applied Soft Computing. <http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6W86-4GXW906-1&_coverDate=01%2F31%2F2007&_alid=504765908&_rdoc=1&_fmt=&_orig=search&_qd=1&_cdi=6646&_sort=d&view=c&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=093caef52bf4a8ef82a950e258f32b3f>.
2. CEBALLOS, J. y TILTON, J. 2005. Análisis del Fondo de Compensación del Cobre de Chile. En: LAGOS, G (Edit.). Minería y Desarrollo. Foro en Economía de Minerales, vol. III.
3. CIUDAD, J. 2005. Determinantes del Precio Spot del Cobre en las Bolsas de Metales. En: COCHILCO (Edits). Mercado del Cobre y Desarrollo Sustentable en la Minería. Santiago, Comisión Chilena del Cobre. pp. 145-187.
4. CLEMENTS, M. y HENDRY D. 2001. Forecasting Non-Stationary Economic Time Series. The MIT Press.
5. COMISION CHILENA DEL COBRE. 2006. [en línea]Anuario: Estadísticas del Cobre y Otros Minerales 1986-2005. <<http://www.cochilco.cl/anm/articlefiles/456-ANUARIO2005-PDF%2811-AGO-06%29.pdf>>.
6. COMISION CHILENA DEL COBRE. [en línea] Precio del Cobre. <<http://www.cochilco.cl>>.

7. CRONE, S. 2005. Stepwise Selection of Artificial Neural Network Models for Time Series Prediction. *Journal of Intelligent Systems* 14 (2-3). pp. 99-122.
8. ENDERS, W. 2004. *Applied Econometric Time Series*. 2ªed., John Wiley & Sons, Inc.
9. ENGEL, E. y VALDES, R. 2002. Prediciendo el Precio del Cobre: ¿Más Allá del Camino Aleatorio?. En: MELLER, P. (Edit.). *Dilemas y Debates en Torno al Cobre*. Santiago, Dolmen Ediciones. pp. 269-290.
10. FARADAY, J. y CHATFIELD, C. 1998. Times Series Forecasting with Neural Networks: A Case Study. *Applied Statistics* (47): 231-250.
11. FAYYAD, U., PIATETSKY-SHAPIRO, G. & SMYTH, P. 1996. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17 (3): 37-54.
12. FIFTH INTERNATIONAL CONFERENCE ON HYBRID INTELLIGENT SYSTEMS: 6 al 9 de Noviembre de 2005. Rio de Janeiro, Brasil. 2005. The Institute of Electrical and Electronics Engineers, Inc.
13. GUJARATI, D. N. 2003. *Basic Econometrics*. 4ªed., McGraw-Hill/Irwin.
14. HANKE, J. y REITSCH, A. 1995. *Estadística Para Negocios*. Irwin Professional Publishing.
15. HILERA, J. y MARTINEZ, V. 1995. *Redes Neuronales Artificiales. Fundamentos, Modelos y Aplicaciones*. Madrid, RA-MA, Addison-Wesley Iberoamericana, S.A.
16. IEEE CONNECTIONS, the Newsletter of the IEEE Computational Intelligence Society. 2003. 1(4).
17. INTELIGENCIA ARTIFICIAL. Modelos, Técnicas y Áreas de Aplicación. 2003. Por Francisco Escolanos "et al". Madrid, Thomson Editores Spain Paraninfo, S.A.
18. INTERNATIONAL MONETARY FUND. 2006. *World Economic Outlook September 2006, Financial Systems and Economic Cycles*. Washington, D.C., USA..
19. ISASI, P. y GALVAN, I. 2004. *Redes de Neuronas Artificiales. Un Enfoque Práctico*. Madrid, Pearson Educación, S.A.

20. KAASTRA, I. y BOYD, M. 1996. Designing a Neural Network for Forecasting Financial and Economic Time Series. *Neurocomputing* 10(3): 215-236.
21. MADDALA, G. S. 1996. *Introducción a la Econometría*. 2ªed., Prentice-Hall Hispanoamericana, S. A.
22. MARSHALL, I. y SILVA, E. 1998. Fluctuaciones del Precio del Cobre. Informe Macroeconómico para la Empresa. N°35: 38-60. Instituto de Economía, Pontificia Universidad Católica de Chile.
23. MELLER, P. 2002. El Cobre Chileno y la Política Minera. En: MELLER, P. (Edit.). *Dilemas y Debates en Torno al Cobre*. Santiago, Dolmen Ediciones. pp. 11-77.
24. NAU, R. 2006. What's the Bottom Line? How to Compare Models. [en línea]. <<http://www.duke.edu/~rnau/compare.htm>>.
25. PHILLIPS, S. y SWISTON, A. 2002. Forecasting Copper Prices in the Chilean context. Chile: Selected Issues. IMF Country Report No.02/163. International Monetary Fund.
26. PINDYCK, R. S. y RUBINFELD D. L. 2001. *Econometría: Modelos y Pronósticos*. 4ªed. México, D.F., McGraw-Hill/Interamericana Editores, S. A.
27. REED, R. D. y MARKS, R. J. II. 1999. *Neural Smithing*. The MIT Press.
28. ROMAGUERA, P. 1991. Las Fluctuaciones del Precio del Cobre y su Impacto en la Economía Chilena. CIEPLAN. Notas técnicas N°143.
29. SILIPO, R. 2003. Neural Netwroks. En: BERTHOLD, M. y HAND, D. (Edits.). *Intelligent Data Analysis*. Nueva York, Springer-Verlag New York, Inc. pp: 269-320.
30. TANG, Z. y FISHWICK, A. 1993. Feed-Forward Neural Nets as Models for Time Series Forecasting. *ORSA Journal of Computing*, 5(4): 374-386.
31. TASHMAN, L. 2000. Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review. *International Journal of Forecasting* (16): 437-450.

32. U.S. DEPARTMENT OF LABOR, Bureau of Labor Statistics. Producer Price Index. Commodity Data. All Commodities Not Seasonally Adjusted. <<http://data.bls.gov/cgi-bin/surveymost>>. [Consulta: 5 de agosto de 2006].
33. ULLOA, A. 2002. Tendencia y Volatilidad del Precio del Cobre. En: MELLER, P. (Edit.). Dilemas y Debates en Torno al Cobre. Santiago, Dolmen Ediciones. pp. 291-337.
34. VIAL, J. 1988. An Econometric Study of the World Copper Market, Ph.D. Dissertation. University of Pennsylvania. CIEPLAN. Notas Técnicas N°112.
35. VIAL, J. 2004. Modeling Commodity Markets in the Global Economy: Familiar Finding and New Strategies. The Earth Institute at Columbia University. Center on Globalization and Sustainable Development. Working Paper N°18.

ANEXO A

TEST DE DICKEY-FULLER Y CORRELOGRAMA DEL PRECIO ANUAL DEL COBRE

Tabla N°29: Test de Dickey-Fuller.

Logaritmo Natural del Precio Anual del Cobre, 1913-2006.

Null Hypothesis: PCU has a unit root
 Exogenous: Constant
 Lag Length: 1 (Fixed)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.311066	0.0172
Test critical values: 1% level	-3.503049	
5% level	-2.893230	
10% level	-2.583740	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(PCU)
 Method: Least Squares
 Date: 08/13/06 Time: 15:45
 Sample (adjusted): 1915 2006
 Included observations: 92 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PCU(-1)	-0.202255	0.061085	-3.311066	0.0013
D(PCU(-1))	0.252183	0.105839	2.382714	0.0193
C	1.015135	0.306022	3.317194	0.0013
R-squared	0.129819	Mean dependent var		0.003627
Adjusted R-squared	0.110264	S.D. dependent var		0.201894
S.E. of regression	0.190438	Akaike info criterion		-0.446913
Sum squared resid	3.227735	Schwarz criterion		-0.364681
Log likelihood	23.55802	F-statistic		6.638775
Durbin-Watson stat	1.828229	Prob(F-statistic)		0.002054

El número de rezagos se fijó en uno dado que en la ecuación del test los rezagos adicionales tenían un estadístico t que no permitía rechazar que sus coeficientes fueran iguales a cero [8].

Figura N°10: Correlograma.

Logaritmo Natural del Precio Anual del Cobre, 1913-2006.

Date: 08/16/06 Time: 16:53

Sample: 1913 2006

Included observations: 94

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.800	0.800	62.148	0.000
		2 0.605	-0.100	98.004	0.000
		3 0.493	0.115	122.14	0.000
		4 0.373	-0.106	136.10	0.000
		5 0.273	0.012	143.68	0.000
		6 0.221	0.039	148.67	0.000
		7 0.186	0.018	152.24	0.000
		8 0.158	0.015	154.85	0.000
		9 0.144	0.022	157.05	0.000
		10 0.123	-0.023	158.68	0.000
		11 0.087	-0.039	159.49	0.000
		12 0.068	0.026	160.01	0.000
		13 0.059	0.004	160.40	0.000
		14 0.046	0.001	160.64	0.000
		15 -0.027	-0.187	160.72	0.000
		16 -0.123	-0.119	162.48	0.000
		17 -0.164	0.026	165.63	0.000
		18 -0.156	0.074	168.53	0.000
		19 -0.204	-0.168	173.54	0.000
		20 -0.241	-0.026	180.61	0.000
		21 -0.201	0.095	185.61	0.000
		22 -0.195	-0.085	190.36	0.000
		23 -0.194	0.017	195.12	0.000
		24 -0.174	-0.014	199.01	0.000
		25 -0.185	-0.062	203.47	0.000
		26 -0.188	0.016	208.17	0.000
		27 -0.223	-0.182	214.90	0.000
		28 -0.266	-0.044	224.55	0.000
		29 -0.274	0.066	234.94	0.000
		30 -0.278	-0.048	245.83	0.000

ANEXO B

DESCRIPCIÓN DEL ALGORITMO BACKPROPAGATION⁷

Paso N°1: Inicialización aleatoria de ponderadores.

La red se inicializa con ponderadores aleatorios y pequeños (se quiere evitar la saturación de las funciones de transferencia)..

Paso N°2: Propagación hacia adelante.

La red calcula su salida de acuerdo a los ponderadores existentes.

- Evaluación de la(s) neurona(s) de la(s) capa(s) oculta(s). Para la neurona j de la capa oculta h:

$$net_{pj}^h = \sum_{i=1}^N w_{ji}^h x_{pi} + \theta_j^h$$

$$y_{pj} = f_j^h(net_{pj}^h)$$

donde,

w_{ji}^h : ponderador de la conexión entre la neurona i y la neurona j de la capa oculta h.

x_{pi} : componente i del vector de entrada p.

θ_j^h : sesgo de la neurona j de la capa oculta h

f_j^h : función de transferencia de la neurona j de la capa oculta h.

- Evaluación de la(s) neurona(s) de la capa de salida. Para la neurona k de la capa de la salida:

$$net_{pk}^o = \sum_{j=1}^L w_{kj}^o y_{pj} + \theta_k^o$$

$$y_{pk} = f_k^o(net_{pk}^o)$$

donde,

w_{kj}^o : ponderador de la conexión entre la neurona j y la neurona k de la capa de salida.

y_{pj} : componente j del vector de entrada p.

θ_k^o : sesgo de la neurona k de la capa oculta h

f_k^o : función de transferencia de la neurona k de la capa de salida.

⁷ Fuente: REED, R. D. y MARKS, R. J. II. 1999. Neural smithing. The MIT Press [27].

Paso N°3: Cálculo del error (neuronas de la capa de salida).

$$E = \frac{1}{2} \sum_p \sum_k (d_{pk} - y_{pk})^2$$

donde,

d_{pk} : salida deseada de la neurona k dado el patrón de entrenamiento p.
 y_{pk} : salida efectiva de la neurona k dado el patrón de entrenamiento p.

El error se calculando sumando los errores de todas las neuronas para todos lo patrones de entrenamiento.

Paso N°4: Backpropagation.

Se calcula la derivada del error con respecto a los ponderadores de las conexiones.

$$\frac{\partial E}{\partial w_{kj}} = \sum_p \frac{\partial E_p}{\partial w_{kj}} = \sum_p \frac{\partial E_p}{\partial net_{pk}} \frac{\partial net_{pk}}{\partial w_{kj}} = \sum_p \delta_k \frac{\partial net_{pk}}{\partial w_{kj}}$$

Para los nodos de la capa de salida:

$$\frac{\partial E}{\partial w_{kj}} = \sum_p \delta_k \frac{\partial net_{pk}}{\partial w_{kj}} = \sum_p \frac{\partial E_p}{\partial y_{pk}} \frac{\partial y_{pk}}{\partial net_{pk}} \frac{\partial net_{pk}}{\partial w_{kj}} = \sum_p -(d_{pk} - y_{pk}) f'_k y_{pj}$$

Para los nodos de la capa oculta, el término δ se obtiene de manera indirecta:

$$\delta_j = \frac{\partial E_p}{\partial net_{pj}} = \sum_k \frac{\partial E_p}{\partial net_{pk}} \frac{\partial net_{pk}}{\partial net_{pj}} = \sum_k \delta_k \frac{\partial net_{pk}}{\partial net_{pj}} = \sum_k \delta_k \frac{\partial net_{pk}}{\partial y_{pj}} \frac{\partial y_{pj}}{\partial net_{pj}} = \sum_k \delta_k w_{kj} f'_j$$

$$\frac{\partial E}{\partial w_{ji}} = \sum_p \delta_j y_{pj}$$

Paso N°5: Actualización de ponderadores. Método de descenso del gradiente.

Por definición error crece más rápido en la dirección del gradiente del error. Para minimizar el error, los ponderadores deben ser ajustados en el sentido opuesto al gradiente.

$$\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}}$$

donde,

η : tasa de aprendizaje.

ANEXO C

DESCRIPCIÓN DEL METODO DE LEVENBERG-MARQUARDT⁸

El desarrollo en series de Taylor, truncado en los términos de segundo orden, para la función de error es:

$$\hat{E}(w) = E_0 + g^T w + \frac{1}{2} w^T H w$$

donde,

g : derivada del error en función del ponderador w .

H : Hessiano, matriz de segundas derivadas del error en función de los ponderadores.

Si H fuera definida positiva y $E(w)$ fuera realmente una función cuadrática, el valor optimo de los ponderadores se podría obtener en un solo paso: $w^* = -H^{-1}g$.

Dado que, en la práctica, E no es exactamente cuadrática se hace necesario iterar:

$$w(t+1) = w(t) - H^{-1}g$$

Esta descripción corresponde al método de Newton.

En el método de Levenberg-Marquardt el ajuste de los ponderadores se hace con una combinación lineal de la dirección del método de Newton y de la dirección de mayor descenso del error (sentido opuesto del gradiente).

$$w(t+1) = w(t) - (H + \lambda I)^{-1}g$$

El algoritmo comienza con un valor de λ elevado, momento en que el método de actualización de ponderadores equivale al del método de descenso del gradiente descrito en el Anexo B. A medida que λ disminuye, acercándose a cero, el método de Levenberg-Marquardt se aproxima al método de Newton. Así se aprovechan las ventajas de ambas estrategias: el método de Newton converge rápidamente cerca del mínimo, pero diverge al no encontrarse en dicha vecindad; el método de descenso del gradiente es menos susceptible a diverger, pero converge muy lentamente.

⁸ Fuente: REED, R. D. y MARKS, R. J. II. 1999. Neural smithing. The MIT Press [27].

ANEXO D

PROGRAMA COMPUTACIONAL PARA LA GENERACIÓN Y EVALUACIÓN DE PROYECCIONES FUERA DE MUESTRA CON MODELOS DE SERIES DE TIEMPO

```
'regresión uno. proyección anual  
workfile F:\Tesis_Final\SerieTiempo\AR1\AR1_dif a 1913 2006  
fetch F:\Tesis_Final\SerieTiempo\pcu  
!inicio=@dtoo("1913")  
!fin=@dtoo("1976")
```

```
series prono1  
series prono2  
series prono3  
series prono4  
series prono5  
series prono6
```

```
series ape1  
series ape2  
series ape3  
series ape4  
series ape5  
series ape6
```

```
series serror1  
series serror2  
series serror3  
series serror4  
series serror5  
series serror6
```

```
scalar u
```

```
scalar mape1  
scalar mape2  
scalar mape3  
scalar mape4  
scalar mape5  
scalar mape6
```

```
scalar desvape1  
scalar desvape2  
scalar desvape3  
scalar desvape4  
scalar desvape5  
scalar desvape6
```

```

scalar rmserror1
scalar rmserror2
scalar rmserror3
scalar rmserror4
scalar rmserror5
scalar rmserror6

u=0

for !i = 1 to 30

    %ini=@otod(!inicio)
    %fini=@otod(!fin+!i-1)

    smpl %ini %fini

        equation reg {!i}.ls d(log(pcu)) c ar(1)

next

for !i = 1 to 30

    if !i>=26 then
        u=u-1
    endif

    %fini2=@otod(!fin+!i)
    %fini3=@otod(!fin+!i+5+u)

    smpl %fini2 %fini3
        reg {!i}.forecast preciofit{!i}

    %z1 = @otod(64+!i)
        prono1(64+!i) = @elem(preciofit{!i},%z1)

    if !i<30 then
        %z2 = @otod(64+!i+1)
        prono2(64+!i+1) = @elem(preciofit{!i},%z2)
    endif

    if !i<29 then
        %z3 = @otod(64+!i+2)
        prono3(64+!i+2) = @elem(preciofit{!i},%z3)
    endif

    if !i<28 then
        %z4 = @otod(64+!i+3)
        prono4(64+!i+3) = @elem(preciofit{!i},%z4)
    endif

endif

```

```

    if !i<27 then
        %z5 = @otod(64+!i+4)
        prono5(64+!i+4) = @elem(preciofit{!i},%z5)

    endif

    if !i<26 then
        %z6 = @otod(64+!i+5)
        prono6(64+!i+5) = @elem(preciofit{!i},%z6)

    endif

    delete preciofit{!i}

next

smp1 1977 2006

graph uno.line pcu prono1 prono2 prono3 prono4 prono5 prono6

ape1=abs(prono1-pcu)/pcu
serror1= (prono1-pcu)^2

ape2=abs(prono2-pcu)/pcu
serror2= (prono2-pcu)^2

ape3=abs(prono3-pcu)/pcu
serror3= (prono3-pcu)^2

ape4=abs(prono4-pcu)/pcu
serror4= (prono4-pcu)^2

ape5=abs(prono5-pcu)/pcu
serror5= (prono5-pcu)^2

ape6=abs(prono6-pcu)/pcu
serror6= (prono6-pcu)^2

mape1 = @mean(ape1)*100
mape2 = @mean(ape2)*100
mape3 = @mean(ape3)*100
mape4 = @mean(ape4)*100
mape5 = @mean(ape5)*100
mape6 = @mean(ape6)*100

desvape1 = (@var(ape1))^0.5*100
desvape2 = (@var(ape2))^0.5*100
desvape3 = (@var(ape3))^0.5*100
desvape4 = (@var(ape4))^0.5*100
desvape5 = (@var(ape5))^0.5*100
desvape6 = (@var(ape6))^0.5*100

```

```

rmserror1 = (@mean(serror1))^0.5
rmserror2 = (@mean(serror2))^0.5
rmserror3 = (@mean(serror3))^0.5
rmserror4 = (@mean(serror4))^0.5
rmserror5 = (@mean(serror5))^0.5
rmserror6 = (@mean(serror6))^0.5

```

```

table(45,4) precision
setcell(precision,1,1,"MAPE, %")
setline(precision,2)
setcell(precision,3,3,"Proy & PCu")
setcell(precision,5,2,"1 paso")
setcell(precision,5,3,mape1)
setcell(precision,6,2,"2 pasos")
setcell(precision,6,3,mape2)
setcell(precision,7,2,"3 pasos")
setcell(precision,7,3,mape3)
setcell(precision,8,2,"4 pasos")
setcell(precision,8,3,mape4)
setcell(precision,9,2,"5 pasos")
setcell(precision,9,3,mape5)
setcell(precision,10,2,"6 pasos")
setcell(precision,10,3,mape6)

```

```

setcell(precision,12,1,"DESV. EST APE")
setline(precision,13)
setcell(precision,16,2,"1 paso")
setcell(precision,16,3,desvape1)
setcell(precision,17,2,"2 pasos")
setcell(precision,17,3, desvape2)
setcell(precision,18,2,"3 pasos")
setcell(precision,18,3, desvape3)
setcell(precision,19,2,"4 pasos")
setcell(precision,19,3, desvape4)
setcell(precision,20,2,"5 pasos")
setcell(precision,20,3, desvape5)
setcell(precision,21,2,"6 pasos")
setcell(precision,21,3, desvape6)

```

```

setcell(precision,23,1,"RMSE")
setline(precision,24)
setcell(precision,25,3,"Proy & PCu")
setcell(precision,27,2,"1 paso")
setcell(precision,27,3,rmserror1)
setcell(precision,28,2,"2 pasos")
setcell(precision,28,3,rmserror2)
setcell(precision,29,2,"3 pasos")
setcell(precision,29,3,rmserror3)
setcell(precision,30,2,"4 pasos")
setcell(precision,30,3,rmserror4)
setcell(precision,31,2,"5 pasos")
setcell(precision,31,3,rmserror5)
setcell(precision,32,2,"6 pasos")
setcell(precision,32,3,rmserror6)

```

```

show precision

```

ANEXO E

PROGRAMA COMPUTACIONAL PARA LA GENERACIÓN Y EVALUACIÓN DE PROYECCIONES FUERA DE MUESTRA CON REDES NEURONALES

% 0. PRECIOS HISTORICOS 1913-2006.

pcu=[
5.38678601454
5.28421844217
5.49306144334
5.79209855570
5.46043621602
5.28776185261
4.91559174541
4.71402459090
4.83469334427
4.90082042809
4.94306997460
4.87443367292
4.89410147784
4.91412439372
4.89858579029
4.99856271286
5.21927415921
5.00058495824
4.71671156072
4.42364830936
4.59713801429
4.55702981066
4.41642806139
4.61115225767
4.88431592742
4.68027765847
4.64535197562
4.85748411460
4.79661665056
4.67842064773
4.63278535302
4.62693167777
4.60916220726
4.69501088999
5.01263329676
4.96074452448
4.91559174541
4.89933122454
4.99991133073
5.18794430907
5.19628464098
5.16364246322
5.50329694723
5.40312773090

4.97120122485
4.85359154449
5.03500265054
5.06827507354
5.00327493969
5.01926462079
5.02322240960
5.43241110103
5.69675785972
5.83422577068
5.52465580876
5.59396760865
5.72489143787
5.65283999592
5.35611446527
5.29781724151
5.68391984556
5.65494230924
5.05624580535
5.13932163506
5.00863329143
4.97742315989
5.23110861685
5.19849703127
4.88280192259
4.69957086141
4.75960653929
4.59208494644
4.62595272517
4.62399194023
4.86136159135
5.19739144796
5.23909800689
5.13756158767
5.00528768777
4.97535347995
4.78331637137
4.96144504991
5.16478597392
4.89260222718
4.88658264543
4.59208494644
4.53474772169
4.62104353514
4.47049528266
4.48074010761
4.56226268498
4.97880057058
5.15848042136
5.61786141215
];

numdatos=94;

% 1A. GENERACION DE MATRICES CORRESPONDIENTES A LOS CONJUNTOS DE ENTRENAMIENTO, TESTEO Y EVALUACION. 3 NEURONAS DE ENTRADA.

% 1A.1. RATIO TRAINING/TESTEO: 90/10.

```
numinput=3;
aux1=1;
aux2=15;
nn_hidden=[1 2 3 4 5 6];
calibra=[30 20 10];

for q=nn_hidden,
    for i=aux1:aux2,
        aux3=1;
        for aux0=calibra,

            datos1=zeros(numdatos+numinput,numinput+1);
            j=numinput+1;
            k=numdatos;

            for m=j:-1:1,

                datos1(m:k+m-1,j+1-m)=pcu;

            end

            datosprep1c=datos1(j:k,:);
            train_fracc=0.9;
            test_fracc=0.1;
            train_conj1=datosprep1c(1:round(train_fracc*(k-aux0-numinput)),:);
            test_conj1=datosprep1c(round(train_fracc*(k-aux0-numinput))+1:(k-aux0-numinput),:);
            eval_conj1=datosprep1c(k-aux0-numinput+1:k-aux0-numinput+10,:);
            train_o1=train_conj1(:,4);
            train_i1=train_conj1(:,1:3);
            test1.T=test_conj1(:,4);
            test1.P=test_conj1(:,1:3);
            eval_o1=eval_conj1(:,4);
            eval_i1=eval_conj1(:,1:3);

        end
    end
end
```

% 1A.2. SE TRASPONEN LAS MATRICES QUE ALIMENTARAN REDES.

```
train_o1=train_o1';
train_i1=train_i1';
test1.T=test1.T';
test1.P=test1.P';
eval_o1=eval_o1';
eval_i1=eval_i1';
```

% 2A. ENTRENAMIENTO Y EVALUACION.

% 2A.1. CAPA DE ENTRADA: REZAGOS 1 A 3.

```
net=newff([4.19 6.02;4.19 6.02;4.19 6.02],[q 1],{'tansig','purelin'},'trainlm');
net.trainParam.show=50;
net.trainParam.epochs=300;
net=init(net);
[net,tr]=train(net,train_i1,train_o1,[],[],test1);
almacen{i,aux3}=net;
ytrain=sim(net,train_i1);
apetrain=abs(exp(train_o1)-exp(ytrain))./exp(train_o1);
auxmapetrain(aux3)=sum(apetrain)/size(train_o1,2);
auxstdapetrain(aux3)=std(apetrain,1);
serrortrain=(exp(train_o1)-exp(ytrain)).^2;
auxrmstrain(aux3)=sqrt(sum(serrortrain)/size(train_o1,2));

numtrain(aux3)=size(train_o1,2);
ytest=sim(net,test1.P);
apetest=abs(exp(test1.T)-exp(ytest))./exp(test1.T);
auxmapetest(aux3)=sum(apetest)/size(test1.T,2);
auxstdapetest(aux3)=std(apetest,1);
serrortest=(exp(test1.T)-exp(ytest)).^2;
auxrmstest(aux3)=sqrt(sum(serrortest)/size(test1.T,2));
numtest(aux3)=size(test1.T,2);

% Proyecciones: 1 a 6 pasos

y=sim(net,eval_i1);
pronosticos{i,aux3}=y;
serroreval=(exp(eval_o1)-exp(y)).^2;
auxrmseval(aux3)=sqrt(sum(serroreval)/size(eval_o1,2));
numeval(1,aux3)=size(eval_o1,2);
apeeval=abs(exp(eval_o1)-exp(y))./exp(eval_o1);
auxmapeeval(1,aux3)=sum(apeeval)/size(eval_o1,2);
auxstdapeeval(1,aux3)=std(apeeval,1);

for aux4=1:5,

    eval_i1(1,:)=eval_i1(2,:);
    eval_i1(2,:)=eval_i1(3,:);
    eval_i1(3,:)=y;
    y=sim(net,eval_i1);

    if aux3<3

        aux5=y;
        eval_o1=datosprep1c(k-aux0-numinput+1+aux4:k-aux0-numinput+10+aux4,4);

    else
```

```

aux5=y(1,1:10-aux4);
eval_o1=datosprep1c(k-aux0-numinput+1+aux4:k-aux0-numinput+10,4)';

end

numeval(aux4+1,aux3)=size(eval_o1,2);
apeeval=abs(exp(eval_o1)-exp(aux5))./exp(eval_o1);
auxmapeeval(aux4+1,aux3)=sum(apeeval)/size(eval_o1,2);
auxstdapeeval(aux4+1,aux3)=std(apeeval,1);

end

aux3=aux3+1;

end

mapetrain(i)=sum(auxmapetrain.*numtrain,2)/sum(numtrain,2);
stdapetrain(i)=(sum((auxstdapetrain.^2).*numtrain,2)/sum(numtrain,2))^0.5;
mapetest(i)=sum(auxmapetest.*numtest,2)/sum(numtest,2);
stdapetest(i)=(sum((auxstdapetest.^2).*numtest,2)/sum(numtest,2))^0.5;
rmseval(i)=sum((auxrmseval.^2).*numeval(1,:),2)/sum(numeval(1,:),2);

mapeeval(1,i)=sum(auxmapeeval(1,:).*numeval(1,:),2)/sum(numeval(1,:),2);
stdapeeval(1,i)=(sum((auxstdapeeval(1,:).^2).*numeval(1,:),2)/sum(numeval(1,:),2))^0.5;
mapeeval(2,i)=sum(auxmapeeval(2,:).*numeval(2,:),2)/sum(numeval(2,:),2);
stdapeeval(2,i)=(sum((auxstdapeeval(2,:).^2).*numeval(2,:),2)/sum(numeval(2,:),2))^0.5;
mapeeval(3,i)=sum(auxmapeeval(3,:).*numeval(3,:),2)/sum(numeval(3,:),2);
stdapeeval(3,i)=(sum((auxstdapeeval(3,:).^2).*numeval(3,:),2)/sum(numeval(3,:),2))^0.5;
mapeeval(4,i)=sum(auxmapeeval(4,:).*numeval(4,:),2)/sum(numeval(4,:),2);
stdapeeval(4,i)=(sum((auxstdapeeval(4,:).^2).*numeval(4,:),2)/sum(numeval(4,:),2))^0.5;
mapeeval(5,i)=sum(auxmapeeval(5,:).*numeval(5,:),2)/sum(numeval(5,:),2);
stdapeeval(5,i)=(sum((auxstdapeeval(5,:).^2).*numeval(5,:),2)/sum(numeval(5,:),2))^0.5;
mapeeval(6,i)=sum(auxmapeeval(6,:).*numeval(6,:),2)/sum(numeval(6,:),2);
stdapeeval(6,i)=(sum((auxstdapeeval(6,:).^2).*numeval(6,:),2)/sum(numeval(6,:),2))^0.5;

i

end

aux1=aux1+15;
aux2=aux2+15;

end

```

```
% 3. RESULTADOS.
```

```
resultados_mape(1,1)=max(mapetrain,[],2);  
resultados_mape(2,1)=min(mapetrain,[],2);  
resultados_mape(3,1)=mean(mapetrain);  
resultados_mape(4,1)=std(mapetrain,1,2);  
resultados_mape(1,2)=max(mapetest,[],2);  
resultados_mape(2,2)=min(mapetest,[],2);  
resultados_mape(3,2)=mean(mapetest);  
resultados_mape(4,2)=std(mapetest,1,2);  
resultados_mape(1,3)=max(mapeeval(1,:),[],2);  
resultados_mape(2,3)=min(mapeeval(1,:),[],2);  
resultados_mape(3,3)=mean(mapeeval(1,:));  
resultados_mape(4,3)=std(mapeeval(1,:),1,2);  
resultados_mape
```

```
min(mapeeval(1,:))  
min(mapeeval(2,:))  
min(mapeeval(3,:))  
min(mapeeval(4,:))  
min(mapeeval(5,:))  
min(mapeeval(6,:))
```

```
save final_3cal
```

ANEXO F

REDES NEURONALES: MAPE EN LOS CONJUNTOS DE ENTRENAMIENTO Y TESTEO

Tabla N°30: Promedio del Error Porcentual Absoluto (MAPE) %.

	Entrenamiento	Testeo
Red(3; 5; 80/20; 14; 3)	12,225	17,642
Red(6; 1; 70/30; 3; 3)	15,965	22,755
Red(6; 2; 70/30; 10; 3)	14,781	25,188
Red(6; 2; 80/20; 9; 3)	13,949	22,925
Red(6; 3; 70/30; 13; 3)	14,756	25,041
Red(9; 1; 80/20; 8; 3)	16,563	18,666
Red(6; 2; 80/20; 9; 6)	13,302	18,148
Red(6; 3; 70/30; 9; 6)	16,251	20,498
Red(6; 3; 80/20; 2; 6)	14,775	19,727
Red(6; 3; 80/20; 15; 6)	13,045	18,731
Red(6; 4; 70/30; 12; 6)	14,395	21,817
Red(9; 2; 70/30; 15; 6)	14,230	20,282
Red(6; 2; 70/30; 4; 10)	13,285	19,798
Red(6; 2; 70/30; 7; 10)	12,588	19,234
Red(6; 2; 70/30; 10; 10)	17,850	22,256
Red(6; 2; 70/30; 13; 10)	15,008	21,484
Red(6; 3; 70/30; 8; 10)	14,496	18,807
Red(6; 3; 70/30; 3; 6)(uvp)	12,161	20,859
Red(3; 2; 80/20; 4; 3)(dep)	17,458	22,981
Red(3; 6; 80/20; 11; 3)(dep)	11,375	16,571
Red(3; 1; 80/20; 13; 3)(dep)	12,525	17,768
Red(6; 3; 70/30; 10; 6)(uvp)	14,084	20,920

Se destacan los mejores resultados conseguidos.

ANEXO G

REDES NEURONALES: DESvíO ESTÁNDAR DEL ERROR PORCENTUAL ABSOLUTO EN LOS MODELOS DE MEJOR DESEMPEÑO (MENOR MAPE) ANTES Y DESPUÉS DE LA INCLUSIÓN DE VARIABLES ADICIONALES.

Tabla N°31: DesvíO Estándar del Error Porcentual Absoluto %.

	a 1 año	a 2 años	a 3 años	a 4 años	a 5 años	a 6 años
Red(3; 5; 80/20; 14; 3)	11,337	14,181	18,906	22,863	22,433	28,413
Red(6; 3; 70/30; 9; 6)	17,702	16,619	15,599	17,132	17,117	17,279
Red(6; 3; 80/20; 15; 6)	11,413	13,748	15,249	19,986	19,091	20,713
Red(6; 2; 70/30; 4; 10)	11,731	13,690	15,571	17,601	18,224	22,949
Red(6; 2; 70/30; 7; 10)	13,270	14,479	17,199	19,515	17,515	20,444
Red(6; 3; 70/30; 8; 10)	11,835	12,383	19,905	25,287	21,867	29,124
Red(3; 1; 80/20; 13; 3)(dep)	12,405	17,386	17,849	17,561	17,690	17,592
Red(3; 2; 80/20; 4; 3)(dep)	11,475	15,207	17,197	19,054	19,765	21,566
Red(3; 6; 80/20; 11; 3)(dep)	10,619	27,007	21,502	18,392	26,765	21,283
Red(6; 3; 70/30; 3; 6)(uvp)	11,375	13,430	18,560	21,060	20,837	18,988
Red(6; 3; 70/30; 10; 6)(uvp)	10,852	14,334	17,707	16,969	17,147	14,074

Se destacan los mejores resultados conseguidos.

Al observar el set de modelos que generan proyecciones con menor MAPE, se observa que la inclusión de variables adicionales permite reducir el DesvíO Estándar del Error Porcentual Absoluto sólo en pronósticos a 1 y 6 años, registrándose disminuciones de 6% y 18,5%, respectivamente (la reducción del desvíO estándar del APE en pronósticos a 4 años es menor al 1%).