



Universidad Nacional Autónoma de México
Programa de Posgrado en Ciencias de la Administración

**Pronóstico del Tipo de Cambio Peso-Dólar Utilizando Redes
Neuronales Artificiales**

T e s i s

Que para optar por el grado de:

Maestro en Finanzas

Presenta:

Marco Antonio Villar Corona

Tutor:

Dr. Arturo Morales Castro

Facultad de Contaduría y Administración

México, D. F., noviembre de 2013.



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

A mi familia, por brindarme su cariño, respaldo y comprensión incondicional sin el cual no podría ser lo que soy. Por todo el amor que nos une y que nos permite atravesar cualquier obstáculo.

A la universidad, por permitirme continuar con mis estudios y en especial a la Coordinación de Estudios de Posgrado ya que sin su apoyo el desarrollo del presente trabajo no hubiera sido posible.

A mis sinodales, Dra. Graciela Bribiesca. M.I. Genoveva Barrera, M.F. Francisco Gerardo, M.F. Gabriel Malpica y a mi tutor Dr. Arturo Morales por sus comentarios.

A mis profesores de la Maestría, especialmente a los maestros Ricardo Morales Pelagio, Scott Da Gamma y a la coordinadora Rosario Higuera Torres por su apoyo constante.

A mis compañeros, por todas las experiencias que pasamos juntos dentro y fuera del salón de clases durante estos casi dos años.

Índice	Pág.
Introducción i
Matriz de Congruencia iii
Parte I Marco Teórico	
Cap. 1 Determinantes del Tipo de Cambio	
1.1 ¿Qué es el Tipo de Cambio? 1
1.2 Paridad del Poder Adquisitivo 7
1.3 Paridad de Tasas de Interés 11
1.4 Enfoque de Activos 14
1.4.1 Modelo Monetario 15
1.4.2 Modelo de Portafolio. 16
1.5 Crítica Meese y Rogoff. 18
Cap. 2 Análisis de Series de Tiempo	
2.1 Componentes de una Serie de Tiempo 21
2.2 Métodos de Suavizamiento 23
2.2.1 Media Móvil 23
2.2.2 Suavizamiento Exponencial Simple 24
2.2.3 Suavizamiento Exponencial Doble 25
2.2.4 Método de Holt –Winters 26
2.2.5 Método de Descomposición 28
2.3 Modelos Autorregresivos Integrados de Medias Móviles (ARIMA). 35
Cap.3 Redes Neuronales	
3.1 ¿Qué son las Redes Neuronales Artificiales? 42
3.2 Perceptrón Multicapa 48
3.2.1 Algoritmo de <i>Backpropagation</i> 49
3.3 Estado del Arte 51
Parte II Estimación de los Modelos para el Pronóstico y Comparación de los Resultados	
Cap. 4 Implementado un modelo ARIMA	
4.1 Métodos de Suavizamiento 56
4.2 Identificación del Modelos 57
4.3 Pruebas de Correcta Especificación 61
4.4 Pronóstico 61
Cap. 5 RNA Univariada y Multivariada	
5.1 Selección de Variables 64
5.2 Conjuntos de Entrenamiento y Prueba 68
5.3 Arquitectura de las Redes 69
5.4 Entrenando a las Redes 70
5.5 Implementación de las Redes 71

Cap. 6 Análisis Comparativo de los Resultados	
6.1 Criterios Dentro de la Muestra 79
6.2 Criterios Fuera de la Muestra 80
Conclusiones 81
Anexo 83
Bibliografía 101

Pronóstico del Tipo de Cambio Peso – Dólar Utilizando Redes Neuronales Artificiales

Síntesis: Se modela y pronostica el tipo de cambio fix mensual en el periodo 2000.01-2013.12 utilizando la metodología de las Redes Neuronales Artificiales implementando el algoritmo de Levenberg-Marquardt se establecen dos tipos de redes. En primer lugar, una univariada con un vector de información compuesto por 12 rezagos del tipo de cambio y en segundo, una multivariada con un vector de información de 19 variables como tasas de interés, agregados monetarios, etc. Se toman como modelos de referencia el enfoque tradicional de series de tiempo, compuesto por los métodos de suavizamiento y la metodología de Box Jenkins con una modelo ARIMA (1, 0, 0).

Palabras Clave: Tipo de Cambio, RNA, Levenberg-Marquardt, Box-Jenkins, Series de Tiempo. Modelo Univariado, Modelo Multivariado.

Clasificación JEL: C22, C45, C53, F31, F37

Introducción

En la actualidad es bastante sencillo pensar en una actividad económica que involucre operaciones con tipo de cambio, las remesas que envían los trabajadores mexicanos laborando en Estados Unidos, un estudiante de intercambio, una empresa exportadora, etc. Cada uno de ellos enfrenta un dilema, al tener una divisa y convertirla a otra a qué razón se debe hacer. En función de sus respectivas posiciones largo en peso (mxn) corto en dólares (usd) o viceversa su utilidad se puede ver afectada de forma positiva o negativa. Si tomamos el ejemplo de la exportadora sus costos están en pesos y sus ingresos en dólares, en el momento en que desee convertir sus ingresos a pesos estará corriendo el riesgo de que cada dólar que cambie valga más pesos (en caso de una depreciación) o que por el contrario cada dólar que ganó valga menos pesos (para una apreciación) esta incertidumbre es lo que se conoce comúnmente como riesgo cambiario.

Al enfrentar la incertidumbre cada agente incorpora la información que considera necesaria para formar una expectativa respecto al comportamiento futuro del tipo de cambio. Por esta razón el estudio de los determinantes y evolución del tipo de cambio ha sido de suma importancia por su impacto en la actividad económica. A nivel microeconómico la formación de precios se puede ver afectada ya que el productor puede traspasar los costos de una importación al consumidor por medio del precio. A nivel macroeconómico puede afectar el flujo de inversión extranjera y por consecuencia afectar la actividad económica de un país, otra afectación se da si el proceso de formación de precios se ve afectado de manera generalizada puede ocasionar niveles más altos de inflación, fenómeno conocido como *pass-trough* del tipo de cambio presente en las economías latinoamericanas.

Los anteriores son algunos ejemplos de la relevancia que tiene este precio dentro de la economía, pero analizar sus factores determinantes no es sencillo. Estimar relaciones de variables de oferta (Inversión Extranjera, Exportaciones, Transferencias, etc.) y de demanda (Importaciones, Especulación, Cobertura, etc.) al tener tantas interconexiones, se convierte en una tarea compleja de realizar.

Y es que, como se observa más adelante en el capítulo 1, a partir de la ruptura del sistema de Bretton Woods en 1973 con la libre flotación de las divisas la teoría no lograba avanzar al mismo ritmo que la realidad. Fue así como el resto de la década de 1970 con el tipo de cambio comportándose con una alta volatilidad, el objetivo principal fue concretar un modelo que lograra explicar el comportamiento del tipo de cambio (*ex-post*) dejando a un lado el pronóstico del mismo (*ex-ante*). Una detrás de otra fueron superadas las teorías existentes en la ortodoxia ante la incapacidad de poder explicar los cambios que se vivían.

Esta complejidad de estudiar y explicar la evolución del tipo de cambio se ha mantenido hasta nuestros días por lo que el investigador ha tratado de incorporar nuevas formas de estudio para lograr entender y estimar de manera adecuada las relaciones del tipo de cambio. Una de estas son las redes neuronales artificiales, las cuales existen desde la década de 1950 pero no fue sino hasta finales de la década de 1980 con el avance de las computadoras que se lograron implementar las primeras redes neuronales artificiales aplicadas a temas económico-financieros con el trabajo de Halbert White (1988) quien buscaba pronosticar el precio de las acciones de IBM.

Las ventajas que ofrecen las redes neuronales son en primera instancia, que no se necesita que la relación entre variables independientes y dependiente sea lineal y en segunda, no se tiene la limitación de seguir una función teórica. Si bien podemos utilizar la teoría para mejorar las variables que conforman nuestra capa de entrada no es limitante para la implementación de una red a un solo cuerpo teórico. Por otro lado tienen la desventaja de que al ser un sistema de caja negra la relación entre variables no se puede interpretar de forma tan clara como la que se obtiene de métodos como la regresión lineal.

En la mayoría de las ocasiones la limitación de no poder interpretar los coeficientes de un sistema de caja negra se ha visto superada por los resultados obtenidos al aplicar las redes neuronales artificiales al tipo de cambio. Como se puede ver en el capítulo 3 existen un gran número de aplicaciones para distintos tipos de cambio y en todos se han logrado resultados positivos. Para el caso de México no existen trabajos que se hayan aplicado por lo que presenta una gran oportunidad para aplicar las RNA en un tipo de cambio que en tiempos recientes ha presentado gran volatilidad.

De esta forma surgen diferentes preguntas que dan origen a esta investigación. Se ha visto que la teoría económica ha sido superada en muchas ocasiones por la evidencia empírica que ofrecen los datos quebrando así, frecuentemente, los paradigmas establecidos. Esto dificulta la toma de decisiones e implementación de estrategias de corto, mediano y largo plazo de inversionistas, empresarios y administradores de riesgos ya que, ante un entorno de volatilidad realizar un pronóstico adecuado de uno de los precios más importantes de la economía se vuelve un proceso muy complejo.

Una vez dicho lo anterior se plantea la pregunta que será el hilo conductor de la investigación con sus correspondientes objetivos e hipótesis los cuales se pueden ver en el siguiente cuadro.

Cuadro I: Matriz de Congruencia

Pregunta Principal	Objetivo General	Hipótesis Principal
¿Será posible diseñar una arquitectura de red neuronal artificial que brinde un mejor pronóstico que los modelos tradicionales de series de tiempo para el tipo de cambio?	Plantear, diseñar e implementar una red neuronal artificial que permita realizar un mejor pronóstico del tipo de cambio peso-dólar que los modelos tradicionales de series de tiempo.	La metodología de las Redes Neuronales Artificiales ajusta mejor los datos fuera de la muestra que los métodos tradicionales de análisis de series de tiempo.
Pregunta Secundaria	Objetivo Secundario	Hipótesis Secundaria
¿Qué tipo de modelo será más eficaz para estudiar el comportamiento y pronosticar el tipo de cambio, el univariado o multivariado?	Diseñar e implementar una RNA univariada y una RNA multivariada y así conocer cuál tiene un mejor ajuste.	La Red Neuronal Artificial Multivariada obtiene un mejor ajuste a los datos dentro de la muestra que una Red Neuronal Artificial Univariada.

Para lograr cumplir con estos objetivos, primero se revisa la evolución del *corpus* teórico a través de los métodos histórico y analítico, se estudian las principales aportaciones teóricas que en su momento se establecieron para determinar el tipo de cambio a partir de la ruptura de Bretton Woods.

De acuerdo con las teorías se seleccionan las principales variables que explican el tipo de cambio, se realizan pruebas estadísticas para obtener la correlación entre variables independientes y el tipo de cambio, también se realizan pruebas de causalidad para determinar las variables independientes. Después de terminar con el proceso de datos se inicia con el diseño de la arquitectura de la red en función de:

- Composición de la Capa de Entrada
- Número de Capas Ocultas
- Numero de Neuronas de la Capa Oculta
- Entrenamiento de la Red
- Implementación

La complejidad de las redes neuronales está en la construcción de las mismas, para el desarrollo de esta investigación se implementa una variante del perceptrón multicapa con *Backpropagation* al implementar el algoritmo de Levenberg-Marquandt de optimización por su eficacia y gran desempeño en comparación con el gradiente descendiente tradicional ver (Abolfazl et al. 2005). Este proceso se aplica tanto para la RNA univariada como multivariada y ambas tienen como *benchmark* la metodología Box-Jenkins de manera que se puedan comparar con la forma tradicional de realizar pronósticos.

Al implementar un modelo de tipo caja negra como lo son las RNA es razonable encontrarse con cuestionamientos acerca de qué ocurre por dentro, qué representan los pesos sinápticos y cómo interpretarlos. Desde luego no son de fácil interpretación como los resultados obtenidos de una regresión lineal clásica con propensiones marginales, elasticidades o semielasticidades pero en realidad no es el objetivo de esta investigación. Al incluir las relaciones provenientes del cuerpo teórico se busca principalmente mejorar el pronóstico del tipo de cambio al incorporar la mejor información más no un análisis tradicional de los determinantes del tipo de cambio obteniendo sus respectivos parámetros para validar una teoría en particular. A continuación se presenta un resumen de los capítulos de este trabajo:

En el primer capítulo se realiza una pequeña descripción de la historia reciente en el sistema financiero internacional y cómo el tipo de cambio se confirmó como un mecanismo de transmisión y las consecuencias de esto en la política monetaria de los países que se han refugiado en distintos regímenes cambiarios para poder controlar los impactos en la economía. También se desarrollan los principales modelos que dan forma al cuerpo teórico que rodea al tipo de cambio desde la paridad del poder adquisitivo cuyos orígenes remontan a 1918 pasando por la paridad de tasas de interés, el modelo monetario tanto con precios flexibles como fijos, el modelo de portafolio, etc.

En el segundo capítulo se desarrollan los modelos que nos sirven como referencia para las redes neuronales. Desde la descomposición clásica de una serie (*i.e.* obtener la tendencia, el ciclo, la estacionalidad y el componente aleatorio), pasando por los métodos de suavizamiento exponencial simple, de Brown y Holt-Winters con una serie generada para ejemplificar estos modelos. Por último se desarrolla la metodología de Box-Jenkins junto con un ejemplo de proceso autorregresivo de primer orden y un proceso de media móvil de primer orden.

El tercer capítulo presenta la definición de las redes neuronales artificiales y junto a esto también se presenta la evolución de las mismas desde McCulloch y Pitts hasta Rumelhart. Se presenta el perceptrón multicapa y el algoritmo *backpropagation*, la tasa de aprendizaje y el momentum. Por último se revisa el estado de arte respecto a la aplicación de las redes neuronales en el pronóstico de distintos tipos de cambio demostrando que a nivel internacional ha obtenido excelentes resultados.

En el capítulo cuatro se realizan las estimaciones del tipo de cambio con los modelos de suavizamiento y se observa que obtienen un gran ajuste dentro de la muestra pero no logran el mismo resultado fuera de la muestra. Los casos de Holt-Winters aditivo y multiplicativo logran capturar los efectos de estacionalidad y generan mejor pronóstico que

los otros 3. En la siguiente parte se aplica la metodología de Box-Jenkins y con base en pruebas de estacionariedad y la función de autocorrelación y autocorrelación parcial se determina que el mejor modelo para estimar el tipo de cambio es un ARIMA (1, 0, 0) después de realizar las pruebas de correcta especificación se realiza el pronóstico para el periodo 2013.01 a 2013.12.

Para el siguiente capítulo se implementan dos redes neuronales artificiales, una red univariada y una multivariada. La red univariada se genera tomando como la información de entrada o *input* 12 rezagos del tipo de cambio con una capa oculta y una de salida. La red multivariada toma 19 variables de entrada las cuales se implementan en función de los modelos del capítulo 1 y por último se realizan pruebas de causalidad para comprobar la relación estadística. Una vez que se tiene la arquitectura de ambas redes se entrenan con el algoritmo Levenberg-Marquardt el cual toma como función de desempeño la media de los errores al cuadrado (MSE) la cual se busca minimizar.

En el último capítulo se establecen una serie de criterios estadísticos para comprobar el ajuste de cada uno de los modelos dentro y fuera de la muestra. Con los resultados obtenidos es claro que el mejor ajuste de todos los modelos es la red neuronal multivariada. Por último se plantean las conclusiones del trabajo

Cap. 1 Determinantes del Tipo de Cambio

“Under the skin of any international economist lies a deep seated belief in some variant of the PPP theory of exchange rates”

Dornbusch y Krugman, 1976, p.540

En este capítulo se presentan los principales modelos que se emplean para el estudio del tipo de cambio. Debido a que la variable de estudio se encuentra en función de la política monetaria de un país o simplemente del mercado sería un supuesto casi imposible de sostener que un modelo explique el tipo de cambio para todos los países o que incluso lo haga para el mismo país de igual manera para cualquier muestra temporal que se señale, razón por la cual este capítulo nos brinda las principales relaciones teóricas que tendremos en cuenta para elegir el vector de información de entrada para la red neuronal a desarrollar en capítulos posteriores.

1.1 ¿Qué es el Tipo de Cambio?

En la actualidad uno de los mercados con mayor eficiencia (*i.e.* menor *spread* de compra-venta) es el mercado Forex (acrónimo de *Foreign Exchange*) un mercado global, electrónico de 24 horas y como lo menciona Shamah “*It is by far the largest market in the world, with an estimated \$1.6 trillion average daily turnover*” (2004, p. 5) en esta misma línea Weithers menciona “*To put the magnitude of FX trading in perspective in one other way, the current volume of foreign exchange transactions corresponds to every person on the planet trading the equivalent of over USD 300 (PER DAY)!*” (2006, p. 72). Es aquí donde los tipos de cambio se determinan todos los días lo que hace surgir la pregunta ¿qué es esto que se compra y vende en cantidades exorbitantes todo el día de lunes a viernes en todo el mundo?

La respuesta más sencilla es que el tipo de cambio representa básicamente el precio de la moneda de un país en términos de la moneda de otro país. Esta definición clásica *per se* no podría explicar el porqué del tamaño del mercado Forex se necesita entender que la importancia y crecimiento de este y otros mercados en las últimas décadas se encuentran íntimamente relacionados con los procesos de globalización económica y financiera.

Claro que el principio de la conversión de divisas es que permite transferir el poder de compra de una moneda a otra, de un país a otro país. En la medida en que exista convertibilidad entre las diferentes monedas de los diferentes países, (*i.e.* que existan mercados libres o al menos relativamente libres de divisas) las transacciones internacionales de bienes y servicios serán viables. Otro de estos mercados que permiten la convertibilidad son: de entrega inmediata también llamado al contado o *spot*, a plazos o de

contratos adelantados también conocido como *forward*, el mercado de opciones sobre divisas y el de *swaps* sobre divisas¹.

En décadas pasadas, particularmente antes de 1970, la importancia del mercado de divisas estaba casi totalmente restringida a las operaciones spot y forwards asociados con transacciones de importación y/o exportación. La inversión extranjera directa y el endeudamiento externo dieron también lugar a importantes movimientos de capitales internacionalmente. No existían por tanto productos derivados sobre divisas, bursátiles o en mercados descentralizados.

Con el quiebre del acuerdo de Bretton Woods, las monedas de los principales países industrializados fluctuaron ampliamente en respuesta a desequilibrios comerciales, las tasas de interés, precios de los *commodities*, guerras e incertidumbre política.

Durante la década de 1990 se observó un período de inestabilidad financiera internacional muy importante generando varias crisis en México, Corea del Sur, Rusia y Brasil conocidos de manera coloquial como el efecto *tequila*, *dragón*, *vodka* y *samba*. Todas estas con características muy similares, en primer lugar una atracción de gran volumen de capital vía fondos de inversión² debido a las siguientes causas: altas tasas de interés internas, un sentido de estabilidad derivada de tipos de cambio rígidos y lo que en el momento parecían ser perspectivas muy optimistas. Lo anterior ayudó a impulsar un auge de sus respectivos mercado de valores y a financiar grandes déficits por cuenta corriente (Edwards *et. al.*, 2003).

En algún momento y por diversas razones, los rendimientos de estos fondos comenzaron a revertirse y al mismo tiempo aumentaba el nivel de incertidumbre y el grado de riesgo de los países. Como resultado ocurrió un abandono de capital masivo en los países en cuestión, las reservas internacionales cayeron a niveles muy bajos y los tipo de cambio reales se sobrevaluaban. Finalmente, el tipo de cambio fijo tenía que ser abandonado, y el país se veía obligado a dejar flotar su moneda.

Tal como lo advierte Stanley Fischer “*Little wonder, then, that policymakers involved in dealing with these crises have warned strongly against the use of pegged rates for countries open to international capital flows. That warning has tended to take the form of advice that intermediate policy regimes between hard pegs and floating are not sustainable. This is the bipolar or two-corner solution view*” (2001, p.3)

Diversos regímenes se implementaron desde entonces cada uno de ellos con diferentes características cuestión que, para quién busca entender y pronosticar el tipo de cambio, es

¹ Un desarrollo más extenso de este tema se puede encontrar en el capítulo 2 de Weithers (ibíd.).

² Para leer del tema con la perspectiva de uno de los principales actores se invita al lector a leer Greenspan (2007) en particular el capítulo titulado *Millenium Fever* donde toca la relación de estas crisis con uno de los principales fondos del período Long Term Capital Management.

fundamental comprender cada uno de ellos y saber distinguir en que países se aplica. Sobre este tema el Fondo Monetario Internacional dentro del informe anual sobre regímenes de cambio y restricciones cambiarias (AREAR por sus siglas en inglés) define categorías y clasifica a los países miembros en función de los principales regímenes cambiarios que existen dentro de los mismos.

Siguiendo a Bubula y Ökter-Robe (2002) los regímenes que considera son los siguientes:

1. Regímenes con otra moneda de curso legal (dolarización):

El país utiliza la moneda de otro país, que circula como única moneda legal. El ejemplo más claro es la dolarización, donde el dólar circula como la única moneda de curso legal es una entrega completa de control de las autoridades de la política monetaria nacional, y como tal, es visto como la forma más dura de un régimen de cambio fijo.

2. Regímenes con la misma moneda de curso legal (uniones monetarias)

El país es miembro de una unión monetaria o cambiaria en la que comparten la misma moneda de curso legal. Al implementar estos regímenes las autoridades también entregar el control sobre la política monetaria interna.

3. Cajas de Conversión o Juntas Monetarias

Un régimen basado en el compromiso legislativo explícito de cambiar moneda nacional por una moneda extranjera especificada a un tipo de cambio fijo, combinado con las restricciones a la autoridad de asegurar el cumplimiento de su obligación legal. Esto implica que la moneda nacional permanezca plenamente respaldada por activos externos, eliminando las funciones tradicionales del banco central, como el control monetario y la de prestamista de última instancia.

4. Tipo de cambio fijo convencional

El país vincula su moneda a un cambio fijo respecto a otra moneda o a una canasta de monedas de los principales socios comerciales y financieros del país, los Derechos Especiales de Giro son un caso de canasta estandarizada. Las autoridades están dispuestas a mantener la paridad fija por intervención directa con la compra y/o venta de divisas en el mercado o de manera indirecta, mediante el uso agresivo de la tasa de interés o la imposición de normas cambiarias. La flexibilidad de la política monetaria, aunque limitada, es mayor que en los tres primeros casos.

5. Paridad con desliz moderado

La divisa se ajusta periódicamente respecto de una moneda única o una canasta de monedas con movimientos de pequeña magnitud, a una tasa fija o en respuesta a cambios de

determinados indicadores, por ejemplo los diferenciales de las tasas de inflación con los principales socios. Se distingue el ajuste retrospectivo o *backward looking* cuando está configurado para generar cambios en la divisa ajustados por inflación y como prospectivo o *forward looking* cuando el tipo de cambio se ajusta a una tasa fija preanunciada, por lo general se emplea para tener un ancla nominal en los precios.

El mantener una paridad móvil creíble limita la política monetaria tal como en el sistema de tipo de cambio fijo, sobre todo en un ajuste prospectivo, ya que se espera que las autoridades intervengan para asegurar el camino específico de depreciación fija. El grado de intervención en un ajuste retrospectivo será menor debido a la falta de compromiso con un camino de depreciación fija y la ausencia de una necesidad de anclar las expectativas.

6. Tipos de cambio fijo dentro de una banda horizontal

Permite el movimiento dentro de ciertos márgenes de fluctuación, las autoridades están dispuestas a defender los límites de la banda a través de la intervención directa o indirecta. Este régimen permite un grado limitado de discrecionalidad en la política monetaria en función de que tan ancha sea la banda.

7. Flotación administrada sin trayectoria predeterminada del tipo de cambio

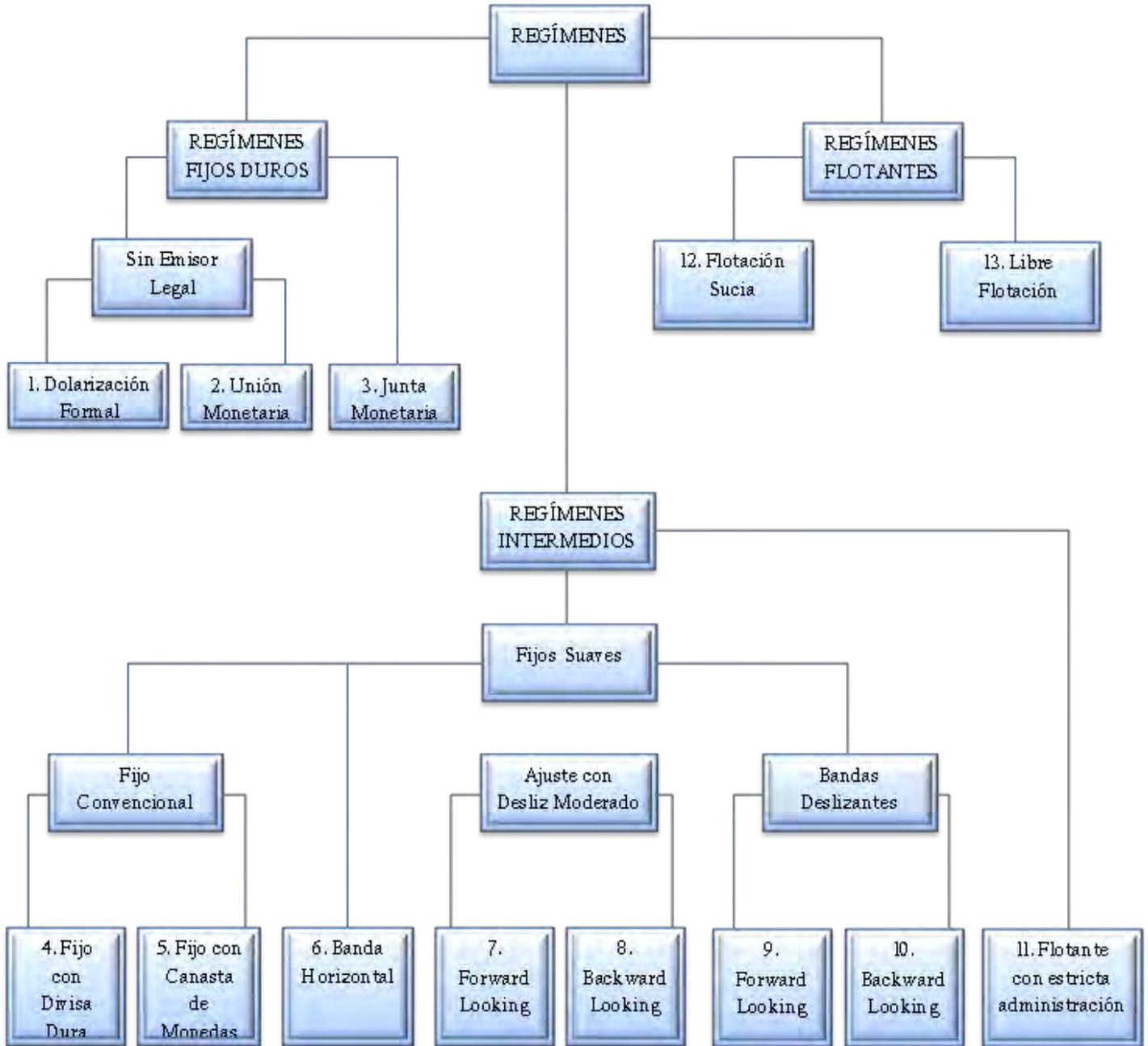
Se influyen los movimientos del tipo de cambio mediante una intervención activa -directa o indirecta- para contrarrestar la tendencia de largo plazo del tipo de cambio. Se interviene sin especificar un objetivo predeterminado de tipo de cambio, dicho de otra manera se le permite flotar de manera controlada por lo que se le conoce como flotación sucia.

8. Libre Flotación

El tipo de cambio es determinado por el mercado, cualquier intervención cambiaria tiene por objetivo moderar la volatilidad y evitar fluctuaciones excesivas del tipo de cambio que no están justificadas por los fundamentales económicos, en lugar de establecer un nivel del tipo de cambio.

La gráfica 1.1 esquematiza la clasificación del IMF pero es necesario saber cómo está dividido el mundo respecto a los regímenes actuales, esta información se presenta en el cuadro 1.1 donde se pueden observar los actuales regímenes de los países miembros del Fondo Monetario Internacional y los respectivos marcos alternativos de política monetaria con el fin de poner de relieve el papel del tipo de cambio en la política económica

Gráfica 1.1: Clasificación de los Regímenes de Tipo de Cambio



Fuente: Elaboración propia siguiendo a Bubula y Ökter-Robe (Ibid.)

Cuadro 1.1: Regímenes cambiarios en el mundo vigentes durante 2012.

Regímenes	Tipo de Cambio como Ancla Nominal ^{7/}				Objetivo en los Agregados Monetarios ^{8/}	Meta Inflacionaria ^{9/}	Otro ^{10/}	Total
	Dólar	Euro	Canasta	Otro				
Sin Moneda de Curso Legal Propia ^{1/}	8	3	0	2	0	0	0	13
Junta Monetaria ^{1/}	8	3	0	1	0	0	0	12
Fijo Convencional ^{1/}	14	19	5	5	0	0	0	43
Fijo Estable ^{2/}	7	1	1	0	2	1	4	16
Paridad con deslíz moderado ^{1/}	1	0	1	0	0	0	1	3
Paridad <i>estilo</i> deslíz ^{3/}	4	1	0	0	4	1	2	12
Fijo con Bandas Horizontales ^{1/}	0	0	1	0	0	0	0	1
Otros ^{4/}	1	0	5	0	9	0	9	24
Flotación ^{5/}	0	0	0	0	14	19	2	35
Libre Flotación ^{6/}	0	0	0	0	0	11	20	31
Total	43	27	13	8	29	32	38	190

Fuente: Elaboración propia con información del Reporte Anual del Fondo Monetario Internacional del 2012 y Habermeier *et. al.* (2009)

Notas:

1/ Es necesaria la confirmación de las autoridades de cada país respecto a que la calidad del régimen sea *de jure* y no sólo *de facto*.

2/ Mercado spot debe mantenerse dentro de un margen de 2% por 6 meses o más y no ser flotante. El margen se puede medir respecto a una divisa o a una canasta de divisas. La clasificación requiere que se cumplan los criterios estadísticos y que se mantenga la estabilidad como resultado de acción oficial.

3/ Mercado spot debe mantenerse dentro de un margen de 2% por 6 meses o más y no ser flotante. Debe mantener un cambio anualizado de 1% de manera monótona y continua.

4/Categoría libre que alberga a las naciones que no cumplen con los criterios de las otras categorías.

5/ En particular, un tipo de cambio que cumple con 2/ ó 3/ será clasificado como tal a menos que sea claro que la causa de la estabilidad no provenga de acciones oficiales.

6/ Para clasificar como libre flotación la intervenciones tienen que ser máximo 3 en los últimos 6 meses con una duración menor a 3 días hábiles para atacar condiciones de desorden del mercado y las autoridades deben de proveer evidencia de no hacerlo se clasifica como Flotación.

7/ El tipo de cambio tanto sirve como ancla nominal o como un objetivo intermedio de la política monetaria, estando asociado con los regímenes cambiarios sin moneda de curso legal, regímenes fijos duros e intermedios.

8/ La autoridad monetaria utiliza sus instrumentos para lograr una tasa de crecimiento fijada como meta para un agregado monetario, como la base monetaria, M1 o M2, y el agregado se convierte en el ancla nominal o el objetivo intermedio de la política monetaria.

9/ Se anuncia una tasa objetivo para la inflación, con el compromiso institucional del banco central de alcanzar dicha meta con el fin de anclar las expectativas. Las decisiones de política monetaria suelen guiarse por la desviación de los pronósticos de inflación futura de la meta de inflación anunciada.

10/ El país no tiene un ancla nominal establecida explícitamente sino que efectúa un seguimiento de diversos indicadores en la política monetaria.

El caso de México es uno de los 11 países –de acuerdo con la información del cuadro 1.1– en el mundo que tienen libre flotación bajo política monetaria de metas de inflación. Con esta información se confirma la visión bipolar de Fischer (*op. cit.*) y nos permite dar el peso necesario a los modelos monetarios que concuerdan con la política de objetivos de inflación para poder establecer los mecanismos de transmisión en la siguiente sección analizamos los

principales modelos de tipo de cambio para poder establecer las variables relevantes para el tipo de cambio.

1.2 Paridad del Poder Adquisitivo

La más antigua teoría de la determinación del tipo de cambio es la paridad de poder adquisitivo (PPP en adelante) la teoría, comúnmente atribuido a Gustav Cassel (1918) a pesar de que-como es habitual- existen algunos precursores en épocas anteriores (sobre este punto véase Oficial, 1982).

Cassel define “*At every moment the real parity between two countries is represented by this quotient between the purchasing power of the money in the one country and the other. I propose to call this parity the purchasing power parity*” (Ibíd, p.413). Actualmente se distinguen dos versiones de PPP, la paridad absoluta y la paridad relativa.

De acuerdo con la versión absoluta, el tipo de cambio entre dos monedas es igual a la razón entre los valores, expresados en las dos monedas consideradas, de la misma canasta básica que contiene la misma cantidad de los mismos productos. Si, por ejemplo, dicha canasta costará \$900 en los Estados Unidos y € 1,000 en la Unión Europea, el tipo de cambio EURUSD será 0.9 dólares por euro.

Como ejemplo de ésta versión *The Economist* creó en 1986 el índice *Big Mac* con la función de ser una herramienta para que la teoría cambiaria fuera más digerible y sin embargo se ha convertido en un estándar global, acuñándose el término de *Burgernomics* incluido en varios libros de texto económicos y en varios trabajos de investigación.³

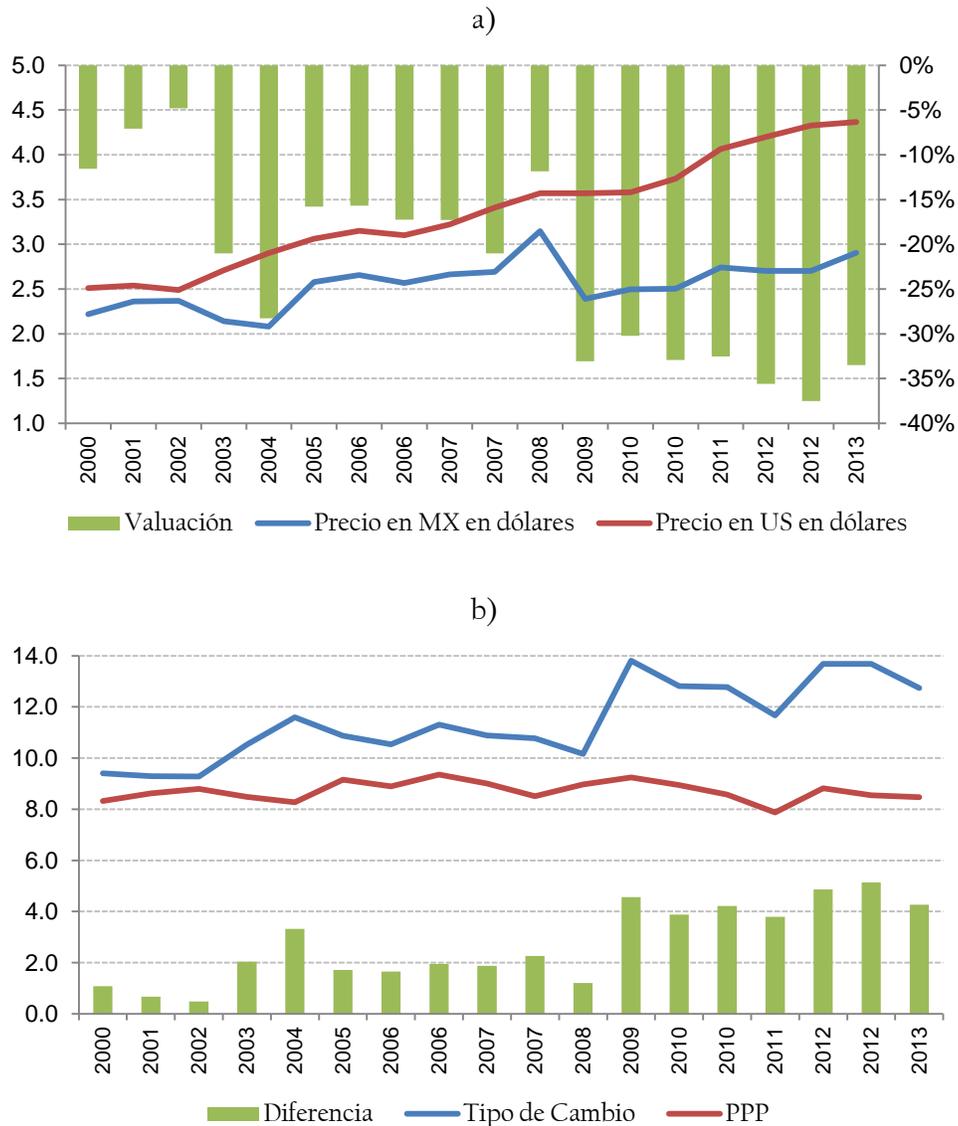
Para nuestro caso empleamos los precios de la hamburguesa para Estados Unidos y en México en dólares tomando en cuenta el tipo de cambio observado en dicho momento y se establece la sub o sobre valuación del peso respecto al dólar como se aprecia en la gráfica 1.2 a el peso ha estado entre 4.8 y 37.5% subvaluado estando en promedio 22.6% subvaluado. De acuerdo a la lógica de largo plazo de PPP el peso se tendrá que depreciar frente al dólar para estar en equilibrio con la relación de largo plazo, dicho lo anterior es claro que enfrentamos un primer problema con el modelo, los costos de producción en países desarrollados e.g. Estados Unidos es mayor derivado del costo laboral a diferencia de los países emergentes donde el mismo costo es mucho menor.

De manera alterna podemos observar en la gráfica 1.2 b la evolución del tipo de cambio observado y el que sugiere la PPP a través del cociente de precio local (i.e. el precio de la hamburguesa en pesos) y entre precio extranjero (e.g. \$4.37 dólares por hamburguesa). En el período observado el rango del tipo de cambio de acuerdo a PPP fue 7.87 a 9.35 y en

³ Se invita al lector a revisar los trabajos de Cumby (1996), Pakko y Pollard (1996 y 2003), Chen, Shen y Wang (2007), Dunn (2007) Mathä (2009) y Clements y Shea (2010).

promedio 8.71 pesos por dólar. Si dejamos de lado el hecho de que PPP es de largo plazo y tomamos estos datos como un pronóstico *naive* tendríamos una subestimación constante ocasionando un error promedio de 2.72 pesos.

Gráfica 1.2. México: Índice Big Mac y Paridad del Poder Adquisitivo. 2000-2013



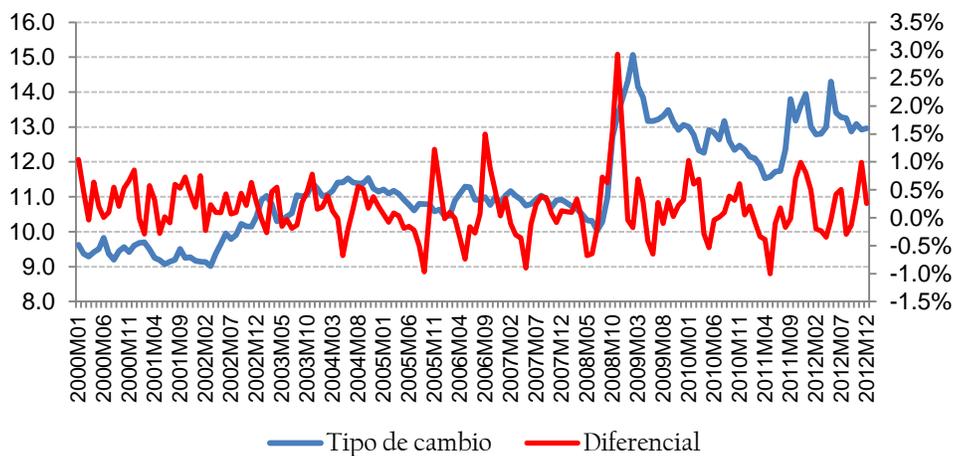
Fuente: Elaboración propia con información de The Economist. Consultada en el sitio: <http://media.economist.com/sites/default/files/media/2013InfoG/databank/EconomistBMI20130202.xls> el día 3 de Febrero de 2013 a las 16:00 hrs.

De acuerdo con la versión relativa, las variaciones porcentuales en el tipo de cambio son equivalentes a las variaciones porcentuales en la relación del nivel de precio de los dos países, es decir, la diferencia entre el porcentaje de variación en los dos niveles de precios, o la inflación diferencial. Para nuestro ejemplo empleamos el índice de precios al consumidor

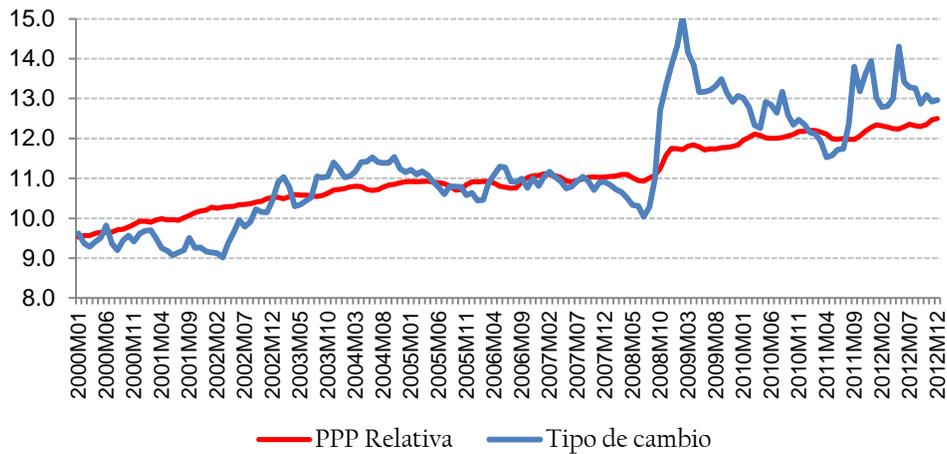
de Estados Unidos *Consumer Price Index* (CPI) y el índice nacional de precios al consumidor (INPC) de México para calcular la inflación mensual y así poder obtener el diferencial de tasas de inflación. Hecho lo anterior el diferencial nos indica la variación porcentual del tipo de cambio, asumiendo un tipo de cambio del pasado como base, en este ejemplo se tomó el dato de diciembre de 1999 como base y se aplicó la respectiva variación al mismo para obtener el tipo de cambio teórico de acuerdo con PPP relativa.

Dentro del periodo de crisis financiera el peso se depreció frente al dólar pasando de 10.9814 en finales de septiembre de 2008 a 12.7125 a finales del siguiente mes, el diferencial de tasas de inflación llegó a 2.9% y aunque fue una cifra considerable no se asemeja a la apreciación del dólar frente al peso de más del 10%. Tal como se observa en la gráfica 1.3 a y b si existe evidencia para sustentar la relación de largo plazo ya que aunque es claro que existen desviaciones del tipo de cambio observado respecto al sugerido por PPP relativa, puede considerarse a este último como una tendencia del primero.

Gráfica 1.3.A México: Paridad del Poder Adquisitivo Relativa. 2000.I-2012.I2



Gráfica I.3.B México: Paridad del Poder Adquisitivo Relativa. 2000.1-2012.12



Fuente: Elaboración propia con información de Banco de Información Económica (BIE), Federal Reserve Economic Data (FRED) y Banco de México (BANXICO). Consulta de los sitios:

<http://www.inegi.org.mx/sistemas/bie/>

<http://research.stlouisfed.org/fred2/series/CPIAUCSL?cid=9>

<http://www.banxico.org.mx/portal-mercado-cambiario/index.html> el día 5 de Febrero de 2013 a las 18:00 hrs.

De la popularidad y controversia de este modelo han surgido una gran cantidad de trabajos de investigación que le mantienen como parte del instrumental análisis básico para quién se dedica a las finanzas internacionales, pero también críticas que dan por tierra todo lo relacionado con la misma. Algunas de las sugerencias hechas al modelo son:

- Emplear índices basado únicamente en bienes que se comercien internacionalmente. Esta razón tiene sustento en la idea de que el arbitraje de *commodities* es lo que está detrás de PPP.
- Emplear índices de costos de producción ya que, la competitividad nacional y el grado de internacionalización de las industrias son la principal fuerza detrás de PPP.
- Emplear las tasa de inflación ya que, en cualquier país la tasa de interés nominal es igual a la suma de la tasa de interés real y la tasa de inflación, condición conocida como la ecuación de Fisher.

Ninguna de estas propuestas está exenta de inconvenientes y cada una ha sido objeto de serias críticas. Por ejemplo, la idea de arbitraje en los productos básicos fue criticada por considerar que supone la libre movilidad de bienes (ausencia de aranceles y otras restricciones al comercio) y una relación constante, dentro de cada país, entre los precios de los bienes comerciados y no comerciados. Todas estas condiciones son claramente inexistentes en el mundo aún en el largo plazo. Además, la ley de un precio supone que los bienes son altamente homogéneos, otra suposición en contradicción con los hechos y por

las nuevas teorías del comercio internacional, que se especializan en el papel de la diferenciación de productos.

1.3 Paridad de Tasas de Interés

La paridad de tasas de interés es un supuesto base de varios modelos razón por la cual desarrollamos ambas versiones. En esencia, la paridad de tasas de interés cubierta (CIP por sus siglas en inglés) es una aplicación de la ley de precio único para los mercados financieros postulando que, cuando el riesgo cambiario está cubierto con el mercado a plazo (con un forward), la tasa de rendimiento sobre un activo nacional debe ser igual a la de un activo extranjero con similares características. Si este no es el caso, entonces el arbitraje con interés cubierto se pone en marcha y continúa hasta que los cambios resultantes en las fuerzas de la oferta y la demanda (para los activos referidos) dan lugar a la restauración de la condición de no arbitraje representado por la CIP.

El arbitraje con interés cubierto es provocado por la violación de la condición de paridad de intereses cubierto, que describe la relación de equilibrio entre el tipo de cambio al contado (spot), el tipo de cambio a plazo (forward), las tasas de interés nacionales y las tasas de interés extranjeras. Considere el caso de un inversionista que cuenta con un capital inicial (K), y se enfrenta a dos alternativas:

- i. Inversión interna, por lo que el inversionista compra los activos locales, ganando la tasa de interés nacional i .
- ii. Inversión externa, por lo que el inversionista debe convertir de moneda nacional a moneda extranjera para comprar activos en el extranjero, obteniendo la tasa de interés externa i^* .

Por un lado la inversión nacional no implica la conversión de moneda por lo que no existe riesgo cambiario. Por otro lado, en la inversión extranjera existe exposición al riesgo de cambio, pero esta exposición puede ser cubierta con la venta forward de moneda extranjera (compra de moneda nacional). El riesgo cambiario se elimina debido a que el tipo de cambio forward se conoce de antemano, por lo tanto, el inversionista conoce de antemano el valor en moneda nacional de la inversión extranjera. Si la posición no está cubierta en el mercado a plazo, el inversionista tiene que esperar hasta el vencimiento y aplicar el tipo de cambio de contado vigente entonces para determinar el valor en moneda nacional de la inversión extranjera.

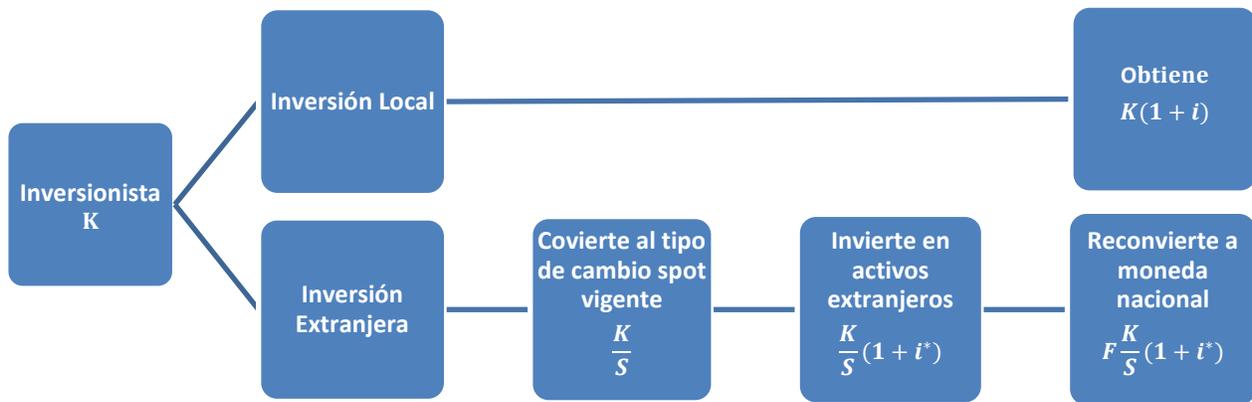
Suponiendo una inversión de un período, la cual inicia con la adquisición de un activo financiero (por ejemplo, un depósito) y terminando en la fecha de vencimiento de este activo.

Cuando el inversionista elige la inversión nacional, el capital invertido se suma a la tasa de interés nacional, y el inversionista termina con el capital inicial más los intereses, es decir $K(1 + i)$.

Si el inversionista elige la inversión extranjera, se convierte el capital inicial a la moneda extranjera al tipo de cambio de contado vigente, obteniendo de K/S unidades de moneda extranjera, donde (S) se mide en unidades de moneda nacional por una unidad de la moneda extranjera.

Si K/S es invertido en activos en el extranjero, este capital se suma a la tasa de interés externa, de manera que el valor de la inversión a vencimiento es $K/S (1 + i^*)$. El valor en moneda nacional de la inversión se obtiene mediante la conversión de este resultado a moneda nacional al tipo de cambio forward (F) para obtener $F(K/S) (1 + i^*)$.

Cuadro 1.2 Paridad Cubierta de Tasas de Interés



Si suponemos que no hay restricciones a los movimientos de capital y que no hay costos de transacción y que los inversionistas son neutrales al riesgo (en el sentido de que son indiferentes entre mantener activos nacionales y extranjeros si estos ofrecen rendimientos iguales). La condición de equilibrio que excluye la posibilidad de arbitraje es que las dos inversiones deben ser igualmente rentables, o dicho de otra manera deben proporcionar la misma cantidad de capital más intereses en moneda local, por lo tanto:

$$K(1 + i) = \frac{K}{S}(1 + i^*)F \tag{1.1}$$

Al expresar la condición en términos de una unidad de la moneda local, obtenemos

$$(1 + i) = \frac{F}{S}(1 + i^*) \tag{1.2}$$

La condición CIP nos dice que el rendimiento bruto nacional es igual al rendimiento bruto cubierto en el exterior. El lado izquierdo de la ecuación (1.2) representa el rendimiento

bruto nacional: es bruto debido a que incluye el monto invertido (una unidad de la moneda local) y los intereses devengados i . Puesto que $F/S = 1 + f$, donde f es el diferencial forward,⁴ se tiene que:

$$(1 + i) = (1 + f) (1 + i^*) \quad (1.3)$$

Al simplificar la ecuación (1.3), ignorando el término i^*f obtenemos la aproximación a la condición CIP:

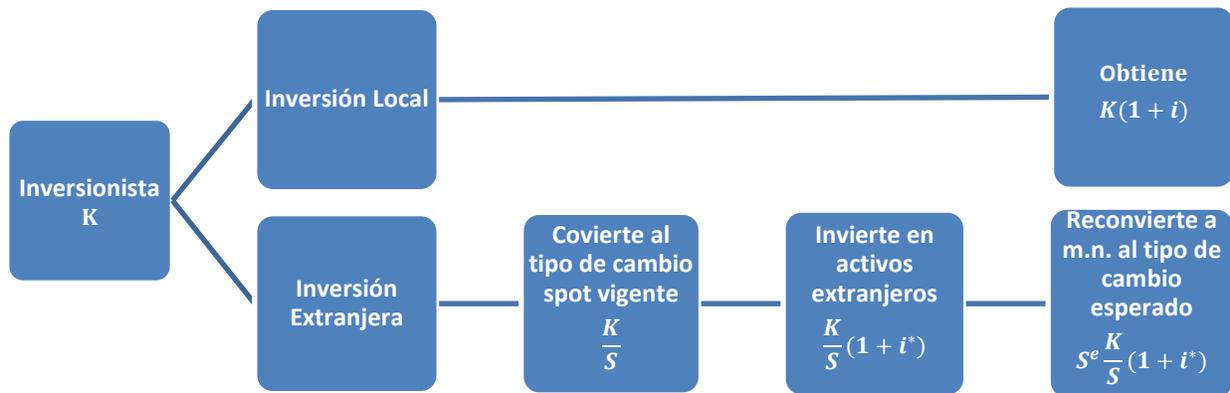
$$i - i^* = f \quad (1.4)$$

Lo que nos dice que en equilibrio, el diferencial de tasas de interés debe ser igual al diferencial forward. La ecuación (1.4) implica que la moneda que ofrece la tasa de interés más alta debe venderse a descuento forward, y viceversa. Esto es porque si $i > i^*$, entonces $f > 0$, lo que significa que la moneda extranjera (que ofrece una tasa de interés más baja) se vende a una prima forward, mientras que la moneda local (que ofrece una tasa de interés más alta) se vende con un descuento forward. Si, por el contrario $i < i^*$, entonces $f < 0$, lo que implica que la moneda extranjera se vende a descuento, mientras que la nacional se vende con prima.

La paridad de tasas de interés descubierta (UIP por sus siglas en inglés) es una operación similar a la anterior con la excepción de que la posición larga no está cubierta con un forward por lo que la conversión toma el tipo de cambio spot vigente a la fecha de vencimiento de la inversión. Naturalmente el tipo de cambio que estará vigente al vencimiento es desconocido por lo que la decisión de entrar en un arbitraje descubierto depende de la expectativa sobre el tipo de cambio al vencimiento S^e . Teniendo de nueva cuenta las dos opciones de inversión entre local y extranjera como se expresa en el cuadro 3 al obtener $K/S(1 + i^*)$ en la fecha de vencimiento de la inversión debe reconvertir a moneda nacional (expuesto al riesgo cambiario) con el tipo de cambio S^e .

⁴ Es la diferencia entre el tipo de cambio spot y el tipo de cambio forward como razón del tipo de cambio spot entonces $f = (F - S)/S$

Cuadro 1.3: Paridad de Tasas de Interés Descubierta



De manera similar a la condición CIP podemos escribir la condición UIP como

$$(1 + i) = \frac{S^e}{S}(1 + i^*) \quad (1.5)$$

La condición UIP nos dice que el rendimiento bruto nacional debe ser igual al rendimiento esperado bruto del exterior. El lado izquierdo de la ecuación (1.5) representa el rendimiento bruto nacional. Al simplificar la ecuación (1.5), obtenemos una aproximación a la condición UIP:

$$i - i^* = \dot{S}^e \quad (1.6)$$

Donde \dot{S}^e es la variación del tipo de cambio⁵ la cual es igual al diferencial de las tasas de interés. Esta condición nos indica entonces que la moneda extranjera que ofrece una tasa de interés más baja debe de esperar una apreciación, mientras que la moneda que ofrece una tasa de interés más alta debe de esperar una depreciación. Lo anterior es parte importante del pronóstico del tipo de cambio.

1.4 Enfoque de Activos

Este enfoque se caracteriza por funcionar con expectativas racionales y tiene dos versiones principales el modelo monetario y el de portafolio. El modelo monetario funciona con el supuesto de que existe perfecta sustitución de capitales lo que implica que no existe una prima de riesgo por lo tanto se cumple la paridad de tasas de interés descubierta. El otro modelo es el de portafolio que a diferencia del primero acepta que no hay perfecta sustitución de capitales lo que invariablemente genera una prima de riesgo por lo tanto se plantea la paridad de tasas de interés cubierta.

⁵ Es la diferencia entre el tipo de cambio spot y el tipo de cambio spot esperado como razón del tipo de cambio spot entonces $\dot{S}^e = (S^e - S)/S$

1.4.1 Modelo monetario

Puede ser visto como una extensión de la PPP ya que lleva la teoría monetaria de la determinación de precios al modelo para explicar las fluctuaciones nominales del tipo de cambio. Movimientos en los agregados monetarios determinan a los precios y por lo tanto el tipo de cambio nominal. Si establecemos la siguiente ecuación como la función de precios

$$P = P_d^\theta + P_f^{1-\theta} \quad (1.7)$$

Donde θ es la parte de costos domesticos y $1 - \theta$ la parte de costos extranjeros. En una pequeña economía abierta la ecuación (1.7) lleva a la ley de precio único

$$P = SP^* \quad (1.8)$$

Donde S es el tipo de cambio nominal P y P^* son los índices de precios nacional y extranjero respectivamente. Con base en PPP el tipo de cambio real puede definirse como

$$Q = S(P^*/P) \quad (1.9)$$

Si expresamos en tasas de crecimiento la ecuación se re-expresa como

$$q = s + p^* - p$$

Se asume que como PPP se mantiene (supuesto criticable del modelo) tenemos

$$s = p - p^* \quad (1.10)$$

Esta ecuación –ya conocida– muestra que el tipo de cambio varía de acuerdo al diferencial inflacionario entre las dos economías. Al considerar el equilibrio en los mercados monetarios doméstico y extranjero

$$\frac{M}{P} = L(i, Y) \quad (1.11)$$

$$\frac{M^*}{P^*} = L^*(i^*, Y^*) \quad (1.12)$$

Para capturar la base del mecanismo de transmisión en el modelo monetario, cualquier desequilibrio en la oferta y demanda de dinero se ve reflejado en las variaciones de los precios los cuales a su vez afectan el nivel y la variación del tipo de cambio nominal. Por lo que para capturar esto se expresa a (1.11) y (1.12) en tasa de crecimiento para obtener

$$p = m - l \quad (1.13)$$

$$p^* = m^* - l^*$$

Al sustituir en (1.10) y asumiendo la demanda de dinero como endógena obtenemos

$$l^* - l = \varphi(y^* - y) + \lambda(i - i^*); \varphi, \lambda > 0 \quad (1.14)$$

Finalmente si se sustituye a (1.14) en (1.13) se obtiene la

$$s = (m - m^*) + \varphi'(y - y^*) + \lambda(i - i^*); \varphi' < 0 \quad (1.15)$$

La ecuación (1.14) es la propuesta por Frenkel (1976), Mussa (1976) y Bilson (1978) conocido como el modelo monetario de precios flexibles y nos indica que los mecanismos de transmisión que mantienen en equilibrio el mercado monetario a través de cambio en el tipo de cambio son los siguientes:

Genera una depreciación en el tipo de cambio (Δs):

- Un incremento en la oferta monetaria (M1 o M2)
- Un incremento en el diferencial de tasas de interés

Genera una apreciación en el tipo de cambio (∇s)

- Un incremento en el diferencial de ingresos reales

Para Dornbusch (1976) la ecuación (1.15) sólo se mantiene en el largo plazo por lo que el diferencial de tasas de interés se sustituye por diferencial de inflación esperada ($\pi - \pi^*$) y en el caso de Frankel (1979) por la inflación subyacente. En el corto plazo el tipo de cambio puede desviarse de su nivel de largo plazo pero con la expectativa del mercado de que regrese a su tendencia de largo plazo. El modelo de Dornbusch es de sobrerreacción (*overshooting*) debido a que el mercado de bienes se ajusta más lento que el mercado de activos y esta diferencia es la que explica la sobrerreacción. En el corto plazo la brecha entre el nivel de corto plazo y su nivel de equilibrio se puede expresar como

$$s - \bar{s} = -\frac{1}{\theta} [(i - \pi)(i^* - \pi^*)] \quad (1.16)$$

Donde θ captura la reacción, si sustituimos (1.16) en la ecuación de largo plazo (1.15) tenemos

$$s = (\bar{m} - \bar{m}^*) + \varphi'(\bar{y} - \bar{y}^*) + \left(\lambda + \frac{1}{\theta}\right) (\bar{\pi} - \bar{\pi}^*) - \frac{1}{\theta}(i - i^*) \quad (1.17)$$

La ecuación (1.17) integra el equilibrio de largo plazo de las variables de oferta monetaria ingreso real e inflación (estado estacionario) y el ajuste de corto plazo vía.

1.4.2 Modelo de Portafolio

La otra parte del enfoque de activos es el modelo de portafolio, que tiene como conceptos seminales la teoría de portafolio de Markowitz (1952) y el teorema de separación de Tobin (1958). Se simplifica la elección entre bonos y acciones a sólo bonos, entonces el

inversionista determina la composición de su cartera de bonos, es decir, la proporción de bonos domésticos y extranjeros basándose en el rendimiento esperado y el riesgo.

El modelo de portafolio para una economía pequeña de Black (1973), Kouri (1976), Branson (1977) establece que al no existir sustitución perfecta entre los activos domésticos y extranjeros la diferencia entre el rendimiento local (i) y extranjero ($i^* - \dot{\hat{s}}/s$) determinará donde se invierte la riqueza (W), si en bonos locales (N) o extranjeros (F) tal y como se vio en el apartado 1.3. Por tener el supuesto de economía pequeña, los bonos nacionales se mantienen únicamente por los residentes, ya que el país es demasiado pequeño para que sus activos sean de interés para los inversionistas extranjeros. Dada esta suposición la inversión se compone como

$$W = N^d + sF^d \quad (1.18)$$

Donde la demanda se expresa de acuerdo a la teoría de selección portafolio

$$N^d = g\left(i - i^* - \frac{\dot{\hat{s}}}{s}\right)W \quad (1.19)$$

$$sF^d = h\left(i - i^* - \frac{\dot{\hat{s}}}{s}\right)W$$

Donde $g(\dots) + h(\dots) = 1$ por (1.18). Imponiendo la condición cantidades demandadas deben ser iguales a las cantidades ofrecidas (o) como condición de equilibrio, obtenemos

$$N^d = N^o, F^d = F^o \quad (1.20)$$

Al sustituir en (1.19) y dividiendo la segunda ecuación entre la primera, se obtiene

$$\frac{sF^o}{N^o} = \varphi\left(i - i^* - \frac{\dot{\hat{s}}}{s}\right) \quad (1.21)$$

Donde $\varphi(\dots)$ es la razón entre la función $h(\dots)$ y la función $g(\dots)$ De la ecuación (1.21) podemos expresar al tipo de cambio en función de otras variables

$$s = \frac{N^o}{F^o} \varphi\left(i - i^* - \frac{\dot{\hat{s}}}{s}\right) \quad (1.22)$$

La ecuación (1.22) muestra que el tipo de cambio puede considerarse como el precio relativo de los dos acervos de activos, ya que está determinado por las cantidades relativas de N^o y F^o dado el diferencial de tasas de interés corregido por las expectativas de variación del tipo de cambio $i - i^* - \frac{\dot{\hat{s}}}{s}$. El tipo de cambio se ajusta de manera instantánea con el fin de mantener a los mercados de activos (internacionales) en equilibrio. Si por ejemplo, se produce un aumento de la oferta de bonos extranjeros a los residentes del país (a cambio de moneda nacional) y las expectativas son estáticas $\dot{\hat{s}} = 0$ el aumento *ceteris*

paribus provoca una apreciación instantánea del tipo de cambio ya que ahora, los residentes estarán dispuestos a mantener (y demandar) una mayor cantidad de bonos extranjeros, sólo si el precio local que tienen que pagar estos bonos (i.e., el tipo de cambio) es menor (Frankel, 1984).

De esta manera los valores de $sF^d = sF^o$ se mantiene sin cambios, ya que todas las valores presentes del lado derecho de la ecuación (1.19) no se modifican, y el mercado de activos externos permanece en equilibrio ($F^d = F^o$) en un nivel mayor F y un nivel más bajo de s .

1.5 Crítica de Meese y Rogoff

A principios de la década de 1980 Nelson y Plosser (1982) encontraron que la mayoría de las series en económicas de Estados Unidos eran “*non-stationary stochastic processes with no tendency to return to a trend line*” (ibíd., p. 139). El año siguiente Richard Meese y Kenneth Rogoff motivados por los resultados recientes y la poca efectividad que tenían los modelos de la década de 1970 para explicar los movimientos del tipo de cambio, realizaron un par de artículos (1983a, 1983b) donde tomando como modelo de referencia un paseo aleatorio compararon los modelos monetarios anteriormente revisados planteando la siguiente forma estructural:

$$s = \alpha_0 + \beta_1(m - m^*) + \beta_2(y - y^*) + \beta_3(i - i^*) + \beta_4(\bar{\pi} - \bar{\pi}^*) + \beta_5\overline{SBC} + \beta_6\overline{SBC}^* + u \quad (1.23)$$

La ecuación (1.23) contiene a los dos modelos monetarios con precios flexibles se espera que $\beta_4 = \beta_5 = \beta_6 = 0$ para el modelo de Dornbusch-Frankel $\beta_5 = \beta_6 = 0$ y por último agregan una parte que considera el saldo de la balanza comercial correspondiente al modelo de Hooper y Morton (1982), para los tres modelos se espera que $\beta_1 = 1$.

El modelo lo estimaron con mínimos cuadrados ordinarios y en sus resultados entregaba buen ajuste dentro de la muestra⁶ aunque con algunos problemas estadísticos. El contraste lo hacían con un autorregresivo largo $AR(M)$ siendo los rezagos determinados como $M = N/\log N$, un paseo aleatorio con deriva⁷ entre otros y en los pronósticos fuera de la muestra en ambos artículo demostró pobres resultados.

Frenkel (1985) en esta línea y a la luz de los resultados anteriores comentaba “*When we thought that the purchasing-power-parity model worked, it collapsed; when we thought that the simple monetary model worked, it failed; when we thought that richer portfolio-balance model worked, it also failed; when we turned to the current-account model, we did not get much help and so on and so forth. In fact, as a first approximation, exchange rates seem to follow a random walk. Therefore, by and large, changes in exchange rate (aside for trends) are unforecastable*” (p. 37).

⁶ Con observaciones mensuales de marzo de 1973 a junio de 1981.

⁷ Ambos modelos entran en la categoría de procesos autorregresivos que se ven a detalle en capítulo 2 dentro de la metodología de Box y Jenkins.

Las respuestas generadas como consecuencia de los resultados obtenidos fueron de dos tipos: la primera, continuar con estos modelos pero empleando métodos de estimación más sofisticados y la segunda, el fallo de los modelos estructurales como un fallo de la teoría económica abandonando el análisis de los fundamentales económicos para dedicarse al análisis técnico puro para estudiar patrones en los datos y realizar pronósticos sin fundamento. A finales de la década de 1980 Frankel y Froot (1988, 1990) se preguntaron quién maneja el mercado e identificaron esta dicotomía en la postura para estudiar el tipo de cambio “*we now turn to the question of how the existence of different forecasting techniques might lead to «excess volatility.»*” (Frankel y Froot, 1990) asegurando que el análisis técnico tiene sus expectativas en el corto plazo y el análisis fundamental en el largo plazo con la teoría económica de sustento.

Para Frankel y Froot (1988) los administradores de fondos tomaban en cuenta de manera ponderada la expectativas tanto de analistas técnicos como fundamentales y actuaban en consecuencia en el mercado donde ellos determinaban el valor del mismo como

$$\Delta s_{t+1}^m = c(\omega_t \Delta s_{t+1}^{af} + (1 - \omega_t) \Delta s_{t+1}^{at}) + Z_t \quad (1.24)$$

Donde ω_t es el peso que se le da a las expectativas de los analistas fundamentales y $1 - \omega$ a los analistas técnicos y Z_t es el vector de información contemporánea. Un factor importante es que los analistas técnicos al observar los mismo datos y gráficas forman expectativas *bandwagon* de manera que generan una mayor volatilidad en el tipo de cambio que los analistas fundamentales ya que, estos últimos ponderan de manera diferente sus expectativas tomando en cuenta un marco teórico en particular.

Ya a principios de la década de 1990 los modelos de factores determinantes del tipo de cambio, por la volatilidad del corto plazo y el pobre rendimiento de las estimaciones habían dirigido su atención a los desequilibrios comerciales y fiscales de los países y así determinar un posible valor de equilibrio que permitiera medir que tan desviado estaba el tipo actual.

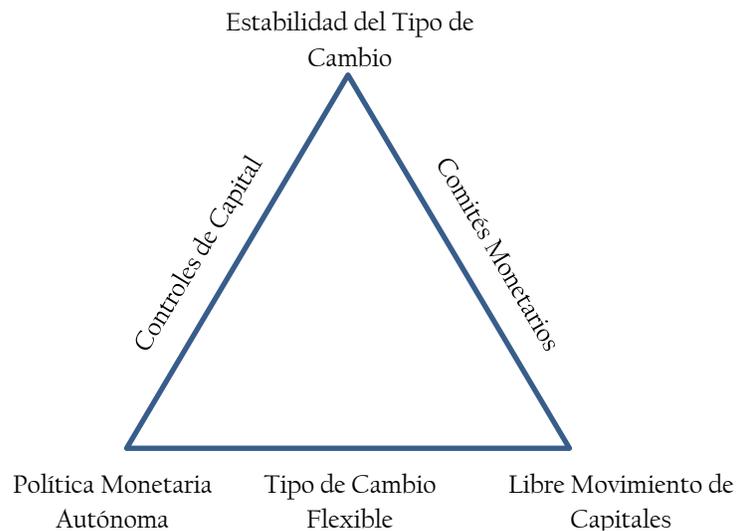
Estos modelos analizan una serie de cuestiones metodológicas que surgen del cálculo del tipo de cambio de equilibrio, identifican como los tipos de cambio efectivos reales que son consistentes con el equilibrio macroeconómico interno y equilibrio externo. Caracterizados en parte por sus nombres FEER (*Fundamental Equilibrium Exchange Rate*), DEER (*Desirable Equilibrium Exchange Rate*), BEER (*Behavioural Equilibrium Exchange Rate*) y el NATREX (*Natural Exchange Rate*). Durante la segunda mitad de la década con las crisis ya mencionadas estos modelos perdieron importancia⁸ dando lugar a modelos de movimientos de capital y especulación. Uno de estos fue realizado por Kaminsky *et. al.* (1998) quienes llevaron a cabo un estudio para determinar que variables estaban detrás o por lo menos

⁸ Si el lector lo desea puede seguir el desarrollo de los modelos FEER y DEER en Bayoumi, T. et. al. (1994).

podría dar señales de una posible crisis y encontraron que “[...] *useful indicators of currency crises include international reserves, the real exchange rate, credit growth, credit to the public sector and domestic inflation. The results also provide support for the trade balance, exports performance, money growth, M2/international reserves, real GDP growth, and the fiscal deficit.*” (ibíd., 1998).

En las economías abiertas se enfrenta un problema de elección conocido como trilema (Obstfeld, *et. al.* 2008) (ver gráfica 1.4) lo que significa que no se pueden alcanzar los tres picos del triángulo sólo combinaciones de dos. Al permitir la liberalización de sus cuentas de capital las economías quedan expuestas y es por eso que la literatura respecto al tipo de cambio en estos días versa más sobre controles de capital, niveles óptimos de reservas (Calvo, *et. al.*, 2012) y la posibilidad de ser víctima de crisis cambiaria por salidas de capital.

Gráfica 1.4: Trilema Político de Economías Abiertas



México ha actuado en consecuencia con su política de incrementar las reservas por intervención directa en el mercado o por venta de opciones *put* y la línea de crédito flexible con el FMI crea una cobertura ante los problemas recién mencionados

Cap. 2 Análisis de Series de Tiempo

“Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful”

George Box⁹

“Don’t never prophesy: If you prophesies right, ain’t nobody going to remember and if you prophesies wrong, ain’t nobody going to let you forget”

Mark Twain

El capítulo presenta las características de las series de tiempo y los métodos de pronóstico que son considerados como punto de referencia para quien realiza dicha actividad. La principal diferencia entre ellos es el tiempo necesario para realizar una estimación adecuada. Los modelos que se consideran dentro del primer bloque brindan la posibilidad de realizar pronósticos de forma rápida y sencilla, razón por la cual siguen teniendo vigencia. Con esto en mente las estimaciones contenidas en este capítulo se pueden realizar en una tradicional hoja de cálculo, si bien existen programas estadísticos y econométricos que cuentan con rutinas para ello, no siempre se tiene acceso a los mismos.

2.1 Componentes de una Serie de Tiempo

Una serie de tiempo se puede definir como una sucesión de observaciones o datos correspondientes a una variable que se generan en distintos momentos de tiempo que por lo general ocurren en intervalos regulares. Así cada serie tiene una periodicidad, variables como el precio diario de una acción, los ingresos trimestrales de una compañía o la población anual de un país son algunos ejemplos. Por lo tanto el análisis de las series de tiempo se realiza para describir la evolución de la variable respecto a su pasado con la finalidad -en muchos casos- de pronosticar sus valores a futuro (Rodríguez, 2000).

Los enfoques empleados en el análisis de series de tiempo son: el análisis clásico o también conocido como enfoque determinista y el moderno o enfoque estocástico. Aunque ambos enfoques se abordan en este capítulo dejamos el enfoque estocástico para más adelante.

De acuerdo al enfoque clásico de series de tiempo toda serie se constituye por 4 componentes:

Tendencia (T_t). Se asocia a la dirección general que presenta la gráfica de una serie de tiempo. La tendencia puede manifestarse en forma de una recta, exponencial o de una función más compleja.

Variaciones cíclicas (C_t). Se refiere a las oscilaciones de larga duración, usualmente mayores a un año, alrededor de la tendencia. En economía los ciclos se asocian a períodos

⁹ Box G., y N. Draper (1987). *In Empirical Model-building and Response Surfaces*. Wiley & Sons

de recesión recuperación y expansión. Por su larga duración se requieren de muchas observaciones para identificar su presencia.

Variaciones estacionales (S_t). Son las regularidades que se presentan en el corto plazo (menor al año) y su frecuencia puede ser de trimestral, mensual, semanal o diaria, suelen distinguirse de manera más clara en variables como el producto interno bruto, las ventas de una empresa o el consumo de energía eléctrica.

Variaciones aleatorias (ε_t). Son respuestas a la presencia de factores no cuantificados o desconocidos que influyen en el comportamiento del fenómeno.

Como lo establecen Makridakis y Wheelright (1983) el modelo básico de una serie de tiempo es $y_t = f(T_t, C_t, S_t, \varepsilon_t)$ y se puede presentar con las siguientes formas funcionales:

$$\text{Aditiva: } y_t = T_t + C_t + S_t + \varepsilon_t \quad (2.1)$$

$$\text{Multiplicativa } y_t = T_t \times C_t \times S_t \times \varepsilon_t \quad (2.2)$$

$$\text{Mixta: } y_t = T_t + C_t \times S_t + \varepsilon_t \quad (2.3)$$

Con la finalidad de ejemplificar el procedimiento de cada uno de los métodos que se tratan en el capítulo se generó una serie de tiempo (ver. Cuadro 2.1) con la intención de contar con una serie donde se puedan apreciar los cuatro componentes de manera más clara.

Cuadro 2.1: Ventas de “Empresa X” 2005.01 – 2012.12
(cifras en miles de pesos)

Mes	2005	2006	2007	2008	2009	2010	2011	2012
Ene	507	1,053	1,371	831	636	1,230	1,506	1,026
Feb	381	930	1,209	648	501	1,089	1,344	792
Mar	705	1,302	1,557	876	885	1,434	1,662	1,008
Abr	942	1,551	1,737	1,071	1,140	1,683	1,902	1,185
May	1,029	1,569	1,713	1,047	1,173	1,695	1,887	1,161
Jun	1,035	1,605	1,689	981	1,218	1,725	1,827	1,170
Jul	1,131	1,647	1,650	1,005	1,266	1,788	1,803	1,134
Ago	1,038	1,530	1,455	870	1,212	1,683	1,647	990
Sep	1,044	1,509	1,344	786	1,200	1,614	1,488	948
Oct	1,326	1,722	1,479	993	1,425	1,839	1,638	1,122
Nov	1,320	1,701	1,341	945	1,458	1,878	1,482	1,116
Dic	2,022	2,331	1,938	1,632	2,163	2,493	2,070	1,734

Fuente: Datos generados por el autor

2.2 Métodos de Suavizamiento o Alisado

2.2.1 Media Móvil (Simple y Doble)

La media móvil es la herramienta más sencilla que tenemos para suavizar los datos históricos además de constituir un elemento necesario para el desarrollo de los siguientes métodos de suavizamiento.

La media móvil simple (S_t) de orden n se expresa como:

$$S_t = \frac{X_{t-n+1} + \dots + X_{t-2} + X_{t-1} + X_t}{n} \quad (2.4)$$

De igual forma la media móvil doble (D_t) de orden m se expresa como:

$$D_t = \frac{S_{t-m+1} + \dots + S_{t-2} + S_{t-1} + S_t}{m} \quad (2.5)$$

Ambas constituyen un promedio aritmético para una muestra que se va recorriendo, eliminando el último y agregando el más reciente de manera que siempre tiene el mismo orden. La diferencia más notoria es que la media móvil simple se aplica a la variable original y la doble se aplica a la simple.

La cuestión más relevante es el criterio para elegir el orden de la media móvil ya que a un orden mayor la serie original se suavizará aún más (*i.e.* cuando $n = t$ la media móvil es el promedio aritmético de toda la serie). Al final el criterio está alineado con las características de la serie tales como la periodicidad (mensual o trimestral), si el factor aleatorio es predominante o lo es el ciclo, etc.

El método lineal de medias móviles se construye con (2.4) y (2.5) como las entradas del modelo que al buscar ajustar la tendencia lineal α_t funge como la ordenada al origen y β_t como la pendiente:

$$\alpha_t = S_t + (S_t - D_t) \quad (2.6)$$

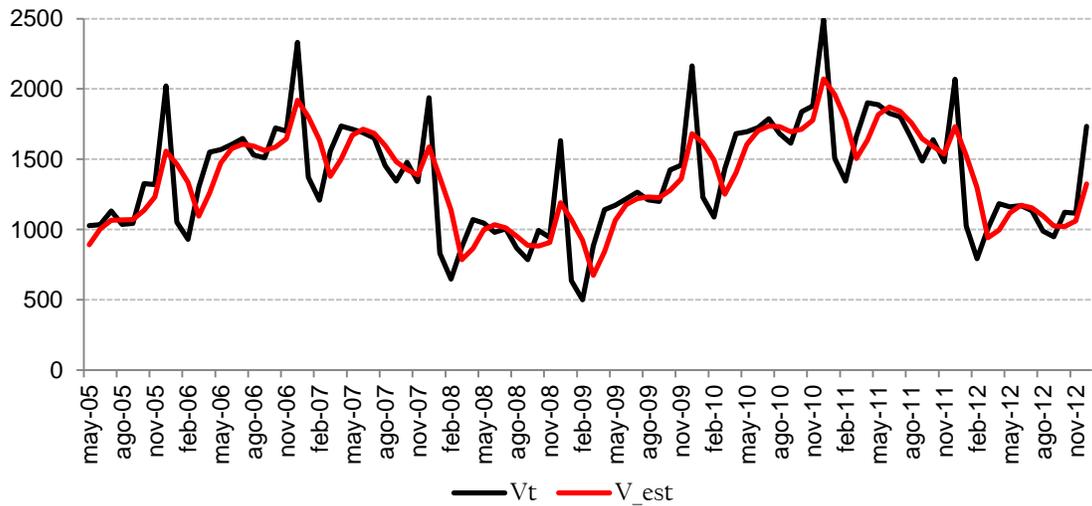
$$\beta_t = \frac{2}{n-1} (S_t - D_t) \quad (2.7)$$

$$F_{t+m} = \alpha_t + \beta_t m \quad (2.8)$$

Empleando la serie de ventas (V_t) realizamos una media móvil simple de orden $n = 3$ y una media móvil doble de orden $m = 3$. Obteniendo $S_1 = \frac{507+381+705}{3} = 531$, dato correspondiente a marzo 2005, el siguiente $S_2 = \frac{381+705+942}{3} = 676$, es la observación para abril 2005 y por último la correspondiente a mayo 2005 $S_3 = \frac{705+942+1029}{3} = 892$. Con los resultados obtenidos de S_1, S_2, S_3 calculamos $D_1 = \frac{531+676+892}{3} = 699.97$. Para obtener las

ventas estimadas para el mes de mayo se sustituyen los valores en (2.6), (2.7) y (2.8) y se resuelve para obtener ventas en mayo de 2005 por 829.7 como se observa en la gráfica (2.1) el modelo subestima las ventas de este mes por casi 130 mil pesos pero sigue adecuadamente el ciclo y tendencia de las ventas.

Gráfica 2.1; Ajuste Histórico del Método de Promedios Móviles Lineales. 2005.5-2012.12



Fuente: Elaboración propia con datos del cuadro 2.1

2.2.2 Suavizamiento Exponencial Simple

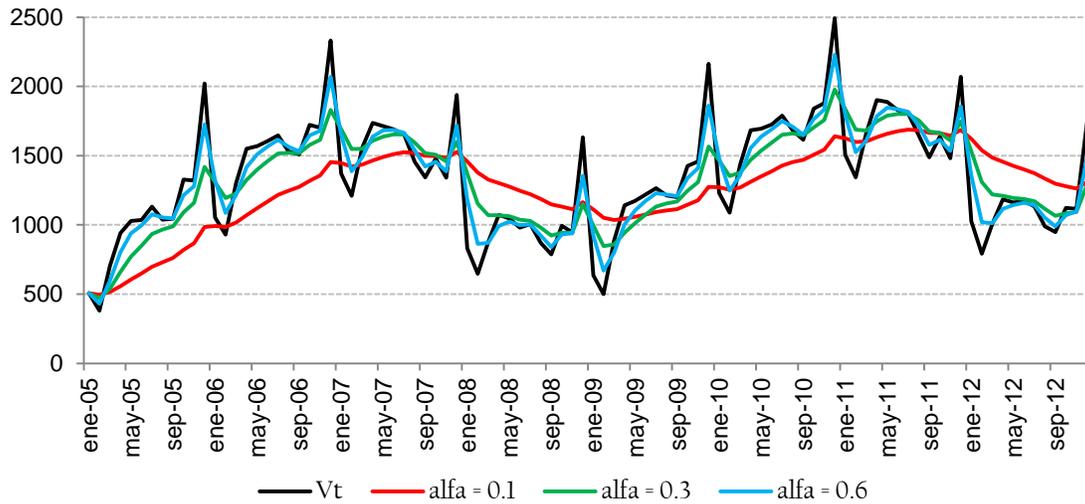
El suavizamiento exponencial posee una característica que las medias móviles no tienen: la ponderación diferente para cada observación. Esto puede representar una ventaja a la hora de emitir un pronóstico ya que los valores más recientes serán los valores con mayor relevancia. Expresando el exponencial simple como:

$$F_t = \alpha Y_t + (1 - \alpha) F_{t-1}; 0 \leq \alpha \leq 1 \quad (2.12a)$$

Por lo tanto la estimación de la serie depende del factor de alisado α y en menor medida de F_1 . Si α es cercano a 1 los valores recientes de la serie tendrán un mayor peso por lo que la estimación se ajustará más a los valores recientes generando menor suavizamiento, tal y como se observa en la gráfica 2.2 con $\alpha = 0.6$.

Respecto a la determinación de F_1 existen distintas maneras de establecer la primera estimación: se puede considerar que sea igual a la media muestral o a la media de las primeras observaciones, también se puede determinar que sea igual al valor de la primera observación de la serie a estimar. Por simplicidad, en nuestro caso se establece que el primer valor de la estimación sea igual al valor observado es decir $F_1 = 507$ para la siguiente estimación $F_2 = (0.1)381 + (1 - 0.1)507 = 494.4$ y así sucesivamente.

Gráfica 2.2; Ajuste Histórico con Suavizamiento Exponencial. 2005.1-2012.12



Fuente: Elaboración propia con datos del cuadro 2.1

2.2.3 Suavizamiento Exponencial Doble:

Brown (1956) en el periodo creciente de la investigación de operaciones buscando aplicaciones de control de inventarios generó una rutina de pronóstico que de igual forma se pudiera aplicar en otras actividades como el proceso de inspección de calidad. Atacando el defecto del suavizamiento exponencial simple de subestimar sensiblemente la serie si ésta presenta una tendencia creciente, el método de Brown utiliza un doble suavizamiento exponencial buscando incorporar la presencia de la tendencia. La diferencia entre las dos series suavizadas es ponderada por dos coeficientes identificados como a_t y b_t . La ecuación (2.12a) la re-expresamos como:

$$F'_t = \alpha Y_t + (1 - \alpha)F'_{t-1}; \quad 0 \leq \alpha \leq 1 \tag{2.12b}$$

El suavizamiento exponencial doble –de igual manera que con las medias móviles– se aplica sobre la serie de exponencial simple como se expresa en la ecuación(2.13):

$$F''_t = \alpha F'_t + (1 - \alpha)F''_{t-1}; \quad 0 < \alpha < 1 \tag{2.13}$$

La predicción adopta la forma de una recta con la ecuación (2.14) como ordenada al origen y la ecuación (2.15) como la pendiente aplicada al número del período que se pretende pronosticar.

$$a_t = F'_t + (F'_t - F''_t) \tag{2.14}$$

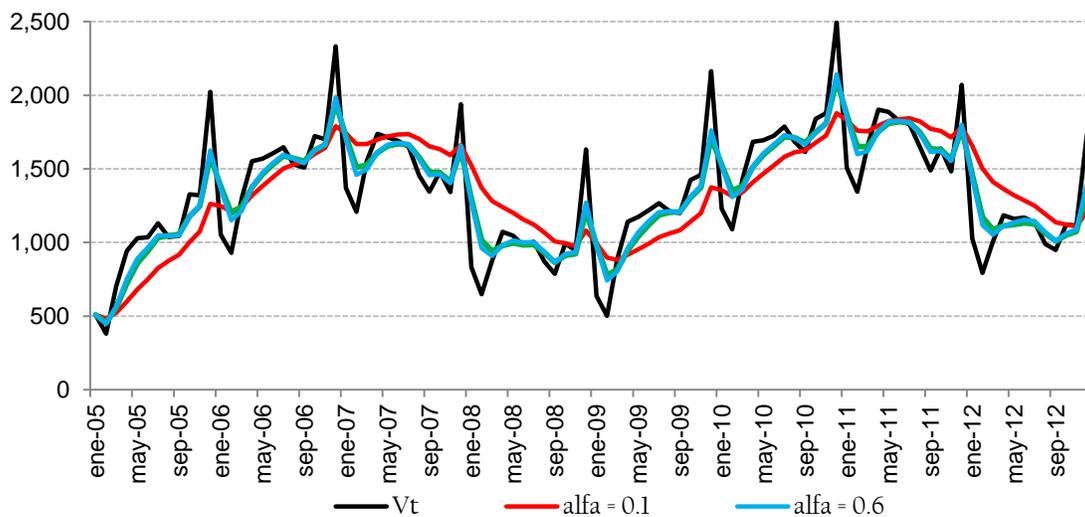
$$b_t = \frac{\alpha}{1-\alpha} (F'_t - F''_t) \tag{2.15}$$

$$F_{t+m} = a_t + b_t m \quad (2.16)$$

Un punto importante de este método es como destaca Brown “*The estimates are not statistically efficient, but they are economically efficient ...*”(ibíd, p.1). Observación que no debe perderse de vista ya que, al decidir que instrumental emplear para realizar el estudio de una serie de tiempo el tiempo y la información con la que se cuenta es determinante para elegir entre modelos más robustos y con mayor complejidad o modelos eficientes y de sencilla aplicación.

Al realizar el suavizamiento exponencial simple cuando α es cercano a 1 da mayor relevancia a datos recientes de la serie original, de la misma forma que lo hace al realizar la el proceso exponencial sobre F'_t . En la gráfica 2.3 se observa que con α mayor se obtiene un mejor ajuste, lo que no se puede generalizar para todos los casos, series que tienen un comportamiento más volátil al final de la muestra podrían no ajustar de igual forma que este caso.

Gráfica 2.3; Ajuste Histórico con Método de Brown. 2005.1-2012.12



2.2.4 Método de Holt-Winters de tres parámetros

A finales de la década de 1950 Charles Holt (1957) realizó un método similar al de Brown con un suavizamiento simple como en la ecuación (2.12) pero para capturar la tendencia lineal utiliza otra exponencial simple sobre la primera con un parámetro adicional β , generando una ecuación de nivel y otra de pendiente.

Pocos años más adelante Peter Winters (1960) alumno de Holt, incorpora una tercera ecuación para para la estacionalidad. Su procedimiento incluye 3 parámetros de alisado α , β , γ para la serie, la tendencia y la estacionalidad. Para este método es importante determinar de qué manera se agregan los componentes de la serie, si de forma aditiva o

multiplicativa. Si suponemos que se agregan de forma aditiva (HWA) se emplean las siguientes cuatro ecuaciones:

$$F_t = L_t + T_t + S_{t-s} \quad (2.17a)$$

$$L_t = \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (2.18a)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (2.19a)$$

$$S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-s} \quad (2.20a)$$

Si de manera opuesta se realiza el supuesto que la serie se agrega de manera multiplicativa (HWM) se emplean las siguientes ecuaciones:

$$F_t = (L_t + T_t)S_{t-s} \quad (2.17b)$$

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (2.18b)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (2.19b)$$

$$S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-s} \quad (2.20b)$$

Sea la opción que sea para poder implementar el método adecuado el usuario debe de brindar los valores iniciales de L_t, T_t, S_t además de elegir los valores de α, β, γ . Para realizar lo anterior Chatfield y Yar (1988) recomiendan que para el valor inicial de L_t se tome el promedio de los dos o tres primeros años de la serie y el valor inicial de T_t sea 0 y establecer S_t como la razón entre $\frac{Y_t}{L_t}$. Respecto a las constantes de suavizamiento reconocen que tanto α como γ generalmente son mayores a 0.3 pero β es regularmente menor a 0.1. Para el ejercicio tomamos el promedio del primer año y al valor inicial de $T_t = 0$ y como constantes de alisado $\alpha = 0.1, \beta = 0.1, \gamma = 0.7$ entonces:

$$L_1 = \frac{507+381+705+\dots+2022}{12} = 1040$$

$$S_1 = \frac{507}{1040} = 0.4875$$

Si asumimos HWA entonces $F_1 = 1040 + 0 + 0.4875 = 1040.49$ y si por el contrario asumimos HWM entonces $F_1 = (1040 + 0) \times 0.4875 = 507$. Para la siguiente observación tenemos desarrollamos HWM

$$L_2 = 0.1 \frac{381}{0.4875} + (1 - 0.1)(1040 + 0) = 1014.15$$

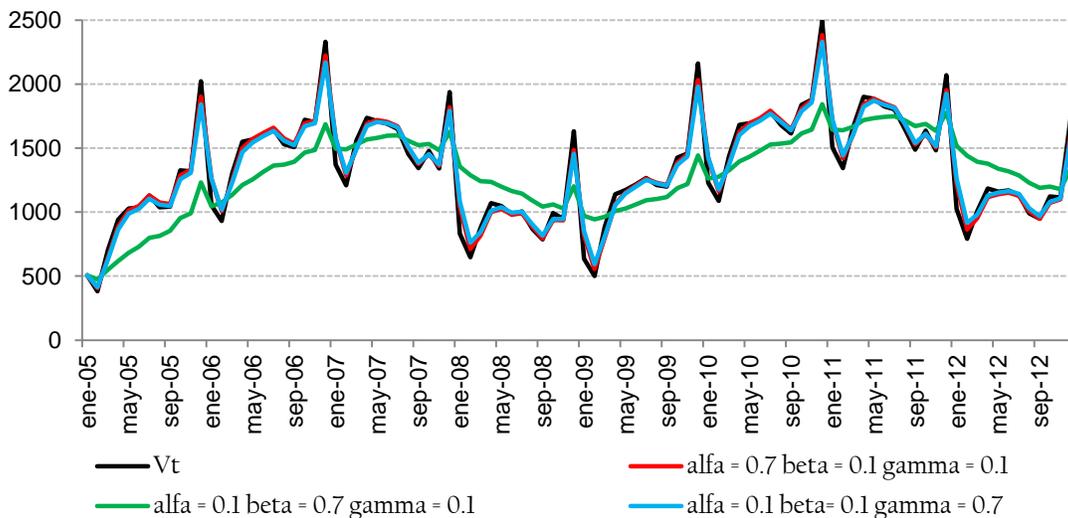
$$T_2 = 0.1(1014.15 - 1040) + (1 - 0.1)0 = -2.58$$

$$S_2 = 0.7 \frac{381}{1014.15} + (1 - 0.7)4875 = 0.4092$$

$$F_2 = (1014.15 - 2.58) \times 0.4092 = 413.96$$

En la gráfica 2.4 se puede observar el ajuste con diferentes combinaciones de constantes de suavizamiento incluida el ajuste con los datos que establecieron. La evidencia favorece a la observación realizada por Chatfield y Yar con α como γ mayores a 0.3 se generó mejor estimación que con β mayor que 0.3. Cuando $\beta = 0.1$ en combinación con α, γ altos se dieron los mejores resultados.

Gráfica 2.4: Ajuste Histórico con Método de Holt Winters. 2005.1-2012.12



Por último, Winters en su artículo -igual que Brown- hace referencia al hecho de que los pronósticos tienen que hacerse en algunas empresas de manera continua y por lo tanto menciona “*There are certain desirable characteristics of these forecasts that are implied by their use: they must be made quickly, cheaply, and easily*” (Ibid., p. 325). Característica con la que cumplen la mayoría de los modelos presentados hasta aquí y razón por la cual se convierten en modelos de referencia para comparar métodos más avanzados.

2.2.5 Método de Descomposición

El método de descomposición aunque *per se* no es un método de suavizamiento, forma parte del enfoque clásico. Como su nombre da a entender consiste descomponer la serie por lo que se necesita identificar cada componente y restarse de la serie original. Se empieza con la tendencia, para después seguir con el componente estacional y se concluye con el componente cíclico. Todo aquello que no entra de los primeros tres corresponde al componente aleatorio.

Para desarrollar el proceso de descomposición clásica empleamos la serie del cuadro (2.1). Primero, para eliminar la tendencia es necesario realizar una regresión lineal de la serie, tomando como variable dependiente las ventas y como variable independiente el tiempo (t), de Enero 2005 ($t = 1$) hasta Diciembre 2012 ($t = 96$) y se obtiene la siguiente ecuación

$$\hat{y}_t = 1210.23 + 2.79 t \quad (2.21)$$

Al sustituir $t = 1$ en el valor estimado \hat{y}_1 obtenemos $1210.23 + 2.79 (1) = 1213.02$ esto se realiza sucesivamente para todas las observaciones. De acuerdo con la forma funcional de la serie, se realiza lo siguiente: si es aditiva, a las ventas se le resta la serie estimada por la regresión $501 - 1213.02 = -706.02$. Por el contrario si es multiplicativo, se divide las ventas entre las ventas estimadas $\frac{501}{1213.02} = 0.42$ de manera que únicamente lo que se realiza es eliminar la tendencia de la serie.

Después de obtener los valores de la serie sin tendencia, para cada uno de los meses se obtiene el promedio, por ejemplo para el mes de Enero tenemos en la aditiva $Ene = \frac{-706.01 - 193.48 + \dots + 92.17 - 421.30}{12} = -310.16$ y en la multiplicativa $Ene = \frac{0.42 + 0.84 + \dots + 1.07 + 0.71}{12} = 0.76$, valores que representan para sus respectivas series, el componente estacional para el mes de Enero. Para obtener el componente cíclico se resta el componente estacional a la serie sin tendencia. Para obtener Enero en el caso aditivo $-706.02 - (-310.16) = -395.86$ y en el caso multiplicativo $Ene = \frac{0.42}{0.76} = 0.55$ y así sucesivamente para el resto de las observaciones.

Por último el componente aleatorio, para obtener este componente necesitamos analizar el componente cíclico, como se puede observar en la gráfica 2.5 (a) se distinguen claramente dos grandes ciclos de 2005.1 a 2008.12 y de 2009.1 a 2012.12 lo que procede después es obtener el promedio de cada observación correspondiente al ciclo 1 y ciclo 2 y restarlo (o dividirlo) del componente cíclico. Para ambos, tanto aditivo como multiplicativo se obtiene el promedio de enero de 2005 y enero de 2009 puntos iniciales de cada ciclo obteniendo para el aditivo $Ene = \frac{-395.86 - 400.73}{2} = -398.30$ para el multiplicativo $Ene = \frac{0.55 + 0.62}{2} = 0.58$. Para calcular la observación aleatoria de Enero se resta (divide) del componente cíclico el promedio.

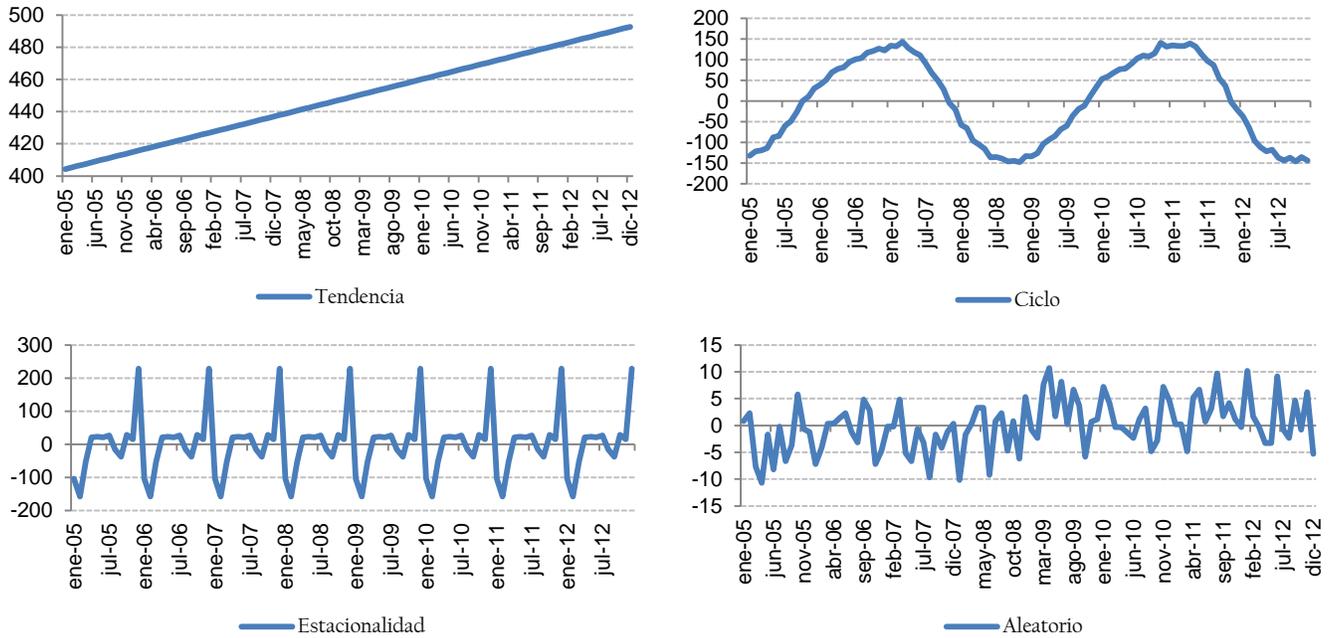
Por último se compone de nueva cuenta la serie para poder realizarlo recordando de las ecuaciones (2.1) y (2.2) el valor estimado para Enero de acuerdo a la composición aditiva es:

$$Ene_{aditivo} = 1213.02 + (-395.86) + (-310.96) + 2.44 = 509.44$$

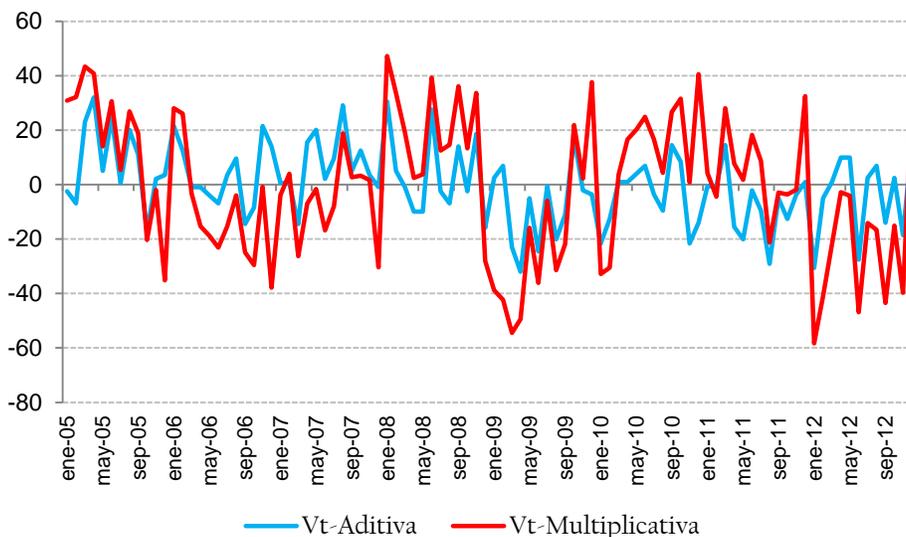
$$Ene_{multiplicativo} = 1213.02 \times 0.5 \times 0.76 \times 0.94 = 476.11$$

Aunque ambos modelos ajustan la serie original muy bien en la gráfica 2.5 (b) se puede ver que los errores del modelo aditivo son menores y tienen una varianza más constante que los del modelo multiplicativo. Para el año de 2009 se observa una subestimación mayor a la del resto de la muestra probablemente por el ajuste del nuevo ciclo que comienza en Enero de 2009. De acuerdo a las irregularidades observadas al momento de realizar el pronóstico todo parece indicar que utilizar el aditivo es la mejor opción.

Gráfica 2.5A: Componentes de Ventas de “Empresa X” 2005.01 - 2012.12



Gráfica 2.5B: Error de la estimación: 2005.01 - 2012.12



Como se mencionó al principio del capítulo el enfoque clásico que ya hemos revisado se caracteriza por analizar las series de tiempo de manera determinística, por otro lado dentro del llamado enfoque moderno son procesos estocásticos. Un proceso estocástico es sucesión de variables aleatorias, en correspondencia con una sucesión de valores de la variable de tiempo $\{Z_t; t \in T\}$ en donde T es el conjunto índice de número reales que representan la variable tiempo y Z_t es la variable aleatoria correspondiente al momento t de T . El proceso es continuo cuando T es un intervalo de número reales (ya sea cerrado o abierto) y es discreto cuando es un conjunto (finito o infinito) numerable (Guerrero, 2003).

Un proceso estocástico es estacionario si su media y varianza son constantes en el tiempo y si el valor de la covarianza entre dos periodos depende sólo de la distancia o rezago entre estos ambos periodos. Estrictamente será estacionaria si todos los momentos de su distribución no varían en el tiempo, aunque es claro que si se trata de un proceso con distribución normal estará completamente especificado por los dos primeros momentos (Gujarati y Porter, 2010) El proceso estacionario tiene las siguientes propiedades:

$$E(Z_t) = \mu \quad (2.22)$$

$$Var(Z_t) = E(Z_t - \mu)^2 = \sigma^2 \quad (2.23)$$

$$Cov(Z_t, Z_{t+h}) = E[(Z_t - \mu)(Z_{t+h} - \mu)] = \gamma_h \quad (2.24)$$

Donde γ_h es la covarianza entre los valores Z_t y Z_{t+h} ambos valores de Z separados h periodos de ahí que se le conoce como autocovarianza, si $h = 0$, γ_0 es la varianza de Z ; si $h = 1$ es la covarianza entre valores adyacentes.

Otra propiedad importante de los procesos estocásticos es que deben tener incrementos estacionarios, si W_t es un proceso que representa los incrementos de Z_t entonces $W_t = Z_t - Z_{t-1}$ también es estacionario, algebraicamente tenemos:

$$\begin{aligned} E(W_t) &= E(Z_t - Z_{t-1}) \\ &= E(Z_t) - E(Z_{t-1}) \\ &= \mu - \mu \\ &= 0 \end{aligned}$$

$$\begin{aligned} Var(W_t) &= Var(Z_t - Z_{t-1}) \\ &= Var(Z_t) + Var(Z_{t-1}) - 2Cov(Z_t, Z_{t-1}) \\ &= \sigma^2 + \sigma^2 - 2\gamma_1 \\ &= 2\sigma^2 - 2\gamma_1 \end{aligned}$$

$$= 2(\sigma^2 - \gamma_1)$$

$$\begin{aligned} Cov(W_t, W_{t+h}) &= E[(W_t - \mu)(W_{t+h} - \mu)] \\ &= E[W_t W_{t+h}] \\ &= E[(Z_t - Z_{t-1})(Z_{t+h} - Z_{t+h-1})] \\ &= E(Z_t Z_{t+h} - Z_t Z_{t+h-1} - Z_{t-1} Z_{t+h} + Z_{t-1} Z_{t+h-1}) \\ &= E(Z_t Z_{t+h}) - E(Z_t Z_{t+h-1}) - E(Z_{t-1} Z_{t+h}) + E(Z_{t-1} Z_{t+h-1}) \\ &= \gamma_h - \gamma_{h-1} - \gamma_{h+1} + \gamma_h \\ &= 2\gamma_h - \gamma_{h-1} - \gamma_{h+1} \end{aligned}$$

Como $Cov(W_t, W_{t+h})$ depende sólo de h y la media es constante por lo tanto el incremento es también un proceso estacionario.

Una manera de probar la estacionariedad es la función de autocorrelación, como lo menciona Gujarati al contar únicamente con una realización de un proceso estocástico (i.e. la muestra) sólo se puede calcular la función de autocorrelación muestral $\hat{\rho}_h$. Para estimar $\hat{\rho}_h$ debemos calcular la covarianza muestral en el rezago h ($\hat{\gamma}_h$), y la varianza muestral, ($\hat{\gamma}_0$) (ibíd., p.749).

$$\hat{\gamma}_h = \frac{\sum_{t=h+1}^T (Z_t - \bar{Z})(Z_{t+h} - \bar{Z})}{T-h} \quad (2.25)$$

$$\hat{\gamma}_0 = \frac{\sum_{t=1}^T (Z_t - \bar{Z})^2}{T} \quad (2.26)$$

$$\hat{\rho}_h = \frac{\hat{\gamma}_h}{\hat{\gamma}_0} \quad (2.27)$$

Sin importar que valor tome h si se obtienen valores de $|\hat{\rho}_h|$ cercanos a 1 se está en presencia de un proceso no estacionario. En orden de saber si los procesos son estacionarios existen más pruebas conocidas como pruebas de raíz unitaria, pero para entenderlas de mejor manera se necesita del operador rezago y diferencia.

Se identifica al operador de retraso por la letra L (del término en inglés *lag* o B de *backward*) y se define como el valor retrasado de la serie indicado por el exponente del operador:

$$LZ_t = Z_{t-1} \quad \forall t \quad (2.28)$$

Al aplicar el operador de manera sucesiva

$$L^2 Z_t = L(LZ_t) = Z_{t-2}$$

... ..

$$L^k Z_t = L(L^{k-1} Z_t) = Z_{t-k}$$

El otro operador importante es el de diferencia (∇ o D) tal como se advierte por el nombre es la diferencia entre el valor correspondiente a t y el valor retrasado k períodos

$$\nabla Z_t = Z_t - Z_{t-1} \tag{2.29}$$

Ya que a la ecuación (2.29) se le puede expresar como $\nabla Z_t = Z_t - LZ_t$ el lado derecho de la ecuación es igual a $(1 - L)Z_t$ que es la relación entre el operador de retraso y de diferencia es $\nabla = (1 - L)$ por lo que:

$$\nabla Z_t = (1 - L)Z_t$$

$$\nabla^2 Z_t = (1 - L)^2 Z_t \tag{2.30}$$

$$= Z_t - 2LZ_t + L^2 Z_t$$

$$= Z_t - 2(Z_{t-1}) + Z_{t-2}$$

Al ∇^k ser un binomio a la potencia k podemos expresar su forma general como

$$\nabla^k Z_t = (1 - L)^k Z_t$$

Retomando nuestra serie de Ventas Z_t construimos dos nuevas series $\{X_t\}$ y $\{Y_t\}$ definimos $X_t = \nabla Z_t$ y $Y_t = \nabla^2 Z_t = \nabla(\nabla Z_t) = \nabla X_t$, entonces siguiendo la ecuación (2.29) para la serie X_t obtenemos los valores de febrero y marzo de 2005 como:

$$X_{feb05} = 381 - 507 = -126$$

$$X_{mar05} = 705 - 381 = 324$$

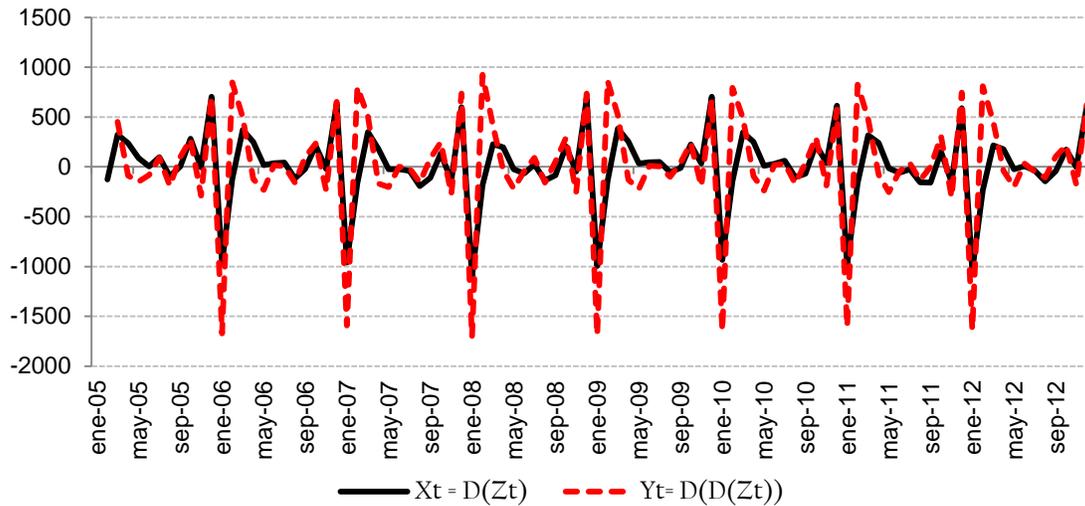
Para la serie Y_t podemos obtener el valor siguiendo la ecuación (2.30)

$$Y_{mar05} = 705 - 2(381) + 507 = 450 \text{ o de manera alternativa } Y_{feb05} = X_{mar05} - X_{feb05}$$

$$Y_{mar05} = 324 - (-126) = 450$$

Al generar la serie X_t se pierde una observación correspondiente a Enero de 2005 como se observa en la gráfica (2.6) de igual forma la serie Y_t pierde 2 observaciones, Enero y Febrero de 2005.

Gráfica 2.6: Primera y Segunda Diferencia de Ventas de la Empresa X



Un proceso estocástico que debe ser diferenciado una o más veces para ser estacionario se conoce como proceso estocástico integrado (Chatfield, 2000). Si el proceso Z_t es un proceso no estacionario (*i.e.* que su media y varianza dependen del tiempo) y se genera $X_t = \nabla Z_t$ siendo X_t estacionaria por lo tanto Z_t es un proceso integrado de primer orden y se denota $Z_t \sim I(1)$.

Si un proceso es estacionario y otro es integrado la combinación lineal será un proceso estocástico integrado. Un caso especial se presenta cuando la combinación lineal de dos procesos estocásticos integrados del mismo orden resulta un proceso integrado del mismo orden con la excepción de que sea una relación de cointegración en tal caso la combinación lineal de dos procesos integrados es un proceso estacionario (Gujarati, *op. cit.*)

Para saber si el proceso estocástico es integrado existen las pruebas de raíz unitaria siendo la prueba de Dickey Fuller (1979) la primera. Esta prueba presenta 3 diferentes posibilidades respecto al proceso:

Que sea un paseo aleatorio

$$y_t = \varphi y_{t-1} + \varepsilon_t \tag{2.31}$$

Un paseo aleatorio con constante, *drift* o deriva

$$y_t = \alpha + \varphi y_{t-1} + \varepsilon_t \tag{2.32}$$

Un paseo aleatorio con deriva y tendencia

$$y_t = \alpha + \beta t + \varphi y_{t-1} + \varepsilon_t \tag{2.33}$$

Las hipótesis son:

$$H_0 : \varphi = 1 \rightarrow y_t \sim I(1)$$

$$H_1 : |\varphi| < 1 \rightarrow y_t \sim I(0)$$

El estadístico que se emplea para realizar las pruebas de contraste fue desarrollado por los mismos autores empleando simulaciones de Monte Carlo. Debido a las limitaciones que presenta la distribución realizaron una versión aumentada con la adición de los valores rezagados de la variable dependiente (*Augmented Dickey Fuller*, ADF) (Dickey y Fuller, 1981).

Además de esta existe una gran variedad de pruebas¹⁰ dentro de las más utilizadas se encuentran las siguientes: otra versión de Dickey-Fuller pero estimada con mínimos cuadrados generalizados (DF-GLS) (Elliott, Rothemberg y Stock, 1996), la prueba de Phillips-Perron (1988) (PP) que emplea métodos estadísticos no paramétricos para evitar la correlación serial en los términos de error y la prueba KPSS (Kwiatkowski, Phillips, Schmidt y Shin, 1992) que tiene como hipótesis nula que el proceso es estacionario a diferencia de las anteriores. De acuerdo a Ayat y Burrige “*the currently most widely employed tests for a unit root the so call augmented version of those developed by D-F*” (2000, p. 74).

2.3 Modelos Autorregresivos Integrados de Medias Móviles (ARIMA)

En 1970 George Box y Gwilym Jenkins presentaron una metodología para identificar, estimar y diagnosticar modelos dinámicos de series de tiempo. El objetivo fue construir un modelo donde la variable exógena estaba constituida por valores retrasados de la misma variable. El argumento principal, la ventaja de ahorrar esfuerzos en la especificación de variables y relaciones adicionales que hasta entonces tenía la econometría tradicional y ortodoxa representada perfectamente por los modelos de la Comisión Cowles, aunque por otra parte al emplear este enfoque se renunciaba a la presencia de otras variables relacionadas con el fenómeno de estudio y la interpretación que le aporta al investigador.

Modelos Autorregresivos

Podemos representar el modelo autorregresivo de orden p con los polinomios de retraso como

$$(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p)(Z_t - \mu) = a_t \quad (2.34)$$

$$\varphi(L)(Z_t - \mu) = a_t$$

¹⁰ Esta gran variedad hace natural preguntarse ¿Qué prueba usar? Al respecto se puede ver el artículo de Elder y Kennedy (2001) para tener un panorama más completo del que se enfoca este trabajo.

Desarrollando el caso más simple, el modelo AR(1) se representa en la siguiente ecuación

$$\hat{Z}_t - \varphi \hat{Z}_{t-1} = a_t \quad (2.35)$$

$$(1 - \varphi L)\hat{Z}_t = a_t$$

Para que este proceso sea estacionario la raíz de la ecuación debe estar fuera del círculo unitario lo que significa que debe cumplir $|\varphi| < 1$ y en forma alternativa se tiene una expresión cuyos términos convergen a cero.

$$\hat{Z}_t = (1 - \varphi L)^{-1}a_t = a_t + \varphi a_{t-1} + \varphi^2 a_{t-2} + \dots$$

Entonces de cumplirse la condición se puede ver que el valor esperado y la varianza de la serie son

$$E(\hat{Z}_t) = E(a_t + \varphi a_{t-1} + \varphi^2 a_{t-2} + \dots) \quad (2.36)$$

$$= E(a_t) + \varphi E(a_{t-1}) + \varphi^2 E(a_{t-2}) + \dots = 0$$

$$Var(\hat{Z}_t) = Var(a_t) + \varphi^2 Var(a_{t-1}) + \varphi^4 Var(a_{t-2}) + \dots$$

$$= \sigma^2(1 + \varphi^2 + \varphi^4 + \dots)$$

$$\gamma_0 = \frac{\sigma_a^2}{(1-\varphi^2)} \quad (2.37)$$

De manera análoga a la ecuación 2.24 obtenemos la autocovarianza del proceso con h rezagos

$$\gamma_h = \frac{\sigma_a^2 \varphi^{|h|}}{(1-\varphi^2)}$$

La autocorrelación igual que en (2.27)

$$\rho_h = \frac{\gamma_h}{\gamma_0} = \frac{\frac{\sigma_a^2 \varphi^{|h|}}{(1-\varphi^2)}}{\frac{\sigma_a^2}{(1-\varphi^2)}} = \varphi^{|h|} \quad (2.38)$$

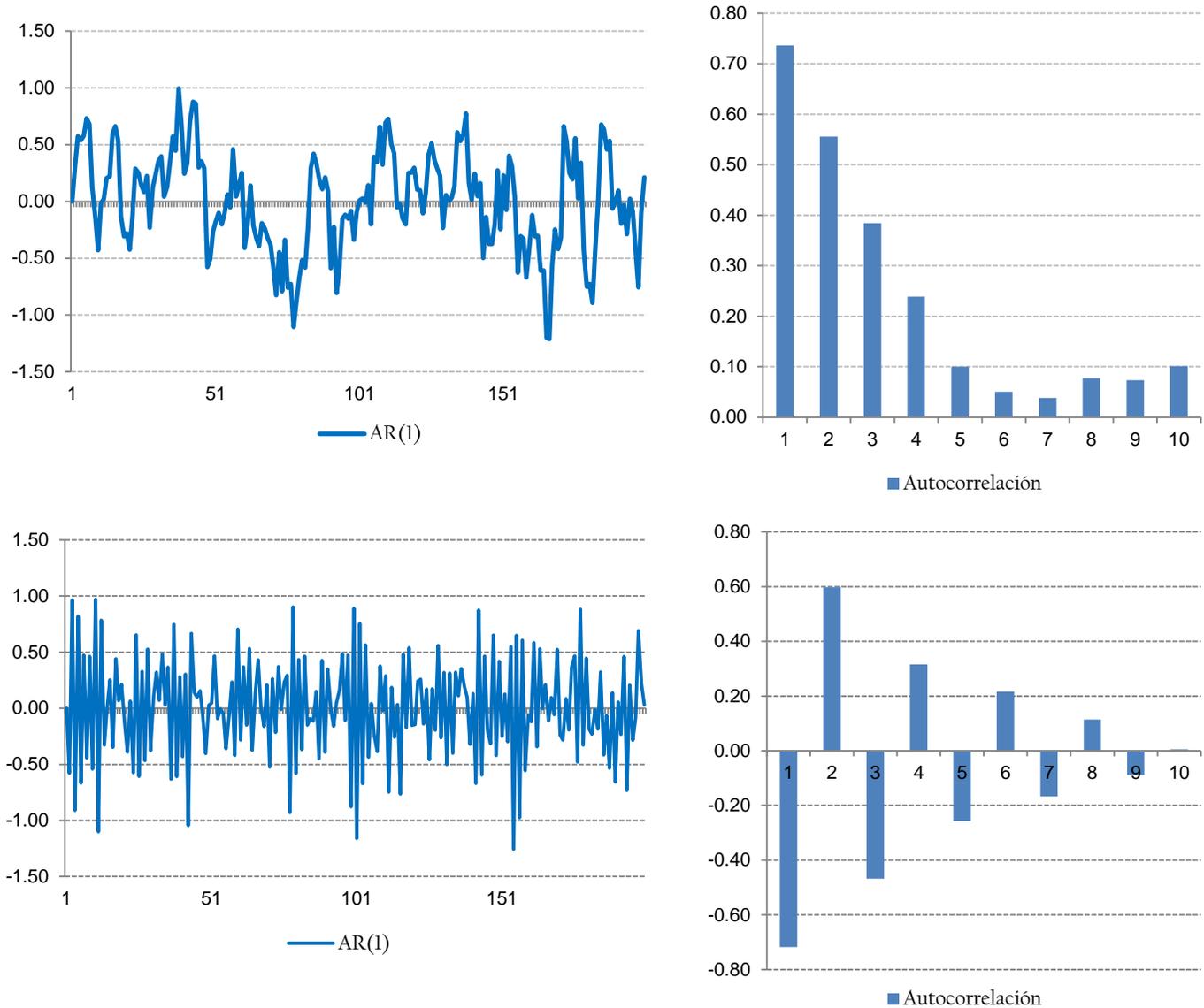
Para ejemplificar el modelo se generan dos series con 200 observaciones por Monte Carlo de un AR(1) con los siguientes parámetros:

$$\varphi = 0.7; \mu = 0; \sigma^2 = 0.6 \quad \text{y} \quad \varphi = -0.7; \mu = 0; \sigma^2 = 0.3$$

La realización de la series se llevó a cabo en Microsoft Excel® 2010 el componente aleatorio se construye con la función estadística $\text{distr.norm.inv}(\text{aleatorio}(), \mu, \sigma^2)$. El valor es igual a

la observación anterior por φ en una celda fija más el valor aleatorio con distribución $N \sim (0,0.6)$ en el primer caso y $N \sim (0,0.3)$ para el segundo caso.

Gráfica 2.7: Modelo Autorregresivo de Primer Orden y Función de Autocorrelación



La parte superior de la gráfica (2.7) muestra la serie con $\varphi = 0.7$ y la parte inferior el AR(1) con $\varphi = -0.7$ la función de autocorrelación (FAC) de ambos confirma que con $h > 0$ crece la FAC tiende a cero y que si $0 < \varphi < 1$ decae exponencialmente y si $-1 < \varphi < 0$ va cambiando los signos.

Modelos de Media Móvil

Podemos representar el modelo de medias móviles de orden q con los polinomios de retraso como

$$\begin{aligned} Z_t - \mu &= (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q) a_t \\ &= \theta(L) a_t \end{aligned} \quad (2.39)$$

Desarrollamos el caso más simple, el modelo MA(1) que se representa en la siguiente ecuación

$$\hat{Z}_t = (1 - \theta L) a_t \quad (2.40)$$

En donde \hat{Z}_t representa la desviación de Z_t respecto a su nivel medio μ y θ es el parámetro de promedio móvil, por lo anterior se puede ver que el valor esperado y la varianza de la serie son

$$E(\hat{Z}_t) = 0$$

$$Var(\hat{Z}_t) = \sigma_a^2 (1 + \theta^2) \quad (2.41)$$

$$Cov(\hat{Z}_t, \hat{Z}_{t-h}) = E[(a_t - \theta a_{t-1})(a_{t-h} - \theta a_{t-h-1})] \quad (2.42)$$

$$\gamma_{h=1} = -\theta \sigma_a^2 ;$$

$$\gamma_{h \geq 2} = 0$$

Por lo tanto la función de autocorrelación es

$$\rho_{h=1} = \frac{\gamma_{h=1}}{\gamma_0} = \frac{-\theta}{1+\theta^2} \quad (2.43)$$

$$\rho_{h \geq 2} = \frac{\gamma_{h \geq 2}}{\gamma_0} = 0$$

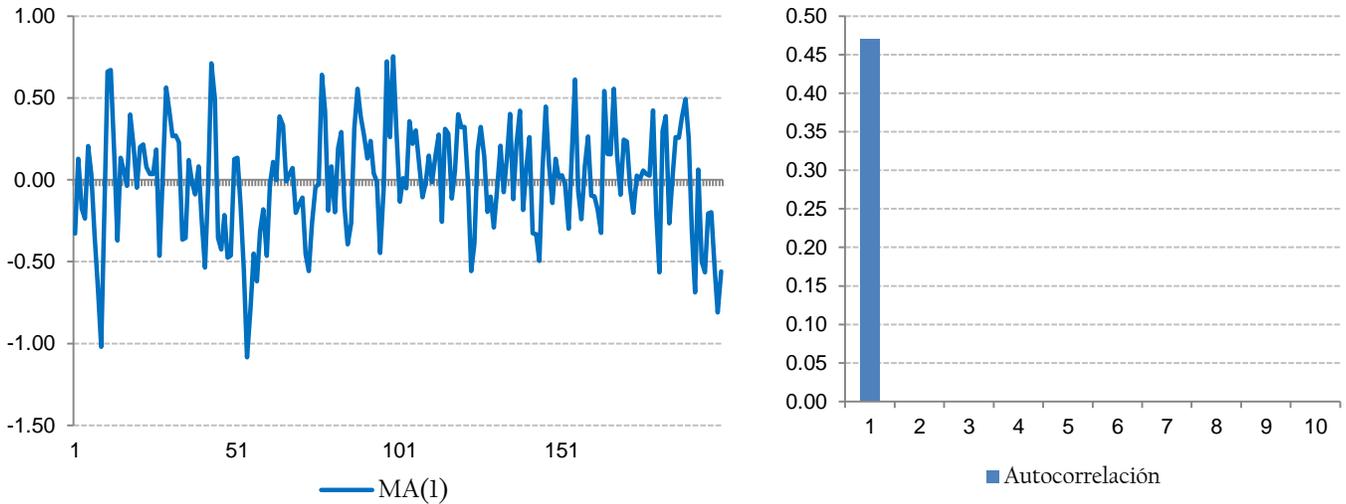
De (2.42) y (2.43) podemos dar cuenta que el proceso únicamente limita su memoria a un periodo atrás y a diferencia del proceso autorregresivo el factor de autocorrelación no podrá ser elevado, generalmente tendrá un valor $|\rho_1| \leq 0.5$.

Otra característica que debe cubrir para ser MA(1) es que sea invertible, para cumplir con esto necesita de igual forma que las raíces del polinomio estén fuera del círculo unitario es decir $|\theta| \leq 1$

Para ejemplificar el modelo se generó una serie con 200 observaciones con los siguientes parámetros: $\theta = -0.7$; $\mu = 0$; $\sigma^2 = 0.3$

Como se observa en la gráfica 2.8 la serie de MA(1) es a simple vista difícil de distinguir respecto a las series AR(1) del ejercicio anterior de ahí la importancia de la Función de Autocorrelación ya que se puede contrastar que la Autocorrelación de primer orden es efectivamente menor a 0.5, la característica de ser invertible se asegura al $\theta = -0.7$.

Gráfica 2.8: Modelo de Media Móvil de Primer Orden y Función de Autocorrelación



Modelo Mixto

La combinación del modelos AR(p) y MA(q) conforma el modelo autorregresivo de media móviles ARMA(p, q) y se representa por medio de polinomios de retraso como

$$\varphi(L)(Z_t - \mu) = \theta(L)a_t \quad (2.44)$$

Si la serie es un proceso estocástico integrado por medio de diferencias se puede representar como proceso estacionario, de manera que elimina el componente de tendencia que presenta la serie. La (d) será el número de diferencias que se realicen a una serie hasta ser un posible modelo ARMA, a este modelo se le llama autorregresivo e integrado de media móvil ARIMA(p, d, q) y se representa con la siguiente ecuación

$$\varphi(L)\nabla^d \hat{Z}_t = \theta(L)a_t \quad (2.45)$$

Ahora el operador combinado de AR es $\varphi(L)(1 - L)^d$ si se sustituye el operador L en esta expresión con una variable x , se puede ver de inmediato que la función $\varphi(x)(1 - x)^d$ tiene d raíces en el círculo unitario indicando que el proceso no es estacionario razón por la cual la diferenciación es requerida.

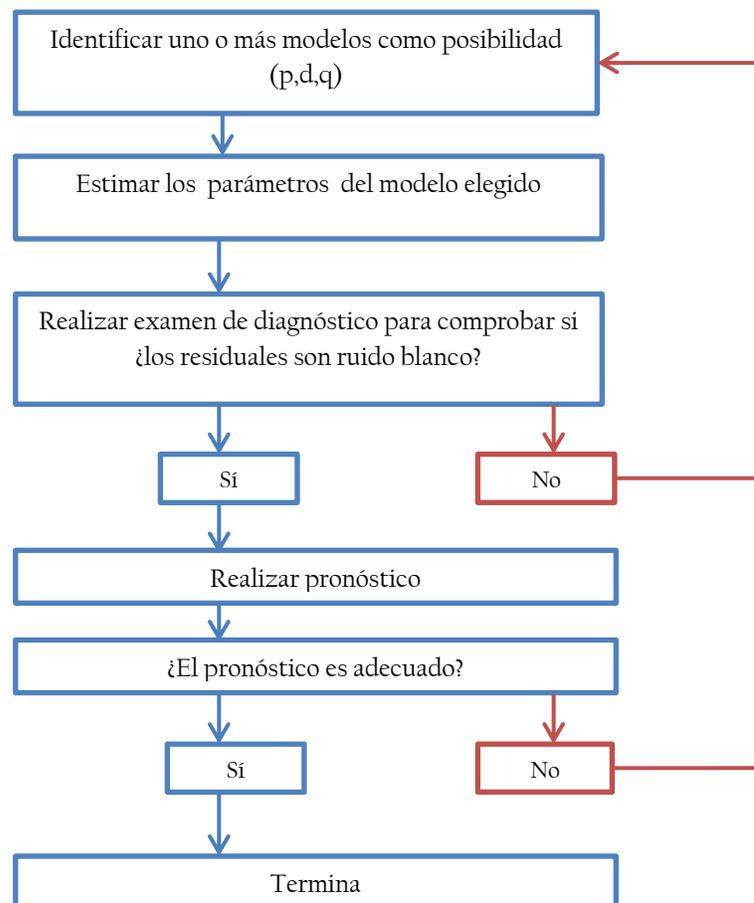
Ya que es necesario tener procesos estacionarios e invertibles, las raíces de $\varphi(x) = 0$ y las raíces de $\theta(x) = 0$ como en los modelos anteriores deben estar fuera del círculo unitario, si consideramos el operador combinado de AR entonces (2.45) se puede re-expresar como

$$\hat{Z}_t = \varphi_1 \hat{Z}_{t-1} + \varphi_2 \hat{Z}_{t-2} + \dots + \varphi_{p+d} \hat{Z}_{t-p-d} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Expresión tradicional que se puede encontrar en la mayoría de los libros de texto y que nos indica que las desviaciones de las diferencia(s) de Z_t respecto a su media están función de su(s) observación(es) pasada(s) y de una suma ponderada de choques aleatorio tipo ruido blanco.

Para los modelos AR y MA la principal dificultad es evaluar el orden del proceso (*i.e.* p y q), con los modelos ARIMA, existe una decisión adicional elegir el orden necesario de diferenciación (*i.e.* d). En definitiva aquí es mayor la relevancia de las pruebas de raíz unitaria para determinar el orden de integración aunque, con la evidencia de las investigaciones se puede asumir que la mayoría de las series con las que se trabaja en economía son de $I(1)$ y en algunos casos extremos $I(2)$.

La metodología Box y Jenkins resulta muy clara por lo tanto es fácil de seguir los siguientes pasos:



Para el paso 1 son fundamentales las pruebas de estacionariedad ya que, como vimos en las gráficas (2.7) y (2.8) la Función de Autocorrelación es vital para poder determinar qué proceso es. Si el correlograma converge a cero es señal de ser un proceso $AR(p)$, si sólo las primeras q autocorrelaciones con distintas de cero entonces es el proceso es $MA(q)$. Otra regularidad un poco más complicada de distinguir es si las primeras q autocorrelaciones son irregulares y después converge a cero entonces será un proceso $ARMA(p, q)$.

Para los pasos 2 y 3 dependerá del paquete estadístico que se utilice para llevar a cabo la estimación, actualmente la mayoría de los mismos cuentan con rutinas para poder estimar modelos de este tipo y por consecuencia también cuentan con las pruebas necesarias para realizar todas las pruebas de diagnóstico que se quiera hacer.

Cap. 3 Redes Neuronales Artificiales

“Typically, neurons are five to six orders of magnitude slower than silicon logic gates, events in a silicon chip happen in the nanosecond range (10^{-9} s) range, whereas neural events happen in the millisecond (10^{-3} s) range”.

Simon Haykin

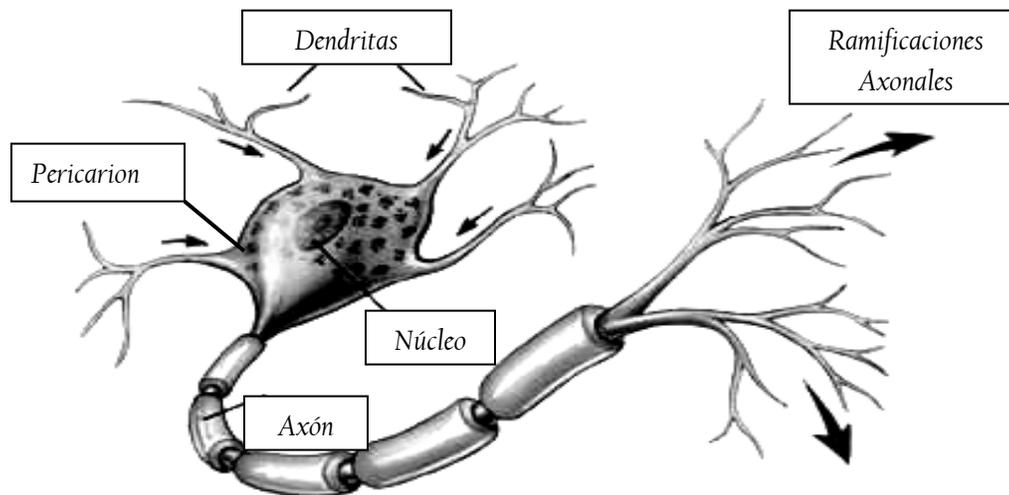
“The brain is a monstrous, beautiful mess. Its billions of nerve cells -called neurons- lie in a tangled web that displays cognitive powers far exceeding any of the silicon machines we have built to mimic it”

William Allman

3.1 ¿Qué son las Redes Neuronales Artificiales?

Las redes neuronales artificiales son modelos matemáticos capaces de adaptar su comportamiento en respuesta a ejemplos presentados por el medio ambiente de manera supervisada o no supervisada. Además la neurona artificial es un procesador paralelo y distribuido conformado por varios procesadores simples que almacenan y utilizan el conocimiento generado de la experiencia (Haykin, 1999). Pero para entenderlas mejor debemos ir a los fundamentos y recordar que finalmente, están inspiradas en las neuronas biológicas así que para empezar hay que repasar como se conforma una neurona.

Gráfica 3.1: Representación básica de una neurona



Primero tenemos las *dendritas*, las cuales funcionan como conexiones de entrada al *pericarion* conduciendo los impulsos nerviosos hacia él. Después tenemos el *pericarion* o también conocido como *soma* en donde se localiza el *núcleo* siendo esta la parte más importante, el centro de síntesis de la neurona, aquí es donde todas las señales que llegan de otras neuronas se procesan y se emiten nuevas señales. Por último el *axón*, el canal transmisor de los impulsos que se generan dentro del *pericarion*, como se aprecia en la gráfica 3.1 al

extremo el axón se ramifica para conectarse con otras neuronas. La conexión con otras neuronas llamada *sinapsis* se inicia con una descarga química que genera una corriente eléctrica en la membrana de la *célula presináptica* o emisora, una vez que el impulso llega al extremo la neurona genera neurotransmisores, los cuales están encargados de excitar o inhibir la acción de la otra célula llamada *post sináptica* o receptora.

En 1943 Warren McCulloch un neurocientífico y Walter Pitts un joven lógico matemático de 20 años fueron los primeros en proponer un modelo matemático tratando de comprender cómo el cerebro podía producir patrones sumamente complejos utilizando muchas neuronas (McCulloch y Pitts, 1943). Este modelo sería el aporte seminal al campo de investigación de las redes neuronales artificiales. Unos cuantos años más tarde Donald Hebb (1949) planteó la idea de que las conexiones neuronales se fortalecen cada vez que se usan, implementando la regla de aprendizaje Hebbiano antecesora de las técnicas modernas de entrenamiento.

Una década más tarde un neurobiólogo de la Universidad de Cornell llamado Frank Rosenblatt realizando una investigación basada en el ojo de una mosca y el procedimiento que ocurre en él cuando se encuentra en peligro activando el vuelo de la mosca. Desarrollo lo que actualmente conocemos como el *Perceptrón Simple* (Rosenblatt, 1958). Dos años después Widrow y Hoff (1960) crearían la regla Delta o también conocida como algoritmo LMS una de las más comunes reglas de aprendizaje y los modelos ADELINE (elementos adaptativos lineales) y MADALINE (múltiples elementos lineales) empleados para eliminar el eco de líneas telefónicas.

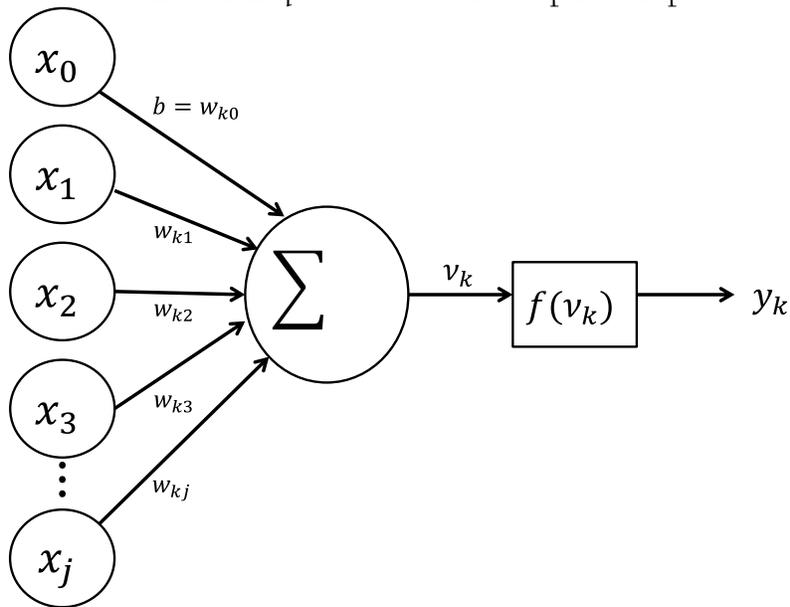
Todos estos avances se vieron minados por un artículo de Minsky y Papert (1969) quienes cuestionaban si existía una simple regla de aprendizaje que garantizara funcionar y solucionar cualquier tipo de problema, en particular si esa regla era la *regla Delta*. Demostrando limitaciones teóricas del *Perceptrón Simple* siendo incapaz de aprender funciones no linealmente separables. Con este trabajo la investigación de las redes neuronales artificiales fue cada vez más escasa, aunque los avances continuaban. En la década de 1970 Paul Werbos (1974) en su tesis de doctorado desarrolló el método de aprendizaje conocido como *Backpropagation*, sin lugar a dudas el algoritmo más utilizado en las aplicaciones de redes neuronales artificiales.

En la siguiente década se daría el resurgimiento de las redes de la mano John Hopfield (1982) al plantear que la inteligencia artificial no sólo debía limitarse a imitar el cerebro humano también debería de usar sus conceptos para construir máquinas que fueran capaces de solucionar problemas dinámicos. Tomando ese planteamiento desarrollo una red recurrente conocida como el modelo Hopfield el cual ayudaba a comprender la memoria humana a través de la memoria asociativa. Por último en 1986 (Rumelhart, *et. al*, 1986)

perfeccionan el algoritmo *backpropagation* ofreciendo una mejor solución para la construcción de redes neuronales artificiales.

Para comenzar el estudio de las redes neuronales artificiales se presenta el modelo propuesto por McCulloch y Pitts (*op. cit*) y desarrollado por Rosenblatt (*op. cit*) conocido como Perceptrón Simple. Es el modelo más sencillo de todas las redes y forma la base de partida para la construcción de muchas otras, el perceptrón simple es un modelo *feedforward*, compuesto únicamente por dos capas de neuronas, una de entrada y una de salida.

Gráfica 3.2 Esquema Básico del Perceptrón Simple



En el perceptrón simple se tiene una serie de entradas $x_j(t)$, pesos sinápticos¹¹ w_{kj} (i.e. la intensidad de interacción entre la neurona pre sináptica j y la neurona post sináptica k) y con frecuencia se añade un parámetro más denominado umbral o *bias* b_k con una entrada de -1 , en este esquema se define que $w_{k0} = b_k$ y $x_0 = 1$.

Después se procesa dicha información con una regla de propagación en este caso

$$v_k(t) = \sum_{j=0}^n w_{kj}x_j \quad (3.1)$$

Por último una función de activación o también conocida como función de transferencia que representa simultáneamente la función de salida

$$y_k(t) = f(v_k(t)) \quad (3.2)$$

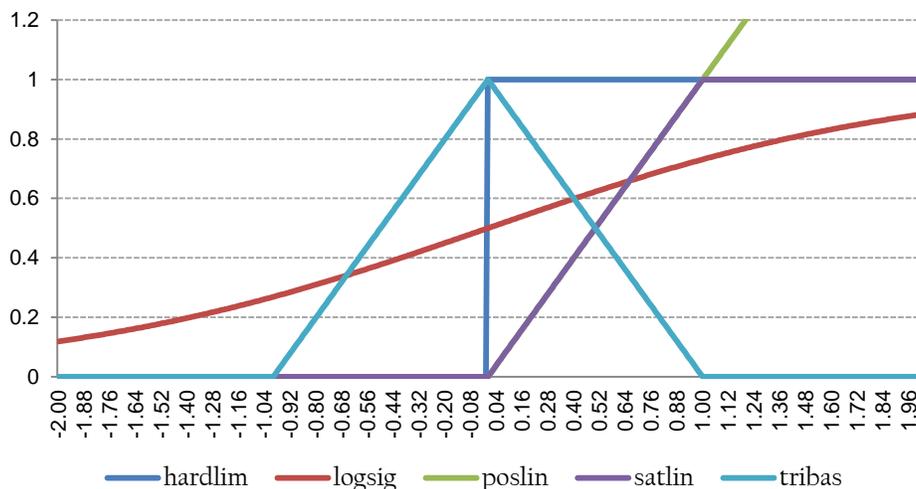
¹¹ Igual que con la neurona biológica se puede inhibir o excitar a la neurona post sináptica, en el modelo de neurona artificial esto se refleja con $w_{kj} < 0$ para inhibir y $w_{kj} > 0$ para excitar.

La elección de la función de activación es un factor tan importante como los pesos sinápticos y las entradas ya que determina la re expresión del potencial sináptico hacia la siguiente capa siendo en la mayoría de los casos funciones de activación distintas para la capa oculta como para la capa de salida. En el cuadro 3.1 y en la gráfica 3.3 se presentan las principales opciones de funciones de activación.

Cuadro 3.1: Funciones de Transferencia más Empleadas

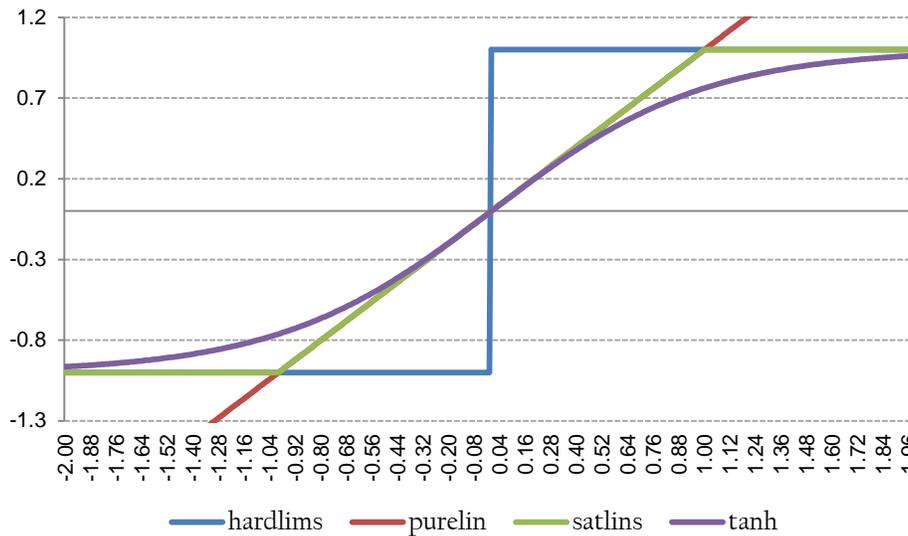
Nombre	Relación Entrada/Salida
Lineal (purelin)	$a = n$
Lineal positiva (poslin)	$a = \begin{cases} 0 & \text{si } n < 0 \\ n & \text{si } n > 0 \end{cases}$
Lineal saturado (satlin)	$a = \begin{cases} 0 & \text{si } n < 0 \\ n & \text{si } 0 \leq n \leq 1 \\ 1 & \text{si } n > 1 \end{cases}$
Lineal saturado simétrico (satlins)	$a = \begin{cases} -1 & \text{si } n < -1 \\ n & \text{si } -1 \leq n \leq 1 \\ 1 & \text{si } n > 1 \end{cases}$
Limitador fuerte (hardlim)	$a = \begin{cases} 1 & \text{si } n \geq 0 \\ 0 & \text{si } n < 0 \end{cases}$
Limitador fuerte simétrico (hardlims)	$a = \begin{cases} 1 & \text{si } n \geq 0 \\ -1 & \text{si } n < 0 \end{cases}$
Triangular (tribas)	$a = \begin{cases} 0 & \text{si } n < -1 \\ 0 & \text{si } n > 1 \\ f(n) & \text{si } a = 1, n = 0 \end{cases}$
Sigmoidal logarítmico (logsig)	$a = \frac{1}{1 + e^{-n}}$
Tangente hiperbólica (tanh)	$a = \frac{e^n - e^{-n}}{e^n + e^{-n}}$

Gráfica 3.3a Funciones de Activación



Villar Corona Marco Antonio

3.3b Funciones de Activación



Al final el punto más importante del perceptrón es que tiene la capacidad de aprender, el cual se realiza modificando los parámetros del mismo ante estímulos. En realidad lo que se busca con el aprendizaje es que la red produzca el tipo de funcionamiento deseado a partir de modificar los pesos sinápticos. Por lo tanto debemos indicar cuál es el funcionamiento adecuado incorporando los resultados deseados d_k .

En cuanto a formas de aprendizaje, existen varios tipos, cada uno aplicable a distintos tipos de redes neuronales específicas con diferentes capacidades para resolver diversos problemas. Las 5 reglas básicas para cada tipo de aprendizaje son:

Aprendizaje Hebbiano

Emplea la regla de aprendizaje más antigua además de celebre y se basa en la siguiente idea: “When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic changes takes place in one or both cells such that A’s efficiency as one of the cells firing B, is increased” Hebb (*op. cit.*, p.62). Al aplicar este principio a la red se traduce en que cuando dos neuronas que comparten una conexión sináptica se activan simultáneamente entonces dicha conexión se fortalece y viceversa.

Aprendizaje competitivo

Las neuronas compiten unas con otras con el fin de llevar a cabo una tarea, al hacer esto se pretende que cuando se presente a la red nueva información de entrada, sólo una de las neuronas de salida de la red se active (*i.e.* alcance su respuesta máxima). Por tanto las neuronas compiten para activarse quedando finalmente una como la neurona vencedora y el resto son forzadas a sus valores de respuesta mínimos.

Aprendizaje basado en la memoria

En este aprendizaje todas o casi todas las experiencias pasadas son explícitamente almacenadas en la memoria donde están clasificados los ejemplos pasados de patrones y la relación del vector de entrada \vec{x}_j y su respectiva de salida deseada d_k . Después la red recibe un vector de prueba ($test = t$) \vec{x}_t y basado en las clasificación de las experiencias anteriores determina su salida en función de las clasificaciones vecinas.

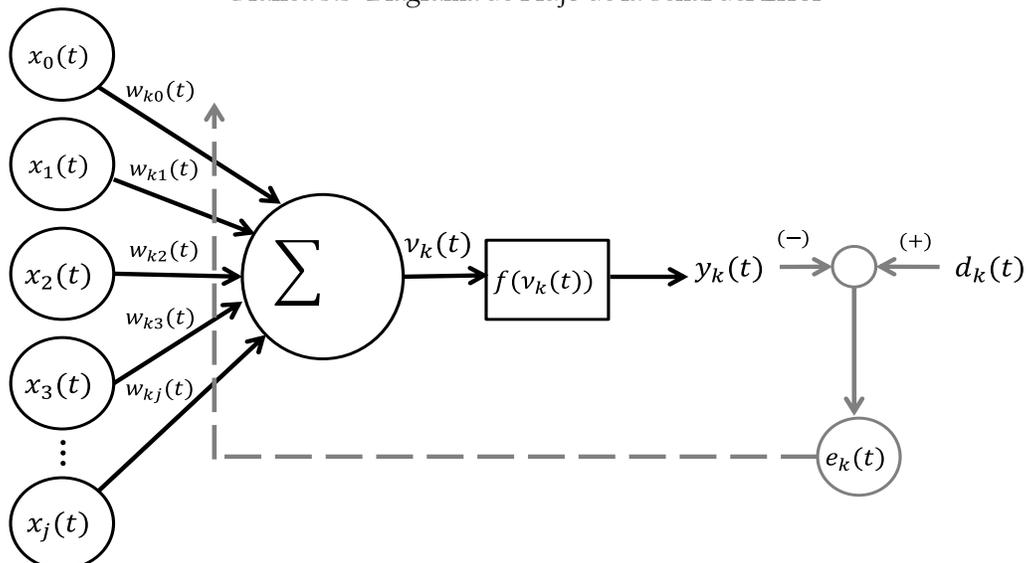
Aprendizaje de Boltzman

Se aplica a las redes neuronales estocásticas conocidas como Máquina de Boltzman y se caracteriza por el estado binario de sus neuronas las cuales se encuentran activas (1) o inactivas (-1). La máquina tiene un estado de energía en función del estado de sus neuronas y calcula una pseudo-temperatura entonces, de manera aleatoria cambia el estado de una de las neuronas y estima el cambio del estado de energía. Este procedimiento se repite continuamente hasta alcanzar el equilibrio térmico.

Aprendizaje por corrección de error

El más importante para este trabajo ya que permite a diferencia de los otros entrenar la red a través de ejemplos los cuales se comparan con la respuestas que esperadas y de existir un error el mismo se corrige para los siguientes ejemplos. Razón por la cual se utiliza en nuestra red. Tomando en cuenta la ecuación (3.2) se comparan los resultados obtenidos con los resultados deseados $d_k(t)$ generando una señal del error $e_k(t) = d_k(t) - y_k(t)$ la cual actúa como mecanismo de control siendo el propósito aplicar pequeñas correcciones o ajustes a los pesos sinápticos de la neurona k los cuales están encaminados a minimizar la distancia entre el resultado deseado y el obtenido.

Gráfica 3.3 Diagrama de Flujo de la Señal del Error



De esta forma el *perceptrón* simple puede representar en un mapa las regiones de decisión creadas en el espacio multidimensional de entrada de la red. En estas regiones se visualiza qué patrones pertenecen a una clase y cuales a otra. Realiza la separación de las regiones del hiperplano y es determinada por los pesos de las conexiones, el valor umbral y la función de activación de la neurona.

Pero como lo señalaba Minsky y Papert (*op. cit.*) su principal limitante era sólo poder representar funciones linealmente separables. Un ejemplo de una función que no es posible tratar con un perceptrón simple es la función XOR. Por esta razón se construyeron redes más poderosas, lo cual se logró al aumentar el número de capas por lo que se les conoce como *perceptrón* multicapa el cual utilizamos en el trabajo más adelante.

3.2 Perceptrón Multicapa

Esta red suele entrenarse con el algoritmo denominado *backpropagation* de errores, motivo por el cual a una arquitectura del tipo perceptrón multicapa con aprendizaje *backpropagation* suele denominarse red *backpropagation*.

El entrenamiento del perceptrón multicapa se basa en la presentación sucesiva de pares de vectores en las capas de entrada y salida. La red crea un modelo ajustando sus pesos en función de los vectores de entrenamiento de forma que a medida que se pasan estos patrones para cada vector de entrada la red producirá un valor de salida más similar al vector de salida esperado.

Se ha demostrado por Funahashi (1989) que un perceptrón multicapa de 3 capas con 1 de entrada, 1 capa oculta y una de salida puede aproximar con gran exactitud dado cualquier conjunto de funciones. En su trabajo plantea en sus *Teoremas 1 y 2* una versión del teorema de aproximación universal, donde

Sea $\phi(x)$ una función no constante, acotada y monótona creciente. Sea K un subconjunto compacto (acotado y cerrado) de \mathbb{R}^n y $f(x_1, \dots, x_n)$ sea una función real continua de K . Entonces para una arbitraria $\varepsilon > 0$, existe un número entero N y constantes reales c_i θ_i ($i = 1, \dots, N$), w_{ij} ($i = 1, \dots, N, j = 1, \dots, n$) tal que

$$\tilde{f}(x_1, \dots, x_n) = \sum_{i=1}^N c_i \phi \left(\sum_{j=1}^n w_{ij} x_j - \theta_j \right) \quad (3.3)$$

Satisface $\max_{k \in K} |f(x_1, \dots, x_n) - \tilde{f}(x_1, \dots, x_n)| < \varepsilon$. En otras palabras, para una arbitraria $\varepsilon > 0$ existe un red de tres capas cuyas funciones de salida para la capa oculta son $\phi(x)$, cuyas funciones de salida para las capas de salida y entrada son lineales y tiene una función entrada-salida $\tilde{f}(x_1, \dots, x_n)$ tal que $\max_{k \in K} |f(x_1, \dots, x_n) - \tilde{f}(x_1, \dots, x_n)| < \varepsilon$. (Ibíd., p. 185).

Si cualquier mapeo $g: x \in K \rightarrow (g_1(x), g_2(x), \dots, g_m(x)) \in \mathbb{R}^m$, con $g_i(x)$ sumables en K , puede ser aproximado en el sentido de la topología en K por el mapeo entrada-salida representado por una red neuronal unidireccional (perceptrón multicapa) de K capas ($K-2$ ocultas) con $f(x)$ como función de transferencia de las neuronas ocultas y funciones lineales para las capas de entrada y salida, si $K = 3$ entonces se tiene un perceptrón de 3 capas (1 de entrada, 1 oculta y una de salida).

Ahora, la cuestión es elegir el número de neuronas ocultas para ajustar el modelo (buscando evitar el problema de sobreajuste). La solución a esta cuestión se logra a prueba y error en la mayoría de los casos.

Entonces como en el perceptrón simple denominamos a x_i como las entradas de la red, h_i a las salidas de la capa oculta y y_k a las de la capa de salida; d_k serán las salidas deseadas. Por otro lado w_{ji} son los pesos de la capa oculta y w'_{kj} los pesos de la capa de salida. La operación de un perceptrón multicapa con una capa oculta (esto por el teorema de Funahashi) y neuronas de salida lineal se expresa matemáticamente de la siguiente manera

$$\begin{aligned} y_k &= \sum_j w'_{kj} h_j \\ &= \sum_j w'_{kj} f\left(\sum_i w_{ji} x_i\right) \end{aligned} \quad (3.4)$$

De nueva cuenta tomamos al umbral de la capa de entrada a la capa oculta como $b_j = w_{j0}$ y el umbral de la capa oculta a la capa de salida como $b_k = w_{k0}$. La función de transferencia es del tipo sigmoideal para la capa oculta y normalmente la función de activación para la capa de salida es lineal por lo tanto $y_k = f(v_k) = \sum_j w'_{kj} h_j$.

3.2.1 Algoritmo *backpropagation*

El algoritmo consiste en propagar el error hacia atrás es decir, de la capa de salida hacia la capa de entrada pasando por las capas ocultas intermedias y ajustando los pesos de las conexiones con el fin de reducir dicho error.

El algoritmo es el siguiente:

Paso 1 Establecer los pesos de la red con valores aleatorios de preferencia entre el intervalo $[0, 1]$.

Paso 2 Presentar vector de observaciones de entrada \vec{x}_i con sus respectivas salidas deseadas las cuales debe generar la red \vec{d}_k .

Paso 3 Calcular la salida actual de la red y_k . Primero, se calculan las entradas netas con la regla de propagación y obtener el potencial sináptico para las neuronas ocultas procedentes de las neuronas de entrada. Para una neurona oculta j tenemos

$$v_j^o = \sum_{i=0}^N w_{ji}^o x_i \quad (3.5)$$

Donde el superíndice o se refiere al número de la capa oculta el subíndice j a la j -ésima neurona oculta. El término b_j puede ser opcional pues actúa como una entrada más, de otra forma quedaría expresado como $\sum_{i=1}^N w_{ji}^o x_i - b_j$

Se calculan las salidas de las neuronas ocultas

$$h_j = f_j^o(v_j^o) \quad (3.6)$$

La función $f_j^o(\cdot)$ será diferenciable y monótona creciente.

Una vez hecho esto se realiza los mismos cálculos para obtener la salida de la neurona de salida

$$y_{kj}^s = \sum_{j=0} w_{kj}^s h_j \quad (3.7)$$

Paso 4 Se determina el error de las capas de salida si la neurona K es una neurona de la capa de salida el valor delta se define como

$$\delta_{kj}^s = (d_{kj} - Y_{kj}) f_k^{s'}(y_{kj}^s) \quad (3.8)$$

Siendo $f'(\cdot)$ la derivada parcial respecto de y_{kj}^s se calcula el error de las capas ocultas como

$$\delta_{ji}^o = f_j^{o'}(h_{ji}^o) \sum_{k=j}^M w_{kj} \delta_{kj}^s \quad (3.9)$$

Para distribuir el error (δ) de cada una de las neuronas de salida hacia todas las neuronas ocultas en proporción de como dichas neuronas contribuyen en la salida.

Paso 5 Se realiza la actualización de los pesos comenzando por las neuronas de salida y trabajando hacia atrás hasta llegar a la capa de entrada. Para actualizar los pesos en la capa de salida

$$w_{kj}^s(t+1) = w_{kj}^s(t) + \eta \delta_{kj}^s h_{pj} + \mu (w_{kj}^s(t) - w_{kj}^s(t-1)) \quad (3.10)$$

Para actualizar los pesos en la capa oculta

$$w_{ji}^o(t+1) = w_{ji}^o(t) + \eta \delta_{kj}^o x_i + \mu (w_{ji}^o(t) - w_{ji}^o(t-1)) \quad (3.11)$$

Donde η es la tasa de aprendizaje, μ el momentum y t es el número de iteración

La tasa de aprendizaje (η) nos indica velocidad con que la red está aprendiendo y el tiempo que requiere para entrenarse esta en función de este parámetro. Un valor pequeño significa que la red tendría que hacer un gran de iteraciones. Si es muy grande, los cambios son muy

grandes, avanzando rápidamente por la superficie de error, con el riesgo de saltar el mínimo y oscilar alrededor de él pero sin poder alcanzarlo, es por esto que lo mejor es que conforme la red va aprendiendo α vaya disminuyendo.

El momentum (μ) es una constante que determina el efecto en $t + 1$ del cambio de los pesos en el instante t y que se emplea para filtrar las oscilaciones de (η) Con el momentum la red converge en un menor número de iteraciones, ya que si en t el incremento de un peso era positivo y en $t + 1$ también, entonces el descenso por la superficie de error en $t + 1$ es mayor. El valor de μ puede mantenerse sólo en el intervalo $[0,1]$ aunque este valor también debe reducirse a medida que la red va aprendiendo.

Paso 6 Calcular la función del error. La función indica la capacidad que tiene la red para adaptarse y deberá progresivamente disminuir hasta niveles deseados. La minimización del indicador del error o función de desempeño puede realizarse con varios métodos de optimización.

$$E_k = \frac{1}{2} \sum_{k=1}^M (d_{kj} - Y_{kj})^2 \quad (3.12)$$

Cabe señalar que al hablar de redes *backpropagation* se hace referencia a un algoritmo de aprendizaje más que a una arquitectura (topología) determinada, es decir, el modelo es un perceptrón multicapa con aprendizaje *backpropagation* o simplemente BP cuya función de activación de las capas ocultas es no lineal y la función de activación de la salida es lineal.

3.3 Estado del Arte

Durante la primera parte de la década de 1980 como se menciona anteriormente se dio la combinación del resurgimiento de las redes y la crítica de los modelos econométricos tradicionales lo que incrementó la probabilidad de realizar estudios sobre variables económicas utilizando como metodología la inteligencia artificial ayudando así a dejar atrás el problema de asumir linealidad en un mundo donde la mayoría de las variables son no lineales.

El trabajo seminal en el uso de las redes neuronales para el pronóstico de series de tiempo fue la realizada por Halbert White (1988) quien fue el primero en utilizar las redes motivado a encontrar regularidades no lineales en los movimientos del precio de las acciones. Básicamente su objetivo fue ilustrar como la búsqueda de estas regularidades podía aplicarse en particular a los rendimientos diarios de las acciones de IBM. Utilizando una sencilla red *feedforward* de tres capas (entrada, oculta y salida) pudo confirmar la potencia de esta metodología para las relaciones no lineales. A partir de esta investigación se comenzó a utilizar a las redes neuronales en el campo de las finanzas en diversos temas.

Tiempo después de trabajo de White (*ibíd.*) Altman, Marco y Varetto (Altman *et al.* 1994) realizaron en Italia pruebas de diagnóstico a empresas medianas y pequeñas para pronosticar dificultades financieras, excluyendo a las empresas con ventas por más de 60 millones de dólares, en un periodo de 1985 a 1992. Utilizaron 404 compañías sanas y 404 con problemas financieros. Empleando análisis discriminante y redes neuronales para pronosticar aquellas empresas que tendrían problemas. Dentro de la muestra ambas tuvieron un desempeño similar, pero en cuanto a pronóstico, las redes neuronales probaron ser mejores que el análisis discriminante alcanzando un 94.7% de acierto.

Otra aplicación de las redes en finanzas es el trabajo hecho por Chiang, Urban y Baldrige (Chiang *et al.* 1996) quienes utilizaron redes para pronosticar el valor neto de los activos al final del año de fondos de inversión comparando sus pronósticos con regresiones lineales y no lineales. Encontraron que las redes neuronales superan de sobremanera a ambos tipos de regresiones, en especial cuando se encuentran con disponibilidad limitada de datos.

En cuanto a exclusivamente pronosticar series, las redes neuronales eran cada vez más socorridas, con la motivación de demostrar si eran superiores a otros métodos Hill, O'Connor y Remus (Hill *et al.* 1996) realizaron un ejercicio comparativo entre el desempeño de las redes neuronales artificiales y otros métodos de pronóstico: suavizamiento exponencial, método de Holt y procesos autorregresivo integrados de media móvil (ARIMA) por mencionar algunos. Utilizaron 104 series de tiempo con diferentes frecuencias: 65 series mensuales, 21 series trimestrales y 18 series anuales, encontraron que para las series mensuales y trimestrales las redes neuronales eran significativamente mejores que el resto. Para las series anuales no existió una gran diferencia, obteniendo así desempeños similares con el resto de los métodos.

Un trabajo de investigación fundamental es el realizado Yu Wang, Huang y Lai (Yu *et al.* 2007) quienes presentan una revisión de la literatura respecto al pronóstico de tipo de cambio utilizando las redes neuronales artificiales. Utilizando cierto criterio seleccionan una muestra de 45 artículos de investigación, de los cuales únicamente en 2 casos las redes neuronales artificiales tuvieron un peor rendimiento que la metodología con la cual se comparaba.

Y es que al tratar de mantener una ventaja competitiva la mayoría de las investigaciones ha incorporado sistemas cada vez más avanzados como las RNA, los algoritmos genéticos y la lógica difusa. Aún con la reacción tan positiva de las finanzas ante el uso de las redes neuronales en América Latina no se le ha dado el seguimiento como en otros países, únicamente se encontraron dos investigaciones hechas en Colombia para tomar como referencia.

En el siguiente cuadro se presenta un resumen de los artículos más destacados en los últimos 20 años, para el pronóstico de tipo de cambio con RNA.

Cuadro 3.2: El Uso de las Redes Neuronales para el Pronóstico del Tipo de Cambio.

Año	Autores	Tipo de Cambio ¹	Frecuencia de los datos	Horizonte de Pronóstico	Modelo Estadístico Comparativo	Indicador para comparar ajustes ²	Resultado comparativo
1995	Wu	twd/usd	Mensual	1 mes	ARIMA	RMSE, MAE y MAPE	Mejor
1996	Poddig y Rehkugler	usd/dem, jpy/dem, usd/jpy	Mensual	6 meses	Regresión y Paseo Aleatorio	MSE	Mejor
1997	Jamal y Sundar	usd/dem y usd/frf	Mensual	1 mes	Regresión	PE	Mejor
1998	Franses y Van Homelen	usd/nlg, cad/nlg, gbp/nlg, jpy/nlg	Diaria	1 día	GARCH	Certeza Direccional	Mejor
1999	El Shazaly y El Shazaly	gbp/usd, dem/usd, jpy/usd, chf/usd	trimestral	1 trimestre	Algoritmos Genéticos	TAFE y MAFE	Mejor
2000	Leung <i>et al.</i>	gbp/usd, cad/usd, jpy/usd	mensual	1 mes	Paseo Aleatorio y ARIMA	RMSE y MAE	Mejor
2001	Vojinovic <i>et al.</i>	usd/nzd	diaria	1, 3 y 5 días	Proceso Autoregresivo y Paseo Aleatorio	RMSE	Mejor
2002	Kaashock y Van Dijk	gbp/usd, nlg/usd, frf/usd, dem/usd, jpy/usd	mensual	1 mes	ARIMA	RMSE	Mejor
2003	Jasic y Wood	usd/dem, usd/jpy, usd/chf, usd/gbp	diaria	1 día	Modelo lineal y Paseo Aleatorio	RMSE y MSE	Mejor
2007	Jalil y Misas	usd/cop	diaria	60 días	Paseo Aleatorio y ARIMA	RMSE y MAPE	Mejor
2008	Zapata y Díaz	cop/usd	diaria	30 días	Retrospectiva	n.p.	n.p.
2011	Philip <i>et al.</i>	eur/ngn	diaria	100 días	Cadena Oculta de Markov	MSE	Mejor

Fuente: Elaboración Propia

Nota: La notación de las divisas es la que establece el estándar internacional ISO 4217: dólar canadiense (cad), franco suizo (chf), peso colombiano (cop), marco alemán (dem), franco francés (frf), libra esterlina (gbp), yen japonés (jpy), naira nigeriana (ngn), florín neerlandés (nlg), dólar neozelandés (nzd), dólar taiwanés (twd) y dólar estadounidense (usd). Se emplearon diversos modelos de referencia o también conocidos como *benchmark* que se mencionan para cada caso. Por último, se señala qué indicadores se emplean para la comparación dentro de cada artículo: error porcentual (PE), error medio absoluto (MAE), error absoluto porcentual de la media (MAPE), error cuadrático de la media (MSE), raíz del error cuadrático de la media (RMSE), error total absoluto del pronóstico (TAFE), error medio absoluto del pronóstico (MAFE).

Como se puede observar en 3.2 la importancia de la aplicación de las RNA al tipo de cambio es que en casi todos los casos se generó un mejor pronóstico que cada uno de los modelos de referencia utilizados. Otro detalle a destacar es la diversidad de comportamiento de cada una de las divisas, el yen, la naira y el peso colombiano cada uno de ellos tiene comportamiento distinto además de tener diferentes horizontes temporales, desde 1 día hasta 100 días. Esta evidencia da un mayor sustento al emplear RNA en el pronóstico de tipos de cambio, como se vio en el apartado 3.2 por la ley de aproximación universal de Funahashi una red neuronal puede aproximar cualquier conjunto de funciones.

A diferencia del plano internacional en el ámbito local no se han realizado investigaciones del tipo de cambio empleando redes neuronales artificiales. Dentro de las tesis realizadas que utilizan redes neuronales en las licenciaturas y posgrados de la UNAM¹² las aplicaciones han sido innumerables, por mencionar algunos casos se han hecho trabajos para diferenciar en un diagnóstico entre demencia vascular y tipo Alzheimer (García, 1996), la detección de enfermedades cardiovasculares (Alquicira, 2001) o también para la clasificación de tutores para el posgrado en ciencias de la administración (Bribiesca, 2006) entre otros temas.

Con lo anterior podemos ver que aunque las aplicaciones de las redes neuronales artificiales no se han limitado a un sólo campo, sí han sido contados los que se dedican al análisis de series de tiempo. Torres (2007), García (2009) y Melchor (2010) realizan pronósticos del futuro de cetes a 91 días con frecuencia diaria, del producto interno bruto con frecuencia trimestral y de la inflación con frecuencia mensual respectivamente.

Los primeros dos trabajos utilizan un modelo ARIMA como referencia para comparar el desempeño de la red neuronal. En ambos casos la red es superior al modelo ARIMA, únicamente en el caso del pronóstico de cetes ocurría que cuando el horizonte de predicción superaba los 60 días la red neuronal sobreestimaba los datos y el modelo ARIMA brindaba un mejor ajuste. En el caso del pronóstico de la inflación Melchor emplea información de junio de 2002 a marzo de 2010 y combina los algoritmos genéticos con las redes neuronales para pronosticar la inflación de junio de 2009 a junio de 2010 logrando un MSE menor de 0.1.

¹² La búsqueda se realizó a través de la dirección general de bibliotecas de la UNAM en su sitio <http://dgb.unam.mx/> dentro del catálogo de tesis conocido como TESIUNAM.

Cap. 4 Implementando un modelo ARIMA¹³

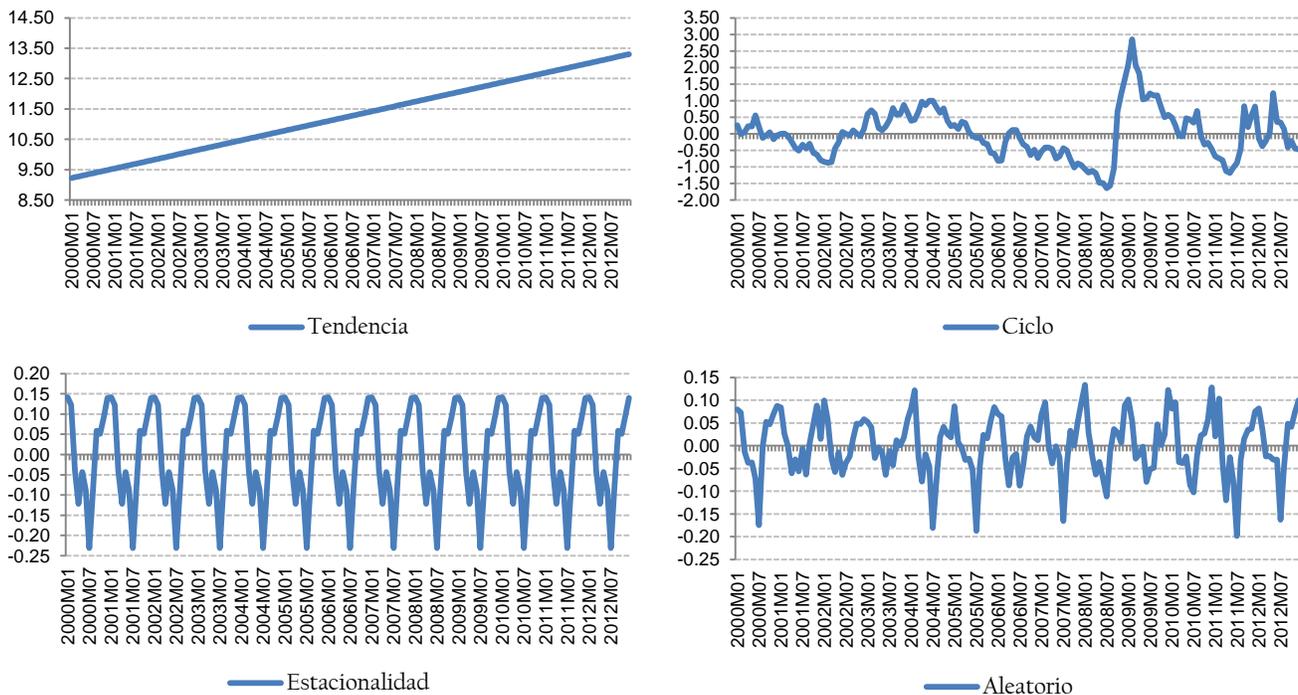
“Prophecy: the art and practice of selling one's credibility for future delivery”

Ambrose Bierce¹⁴

Los métodos tradicionales del estudio de series de tiempo se aplican para realizar el pronóstico *ex post* del tipo de cambio fix promedio mensual tomando en cuenta el período Enero de 2000 a Diciembre de 2012. Con la metodología de Box Jenkins y los procesos de filtrado o suavizamiento se realiza un ajuste en todos los casos mayor del 90%. Por último se genera un pronóstico *ex ante* para los datos de Enero a Diciembre de 2013.

Antes de implementar metodología Box Jenkins se obtienen los principales componentes del tipo de cambio y como se observa en la gráfica 4.1 la evolución dentro del período ha estado caracterizada por una tendencia creciente con una pendiente de \$0.0263 por mes. El ciclo y la estacionalidad presentan los componentes más interesantes por un lado se puede apreciar que de Enero de 2000 a Junio de 2006 el ciclo presentaba una regularidad mayor que la que se presentó en adelante ya que los valles se hicieron más profundos y en general con una mayor volatilidad de manera notoria en los años 2008 y 2009, por otro lado la estacionalidad revela que los meses de Julio y Diciembre se presenta la mayor apreciación (\$0.2309) y depreciación (\$0.1401) del peso. Por último el componente aleatorio no parece tener tendencia ni modificar su varianza con el tiempo

Gráfica 4.1: Componentes del Tipo de Cambio 2000.01 – 2012.12



¹³ Las estimaciones realizadas en este capítulo se llevaron a cabo en los programas Eviews 5 y SPSS 20.

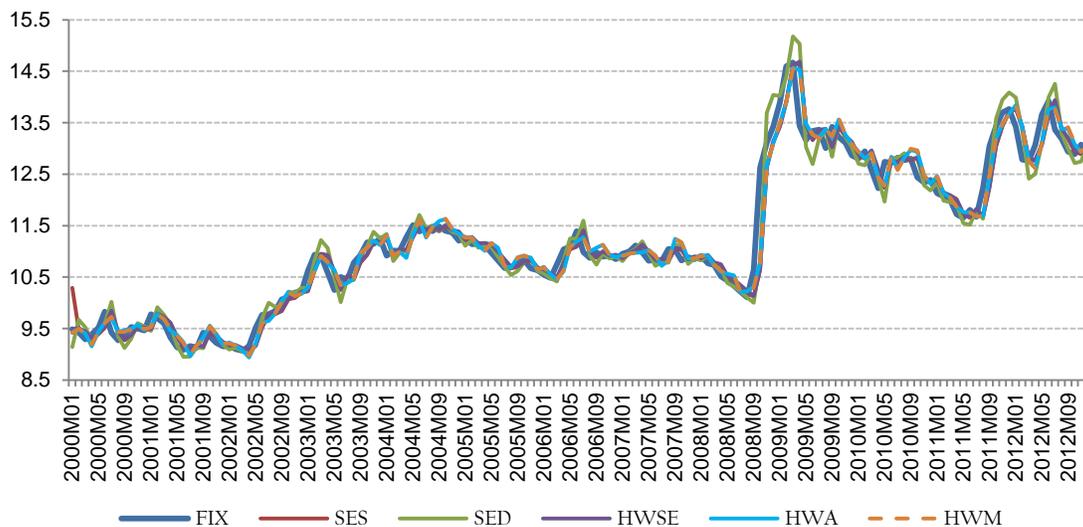
¹⁴ Bierce A. (1993). *The Devil's Dictionary*. Dover Publications

4.1 Métodos de Suavizamiento

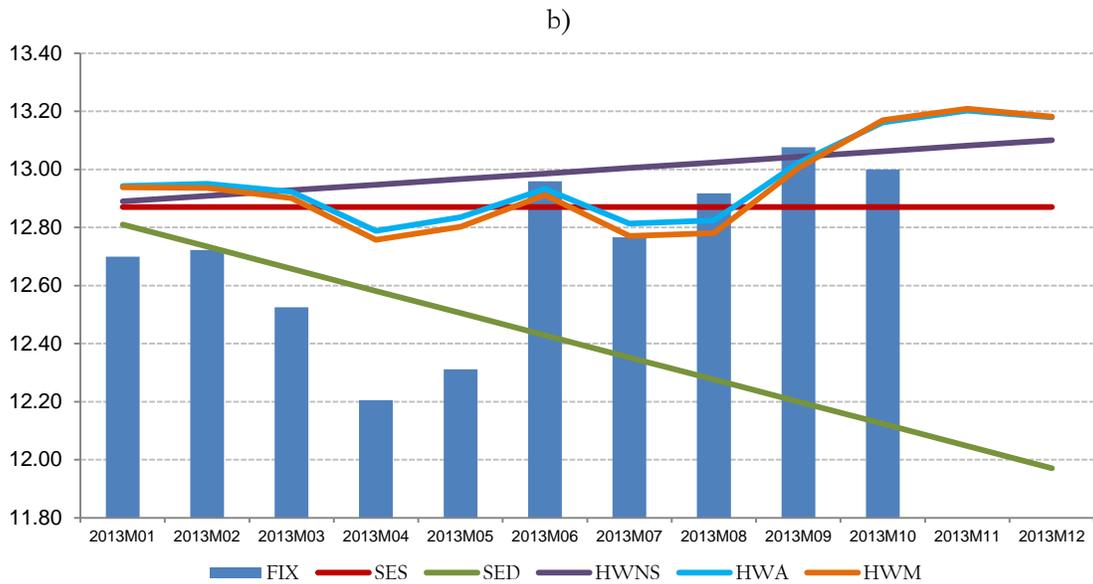
Estos modelos como lo menciona Brown (*op. cit.*) se caracterizan por ser económicos y eficientes ya que el proceso de implementación es muy sencillo de realizar y brindan una gran aproximación a la serie original. En lo que respecta a la selección de parámetros α , β , γ respectivamente se empleó como criterio la suma de los errores al cuadrado (SSR) eligiendo la combinación de parámetros que minimizara la SSR. En el caso del ajuste histórico para los 5 casos¹⁵ α fue superior a 0.7 y tanto β como γ menores a 0.01 logrando una estimación muy cercana dentro de la muestra.

El problema de los modelos de suavizamiento es el pronóstico fuera de la muestra ya que como se puede observar en la gráfica 4.2b tanto el suavizamiento exponencial simple (SES) como el doble (SED) y el modelo Holt - Winters sin estacionalidad (HWSE) son prácticamente tendencias hecho que no conviene del todo en nuestro tema, el modelo aditivo (HWA) y multiplicativo (HWM) logran capturar la estacionalidad lo que de inmediato acentúa la ventaja sobre los otros tres. Aún con los episodios de volatilidad a principio de 2013 tanto HWA como HWM captan la evolución estacional de acuerdo con la cual tendríamos una ligera apreciación del peso en julio y agosto cerrando el año con un dólar cotizando en \$13.0479 en promedio.

Gráfica 4.2 Ajuste Histórico y Pronóstico con Filtros de Suavizamiento, 2013.01 – 2013.12
a)



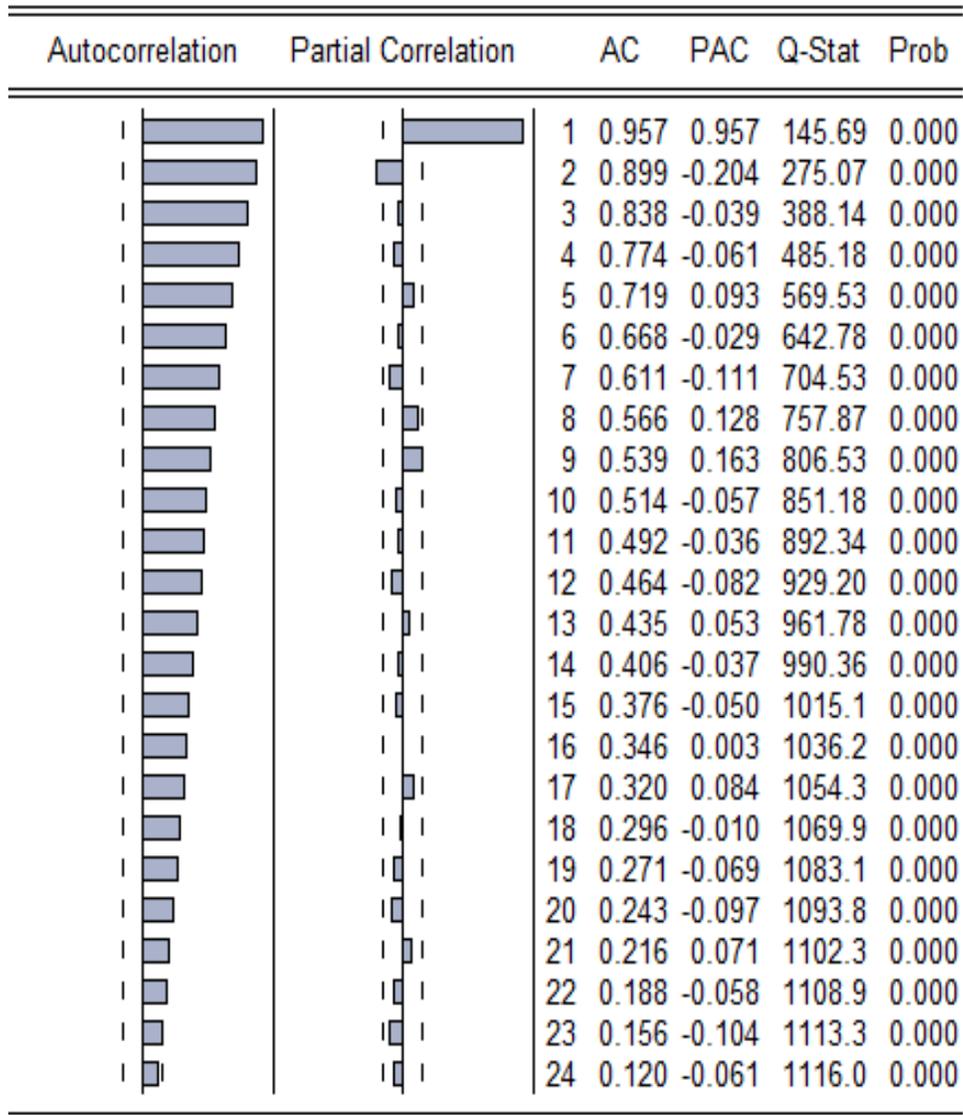
¹⁵ La evidencia se encuentra en el Anexo en los cuadros A1 a A5.



4.2 Identificación de Modelos

En la fase de identificación la naturaleza de la correlación entre los valores de los residuos presentes y sus valores pasados es lo que guía en la especificación de un modelo ARIMA. Tal como se mencionó en el capítulo 2 las autocorrelaciones son los valores de correlación entre el valor actual de la serie y su rezago de cierto número de períodos, esta función junto con la autocorrelación parcial nos indican que modelo emplear por ejemplo si la autocorrelación disminuye geoméricamente y la autocorrelación parcial se vuelve cero después de un rezago el modelo que se debe emplear es un autorregresivo de primer orden o si parecen tener un componente estacional se debe implementar un ARMA. En el presente caso la función de autocorrelación y la autocorrelación parcial nos muestran un claro ejemplo de un modelo autorregresivo de primer orden.

Gráfica 4.3: Función de Autocorrelación y Autocorrelación Parcial del Tipo de Cambio con 24 rezagos



Otro factor que se debe de revisar es el orden de integración de la serie para saber si la serie cuenta con raíz unitaria se aplicaron cuatro pruebas a cuatro transformaciones del tipo de cambio en niveles, logaritmo, primera diferencia y rendimiento los resultados se presentan en el cuadro 4.1

Cuadro 4.1 Estadísticas Descriptivas y Pruebas de Raíz Unitaria

		fix	Δ fix	fix	Δ fix
Media		11.2425	0.0218	2.4122	0.0020
Mediana		10.9830	-0.0248	2.3963	-0.0022
Máximo		14.6695	1.9877	2.6858	0.1712
Mínimo		9.0707	-1.2328	2.2050	-0.0878
Desviación Estándar		1.3878	0.2993	0.1224	0.0251
Asimetría		0.3443	1.5641	0.1419	1.8643
Curtosis		2.2264	15.6086	2.1301	15.8407
ADF	1	0.4977	-8.9620	0.6224	-9.0159
	2	-1.8829	-8.9694	-1.8786	-9.0301
	3	-3.8265	-8.9398	-3.5883	-9.0012
DF GLS	1	-	-	-	-
	2	-0.8746	-8.6546	-0.8792	-8.5816
	3	-3.8019	-8.8823	-3.5936	-8.9050
PP	1	0.6027	-8.9338	0.9056	-8.9155
	2	-1.6551	-8.8555	-1.5278	-8.9243
	3	-3.1956	-8.8243	-3.0862	-8.8939
KPSS	1	-	-	-	-
	2	1.2734	0.0404	1.2402	0.0427
	3	0.0565	0.0402	0.0614	0.0415

Nota: El modelo 1 es sin emplear ni intercepto ni tendencia. El modelo 2 emplea intercepto. El modelo 3 emplea intercepto y tendencia. Los estadísticos en negritas representan el rechazo de la hipótesis nula al 5%.

Respecto a los resultados de las pruebas de raíz unitaria se presenta el dilema de que prueba es más adecuada analizando cada uno de los modelos de las 4 pruebas más empleadas es posible concluir que las series en niveles y en logaritmos presentan evidencia de ser procesos estocásticos no estacionarios ya que de las 4 pruebas sólo en tres se encontró prueba de lo contrario pero, en la presente investigación concluimos que se puede tratar a ambas series como procesos estacionarios debido a que los modelos que confirman la estacionariedad son los modelos más robustos y con los mejores criterios de información de Akaike y Schwarz.¹⁶

Con la evidencia de las funciones de autocorrelación donde se nota la disminución geométrica y la autocorrelación parcial que después de un rezago se vuelve cercana a cero. Si a lo anterior añadimos las pruebas de estacionariedad es claro que el tipo de cambio se comporta como un modelo ARIMA (1, 0, 0) tal y como Meese y Rogoff destacaron.

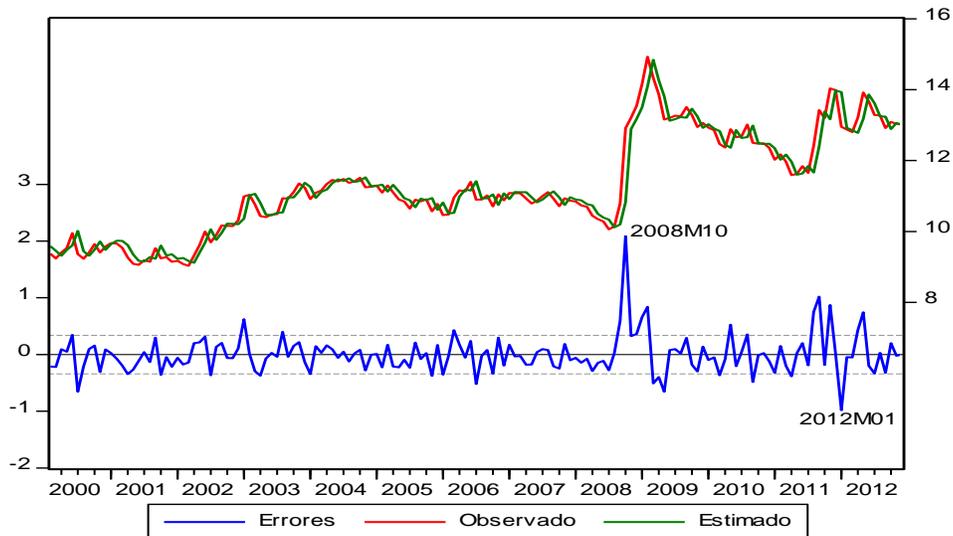
¹⁶ La evidencia se encuentra en el Anexo en los cuadros A.6 a A.15

Cuadro 4.2: Estimación Modelo ARIMA (1, 0, 0)

Dependent Variable: TC
 Method: Least Squares
 Sample (adjusted): 2000M02 2012M12
 Included observations: 155 after adjustments
 Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12.02315	1.053008	11.41792	0.0000
AR(1)	0.970506	0.019459	49.87563	0.0000
R-squared	0.942058	Mean dependent var	11.28059	
Adjusted R-squared	0.941679	S.D. dependent var	1.405126	
S.E. of regression	0.339333	Akaike info criterion	0.689150	
Sum squared resid	17.61748	Schwarz criterion	0.728420	
Log likelihood	-51.40911	F-statistic	2487.578	
Durbin-Watson stat	1.689235	Prob(F-statistic)	0.000000	
Inverted AR Roots	.97			

Gráfica 4.4: Ajuste Histórico con el Modelo ARIMA (1, 0, 0), 2000.01 – 2012.12



De la regresión podemos destacar que el proceso es invertible debido a que el inverso de la raíz del autorregresivo está dentro del círculo unitario. El ajuste de la regresión es elevado pero para saber si es útil se deben hacer las pruebas necesarias para contrastar la adecuada especificación.

4.3 Pruebas de Correcta Especificación

En orden para poder realizar el pronóstico es necesario comprobar que el modelo tiene una correcta especificación, para verificar lo anterior se realizaron las pruebas¹⁷ de Breusch Godfrey para la Autocorrelación y el estadístico de Robert Engle para comprobar la ausencia de Heterocedasticidad Condicional Autorregresiva (ARCH) y debido a que las pruebas de estabilidad tradicionales no se pueden aplicar a modelos ARIMA utilizamos la prueba de pronóstico de Chow para saber si tiene estabilidad estructural o si existe algún punto de quiebre.

Cuadro 4.3 Resumen de Pruebas de Correcta Especificación

Prueba	Estadístico	L1	L2	L3
BG	F	3.93	1.96	1.56
	Prob.	0.05	0.15	0.21
ARCH	F	0.93	0.58	0.63
	Prob.	0.34	0.56	0.6
CHOW	F	1.45	2012M1-2012M12	
	Prob.	0.12		

Después de ajustar el modelo y comprobar su correcta especificación ya que al contrastar las pruebas con sus respectivos estadísticos F podemos afirmar que no hay autocorrelación serial de primer, segundo o tercer orden así como tampoco existe un componente heterocedástico y por último existe estabilidad estructural lo que significa que hay un solo modelo y no ha existido un cambio estructural para el último año se puede proceder a la generación del pronóstico.

4.4 Pronóstico

Para realizar el pronóstico fuera de la muestra empleamos el método estático de un paso adelante o *rolling forecast* ya que el método dinámico en un modelo ARIMA(1,0,0) converge geométricamente a su media de manera monótonica si el parámetro es positivo y oscilatorio si es negativo (Kirchgaässner and Wolters, *op. cit.*) de hacer el pronóstico dinámico el mismo debe converger geométricamente a 11.2806, por lo que en este caso no se aplica.

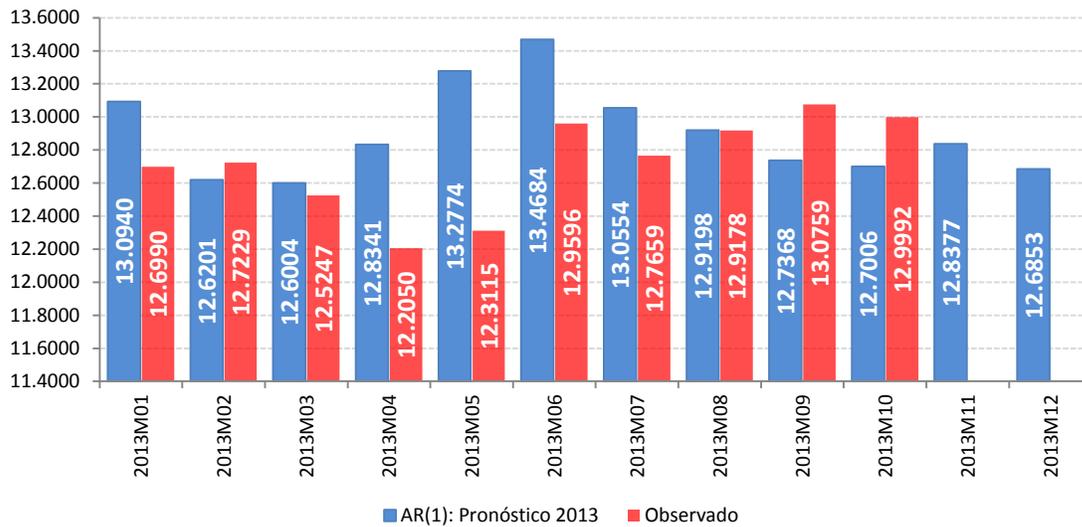
La especificación del modelo servirá para poder generar datos ex ante, en primer lugar se necesita ajustar el rango para incluir todo el 2013 de manera que al pronosticar se genere la observación de Enero de 2013 manteniendo 155 observaciones, al hacer esto se pierde el dato de Febrero de 2000. Para generar el segundo dato fuera de la muestra no es necesario realizar pruebas de correcta especificación para cada observación ya que al ser

¹⁷ Las respectivas regresiones se encuentran en el Anexo en los cuadros A.16 a A.23.

marginalmente la misma muestra no se generan cambios relevantes en las pruebas ni en los estadísticos.

Como se muestra en la gráfica 4.2 el pronóstico no capturo la apreciación del peso que se vivió en Abril y Mayo y se esperaba un depreciación lo que llevo a tener un error cercano a la unidad para Junio alcanza el punto más alto de depreciación que de hecho ocurre pero sobrestima por casi \$0.50. De acuerdo con el modelo ARIMA en Julio se inicia una tendencia de apreciación del peso para cerrar en Diciembre alrededor \$12.70.

Gráfica 4.2 Pronóstico con el Modelo ARIMA (1, 0, 0), 2013.01 – 2013.12



Capítulo 5 Construcción de las Redes Neuronales Artificiales¹⁸

“An economist is an expert who will know tomorrow why the things he predicted yesterday didn't happen today”

Evan Esar

“If it were possible to calculate the future state of the market, the future would not be uncertain. There would be neither entrepreneurial loss nor profit. What people expect from the economists is beyond the power of any mortal man”

Ludwig Von Mises

Dentro de las ciencias sociales es destacable que la única disciplina que constantemente ha buscado una rendija al futuro es la economía. Normalmente se concibe a la econometría como algo que nos aleja de la economía misma a través de abstracciones y supuestos. En este capítulo tomamos en cuenta dos puntos de vista: el primero determina que el tipo de cambio sólo está en función de su pasado y el segundo, que existen relaciones que pueden causar modificaciones en el tipo de cambio.

El objetivo tanto del capítulo como de la tesis es generar un mejor aproximamiento al tipo de cambio a través de las redes neuronales artificiales para poder generalizar de mejor forma y poder establecer mejores pronósticos desde luego buscando que el avance que se logre en las estimaciones represente un intercambio de costo beneficio razonable ante los modelos tradicionales de fácil implementación.

Para llevar a cabo la implementación de la red neuronal artificial univariada y la multivariada se siguen los 8 pasos que se presentan en cuadro 5.1

Cuadro 5.1. Ocho Pasos a Seguir en la Construcción de una RNA

Paso 1	Elección de Variable	
Paso 2	Recolección de Datos	
Paso 3	Pre-proceso de Datos	
Paso 4	Sets de Entrenamiento, Validación y Prueba	
Paso 5	Arquitectura de las Redes Neuronales	Número de Capas Ocultas
		Número de Neuronas Ocultas
		Funciones de Activación
Paso 6	Criterios de Evaluación	
Paso 7	Entrenamiento de la Red Neuronal	Actualización de Levenberg- Marquardt
		Criterios para detener el entrenamiento
Paso 8	Implementación	

¹⁸ Las estimaciones realizadas en este capítulo se llevaron a cabo en los programas EVIEWS 5, SPSS 20 y MATLAB R2012a.

Paso 1 y 2. Elección de Variables y Recolección de Datos

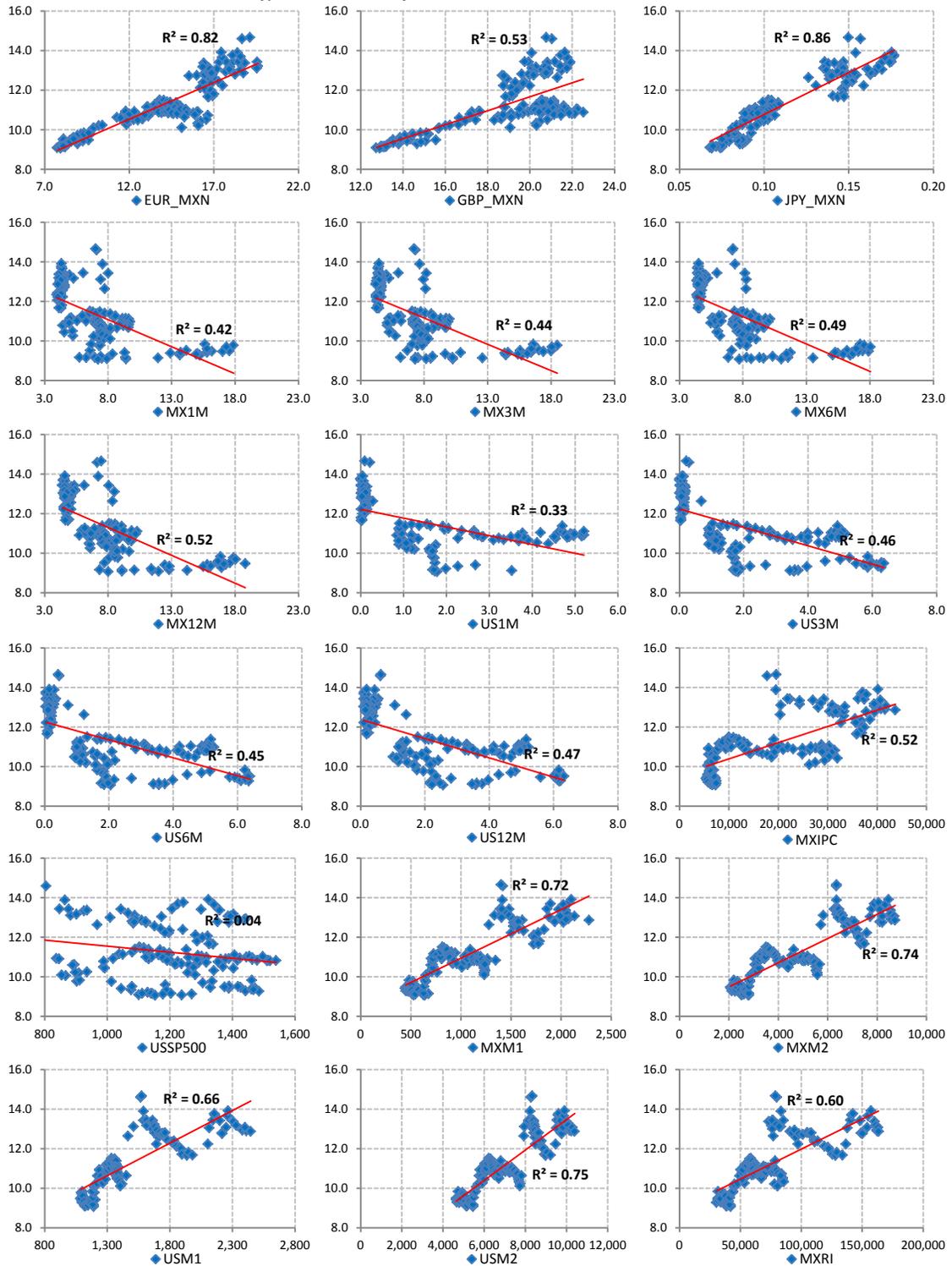
La elección de variables se realizó en función del cuerpo teórico desarrollado en el capítulo 1. El vector de información para la red multivariada es:

$$\mathbf{X} = \left[\begin{array}{l}
 \text{eurmxn}_t = \text{pesos por 1 euro} \\
 \text{gbpmxn}_t = \text{pesos por 1 libra esterlina} \\
 \text{jpymxn}_t = \text{pesos por 1 yen japones} \\
 \text{mx1m}_t = \text{Cetes 28 días} \\
 \text{mx3m}_t = \text{Cetes 91 días} \\
 \text{mx6m}_t = \text{Cetes 182 días} \\
 \text{mx12m}_t = \text{Cetes 364 días} \\
 \text{us1m}_t = T - \text{Bill 1 mes} \\
 \text{us3m}_t = T - \text{Bill 3 meses} \\
 \text{us6m}_t = T - \text{Bill 6 meses} \\
 \text{us12m}_t = T - \text{Bill 12 meses} \\
 \text{mxipc}_t = \text{Indice de Precios y Cotizaciones} \\
 \text{ussp500}_t = \text{Standard \& Poors 500} \\
 \text{mxm1}_t = \text{M1 en México} \\
 \text{mxm2}_t = \text{M2 en México} \\
 \text{usm1}_t = \text{M1 en EEUU} \\
 \text{usm2}_t = \text{M2 en EEUU} \\
 \text{mxri}_t = \text{Reservas Internacionales} \\
 \text{mxsbc}_t = \text{Saldo de Balanza Comercial}
 \end{array} \right] \quad (5.1)$$

La recolección de las series se obtuvo de los sitios de divulgación oficiales. Para el caso de las tasas de interés cetes, los agregados monetarios M1 y M2, reservas internacionales, saldo de la balanza comercial y lo tipos de cambio se obtuvieron de las estadísticas del Banco de México. Por su parte las tasas de interés T-bills, los agregados monetarios y el índice Standard & Poors se obtuvieron del sistema de datos económicos de la Reserva Federal (Federal Reserve Economic Data, FRED).

Para observar si existe una relación entre las variables propuestas en el vector de información y el tipo de cambio se realizan los diagramas de dispersión (ver Gráfica 5.1) y se obtienen los coeficientes de determinación para comprobar que tan fuerte es la relación entre la variable dependiente y el resto de las variables independientes.

Gráfica 5.1 Diagrama de Dispersión Fix v.s. Vector de Información



Se puede observar que existe una fuerte relación positiva entre el tipo de cambio y las otras tres divisas, de acuerdo a los datos correspondientes a la muestra cada que el peso a perdido terreno frente a una moneda se deprecia frente a las otras. Las tasas de interés por

otra parte presentan comportamientos de acuerdo a la teoría en algunos casos, tomando todo lo demás constante, si las tasas de interés de México aumentan atraen inversión extranjera lo cual significa que se demandan más pesos y se ofrecen más divisas lo que claramente es una relación negativa entre tasas de interés y tipo de cambio lo cual se cumple para los cetes pero no así con las tasas de interés de Estados Unidos, las cuales también muestran una relación negativa. Los agregados monetarios también se comportan de acuerdo con la teoría una relación positiva ante un incremento en la oferta monetaria se espera una depreciación del peso. Por último se aprecia una relación positiva con el índice de precios y cotizaciones y con las reservas internacionales.

Una cuestión importante es que aun teniendo estos coeficientes de determinación no sabemos si existe algo más que una correlación de las variables independientes con la dependiente por eso es necesario realizar prueba de causalidad para saber si efectivamente existe la relación entre las variables.

Cuadro 5.2 Pruebas de Causalidad de Granger

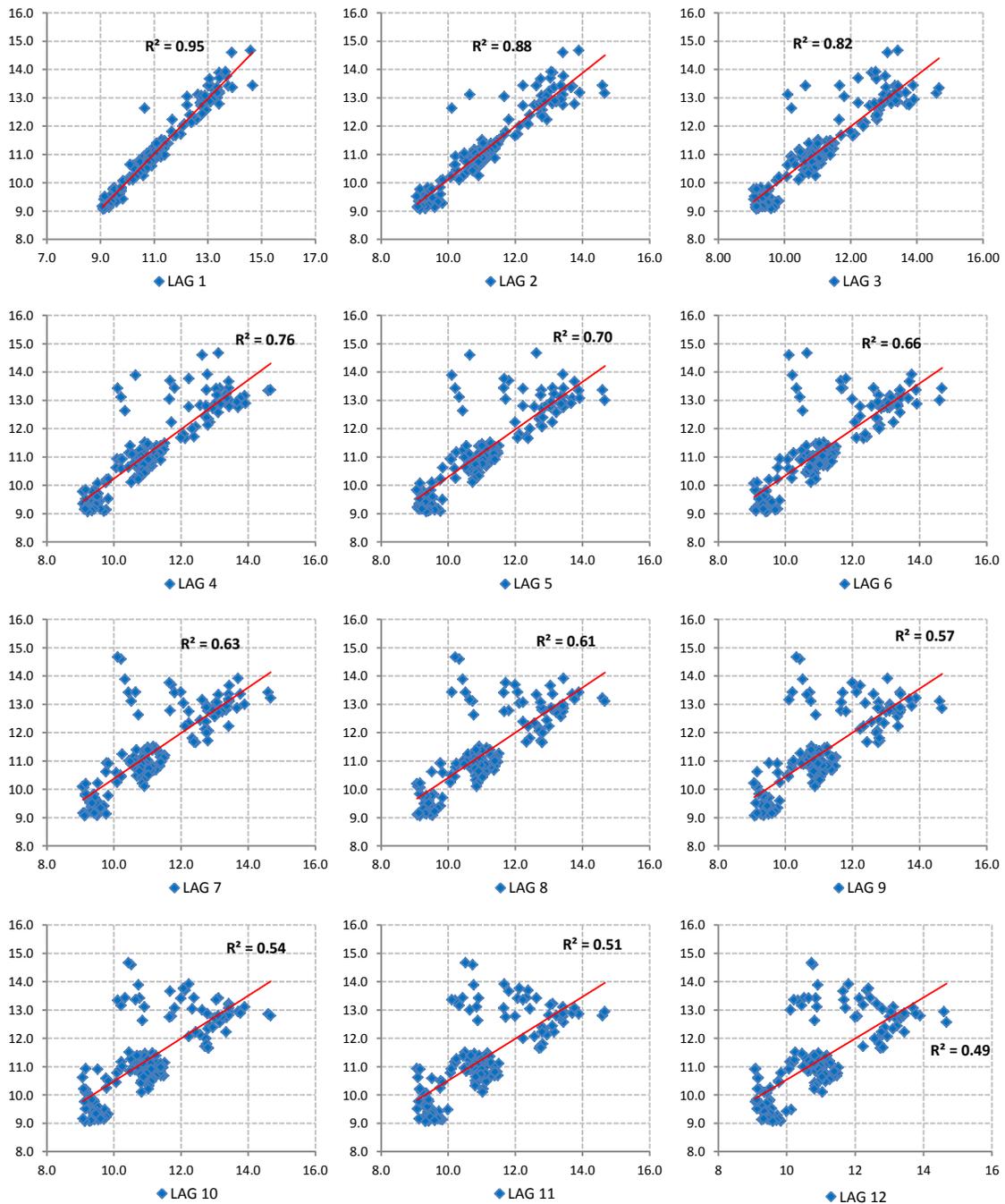
	L (1)		L (2)		L (3)		L (4)		L (5)		L (6)	
	F stat	Prob.	F stat	Prob.	F stat	Prob.	F stat	Prob.	F stat	Prob.	F stat	Prob.
EURMXN --/--> FIX	2.1659	0.143	3.1209	0.047	1.7989	0.150	2.3235	0.059	2.9415	0.015	2.3171	0.037
GBPMXN --/--> FIX	0.0065	0.936	0.0692	0.933	2.6370	0.042	2.1376	0.079	1.7416	0.129	1.7000	0.125
JPYMXN --/--> FIX	1.1018	0.296	3.2024	0.043	2.0800	0.105	1.5417	0.193	1.5045	0.192	1.0426	0.401
MX1M --/--> FIX	3.3358	0.070	3.4033	0.036	2.3165	0.078	1.7721	0.138	1.5003	0.194	1.4934	0.185
MX3M --/--> FIX	3.4846	0.064	3.6595	0.028	2.2860	0.081	1.7254	0.148	1.5089	0.191	1.6164	0.147
MX6M --/--> FIX	3.8224	0.052	4.3601	0.014	2.5826	0.056	1.9611	0.104	1.6773	0.144	1.7708	0.110
MX12M --/--> FIX	3.6456	0.058	3.9293	0.022	2.4911	0.063	1.9359	0.108	1.7514	0.127	1.7144	0.122
US1M --/--> FIX	3.9250	0.047	1.5438	0.217	1.2757	0.286	1.2457	0.295	1.9268	0.095	1.9902	0.072
US3M --/--> FIX	3.9553	0.049	2.2436	0.110	1.3865	0.249	1.1919	0.317	2.1454	0.064	2.3160	0.037
US6M --/--> FIX	3.2535	0.073	2.0025	0.139	1.1807	0.319	0.9406	0.442	2.9342	0.015	2.5420	0.023
US12M --/--> FIX	3.1445	0.078	1.9297	0.149	1.1093	0.347	0.9446	0.440	2.7881	0.020	2.4338	0.029
MXM1 --/--> FIX	2.6605	0.105	3.8805	0.023	2.1814	0.093	1.8575	0.121	3.1652	0.010	2.4510	0.028
MXM2 --/--> FIX	3.7307	0.055	4.6013	0.012	4.2137	0.007	3.1471	0.016	5.9969	0.000	4.5986	0.000
USM1 --/--> FIX	2.4067	0.123	7.0321	0.001	5.9632	0.001	6.2742	0.000	5.1706	0.000	4.4210	0.000
USM2 --/--> FIX	5.4709	0.021	6.0856	0.003	3.7510	0.012	3.0456	0.019	3.1398	0.010	2.6567	0.018
MXIPC --/--> FIX	0.5828	0.446	6.3504	0.002	4.6544	0.004	4.3819	0.002	7.6862	0.000	6.1657	0.000
USSP500 --/--> FIX	0.3979	0.529	0.3402	0.712	0.2062	0.892	2.5609	0.041	2.3508	0.044	2.5296	0.024
MXRI --/--> FIX	2.4475	0.120	2.7736	0.066	1.7039	0.169	3.6796	0.007	3.5772	0.004	3.6120	0.002
MXSBC --/--> FIX	16.3740	0.000	4.2923	0.015	3.9623	0.009	3.3574	0.012	2.8480	0.018	2.1629	0.050

De acuerdo con el cuadro 5.2 podemos ver que existe una relación de causalidad entre las variables independientes y la dependiente si bien no en el mismo orden de rezago pero si en alguno de ellos. Los casos del IPC, el SBC y el M2 de México y de EE.UU. presentan

pruebas estadísticas de causalidad en varios rezagos lo que nos puede indicar que probablemente sean variables del vector de información muy importantes.

Para la red univariada normalmente se recomienda emplear 10 rezagos (Faraway y Chatfield, 1998) pero como se observa en la gráfica 5.2 hasta 12 rezagos existe una relación importante por lo que se propone que la capa de entrada se conforme con los 12 rezagos del tipo de cambio.

Gráfica 5.2 Diagrama de Dispersión Fix v.s. 12 Rezagos



Paso 3. Pre-procesamiento de Datos

Antes de proceder con el entrenamiento es muy útil re escalar las variables para que se encuentren dentro de un rango específico. Una de las transformaciones es la estandarización la cual requiere sustraer la media de la serie y dividir entre la desviación estándar de manera que la nueva serie generada tiene media 0 y varianza 1. Otra transformación es la normalización estándar la cual consiste en sustraer el valor mínimo y dividir entre el rango, con esta transformación la serie generada tiene su rango entre [0 y 1]. Por último la transformación que aplicamos a los datos la normalización ajustada la cual se puede obtener por medio de la ecuación 5.2a o 5.2b y su rango se encuentra entre [-1 y 1].

$$y_t = \frac{2*(x_t - x_{min})}{x_{max} - x_{min}} - 1 \quad (5.2a)$$

$$y_t = \frac{2 - (x_{max} + x_{min})}{x_{max} - x_{min}} \quad (5.2b)$$

Paso 4. Conjuntos de Entrenamiento, Validación y Prueba

Una vez que se han recolectado las variables y se han procesado a través de una normalización ajustada se dividen la matriz de variables independientes y el vector de objetivos en los siguientes tres conjuntos:

Conjunto de entrenamiento

Toma el 70% de los datos de la serie los cuales se seleccionan de forma ordenada de Enero de 2000 a Octubre de 2009. Este conjunto se presenta a la red durante el entrenamiento y la red se ajusta de acuerdo al error.

Conjunto de validación

Toma el siguiente 20% de los datos de la serie que conforman el periodo de Noviembre de 2009 a Julio de 2012. Este conjunto se utiliza para medir la capacidad de generalización y para detener el entrenamiento cuando la generalización deja de mejorar.

Conjunto de prueba

Toma el último 10% de los datos de la serie que conforman de Agosto de 2012 a Diciembre de 2013. Este es el conjunto más importante ya que conforma el pronóstico de la red ya que, la red emplea los conjuntos de entrenamiento y validación para generar la estimación y con los pesos obtenidos genera la parte que corresponde al conjunto de prueba sólo después de generalizar con los dos conjuntos anteriores.

Paso 5. Arquitectura de la Red Neuronal

La capa de entrada de las redes está en función del vector de información, tal como lo establecimos en la elección de variables del paso 1, la capa de entrada de la red multivariada estará compuesta por 19 entradas y la capa de entrada de la red univariada estará compuesta por 12 entradas.

Para establecer el número de capas ocultas tomamos en cuenta a Funahashi (op. cit.) quien establece que una red compuesta por tres capas es capaz de aproximar cualquier función, además de que con mayor número de capas ocultas se puede generar un problema de sobreajuste de la red.

Determinar el número de neuronas dentro de la capa oculta es un poco más complicado ya que aunque existen algunas propuestas como la ecuación 5.3¹⁹ o la ecuación 5.4 si establecemos pocas neuronas se impide el correcto mapeo y por el contrario si se determina un gran número se puede obtener un error menor pero generando problemas de generalización.

$$h = \sqrt{i \times o} \quad (5.3)$$

$$h = \frac{2}{3}(i + o) \quad (5.4)$$

Tomando en cuenta lo anterior se establece tanto para la univariada como para la multivariada una red de tres capas donde la capa oculta cuente con 10 neuronas ocultas y la capa de salida cuente únicamente con una neurona.

Función de Activación

Como se mencionó en el capítulo 3 las funciones de transferencia o activación son generalmente la sigmoideal, tangente hiperbólica y lineal. La función tangente hiperbólica (eq 5.5) se construye como el cociente del seno hiperbólico²⁰ y el coseno hiperbólico²¹ y está acotada entre $[-1, 1]$.

$$\tanh = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (5.5)$$

Un equivalente matemático de la función tangente hiperbólica es la función que se presenta en la ecuación 5.6 la cual se emplea debido a que incrementa la velocidad de estimación y tiene muy pequeños cambios en los valores respecto a la función tangente hiperbólica lo que representa una buena razón de intercambio en las redes neuronales.

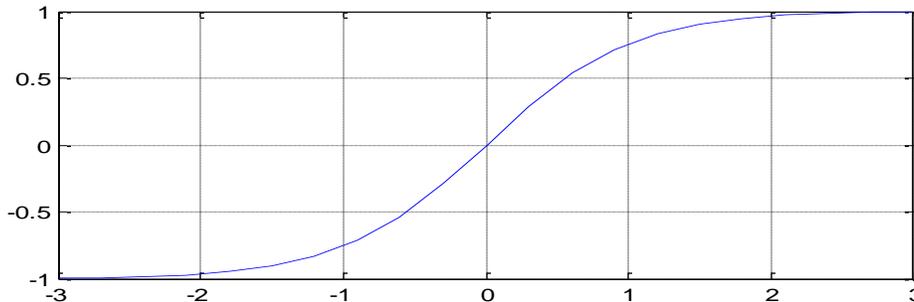
¹⁹ i Es el número de neuronas de la capa de entrada, h es el número de neuronas de la capa oculta y o es el número de neuronas de la capa de salida

²⁰ $\sinh = e^x - e^{-x} / 2$

²¹ $\cosh = e^x + e^{-x} / 2$

$$tansig = \frac{2}{(1+e^{-2x})-1} \quad (5.6)$$

Gráfica 5.5: Función de Activación Tangente Hiperbólica Sigmoidal



Paso 6 Criterio de Evaluación

El criterio que se emplea para evaluar el desempeño o *performance* de la red es la media de los errores al cuadrado (MSE) donde el objetivo ideal sería que fuera igual a cero. Básicamente se calcula la suma de los errores al cuadrado y se obtiene su media donde los errores son la diferencia del valor objetivo y el valor de salida.

$$MSE = \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n} \quad (5.7)$$

Este criterio es fundamental para el siguiente paso, el entrenamiento como veremos a continuación es la función que nos indica si continua entrenando o se ha logrado minimizar.

Paso 7. Entrenamiento de la Red Neuronal Artificial

La función de entrenamiento que ocupa la optimización de Levenberg-Marquandt es uno de los algoritmos más rápidos de *backpropagation*. Emplea el vector de validación para detener el entrenamiento si el desempeño de la red no logra mejorar o si se mantiene igual por un número determinado de iteraciones seguidas (*max_fail_epochs in a row*)

El algoritmo *backpropagation* se usa para calcular el jacobiano de la función de desempeño respecto a los pesos. Cuando la función de desempeño tiene la forma de la suma de cuadrados de los errores se puede aproximar la matriz Hessiana como

$$H = J^T J \quad (5.8)$$

El algoritmo emplea esa aproximación a la matriz Hessiana en la actualización del tipo Newtoniana:

$$\Delta x = [J^T J + \mu I]^{-1} J^T e \quad (5.9)$$

Cuando el escalar μ es zero se convierte en el método de Newton y a mayor dimensión del escalar μ se convierte en el método de gradiente descendente con un pequeño step size. Al ser más eficiente el método de Newton cerca del error mínimo μ decrece después de cada reducción en la función de desempeño y se incrementa cuando tentativamente un paso más incrementaría la función de desempeño de manera que la función de desempeño se reduce cada iteración del algoritmo.

Los pasos a seguir al entrenar una red neuronal artificial en el modo de *batch training* utilizando el algoritmo Levenberg-Marquardt son los siguientes:

1. Presentar todos los inputs a la red y estimar los correspondientes outputs y errores. Calcular la media del error cuadrático tal como se presenta en la ecuación 5.7
2. Obtener la matriz Jacobiana, $J(x)$ donde x representa los pesos y sesgos de la red.
3. Resolver la ecuación de Levenberg-Marquardt de actualización de pesos para obtener Δx . (ecuación 5.9)
4. Volver a calcular el error usando $x + \Delta x$. Si el nuevo error es más pequeño que el estimado en el paso 1 se reduce μ en μ_{dec} unidades, se deja $x = x + \Delta x$, y se vuelve al paso 1. Si el error no se reduce, entonces se incrementa μ en μ_{inc} unidades y se vuelve al paso 3. μ_{dec} y μ_{inc} se definen como $\mu_{inc} = 10$ y $\mu_{dec} = 0.1$.
5. El algoritmo se asume que converge cuando la norma del gradiente es menor que algún valor predetermine o cuando el error ha sido reducido a algún valor meta normalmente 0.

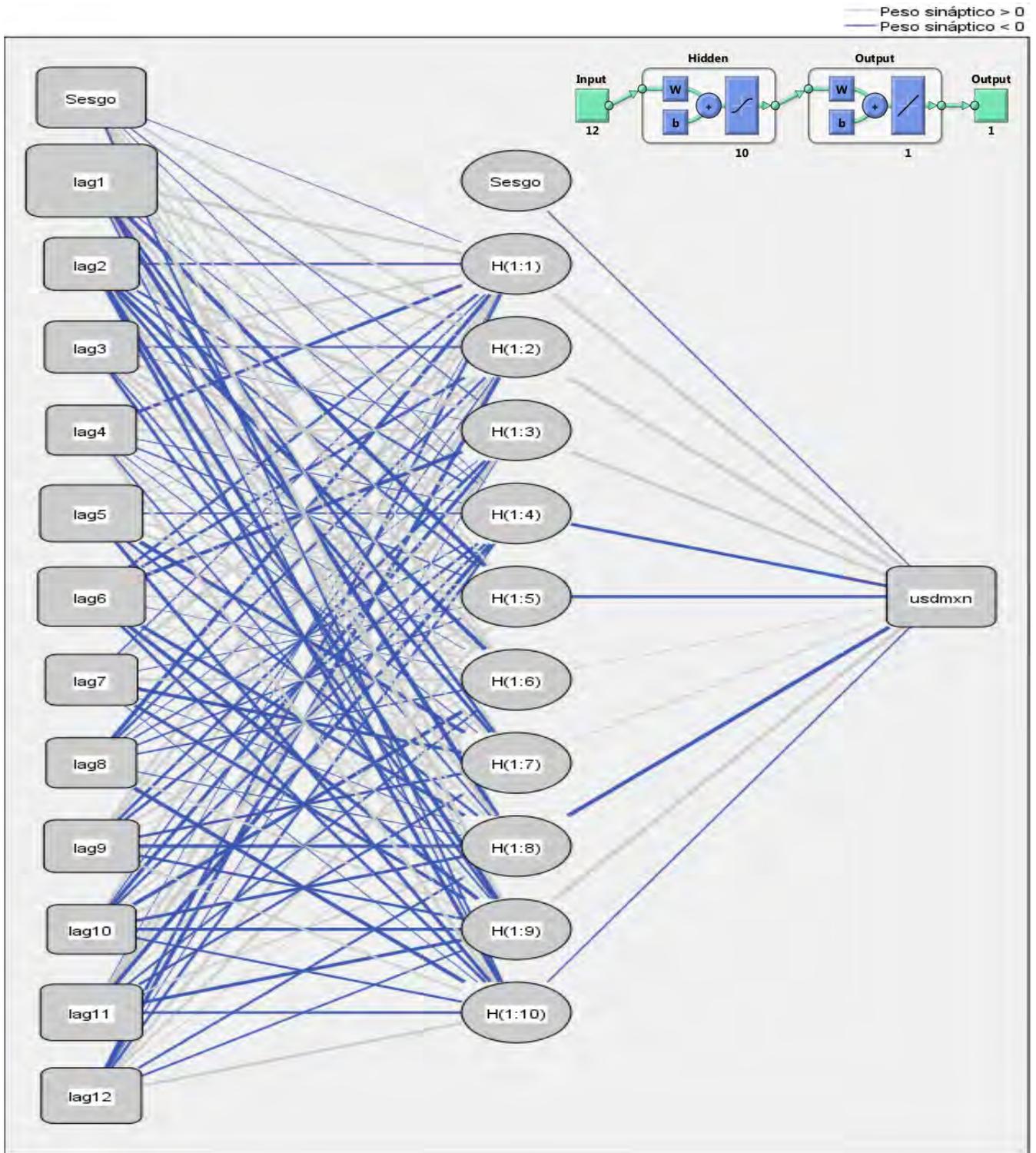
Este entrenamiento se conoce como entrenamiento por lote (*batch training*) y funciona actualizando los pesos sinápticos de la red hasta que se presentan todos los datos que forman parte del entrenamiento actualizando los pesos en varias ocasiones hasta que se alcanza un criterio para detener el algoritmo, los cuales son:

1. Alcanzar el número máximo de iteraciones
2. Alcanzar el tiempo máximo de estimación predeterminado.
3. La función de desempeño es minimizado y alcanza la meta.
4. El gradiente de la función de desempeño alcanza el mínimo (*min_grad*)
5. μ excede el valor máximo (*mu_max*)
6. El desempeño para la validación se ha incrementado más veces que el establecido (*max_fail*) desde la última vez que disminuyo.

Paso 8. Implementación

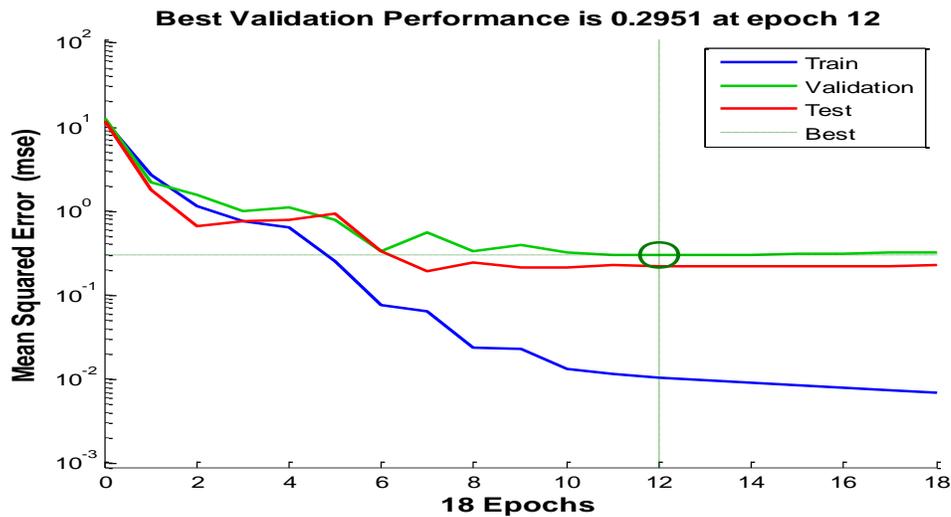
Una vez que se establecieron los criterios anteriores se implementa tanto la red neuronal univariada como la red multivariada. En primer lugar se presentan los resultados de la red univariada y su pronóstico para después presentar los propios de la red multivariada.

Gráfica 5.6: Arquitectura de la Red Neuronal Artificial Univariada



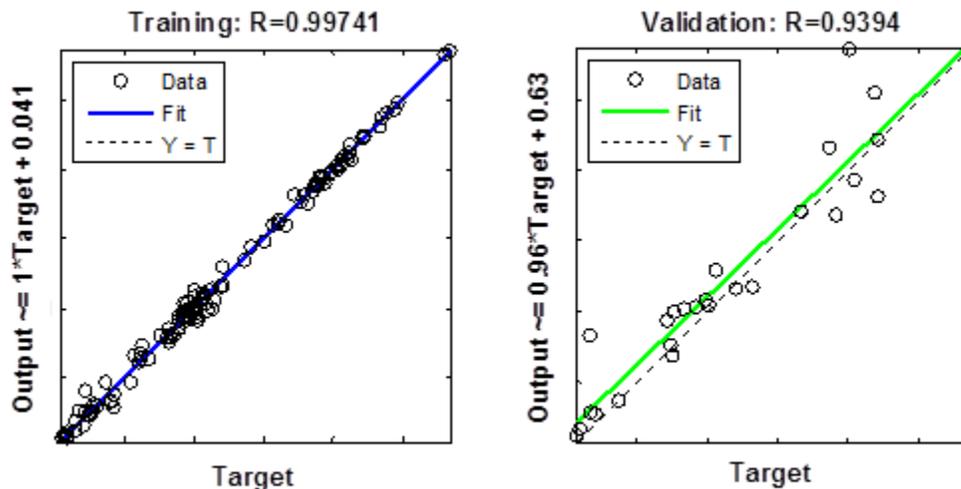
Ya que contamos con la red podemos observar en la gráfica 5.7 que al implementar el entrenamiento por lotes con el algoritmo de Levenberg-Marquardt la función de desempeño para el conjunto de validación encontró el mejor punto después de 12 iteraciones. De la misma forma se observa cómo se disminuye el error conforme ocurren más iteraciones tanto el conjunto de entrenamiento como el de validación y por último el conjunto de prueba que demuestra como la red realizó una adecuada generalización y no tiene sobreajuste de los datos.

Gráfica 5.7: Red Neuronal Artificial Univariada: Validación Cruzada

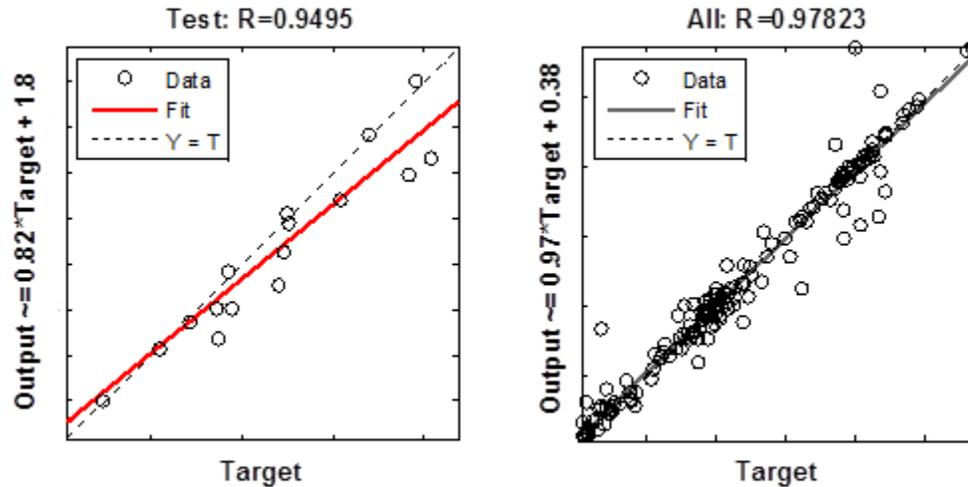


Para evaluar la bondad de ajuste de la serie estimada a través del entrenamiento supervisado y los objetivos se establecen los coeficientes de determinación para cada uno de los conjuntos y el global de los datos. Como se observa en la gráfica 5.8 existe un ajuste mejor de los tres conjuntos tanto individual como grupalmente respecto al modelo de Box-Jenkins ARIMA (1,0,0) teniendo un ajuste global de 0.98.

Gráfica 5.8: Red Neuronal Artificial Univariada: Bondad de Ajuste

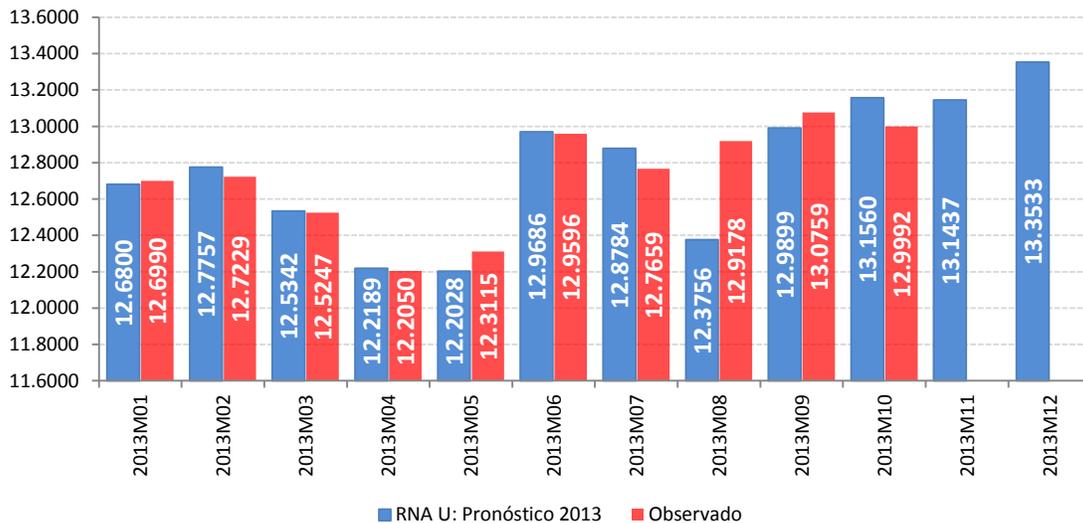


Gráfica 5.8: Continuación



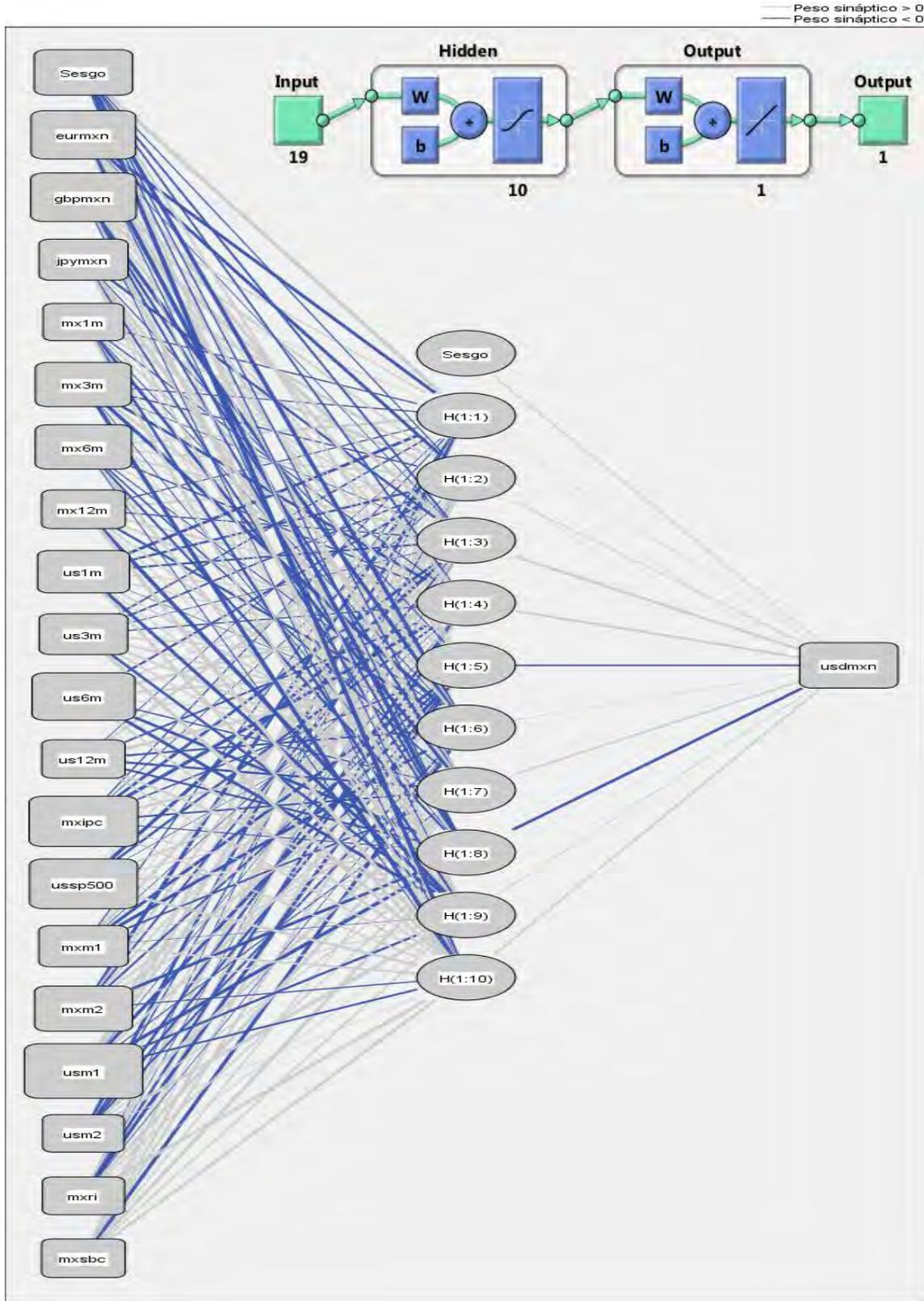
Tomando en cuenta los datos generados para 2013 se puede observar una considerable diferencia respecto al modelo ARIMA (1,0,0) el cual no podía capturar el efecto de la apreciación del peso en los meses de abril y mayo cosa que la red logra. El punto que debe llamar la atención es que de acuerdo a los pronósticos se espera a partir de septiembre un período de depreciación del peso cerrando en Diciembre en \$13.35 lo cual podría generar algunas presiones inflacionarias sobre los productos importados para el fin de año.

Gráfica 5.9 Pronóstico con la RNA Univariada, 2013.01 – 2013.12



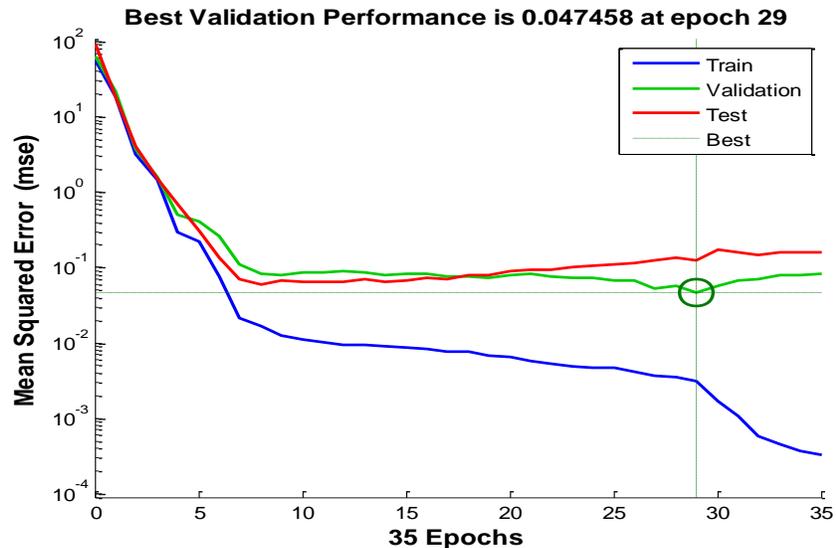
A continuación se realiza la implementación de la red neuronal multivariada con la arquitectura que se determinó en el paso 5 con un vector de información compuesto por 19 variables 1 capa oculta compuesta por 10 neuronas como se presenta en el gráfico 5.10.

Gráfica 5.10: Arquitectura de la Red Neuronal Artificial Multivariada



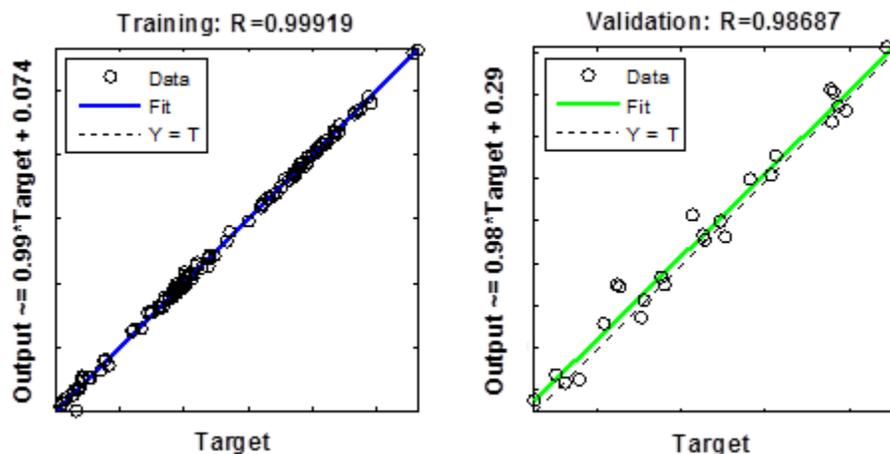
Con la actualización Δx vemos como en la gráfica 5.11 dentro de las primeras 10 iteraciones hay una gran disminución sobre la superficie del error y desde ahí encuentra el mínimo en la iteración 29 para el conjunto de validación, del conjunto de entrenamiento todavía baja hasta el orden de 10^{-3} . Esta red ha logrado minimizar aún más el error que la red neuronal univariada lo que significa que con el vector de información propuesto se ha alcanzado una mejor generalización.

Gráfica 5.11: Red Neuronal Artificial Multivariada: Validación Cruzada



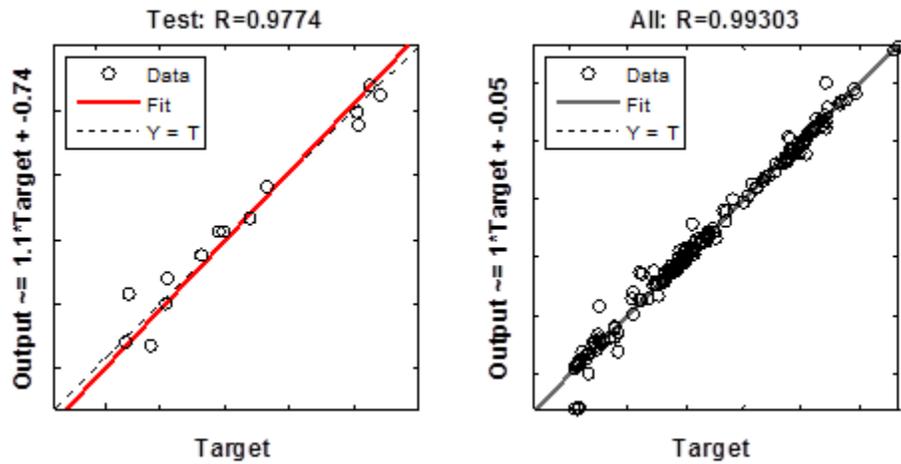
Además de lo anterior podemos comprobar en la gráfica 5.12 la bondad de ajuste de la red neuronal multivariada que dentro de los tres conjuntos alcanza un mejor ajuste respecto a la red univariada y el modelo ARIMA(1,0,0) logrando un ajuste global de 0.99. Si bien la estimación realizada para el conjunto de prueba fue menor que los dos anteriores se tiene el ajuste más alto de los tres modelos.

Gráfica 5.12: Red Neuronal Artificial Multivariada: Bondad de Ajuste



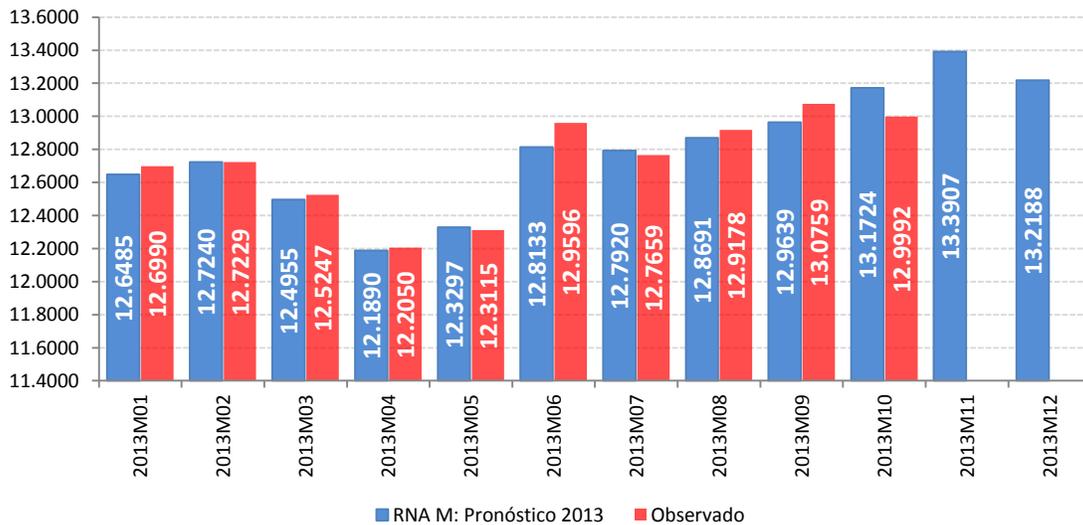
Villar Corona Marco Antonio

Gráfica 5.12: Continuación



Sobre el pronóstico para 2013 podemos observar que tiene un gran ajuste capturando la apreciación del peso que comenzó en Febrero y termino en Abril incluso con un error de \$0.02. Otro punto importante es que la red mostró una tendencia de Julio a Noviembre de depreciación del peso lo que se comprobó con los datos observados de hasta Septiembre con la diferencia de Octubre colocándose en los \$13 pesos cambiando la tendencia anterior. Será importante evaluar la evolución de los primeros días de Noviembre para confirmar el cambio de tendencia.

Gráfica 5.13 Pronóstico con la RNA Multivariada, 2013.01 – 2013.12



Capítulo 6 Análisis Comparativo

“Prediction is very difficult, especially if it’s about the future”

Niels Bohr

Una vez que se generaron todos los pronósticos del tipo de cambio dentro y fuera de la muestra establecemos los siguientes criterios para realizar la evaluación de cada uno de los modelos desarrollados en los dos capítulos anteriores y poder comparar el ajuste que se logra con cada uno de ellos. Para realizar dicha tarea tomamos 6 criterios que se emplean en la mayoría de artículos que giran alrededor de los pronósticos por lo que han demostrado ser de gran utilidad.

El primer criterio del que partimos es la suma de los errores al cuadrado (*sum of squared residuals*, SSR) es la medida más básica de discrepancia entre dos series una observada y la otra estimada a menor nivel de SSR se tiene un mejor ajuste.

$$SSR = \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 \quad (6.1)$$

El segundo criterio se deriva de la suma de los errores al cuadrado, es la media de los errores al cuadrado (*mean squared error*, MSE) por lo que se puede caracterizar como la varianza de los errores. En particular este criterio es la función de desempeño a minimizar más popular, se empleó dentro de las estimaciones de los suavizamientos exponencial, de Holt Winters y de las RNA.

$$MSE = \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n} \quad (6.2)$$

El tercer criterio se obtiene de estimar la raíz cuadrada de MSE (*root mean squared error*, RMSE) y si al anterior lo podemos caracterizar como la varianza de los errores este criterio representa la desviación estándar de los errores por lo tanto está expresada en las mismas unidades que el error (*i.e.* pesos)

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n}} \quad (6.3)$$

El cuarto criterio se obtiene de medir la distancia promedio de discrepancia entre el valor objetivo y el estimado, entonces el error absoluto medio (*mean absolute error*, MAE) no toma en cuenta si se subestima o sobrestima sino la magnitud de los errores.

$$MAE = \frac{\sum_{t=1}^n |Y_t - \hat{Y}_t|}{n} \quad (6.4)$$

El siguiente criterio es la relación en términos absolutos entre el valor observado y el valor estimado dividido por el primero. Se le conoce como el error absoluto porcentual promedio (*mean absolute percentage error*, MAPE) es el error absoluto medio en tanto por uno y al multiplicarlo por 100 nos da el error porcentual. La desventaja que presenta este criterio

respecto a los otros se presenta cuando los valores estimados y observados son cercanos a cero o con numeradores muy pequeños.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \quad (6.5)$$

Por último el error absoluto porcentual promedio simétrico (*symmetric mean absolute percentage error*, sMAPE) se utiliza debido a que busca contemplar de manera simétrica los errores. Se establecen dos versiones del mismo criterio debido a que 6.6a puede presentar problemas al medir de la misma manera la sobre estimación que la sub estimación.

$$sMAPE = \frac{1}{n} \sum_{t=1}^n \frac{|\hat{Y}_t - Y_t|}{(Y_t + \hat{Y}_t)/2} \quad (6.6a)$$

$$sMAPE = \frac{\sum_{t=1}^n |\hat{Y}_t - Y_t|}{\sum_{t=1}^n (Y_t + \hat{Y}_t)} \quad (6.6b)$$

Al estimar los criterios dentro de la muestra se observa que la capacidad de ajuste de los métodos de suavizamiento es bastante adecuada logrando resultados similares a los de la RNA univariada o la del modelo ARIMA (1,0,0). Por mencionar un criterio en particular el error porcentual fue menor al 2% en todos los casos y el error porcentual simétrico incluso fue menor del 1%. De esos hallazgos podemos contrastar la eficiencia de estos métodos tan sencillos de implementar lo cual en principio representa una relación costo-beneficio favorable a los primeros

Cuadro 6.1: Resumen de Criterios de Valuación Dentro de la Muestra 2000.1 – 2012.12

	SES	SED	HWNS	HWA	HWM	ARIMA(1,0,0)	RNA_U_12L	RNA_M
SSR	14.5225	16.8675	13.7991	12.5139	12.4502	12.1165	13.8801	4.6175
MSE	0.0931	0.1081	0.0885	0.0802	0.0798	0.0777	0.0890	0.0296
RMSE	0.3051	0.3288	0.2974	0.2832	0.2825	0.2787	0.2983	0.1720
MAE	0.1977	0.2183	0.1941	0.1912	0.1913	0.1893	0.1656	0.0925
MAPE	0.0173	0.0189	0.0169	0.0166	0.0166	0.0165	0.0146	0.0088
SMAPE	0.0172	0.0189	0.0168	0.0165	0.0165	0.0165	0.0146	0.0088
SMAPE 2	0.0088	0.0097	0.0086	0.0085	0.0085	0.0085	0.0074	0.0041

En cuanto al desempeño fuera de la muestra los resultados positivos que se acaban de mencionar no se replican, lo único consistente es que el suavizamiento exponencial doble realiza el peor ajuste de acuerdo con los 6 criterios y la RNA multivariada se mantiene con los mejores resultados. En promedio se genera un error porcentual de 2.19% de los métodos de suavizamiento comparado con el 0.68% de las RNA's.

Cuadro 6.2: Resumen de Criterios de Valuación Fuera de la Muestra 2013.1 – 2013.10

	SES	SED	HWNS	HWA	HWM	ARIMA(1,0,0)	RNA_U_12L	RNA_M
SSR	1.0066	2.6072	1.2880	0.9265	0.8452	2.0478	0.3539	0.0710
MSE	0.0953	0.2415	0.1216	0.0872	0.0797	0.1907	0.0326	0.0065
RMSE	0.3087	0.4914	0.3487	0.2953	0.2823	0.4367	0.1805	0.0803
MAE	0.2465	0.4163	0.2645	0.2362	0.2301	0.3606	0.1110	0.0621
MAPE	0.0192	0.0338	0.0204	0.0183	0.0179	0.0276	0.0088	0.0048
SMAPE	0.0194	0.0331	0.0208	0.0186	0.0181	0.0281	0.0087	0.0048
SMAPE 2	0.0096	0.0165	0.0103	0.0092	0.0090	0.0141	0.0044	0.0024

Conclusiones

“I hear and I forget. I see and I remember. I do and I understand”

Confucius

En el presente trabajo se buscó implementar una metodología que fuera eficiente al estudiar una variable tan importante y con las características propias del tipo de cambio. Al estudiar a la variable en sí, se encontraron algunos patrones importantes como la estacionalidad que presenta la serie, o el ciclo que se modificó después de Abril de 2006 de acuerdo a la descomposición clásica del tipo de cambio.

De acuerdo con nuestra matriz de congruencia y con los resultados obtenidos podemos afirmar que nuestras hipótesis fueron verdaderas y se consiguió obtener un mejor ajuste dentro de la muestra y un mejor pronóstico del tipo de cambio peso – dólar por parte de las RNA, en particular de la red multivariada respecto con respecto al resto de los modelos. Dicho lo anterior es importante señalar que aunque se generaron mejores resultados con las RNA sería un error extrapolar este resultado a todas las demás series de tiempo, en particular las financieras que se caracterizan por tener tan amplia gama de peculiaridades.

Un resultado para destacar fue la eficiencia de los métodos tradicionales: el suavizamiento exponencial simple o doble y los diferentes tipos de Holt-Winters ya que por un bajo costo de implementación se logra un buen ajuste de las series. Teniendo en cuenta que el factor a considerar su pertinencia puede radicar en generar pronósticos de muy corto plazo lo que significa pocas observaciones fuera de la muestra.

Con respecto a la metodología de Box Jenkins es importante rescatar que el hallazgo de Meese y Rogoff se mantiene en el sentido de que actualmente sigue siendo posible establecer al tipo de cambio como un proceso autorregresivo de primer orden, desde luego no con el mejor ajuste posible. La simpleza y eficacia del modelo son difíciles de lograr con otro tipo de variables económicas y que se mantenga como tal. Su implementación es un poco más complicada que los modelos de suavizamiento pero su relación de costo-beneficio es destacable por lo cual es la primera opción al enfrentar escaso tiempo disponible y la necesidad de generar un pronóstico satisfactorio.

El hecho de que la RNA multivariada haya obtenido el mejor ajuste dentro y fuera de la muestra nos indica la importancia de los fundamentos teóricos y como al combinarlos con herramientas como las RNA se obtienen mejores resultados, siendo de esta forma complementarios y no excluyentes. Esa parte del análisis fundamental que corresponde a preguntarse ¿cuáles son las relaciones que nos pueden explicar el comportamiento del tipo

de cambio en el largo plazo? mezclada con un instrumental especializado en encontrar patrones y por medio de los mismos aprender logran una perfecta combinación.

Si bien en el caso de las RNA presentadas en este trabajo superan a los métodos tradicionales no se debe ver como algo definitivo e invariable que debe sustituir a los anteriores por el contrario, se debe ver como un complemento y por lo tanto una herramienta más para el investigador, el cual en todo momento debe ponderar la relación de costo de implementación *versus* mejora en la estimación.

Contrastando los resultados obtenidos con la lectura del mercado (*i.e.* sus expectativas) tomando como referencia la última encuesta sobre las expectativas de los especialistas en economía del sector privado de Octubre publicada por el Banco de México. De acuerdo a lo expresado por los especialistas se espera que para final de año el tipo de cambio cierre en \$12.79. Esta cifra se ha estado revisando continuamente al alza por lo que no debería de extrañar si en la encuesta del mes de Noviembre se realice lo mismo.

Tomando en cuenta el dato anterior el único pronóstico que se encuentra en esa línea es el del modelo ARIMA (1,0,0) con \$12.6853 lo que va de acuerdo con el pronóstico de comportamiento estable del tipo de cambio de aquí a fin de año. Por otra parte las RNA's en ambos casos muestran una tendencia de depreciación del peso hacia el cierre de año para quedar en niveles mayores a los \$13.20 con una diferencia de \$0.40 respecto a la expectativa del mercado. Lo más conveniente es revisar la evolución del mes de Octubre y observar si se confirma la tendencia de depreciación del peso o en función de los resultados realizar una revisión a las redes.

Como se mencionó en un principio ante la creciente complejidad de las relaciones económicas se fueron desarrollando nuevos modelos que pudieran dar explicación de las mismas lo que generó la implementación de RNA en el campo de las finanzas. Una vez que se ha logrado desarrollar un pronóstico adecuado con el perceptrón multicapa con aprendizaje por corrección de error queda como posibilidad de futuras líneas de investigación la aplicación de las RNA para análisis de carteras de crédito (*credit rating score*) o modelos de riesgo de crisis cambiarias en la línea de Kaminsky, Lizondo y Reinhart (*op. cit.*).

Anexo

Cuadro A.1: Suavizamiento Exponencial Simple

Sample: 2000M01 2012M12
 Included observations: 156
 Method: Single Exponential
 Original Series: FIX
 Forecast Series: FIXSM1

Parameters: Alpha	0.9990
Sum of Squared Residuals	14.52250
Root Mean Squared Error	0.305111

End of Period Levels: Mean	12.87070
----------------------------	----------

Cuadro A.2: Suavizamiento Exponencial Doble

Sample: 2000M01 2012M12
 Included observations: 156
 Method: Double Exponential
 Original Series: FIX
 Forecast Series: FIXSM2

Parameters: Alpha	0.7300
Sum of Squared Residuals	16.86747
Root Mean Squared Error	0.328823

End of Period Levels: Mean	12.88669
Trend	-0.076298

Cuadro A.3: Suavizamiento Holt Winter sin Estacionalidad

Sample: 2000M01 2012M12
 Included observations: 156
 Method: Holt-Winters No Seasonal
 Original Series: FIX
 Forecast Series: FIXSM3

Parameters: Alpha	1.0000
Beta	0.0000
Sum of Squared Residuals	13.79914
Root Mean Squared Error	0.297416

End of Period Levels: Mean	12.87050
Trend	0.019205

Cuadro A.4: Suavizamiento Holt Winter Aditivo

Sample: 2000M01 2012M12

Included observations: 156

Method: Holt-Winters Additive Seasonal

Original Series: FIX

Forecast Series: FIXSM4

Parameters:	Alpha	1.0000
	Beta	0.0000
	Gamma	0.0000
	Sum of Squared Residuals	12.51389
	Root Mean Squared Error	0.283226

End of Period Levels:	Mean	12.79767
	Trend	0.025727
	Seasonals:	
	2012M01	0.119721
	2012M02	0.100918
	2012M03	0.048645
	2012M04	-0.112336
	2012M05	-0.090471
	2012M06	-0.018874
	2012M07	-0.163770
	2012M08	-0.179128
	2012M09	-0.006801
	2012M10	0.107257
	2012M11	0.122007
	2012M12	0.072834

Cuadro A.5: Suavizamiento Holt Winter Multiplicativo

Sample: 2000M01 2012M12

Included observations: 156

Method: Holt-Winters Multiplicative Seasonal

Original Series: FIX

Forecast Series: FIXSM5

Parameters:	Alpha	1.0000
	Beta	0.0000
	Gamma	0.0000
	Sum of Squared Residuals	12.45020
	Root Mean Squared Error	0.282505

End of Period Levels:	Mean	12.77964
	Trend	0.025727
	Seasonals:	
	2012M01	1.010375
	2012M02	1.008173
	2012M03	1.003419
	2012M04	0.990314
	2012M05	0.991769
	2012M06	0.998304
	2012M07	0.985388
	2012M08	0.984232
	2012M09	0.999569
	2012M10	1.010171
	2012M11	1.011177
	2012M12	1.007110

Prueba de Dickey Fuller Aumentada (**ADF**)

Cuadro A.6: Prueba Raíz Unitaria ADF con Nada

Null Hypothesis: FIX has a unit root

Exogenous: None

Lag Length: 1 (Automatic based on SIC, MAXLAG=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	0.497320	0.8218
Test critical values: 1% level	-2.579495	
5% level	-1.942830	
10% level	-1.615384	

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

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(FIX)

Method: Least Squares

Sample (adjusted): 2000M03 2012M12

Included observations: 154 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FIX(-1)	0.000988	0.001987	0.497320	0.6197
D(FIX(-1))	0.310678	0.077083	4.030439	0.0001
R-squared	0.091508	Mean dependent var		0.021816
Adjusted R-squared	0.085758	S.D. dependent var		0.298257
S.E. of regression	0.285181	Akaike info criterion		0.341036
Sum squared resid	12.84985	Schwarz criterion		0.379475
Log likelihood	-25.28285	Durbin-Watson stat		1.883193

Cuadro A.7: Prueba Raíz Unitaria ADF con Intercepto

Null Hypothesis: FIX has a unit root

Exogenous: Constant

Lag Length: 1 (Automatic based on SIC, MAXLAG=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.882944	0.3397
Test critical values: 1% level	-3.471454	
5% level	-2.879494	
10% level	-2.576422	

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

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(FIX)

Method: Least Squares

Sample (adjusted): 2000M03 2012M12

Included observations: 154 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FIX(-1)	-0.030705	0.016307	-1.882944	0.0616
D(FIX(-1))	0.322787	0.076651	4.211139	0.0000
C	0.362902	0.185355	1.957871	0.0520
R-squared	0.113161	Mean dependent var		0.021816
Adjusted R-squared	0.101863	S.D. dependent var		0.298257
S.E. of regression	0.282658	Akaike info criterion		0.329413
Sum squared resid	12.54359	Schwarz criterion		0.387073
Log likelihood	-23.35306	F-statistic		10.01660
Durbin-Watson stat	1.887961	Prob(F-statistic)		0.000081

Cuadro A.8: Prueba Raíz Unitaria ADF con Intercepto y Tendencia

Null Hypothesis: FIX has a unit root

Exogenous: Constant, Linear Trend

Lag Length: 1 (Automatic based on SIC, MAXLAG=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.826483	0.0176
Test critical values: 1% level	-4.016433	
5% level	-3.438154	
10% level	-3.143345	

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

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(FIX)

Method: Least Squares

Sample (adjusted): 2000M03 2012M12

Included observations: 154 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FIX(-1)	-0.118345	0.030928	-3.826483	0.0002
D(FIX(-1))	0.373939	0.075950	4.923525	0.0000
C	1.104814	0.287999	3.836167	0.0002
@TREND(2000M01)	0.003027	0.000918	3.297552	0.0012
R-squared	0.170949	Mean dependent var		0.021816
Adjusted R-squared	0.155006	S.D. dependent var		0.298257
S.E. of regression	0.274168	Akaike info criterion		0.274531
Sum squared resid	11.72623	Schwarz criterion		0.351410
Log likelihood	-17.96249	F-statistic		10.72232
Durbin-Watson stat	1.937490	Prob(F-statistic)		0.000002

Prueba de Dickey Fuller por Mínimos Cuadrados Generalizados (**DF-GLS**)

Cuadro A.9: Prueba Raíz Unitaria DF-GLS con Intercepto

Null Hypothesis: FIX has a unit root

Exogenous: Constant

Lag Length: 1 (Automatic based on SIC, MAXLAG=13)

	t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic	-0.874590
Test critical values: 1% level	-2.579495
5% level	-1.942830
10% level	-1.615384

*MacKinnon (1996)

DF-GLS Test Equation on GLS Detrended Residuals

Dependent Variable: D(GLSRESID)

Method: Least Squares

Sample (adjusted): 2000M03 2012M12

Included observations: 154 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GLSRESID(-1)	-0.010532	0.012042	-0.874590	0.3831
D(GLSRESID(-1))	0.320419	0.077198	4.150630	0.0001
R-squared	0.094470	Mean dependent var		0.021816
Adjusted R-squared	0.088738	S.D. dependent var		0.298257
S.E. of regression	0.284716	Akaike info criterion		0.337770
Sum squared resid	12.80796	Schwarz criterion		0.376210
Log likelihood	-25.02162	Durbin-Watson stat		1.883478

Cuadro A.10: Prueba Raíz Unitaria DF-GLS con Intercepto y Tendencia

Null Hypothesis: FIX has a unit root

Exogenous: Constant, Linear Trend

Lag Length: 1 (Automatic based on SIC, MAXLAG=13)

	t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic	-3.801897
Test critical values:	
1% level	-3.508000
5% level	-2.970000
10% level	-2.680000

*Elliott-Rothenberg-Stock (1996, Table 1)

DF-GLS Test Equation on GLS Detrended Residuals

Dependent Variable: D(GLSRESID)

Method: Least Squares

Sample (adjusted): 2000M03 2012M12

Included observations: 154 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GLSRESID(-1)	-0.115672	0.030425	-3.801897	0.0002
D(GLSRESID(-1))	0.371274	0.075395	4.924380	0.0000
R-squared	0.169142	Mean dependent var		-0.001438
Adjusted R-squared	0.163883	S.D. dependent var		0.298257
S.E. of regression	0.272724	Akaike info criterion		0.251709
Sum squared resid	11.75179	Schwarz criterion		0.290148
Log likelihood	-18.13670	Durbin-Watson stat		1.933597

Prueba de Phillips Perron (PP)

Cuadro A.II: Prueba Raíz Unitaria PP con Nada

Null Hypothesis: FIX has a unit root

Exogenous: None

Bandwidth: 1 (Newey-West using Bartlett kernel)

	Adj. t-Stat	Prob.*
Phillips-Perron test statistic	0.602708	0.8456
Test critical values: 1% level	-2.579404	
5% level	-1.942818	
10% level	-1.615392	

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

Residual variance (no correction)	0.088033
HAC corrected variance (Bartlett kernel)	0.114486

Phillips-Perron Test Equation

Dependent Variable: D(FIX)

Method: Least Squares

Sample (adjusted): 2000M02 2012M12

Included observations: 161 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FIX(-1)	0.001522	0.002064	0.737101	0.4621
R-squared	-0.001844	Mean dependent var		0.021471
Adjusted R-squared	-0.001844	S.D. dependent var		0.297355
S.E. of regression	0.297629	Akaike info criterion		0.420256
Sum squared resid	14.17331	Schwarz criterion		0.439395
Log likelihood	-32.83058	Durbin-Watson stat		1.367621

Cuadro A.12: Prueba Raíz Unitaria PP con Intercepto

Null Hypothesis: FIX has a unit root

Exogenous: Constant

Bandwidth: 2 (Newey-West using Bartlett kernel)

	Adj. t-Stat	Prob.*
Phillips-Perron test statistic	-1.655102	0.4521
Test critical values: 1% level	-3.471192	
5% level	-2.879380	
10% level	-2.576361	

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

Residual variance (no correction)	0.086772
HAC corrected variance (Bartlett kernel)	0.121106

Phillips-Perron Test Equation

Dependent Variable: D(FIX)

Method: Least Squares

Sample (adjusted): 2000M02 2012M12

Included observations: 161 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FIX(-1)	-0.024042	0.016943	-1.419025	0.1578
C	0.292653	0.192527	1.520061	0.1305

R-squared	0.012506	Mean dependent var	0.021471
Adjusted R-squared	0.006295	S.D. dependent var	0.297355
S.E. of regression	0.296418	Akaike info criterion	0.418251
Sum squared resid	13.97029	Schwarz criterion	0.456529
Log likelihood	-31.66917	F-statistic	2.013631
Durbin-Watson stat	1.353499	Prob(F-statistic)	0.157849

Villar Corona Marco Antonio

Cuadro A.13: Prueba Raíz Unitaria PP con Intercepto y Tendencia

Null Hypothesis: FIX has a unit root

Exogenous: Constant, Linear Trend

Bandwidth: 3 (Newey-West using Bartlett kernel)

	Adj. t-Stat	Prob.*
Phillips-Perron test statistic	-3.195559	0.0890
Test critical values: 1% level	-4.016064	
5% level	-3.437977	
10% level	-3.143241	

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

Residual variance (no correction)	0.084161
HAC corrected variance (Bartlett kernel)	0.125656

Phillips-Perron Test Equation

Dependent Variable: D(FIX)

Method: Least Squares

Sample (adjusted): 2000M02 2012M12

Included observations: 161 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FIX(-1)	-0.084986	0.032217	-2.637945	0.0092
C	0.808667	0.300835	2.688075	0.0080
@TREND(2000M01)	0.002116	0.000956	2.213963	0.0283

R-squared	0.042219	Mean dependent var	0.021471
Adjusted R-squared	0.030095	S.D. dependent var	0.297355
S.E. of regression	0.292847	Akaike info criterion	0.400121
Sum squared resid	13.54993	Schwarz criterion	0.457539
Log likelihood	-29.20978	F-statistic	3.482338
Durbin-Watson stat	1.317034	Prob(F-statistic)	0.033115

Prueba de Kwiatkowski Phillips Schmidt Shin (KPSS)

Cuadro A.14: Prueba Raíz Unitaria KPSS con Intercepto

Null Hypothesis: FIX is stationary

Exogenous: Constant

Bandwidth: 10 (Newey-West using Bartlett kernel)

	LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic	1.273358
Asymptotic critical values*:	
1% level	0.739000
5% level	0.463000
10% level	0.347000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	1.906279
HAC corrected variance (Bartlett kernel)	18.03711

KPSS Test Equation

Dependent Variable: FIX

Method: Least Squares

Sample (adjusted): 2000M01 2012M12

Included observations: 162 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.28975	0.108813	103.7538	0.0000

R-squared	0.000000	Mean dependent var	11.28975
Adjusted R-squared	0.000000	S.D. dependent var	1.384962
S.E. of regression	1.384962	Akaike info criterion	3.495376
Sum squared resid	308.8171	Schwarz criterion	3.514435
Log likelihood	-282.1254	Durbin-Watson stat	0.046051

Cuadro A.15: Prueba Raíz Unitaria KPSS con Intercepto y Tendencia

Null Hypothesis: FIX is stationary

Exogenous: Constant, Linear Trend

Bandwidth: 9 (Newey-West using Bartlett kernel)

	LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic	0.056482
Asymptotic critical values*:	
1% level	0.216000
5% level	0.146000
10% level	0.119000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	0.510985
HAC corrected variance (Bartlett kernel)	3.175155

KPSS Test Equation

Dependent Variable: FIX

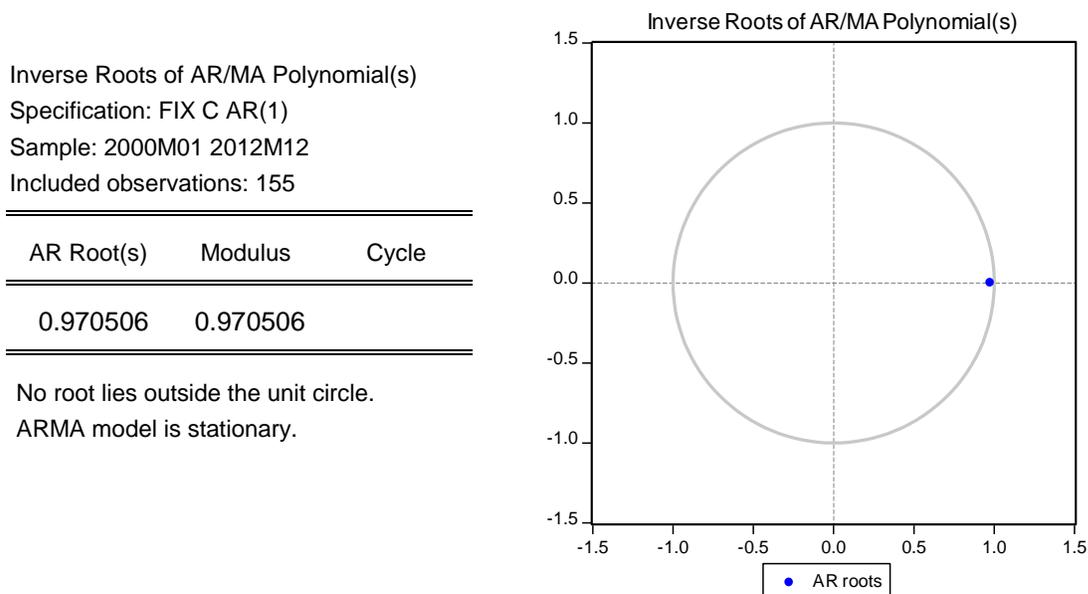
Method: Least Squares

Sample (adjusted): 2000M01 2012M12

Included observations: 162 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.256397	0.112504	82.27639	0.0000
@TREND(2000M01)	0.025259	0.001208	20.90203	0.0000
R-squared	0.731946	Mean dependent var		11.28975
Adjusted R-squared	0.730271	S.D. dependent var		1.384962
S.E. of regression	0.719286	Akaike info criterion		2.191154
Sum squared resid	82.77965	Schwarz criterion		2.229273
Log likelihood	-175.4835	F-statistic		436.8948
Durbin-Watson stat	0.170930	Prob(F-statistic)		0.000000

Cuadro A.16: Raíces inversas del Proceso



Cuadro A.17: Prueba Breusch-Godfrey para Autocorrelación con 1 rezago

F-statistic	3.931042	Probability	0.049205	
Obs*R-squared	3.907570	Probability	0.048069	
Test Equation:				
Dependent Variable: RESID				
Method: Least Squares				
Presample missing value lagged residuals set to zero.				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.245672	1.050398	-0.233885	0.8154
AR(1)	-0.009474	0.019858	-0.477077	0.6340
RESID(-1)	0.163583	0.082506	1.982685	0.0492
R-squared	0.025210	Mean dependent var	1.66E-17	
Adjusted R-squared	0.012384	S.D. dependent var	0.338230	
S.E. of regression	0.336129	Akaike info criterion	0.676520	
Sum squared resid	17.17334	Schwarz criterion	0.735425	
Log likelihood	-49.43027	F-statistic	1.965521	
Durbin-Watson stat	1.994723	Prob(F-statistic)	0.143627	

Cuadro A.18: Prueba Breusch-Godfrey para Autocorrelación con 2 rezagos

F-statistic	1.958088	Probability	0.144695	
Obs*R-squared	3.918295	Probability	0.140979	
Test Equation:				
Dependent Variable: RESID				
Method: Least Squares				
Presample missing value lagged residuals set to zero.				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.232170	1.061872	-0.218642	0.8272
AR(1)	-0.008955	0.020544	-0.435877	0.6635
RESID(-1)	0.164402	0.083153	1.977109	0.0499
RESID(-2)	-0.008681	0.083852	-0.103533	0.9177
R-squared	0.025279	Mean dependent var	1.66E-17	
Adjusted R-squared	0.005914	S.D. dependent var	0.338230	
S.E. of regression	0.337228	Akaike info criterion	0.689352	
Sum squared resid	17.17212	Schwarz criterion	0.767892	
Log likelihood	-49.42477	F-statistic	1.305392	
Durbin-Watson stat	1.996191	Prob(F-statistic)	0.274831	

Cuadro A.19: Prueba Breusch-Godfrey para Autocorrelación con 3 rezagos

F-statistic	1.558959	Probability	0.201772	
Obs*R-squared	4.686648	Probability	0.196234	
Test Equation:				
Dependent Variable: RESID				
Method: Least Squares				
Presample missing value lagged residuals set to zero.				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.343987	1.070337	-0.321382	0.7484
AR(1)	-0.013378	0.021172	-0.631896	0.5284
RESID(-1)	0.169444	0.083416	2.031312	0.0440
RESID(-2)	-0.015818	0.084311	-0.187610	0.8514
RESID(-3)	0.073478	0.083913	0.875644	0.3826
R-squared	0.030236	Mean dependent var	1.66E-17	
Adjusted R-squared	0.004376	S.D. dependent var	0.338230	
S.E. of regression	0.337489	Akaike info criterion	0.697156	
Sum squared resid	17.08479	Schwarz criterion	0.795331	

Villar Corona Marco Antonio

Log likelihood	-49.02963	F-statistic	1.169220
Durbin-Watson stat	1.980148	Prob(F-statistic)	0.326728

Cuadro A.20: Prueba ARCH con 1 rezago

F-statistic	0.927896	Probability	0.336940
Obs*R-squared	0.934401	Probability	0.333722

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Sample (adjusted): 2000M03 2012M12

Included observations: 154 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.105196	0.032102	3.276896	0.0013
RESID^2(-1)	0.077909	0.080879	0.963273	0.3369

R-squared	0.006068	Mean dependent var	0.114109
Adjusted R-squared	-0.000471	S.D. dependent var	0.381384
S.E. of regression	0.381473	Akaike info criterion	0.923351
Sum squared resid	22.11934	Schwarz criterion	0.962792
Log likelihood	-69.09803	F-statistic	0.927896
Durbin-Watson stat	2.005906	Prob(F-statistic)	0.336940

Cuadro A.21: Prueba ARCH con 2 rezagos

F-statistic	0.575177	Probability	0.563841
Obs*R-squared	1.164432	Probability	0.558659

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Sample (adjusted): 2000M04 2012M12

Included observations: 153 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.101384	0.033536	3.023163	0.0029
RESID^2(-1)	0.074641	0.081599	0.914731	0.3618
RESID^2(-2)	0.039794	0.081614	0.487586	0.6266

R-squared	0.007611	Mean dependent var	0.114539
-----------	----------	--------------------	----------

Villar Corona Marco Antonio

Adjusted R-squared	-0.005621	S.D. dependent var	0.382599
S.E. of regression	0.383672	Akaike info criterion	0.941358
Sum squared resid	22.08068	Schwarz criterion	1.000778
Log likelihood	-69.01387	F-statistic	0.575177
Durbin-Watson stat	2.005332	Prob(F-statistic)	0.563841

Cuadro A.22: Prueba ARCH con 3 rezagos

F-statistic	0.632478	Probability	0.595182
Obs*R-squared	1.924049	Probability	0.588319

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Sample (adjusted): 2000M05 2012M12

Included observations: 152 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.094733	0.034812	2.721231	0.0073
RESID^2(-1)	0.071533	0.081989	0.872474	0.3844
RESID^2(-2)	0.034191	0.082165	0.416129	0.6779
RESID^2(-3)	0.071809	0.082006	0.875660	0.3826

R-squared	0.012658	Mean dependent var	0.115237
Adjusted R-squared	-0.007355	S.D. dependent var	0.383766
S.E. of regression	0.385174	Akaike info criterion	0.955723
Sum squared resid	21.95719	Schwarz criterion	1.035298
Log likelihood	-68.63491	F-statistic	0.632478
Durbin-Watson stat	2.021898	Prob(F-statistic)	0.595182

Cuadro A.23: Prueba de Estabilidad: Pronóstico de Chow de 2012.01 a 2012.12

F-statistic	1.450605	Probability	0.117388	
Log likelihood ratio	27.35355	Probability	0.072599	
Test Equation:				
Dependent Variable: FIX				
Method: Least Squares				
Sample: 2000M02 2011M12				
Included observations: 143				
Convergence achieved after 4 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	13.05322	2.952537	4.421020	0.0000
AR(1)	0.985003	0.018620	52.90031	0.0000
R-squared	0.952032	Mean dependent var	11.09093	
Adjusted R-squared	0.951692	S.D. dependent var	1.315494	
S.E. of regression	0.289135	Akaike info criterion	0.370040	
Sum squared resid	11.78745	Schwarz criterion	0.411478	
Log likelihood	-24.45785	F-statistic	2798.443	
Durbin-Watson stat	1.359774	Prob(F-statistic)	0.000000	
Inverted AR Roots	.99			

Bibliografía

1. Abolfazl, A., M. Bagher y A. Hoseinabadi. (2005). “Modified Levenberg-Marquardt Method for Neural Networks Training”. *Proceedings of World Academy of Science, Engineering and Technology*, Vol. 6. pp. 46-48.
2. Alquicira, S. (2001). *Redes Neuronales Artificiales aplicadas a la detección de enfermedades cardiovasculares (rotula ventricular aguda y angina de pecho)*. México: UNAM, Facultad de Contaduría y Administración. Tesis de Licenciatura en Informática.
3. Altman, E., G. Marco y F. Varetto. (1994). “Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (The Italian Experience)”. *Journal of Banking and Finance*, Vol. 18, Núm. 3. pp. 505-529.
4. Ayat, L. y P. Burridge (2000). “Unit Root Test in the Presence of Uncertainty about non-stochastic Trends”. *Journal of Econometrics* Vol. 95, No. 1, pp. 71-96.
5. Bayoumi, T., P. Clark, S. Symansky y M. Taylor (1994). “Robustness of Equilibrium Exchange Rate Calculations to Alternative Assumptions and Methodologies”. *IMF Working Paper* 17.
6. Black, S. (1973). “International money markets and flexible exchange rates”. *Studies in International Finance*, No. 25. Princeton University.
7. Box, G. y G. Jenkins (1970). *Time Series Analysis, forecasting and Control*. Holden Day.
8. Branson, W. (1977) “Asset markets and relative prices in exchange rate determination”. *Sozialwissenschaftliche Annalen*. No. 1 pp.69-89.
9. Bribiesca, G. (2006). *Redes Neuronales Artificiales para la toma de decisiones: El Caso de tutores del Programa de Posgrado en Ciencias de la Administración*. México: UNAM, Programa de Posgrado en Ciencias de la Administración. Tesis de Doctorado en Ciencias de la Administración.
10. Brown, R. (1956). “Exponential Smoothing for Predicting Demand”. Presentado en 10th *National Meeting for the Operations Research Society of America (ORSA)*, 16 de Noviembre de 1956.
11. Bubula A. y I. Ökter-Robe (2002). “The Evolution of Exchange Rate Regimes since the 1990: Evidence from De Facto Policies”. *IMF Working Paper*
12. Calvo, G., A. Izquierdo, y R. Loo-Kung (2012). “Optimal Holdings of International Reserves: Self –Insurance Against Sudden Stop”. *NBER Working Paper* 18219.
13. Cassel, G. (1918). “Abnormal Deviations in International Exchanges”. *Economic Journal*, 28, pp. 413-415
14. Chatfield, C. (2000). *Time-Series Forecasting*. Chapman & Hall, New York.
15. _____ y M. Yar (1988). “Holt-Winters Forecasting: Some Practical Issues”. *The Statistician*, Vol. 37, No. 2, pp. 129-140
16. Chen, C., C. Shen y C.Wang (2007). “Does PPP Hold for Big Mac Price or Consumer Index Price? Evidence from Panel Cointegration”. *Economics Bulletin*, Vol. 6, No.16, pp. 1-15.
17. Chiang, W., T. Urban y G. Baldrige. (1996). “A Neural Network Approach to Mutual Fund Net Asset Value Forecasting”. *International Journal of Managment Science*, Vol. 24, Núm. 2 pp. 205-215.
18. Clements, K. y S. Shea (2010). “The Big Mac Index Two Decades on Evaluation of Burgernomics” *The University of West Australia Business School Discussion Paper* 10.14

19. Cumby, R (1996). "Forecasting Exchange Rates and Relative Prices with the Hamburger Standard: Is What you Want What you Get with McParity". *NBER Working Papers* W5675
20. Dornbusch, R. (1976). "Expectations and Exchange Rate Dynamics". *Journal of Political Economy*, 84, pp. 11161-1176.
21. _____ y P. Krugman (1976). "Flexible Exchange Rates in the Short Run". *Brooking Papers on Economic Activity*. Vol 7. No 3 pp.537-584.
22. Dunn, R. (2007). "Does the Big Mac Predicts Exchange Rates?". *Challenge*, Vol. 50, pp. 113-122
23. Dickey, D. y W. Fuller. (1979). "Distribution on the Estimators for Autoregressive Time Series with a Unit Root". *Journal of American Statistical Association*. Vol. 74, pp. 427-431
24. _____ (1981). "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root". *Econometrica*. Vol. 49 No. 4, pp. 1057-1072
25. Edwards, S., D. Cavallo, A. Fraga y J. Frenkel (2003). "Exchange Rate Regimes". En *Economic and Financial Crises in Emerging Market Economies* de Feldstein, M. University of Chicago Press.
26. Elder, J. y P. Kennedy (2001) "Testing for Unit roots: What Should be Taught?". *Journal of Economic Education*. Vol. pp. 137-146.
27. Elliott, G., T. Rothenberg y J. Stock (1996). "Efficient Tests for an Autoregressive Unit Root". *Econometrica*. Vol. 64, pp.43-67.
28. El Shazly, M. y H, El Shazly. (1999). "Forecasting Currency Prices Using a Genetically Evolved Neural Network Architecture". *International Review of Financial Analysis*. Vol. 8 Núm. 1 Págs 67-82.
29. Faraway, J. y C. Chatfield. (1998). "Time Series Forecasting with Neural Networks: A Comparative Study Using the Airline Data". *Journal of the Royal Statistical Society. Series C (Applied Statistics)*. 47, pp. 231-250.
30. Fischer, S. (2001). "Exchange Rate Regimes: Is the Bipolar View Correct?". *The Journal of Economic Perspectives*, Vol. 15, Núm.2 pp. 3-24,
31. Frankel, J. (1979). "On the Mark: A Theory of Floating Exchange Rate Based on Real Interest Differentials". *American Economic Review*, 69, pp. 610-622
32. Franses, P. y P. Van Homelen (1998). "On Forecasting Exchange Rates Using Neural Networks". *Applied Financial Economics*. Núm. 8 pp. 589-596.
33. Frenkel, J. (1976). "A Monetary Approach to the Exchange Rate: Doctrinal Aspects and empirical evidence". *Scandinavian Journal of Economics*, 78, pp. 169-191
34. _____ (1985) "Comment". En "Cause of Appreciation and Volatility of the Dollar" de W. Branson. *NBER Working Paper Series*. Núm. w17771. National Bureau of Economic Research.
35. Funahashi, K. (1989): "On the approximate realization of continuous mapping by neural networks". *Neural Networks*, 2. pp. 183-192
36. García, M. (1996). *Redes Neuronales Artificiales: Aplicación en el diagnóstico diferencial de la demencia vascular y tipo Alzheimer*. México: UNAM, Facultad de Ingeniería División de Estudios de Posgrado. Tesis de Maestría en Ingeniería Eléctrica.
37. García, J. (2009). *Redes Neuronales Artificiales para el Pronóstico del PIB: México, 1980.1-2008.3*. México: UNAM, Facultad de Economía División de Estudios de Posgrado. Tesis de Maestría en Economía.
38. Greenspan, A. (2007). *Age of Turbulence: Adventures in a New World*. Penguin Group. New York
39. Guerrero, V. (2003). *Análisis Estadístico de Series de Tiempo Económicas*. Thompson, México.

40. Gujarati, D. y D. Porter (2010). *Econometría*. McGraw Hill, México.
41. Habermeier, K., A.Kokenye, R. Veyrune y H. Anderson (2009) “Revised System for the Classification of Exchange Rate Arrangements” *IMF Working Paper* 211.
42. Haykin, S. (1999). *Neural Networks. A Comprehensive Foundation*. Prentice Hall. Segunda Edición.
43. Hebb, D. (1949). *The Organization of Behavior*. John Wiley & Sons.
44. Hill, T., M. O’Connor y W. Remus. (1996). “Neural Network Models for Time Series Forecasts”. *Management Science*, Vol. 42, Núm. 7. pp. 1082-1092.
45. Holt, C. (1957). “Forecasting Seasonal and Trends by Exponentially Weighted Moving Averages”. *Office of Naval Research*, No. 52
46. Hooper, P. y Morton, J. (1982). “Fluctuations in the Dollar. A Model of Nominal and Real Exchange Rate Determination”, *Journal of International Money and Finance*, 1, pp. 39-56.
47. Hopfield, J. (1982). “Neural Networks and Physical Systems with Emergent Collective Computational Abilities”. *Proceedings of the National Academy of Sciences*, Vol. 79. pp. 2554-2558.
48. Jalil, M. y M. Misas. (2007). “Evaluación de Pronósticos del Tipo de Cambio Utilizando Redes Neuronales y Funciones de Pérdida Asimétricas”. *Revista Colombiana de Estadística*, Vol. 30, Núm. 001, pp. 143-161
49. Jamal, A. y C. Sundar (1997). “Modeling Exchange Rates with Neural Networks”. *Journal of Applied Business Research*. Vol. 14 Núm. 1 pp. 1-5
50. Jasic, T. y D. Wood. (2003). “Neural Network Protocols and Model Performance”. *Neurocomputing*. Vol. 55 pp. 747-753.
51. Kaashock, J. y H. Can Dijk. (2002). “Neural Network Pruning Applied to Real Exchange Rate Analysis”. *Journal of Forecasting*. Núm. 21 pp. 559-577.
52. Kaminsky, G, S. Lizondo y C. Reinhart (1998). “Leading Indicators of Currency Crises”. *IMF Staff Papers* Vol. 5 No. 1 pp. 1-48
53. Kirchgässner, G. y J. Wolters (2007). *Introduction to Modern Time Series Analysis*. Springer, New York.
54. Kouri, P. (1976). “The exchange rate and the balance of payments in the short run and in the long run: A monetary approach”. *Scandinavian Journal of Economics* 78, pp. 280-308
55. Kwiatkowski, D., P. Phillips, P. Schmidt y Y. Shin (1992). “Testing the Null Hypothesis of Stationary against the Alternative Unit Root”, *Journal of Econometrics*, Vol. 54 pp. 159-178.
56. Leung, M., A. Chen y H. Daouk. (2000). “Forecasting Exchange Rates Using General Regression Neural Networks”. *Computers & Operations Research*. Núm. 27. pp. 1093-1110
57. Makridakis, S. y S. Wheelwright (1983). *Forecasting: Methods and Applications*. Wiley, New York.
58. Mathä, T. (2009) “Regional McParity: Do Common Pricing Points Reduce Deviations from the Law of One Price?”. *Review of World Economics*, Vol. 145, No.1, pp. 155-166
59. Markowitz, H. (1952), “Portfolio Selection” *Journal of Finance*, Vol. 7 No. 1, pp.77-91
60. McCulloch, W. y W. Pitts. (1943). “A Logical Calculus of Ideas Immanent in Nervous Activity”. *Bulletin of Mathematical Biophysics*. Vol. 5. pp. 115-133
61. Meese, R. y K. Rogoff (1983 a). “Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample”. *Journal of International Economics*, Vol. 14 pp. 3-24

62. _____ (1983 b). "The Out of Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification?". En *Exchange Rates and International Economics* de Frenkel J. University of Chicago Press.
63. Melchor, A. (2010). *Uso de las Redes Neuronales Artificiales para el Pronóstico de la Inflación*. México: UNAM, Facultad de Ciencias. Tesis de Actuaría.
64. Mussa, M. (1976). "The Exchange Rate, the Balance of Payments and Monetary and Fiscal Policy under a Regime of Controlled Floating". *Scandinavian Journal of Economics*, 78, pp. 229-248
65. Nelson, C. y C. Plosser. (1982) Trends and Random Walks in Macroeconomic Time Series. Some Evidence and Implications. *Journal of Monetary Economics*. Núm. 10, pp. 139-162.
66. Obstfeld, M., J. Shambaugh, y A. Taylor (2008). "Financial Stability, the Trilemma, and International Reserves" *NBER Working Paper 14217*.
67. Pakko, M. y Pollard, P. (1996). "For Here or To Go? Purchasing Power Parity and the Big Mac". *Federal Reserve Bank of St. Louis Review*, pp. 3-21.
68. _____ (2003), "Burgernomics: A Big Mac Guide to Purchasing Power Parity". *Federal Reserve Bank of St. Louis Review*, pp. 9-27
69. Philip, A., A. Taofiki y A. Bidemi. (2011) "Artificial Neural Network Model for Forecasting Foreign Exchange Rate". *World of Computer Science and Information Technology Journal*. Vol. 1 Núm. 3 pp. 110-118.
70. Phillips, P. y P. Perron (1988). "Testing for Unit Root in Time Series Regression". *Biometrika*. Vol. 75, pp. 335-346.
71. Poddig, T. y H. Rehkugler (1996). "A "World" Model of Integrated Financial Markets Using Artificial Neural Networks". *Neurocomputing*. Núm 10 pp. 251-273
72. Rodríguez, C. (2000). *Análisis de Series Temporales*. La Muralla, Salamanca.
73. Rosenblatt, R. (1958). "The Perceptron. A Probabilistic Model for Information Storage and Organization in the Brain". *Psychological Review*. Vol. 65, Núm. 6 pp. 386-408.
74. Rumelhart, D. E. Hinton y P. Williams (1986). "Learning Representations by Backpropagation Errors". *Nature*, Vol. 323 pp. 533-536.
75. Shamah, S. (2004). *A Currency Option Primer*. Wiley, New York
76. Tobin J. (1958). "Liquidity Preference as Behavior Towards Risk". *The Review of Economic Studies*, 25, pp. 65-86.
77. Torres, I. (2007). *Pronósticos en el mercado de derivados utilizando redes neuronales y modelos ARIMA: Una aplicación al Cete de 91 días en el MexDer*. México: UNAM, Programa de Posgrado en Ciencias de la Administración. Tesis de Maestría en Finanzas.
78. Vojinovic, Z., V. Kecman y R. Seidel. (2001). "A Data Mining Approach to Financial Time Series Modeling and Forecasting". *International Journal of Intelligent Systems in Accounting, Finance & Management*. Núm. 10 pp. 225-239.
79. Weithers, T. (2006). *Foreign Exchange. A Practical Guide to the FX Markets*. Wiley, New York
80. Werbos, P. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavior Sciences*. PhD thesis. Harvard University.
81. White, H. (1988). "Economic Prediction Using Neural Networks: The Case of IBM Daily Stock Returns". *Proceedings of the IEEF International Conference on Neural Networks*. Vol. 2, pp. 451-458

82. Widrow, B. y M. Hoff (1960). "Adaptative Switching Circuits". *IRE WESCON Convention Record*. Núm. 4 pp. 96-104.
83. Winters, P. (1960). "Forecasting Sales by Exponentially Weighted Moving Averages". *Management Sciences*. No. 6 pp. 324-345.
84. Wu, B. (1995). "Model-free Forecasting for Nonlinear Time Series (With Application to Exchange Rates)". *Computational, Statistics and Data Analysis*. Núm. 19 pp. 433-459
85. Yu, L., S. Wang, W. Huang y K. Lai (2007). "Are Foreign Exchange Rates Predictable? A Survey From Artificial Neural Networks Perspective?" *Scientific Inquiry*, Vol. 8 Núm. 2 pp. 207-228.
86. Zapata, L. y H. Díaz (2008) "Predicción del Tipo de Cambio Peso-Dólar Utilizando Redes Neuronales Artificiales (RNA)". *Pensamiento & Gestión*. Núm. 24. pp. 29-42.