



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ECONOMIA

**REDES NEURONALES ARTIFICIALES PARA EL
PRONÓSTICO DEL PIB; MÉXICO, 1980.1-2008.3**

T E S I S

QUE PARA OBTENER EL GRADO DE

MAESTRO EN ECONOMIA

PRESENTA:

JESÚS MANUEL GARCÍA RAMOS

DIRECTOR DE TESIS: EDUARDO LORÍA



MÉXICO, D.F., 2009



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.

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ECONOMÍA

**REDES NEURONALES ARTIFICIALES PARA EL PRONÓSTICO DEL PIB:
MÉXICO, 1980.1-2008.3¹**

**TESIS QUE PARA OBTENER
EL GRADO DE MAESTRO EN ECONOMÍA
PRESENTA:**

JESÚS MANUEL GARCÍA RAMOS

DIRECTOR DE TESIS: EDUARDO LORÍA

MÉXICO, D. F., 2009

¹ Esta tesis se llevó a cabo gracias a la *Beca para Estudios de Posgrado* que otorga el Consejo Nacional de Ciencia y Tecnología (CONACyT) durante los semestres 2008-I-2009-II.

RESUMEN

En el presente trabajo se modela y pronostica el Producto Interno Bruto, 1980.1-2008.3, a través del *Perceptrón Multicapa* (Rumelhart *et al.*, 1986), la arquitectura Red Neuronal Artificial más utilizada en los últimos años para la modelación económica univariada y multivariada, debido a su gran capacidad para la resolución de problemas tanto lineales como no lineales. La orientación es teórica y empírica, de tal forma que resulte una guía básica, pudiendo ser entendida por quien cuente con una formación matemática y económica básica.

Palabras clave: Redes Neuronales Artificiales, Perceptrón Multicapa, *Backpropagation*, Modelos Univariados, Modelos Multivariados, Series de Tiempo, Econometría Aplicada, Pronóstico Económico, Producto Interno Bruto, México

JEL: C45, C5, C22, E27

ÍNDICE

INTRODUCCIÓN

Más allá del Análisis de Regresión Lineal.....	i
Redes Neuronales Artificiales: un ingrediente extra.....	iii
Plan de capítulos.....	vi

I. REDES NEURONALES ARTIFICIALES

1.1. Breve historia de las Redes Neuronales Artificiales.....	7
1.2. La Neurona Artificial.....	10
1.2.1. La función de activación-salida.....	13
1.2.2. La regla de aprendizaje.....	15
1.3. El Perceptrón Multicapa.....	16
1.3.1. El Algoritmo Backpropagation.....	20
1.4. Estadísticos de bondad de ajuste y correcta especificación.....	23
1.4.1. Número de neuronas en la capa oculta.....	23
1.4.2. Validación cruzada.....	23
1.4.3. Criterios dentro de la muestra.....	25
1.4.4. Criterios fuera de la muestra.....	26

II. REDES NEURONALES ARTIFICIALES PARA EL PRONÓSTICO DEL PIB: MÉXICO, 1980.1-2008.3

2.1. El vector de información.....	28
2.2. Dos modelos de Redes Neuronales Artificiales.....	36
2.2.1. RNA Univariada.....	36
2.2.2. RNA Multivariada.....	41

III. RESULTADOS

3.1. Evaluación y comparación de modelos.....	46
3.1.1. Pronósticos dentro de la muestra.....	46
3.1.2. Pronósticos fuera de la muestra.....	47
3.2. Pronósticos del PIB, 2008.1-2008.3.....	48

CONCLUSIONES.....	50
-------------------	----

ANEXO.....	52
------------	----

BIBLIOGRAFÍA.....	54
-------------------	----

INTRODUCCIÓN

Everything should be made as simple as possible, but not simpler.

Albert Einstein

I have no data yet. It is a capital mistake to theorise before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit fact.

Sir Arthur Conan Doyle,
A scandal in Bohemia

Más allá del Análisis de Regresión Lineal

Bajo los supuestos de que los agentes económicos son racionales y que toman sus decisiones de forma óptima,¹ los investigadores construyen modelos que muestren *regularidades* en los hechos económicos. La técnica más usada para ello es el Análisis de Regresión Lineal (ARL), basado en el método de Mínimos Cuadrados Ordinarios (MCO). Su popularidad radica en que su aplicación es sencilla y los resultados son fáciles de interpretar, en términos de *propensiones marginales* o de *elasticidades*; por estas razones, el ARL constituye la base de otras metodologías más avanzadas, como los modelos de Vectores Autoregresivos (VAR).

Sin embargo, estas técnicas se vuelven poco prácticas cuando el investigador se enfrenta a problemas demasiado complejos para ser simplificados. Algunos estudios (Granger, 1991) muestran que un número considerable de relaciones económicas son no lineales en los parámetros o, incluso, caóticos. Por esta razón, en los últimos diez años se han desarrollado técnicas estadísticas y matemáticas que permiten la modelación y pronóstico de problemas complejos e imprecisos. Muchas de estos métodos alternativos tienen su origen en la Inteligencia Artificial, como los Algoritmos Genéticos, los Sistemas de Clasificación Inteligentes y las Redes Neuronales Artificiales (RNA's).

El uso de RNA's para el modelado y pronóstico económico permite relajar el supuesto de que los agentes conocen el modelo real de la economía. Además, gracias al desarrollo computacional de los últimos años, las RNA's encarnan una atractiva herramienta para los investigadores, puesto que el costo de oportunidad de uso y aplicación

¹ Es decir, los agentes conocen el modelo *real* de la economía.

es muy bajo, y, a diferencia de los modelos basados en técnicas econométricas tradicionales, las RNA's son sistemas adaptativos; es decir, aprenden las formas funcionales a través de los datos incluso si las relaciones son desconocidas o son difíciles de describir. Gracias a su flexibilidad, las RNA's pueden ser aplicadas en diferentes campos de conocimiento, como complementos de los métodos lineales existentes o como metodologías alternativas. Los problemas básicos que se abordan a través de ellas son los siguientes:

- *Clasificación de patrones*: el objetivo es asignarle a un patrón de entrada una clase pre-especificada. Las aplicaciones mejor conocidas son el reconocimiento del habla, la clasificación de cardiogramas y la clasificación de tipos sanguíneos.
- *Clustering*: también conocido como *clasificación no supervisada de patrones*, explora la semejanza entre patrones en un cluster. Las aplicaciones mejor conocidas incluyen el *data mining*, la compresión de datos y el análisis exploratorio de datos.
- *Aproximación de funciones*: supóngase un conjunto de n patrones $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$, generados a partir de una función desconocida $\mu(\mathbf{x})$ con ruido. El objetivo de la aproximación de funciones es encontrar un estimado, $\hat{\mu}$, de la función desconocida μ .
- *Pronóstico*: dado un conjunto de m muestras $\{y_1, y_2, \dots, y_t\}$, donde $t = 1, \dots, m$; el objetivo es calcular y_{t+1} . El pronóstico tiene impacto significativo en las decisiones de negocios, ingeniería, etc.; la predicción económica y climática son aplicaciones típicas.
- *Optimización*: una gran variedad de problemas en matemáticas, estadística, ingeniería, medicina y economía son de optimización. El objetivo es encontrar una solución que satisfaga un conjunto de restricciones.

Básicamente, en el área de las ciencias económicas las RNA's son una eficaz herramienta para detectar y reproducir relaciones lineales y no lineales entre un conjunto de variables (Zhang *et al.*, 1997). En general, cualquier tipo de regresión o modelo multivariante puede estructurarse como una RNA; algunos ejemplos son: la Regresión

Lineal, la Regresión Polinómica, la Regresión Logística, el Análisis Factorial y el Análisis de Componentes Principales.

Algunas de las ventajas de la implementación de las RNA's en el modelado y pronóstico de series económicas son: 1) no es necesario elaborar modelos *a priori*, 2) no es necesario especificar funciones de distribución de probabilidad, 3) generalización de procesos, 4) tolerancia a información incompleta o no lineal, 5) operación de datos en tiempo real.

Redes Neuronales Artificiales: un ingrediente extra

Una pregunta fundamental en el análisis económico contemporáneo es: ¿qué pasa si los agentes tienen que “aprender” la naturaleza de los shocks, los objetivos de política y retroalimentar las reglas de las instituciones a sus propias decisiones; en otras palabras, qué pasa si los agentes no conocen el modelo real de la economía y tienen que hacer “aproximaciones” para capturar y aprender la forma en que las variables clave de la economía se comportan?

Este tipo de cuestionamientos, de gran trascendencia teórica y empírica, no son triviales, pero tampoco son recientes. El objetivo de R. E. Lucas (1978) al publicar el artículo *Econometric Policy Evaluation: A Critique* era indicar que, bajo la hipótesis de expectativas racionales, los modelos econométricos no representaban adecuadamente la realidad económica debido a su incapacidad para 1) incorporar en sus modelos choques económicos aleatorios y 2) los ajustes de los agentes económicos ante cambios de política económica.² Por lo tanto, la econometría (clásica) no podría usarse para formulación de políticas económicas.

La crítica argumentaba que los parámetros de los modelos tienen una tendencia a ser inestables en el tiempo, debido a que los cambios en las políticas económicas van a llevar a los agentes económicos a revisar estos parámetros; de tal manera que, si se tenía un conjunto de parámetros estimados y de pronto se produce un cambio, esos parámetros (considerados invariables) no serían más aplicables.

En una primera lectura, la *crítica de Lucas* parece invalidar todo esfuerzo por modelar la economía, sea a nivel macro o microeconómico; sin embargo, no es así. Los

² Lo cual significa que entre los agentes existe 1) información asimétrica y 2) relaciones no lineales.

contra-argumentos asociados al trabajo de Lucas han mejorado las técnicas econométricas y matemáticas de los últimos años. Por ejemplo, Sims (1980) desarrolló la metodología VAR para demostrar que la *crítica* sólo opera en casos excepcionales; por otro lado, algunos teóricos (en su mayoría matemáticos y estadistas), bajo el hecho de que la realidad económica es compleja y caótica y no simple y ordenada, han desarrollado diversas técnicas para plantear y evaluar modelos con parámetros no lineales o que varían en el tiempo, como los Modelos Dinámicos Lineales³ y las RNA's.

Sin embargo, el uso de este tipo de modelos, clasificados como sistemas de *caja negra*,⁴ también ha recibido numerosas críticas por parte de los investigadores más ortodoxos, debido a que muchos de los parámetros estimados no pueden ser justificados o interpretados económicamente. Y es aquí donde radica su desaprobación: en el deseo de que toda estimación empírica tenga una lectura económica directa o esté respaldada por la teoría económica. Pero muchas veces lo que dictan las teorías es imposible de verificar en la práctica, ya que hay una brecha enorme entre lo que *es* y lo que *debería*; el mismo Lucas (*op. cit.*) indica: “la construcción de modelos econométricos está sujeta a una gran cantidad de críticas mal informadas y casuales. Así, los modelos están condenados por ser *demasiado grandes* (con igual introspección, supongo que uno podría culpar a modelos con menor número de ecuaciones por ser *demasiado pequeños*), demasiado complicados, demasiado simplistas (es decir, no demasiado complicados), y, el golpe final, inferiores a los modelos ingenuos”.

Al parecer, la regla más sensata es: “anteponer la nitidez de los datos a la divinidad econométrica” (Kennedy, 2002). De esta forma, si bien la compleja forma funcional de un sistema *caja negra* como las RNA's hace difícil (pero no imposible) interpretar económicamente los parámetros estimados, su potencia no consiste necesariamente en la *elegancia* de una solución simple, sino en la *generalización* de una respuesta concreta a partir de información de la que se desconoce un modelo *a priori*; además, en vez de considerar que los agentes actúan de forma racional y conocen a la perfección los mecanismos de la economía, el modelado a través de RNA's supone que esto no se cumple;

³ Conocidos erróneamente como *Filtro de Kalman*.

⁴ El término *caja negra* se refiere a un sistema del que se desconocen (o no son de interés de análisis) los elementos o mecanismos internos de solución, pero se sabe que a determinadas entradas (inputs) corresponden salidas bien delimitadas (outputs).

es más, se permite que los agentes aprendan a través del tiempo y reajusten sus errores (expectativas racionales) de forma dinámica.

Por otro lado, también es cierto que el uso de RNA's y otras técnicas de *caja negra* transgrede la regla "modelos simples deben ser preferidos a modelos complicados" (basada en la *navaja de Ockham* o *ley de la Parsimonia*).⁵ Sin embargo, Karl Menger (1960), fundador de la Escuela Austríaca de Economía, al saber que los economistas son demasiado parsimoniosos (y renuentes) en lo que respecta al uso de nuevos modelos (y técnicas) matemáticos y/o estadísticos, formuló una *anti-navaja de Ockham*:⁶ "No tiene caso tratar de hacer con menos lo que requiere más". El objetivo del economista no es simplificar hasta sus últimas consecuencias la realidad, sino proporcionar información de mayor calidad, así como reducir la incertidumbre de los tomadores de decisiones; además, debido a la creciente complejidad de los problemas a los que se enfrentan los investigadores, es preciso enriquecer los métodos y técnicas de análisis para el procesamiento de información y, de esta forma, obtener soluciones más rápidas y eficaces. Loría (2007) afirma: "(...) en la medida que los modeladores incorporen sistemáticamente argumentos estadísticos y económicos nuevos y más sólidos, así como mayor sapiencia de lo que ocurre en la realidad extraeconómica, la brecha que necesariamente existirá entre la realidad económica y su correcta modelación podrá mantenerse dentro de márgenes aceptables". Un modelo econométrico, estadístico o matemático debe ser simple, pero no más simple.

Aprovechar las sinergias entre las técnicas convencionales y las nuevas metodologías para la modelación y pronóstico permitirá, entre otras cosas, evitar los Errores Tipo III,⁷ ya que es mejor una respuesta *aproximada* a la pregunta "correcta", la cual es a menudo imprecisa, que una respuesta *exacta* a la pregunta "incorrecta", la cual con mucha frecuencia puede formularse con exactitud (Tukey, 1962).

Las RNA's son un ingrediente extra en el análisis de datos, en la búsqueda de evidencia adicional.

⁵ *Frustra fit per plura quod potest fieri per pauciora* o "No tiene caso hacer con más lo que se puede hacer con menos". Frase atribuida a Guillermo de Ockham, fraile inglés del siglo XIV. Para una discusión más amplia, véase Rodríguez-Fernández (1999).

⁶ Existen otras *anti-navajas de Ockham*; por ejemplo: el *Principio de Plenitud* de Gottfried Wilhelm von Leibniz, la *Crítica de la Razón Pura* de Immanuel Kant, la *Regla Zellner* de Arnold Zellner.

⁷ Un Error Tipo III ocurre cuando el investigador presenta la respuesta correcta a la pregunta incorrecta. Véase Kimball (1957).

Plan de capítulos

En el presente trabajo se modela y pronostica el Producto Interno Bruto, 1980.1-2008.3, a través del *Perceptrón Multicapa* (Rumelhart *et al.*, 1986), la arquitectura Red Neuronal Artificial más utilizada en los últimos años para la modelación económica univariada y multivariada, debido a su gran capacidad para la resolución de problemas tanto lineales como no lineales. La orientación es teórica y empírica, de tal forma que resulte una guía básica, pudiendo ser entendida por quien cuente con una formación matemática y económica básica.

En el primer capítulo se definen los conceptos básicos y la técnica fundamental del Perceptrón Multicapa, así como las características principales de la *regla de aprendizaje* para el aprendizaje de este tipo de arquitecturas: el *Algoritmo Backpropagation* — Rumelhart (*et al.*, op. cit.), Levenberg (1944) y Marquardt (1963)—.

En la primera parte del segundo capítulo se presentan las características estadísticas y económicas del vector de información utilizado; posteriormente, en las secciones siguientes se especifican dos modelos RNA para el pronóstico del PIB: uno univariado y otro multivariado.

El tercer capítulo presenta los resultados obtenidos, así como la evaluación gráfica y estadística de las arquitecturas neuronales, destacándose las principales características, fortalezas y debilidades de los modelos.

La última sección recoge las conclusiones y recomendaciones derivadas de los capítulos anteriores.

I. REDES NEURONALES ARTIFICIALES

*Sou um técnico, mas tenho técnica só dentro da técnica.
Fora disso sou doido, com todo o direito a sê-lo.
Com todo o direito a sê-lo, ouviram?
Álvaro de Campos, *Lisborn Revisited**

Las Redes Neuronales Artificiales (RNA's), inspiradas en la existencia de una organización celular horizontal en capas en el *córtex cerebral*,¹ son modelos matemáticos que mimetizan las habilidades básicas de las Redes Neuronales Biológicas para procesar grandes cantidades de información. Las RNA's intentan simular dos de las características más importantes del cerebro: 1) capacidad de *aprendizaje* o generalización y 2) procesamiento de información *incompleta* y/o no lineal.

Gracias a su flexibilidad, las RNA's son aplicadas en diferentes campos de conocimiento, como complementos de los métodos existentes o como metodologías alternativas. En los últimos años, las RNA's se han vuelto sumamente populares en economía y finanzas (Trippi y Turban, 1993), ya que estos modelos pueden utilizarse como aproximadores universales de funciones para el reconocimiento de patrones, la clasificación y el pronóstico.

Sin embargo, al igual que los métodos econométricos, las RNA's son un conjunto de técnicas difícil de abordar en su totalidad; por esta razón, en este trabajo nos enfocaremos en la metodología de la RNA más utilizada en economía: el *Perceptrón Multicapa* (Rumelhart *et al.*, 1986). El planteamiento teórico pertenece a diversos autores que se señalarán debidamente, la notación y el desarrollo matemático son propios.

1.1. Breve historia de las Redes Neuronales Artificiales

El cerebro humano está formado por millones de neuronas conectadas entre sí que transmiten y procesan la información percibida a través de los sentidos (vista, tacto, oído, gusto, olfato), para dar una respuesta adecuada a cada uno de los diferentes estímulos. Cada neurona se conforma fundamentalmente de cuatro partes básicas (gráfica 1.1):

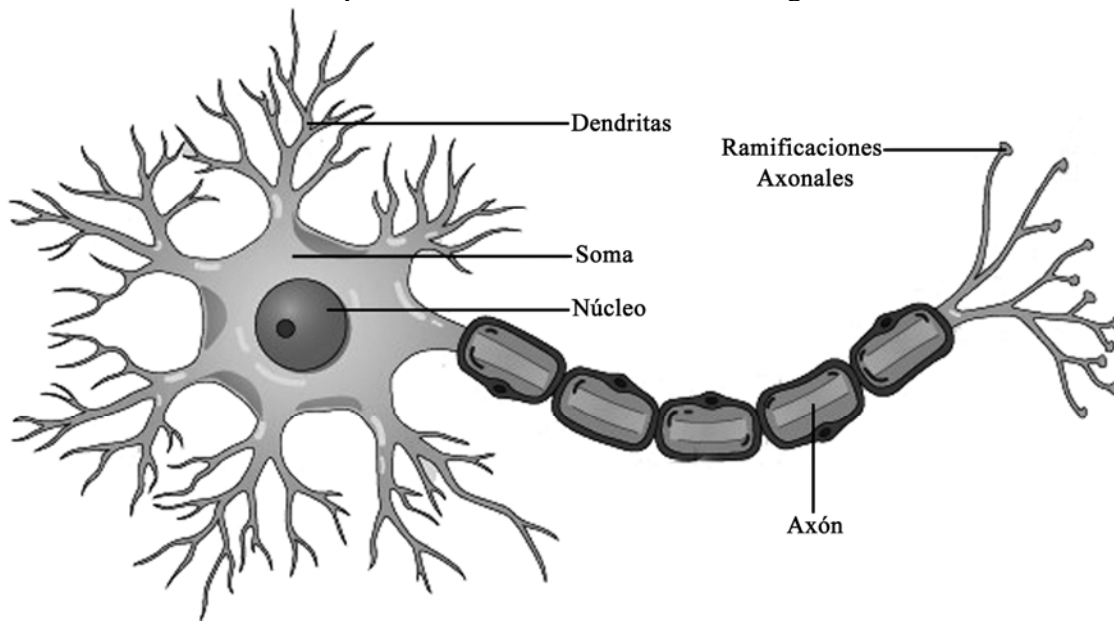
1. *Dendritas*: ramas fibrosas que actúan como conexiones de entrada de la *soma*.

Las dendritas conducen los impulsos nerviosos hacia el cuerpo de la célula.

¹ El *córtex* o corteza cerebral contiene cerca de cien mil millones de neuronas, las cuales son aproximadamente el número de estrellas en la Vía Láctea. Brunak, S. y B. Lautrup (1990)

2. *Soma*: cuerpo celular o estructura de entrada.
3. *Axón*: canal transmisor de los impulsos generados por la soma. En el extremo, la fibra principal se ramifica para conectarse con dendritas de otras neuronas o directamente a los músculos del cuerpo.
4. *Núcleo*: es el centro de síntesis de información de la célula; en él se procesan las señales que llegan de otras neuronas, en forma de impulsos, para generar un nuevo impulso si se cumplen ciertas condiciones.

Gráfica 1.1
Esquema básico de una neurona biológica



La conexión entre neuronas se realiza por medio de uniones especiales llamadas *sinapsis*, las cuales determinan la fuerza y tipo de relación entre ellas. A la intensidad de una sinapsis se le llama *peso sináptico* y se determina biológicamente por el proceso de *aprendizaje*. El aprendizaje de las Redes Neuronales Biológicas (RNB's) se produce mediante la variación de las sinapsis; es decir, en el cambio de la influencia o *peso* que unas neuronas ejercen sobre otras; de esta forma, con el paso del tiempo, el sistema neuronal se modifica para establecer nuevas conexiones, romper otras, alterar las intensidades sinápticas, etc. Estas transformaciones en la intensidad de las conexiones constituyen la idea de trasfondo de las RNA's.

Históricamente, el trabajo del neurofisiólogo Warren McCulloch y del matemático Walter Pitts es el primero en proponer un modelo matemático que imita el comportamiento

básico simplificado de las RNB's (McCulloch y Pitts, 1943). Los años, autores y aportaciones más importantes en la historia de las RNA's puede consultarse en el siguiente cuadro:

Cuadro 1.1
Breve historia de las RNA's

Año	Autor	Aportación
1943	McCulloch y Pitts	Desarrollan el primer modelo de Neurona Artificial.
1949	Hebb	Desarrolla la <i>regla de Hebb</i> , antecesora de las técnicas modernas de entrenamiento de las RNA's.
1958	Rosenblatt	A partir de un trabajo de Minsky (1954), desarrolla el <i>Perceptrón</i> , un modelo simple que aprende a conectar y asociar datos de entrada dados con una unidad de salida aleatoria.
1960	Widrow y Hoff	Crean una variación de la regla de Hebb. La denominada <i>ley de Widrow-Hoff</i> da lugar a las primeras RNA's aplicadas a problemas reales: <i>ADALINE</i> (ADaptive LINear Elements) y <i>MADALINE</i> (Multiple ADALINE).
1969	Minsky y Papert	Demuestran las limitaciones teóricas en el aprendizaje del Perceptrón, incapaz de aprender funciones no linealmente separables.
1974	Werbos	Desarrolla el método de aprendizaje conocido como <i>Backpropagation</i> , el algoritmo más utilizado en las aplicaciones de RNA's.
1977	Anderson <i>et al.</i>	Desarrolla el <i>Asociador Lineal</i> y su extensión <i>Brain-State-in-a-Box</i> , que modela funciones complejas.
1980	Fukushima	Desarrolla diversas redes autoasociativas: la red <i>Neocognitron</i> se especializa en la detección de clusters.
1982	Hopfield	Desarrolla una variante del Asociador Lineal conocido como <i>red de Hopfield</i> , la cual se utiliza en la reconstrucción de patrones y optimización.
1984	Hinton <i>et al.</i>	Desarrollan la <i>Máquina de Boltzman</i> , que reconoce patrones y optimiza funciones simples.
1985	Instituto Americano de Física (IEEE)	Primera reunión anual del <i>Neural Networks for Computing</i> .
1986	Carpenter y Grossberg	Desarrollan el modelo <i>ART</i> (Adaptative Resonance Theory), el cual da respuesta a dos interrogantes fundamentales: cómo puede aprender una RNA nuevos patrones y cómo puede retener los patrones previamente aprendidos.
	Rumelhart <i>et al.</i>	Perfeccionan el algoritmo <i>Backpropagation</i> , ofreciendo una poderosa solución para la construcción de RNA's más complejas.

Cuadro 1.1, continuación

1987	Grossberg <i>et al.</i>	Creación de la primera conferencia internacional sobre RNA's: <i>International Neural Network Society (INNS)</i> .
	Varios	Creación de la reunión anual del <i>Neural Information Processing Systems (NIPS)</i> .
1988	IEEE e INNS	Creación del <i>International Joint Conference on Neural Networks (IJCNN)</i> .
1990	Kohonen	Formaliza el algoritmo de aprendizaje denominado <i>Self-Organizing Map (SOM)</i> , base de las RNA's autoorganizadas de tipo competitivo.
1991	Sociedad Europea de Redes Neuronales	Creación del <i>International Conference on Artificial Neural Networks (ICANN)</i> .
1997	Kohonen	Formaliza el algoritmo de aprendizaje denominado <i>Self-Organizing Feature Map (SOFM)</i> , el cual permite la formación de mapas topológicos multidimensionales.

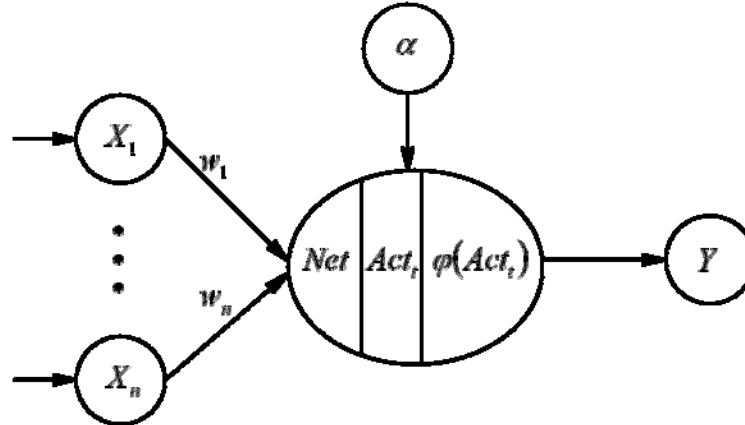
1.2. La Neurona Artificial

Una Neurona Artificial es la unidad básica de procesamiento de información en la operación de una RNA. El modelo, propuesto por Mc Culloch y Pitts (*op. cit.*) y desarrollado por Rosenblatt (1958), se conoce como *Perceptrón Simple (PS)*; sin embargo, lo más frecuente es darle el nombre de nodo o elemento de procesamiento (Freeman y Skapura, 1993).²

Gráficamente, un nodo puede representarse por la gráfica 1.2 y sus elementos constitutivos son: 1) *señales de entrada* o *inputs*, X_1, \dots, X_n ; 2) magnitudes asociadas a las señales de entrada llamadas *pesos sinápticos* o *intensidades de conexión*, w_1, \dots, w_n ; 3) un término de *sesgo* o *umbral*, α , el cual tiene el efecto de incrementar o disminuir la entrada neta dependiendo si éste es positivo o negativo; 4) una *función de propagación* o *entrada neta*, Net , que es la suma ponderada de las señales de entrada con sus respectivos pesos sinápticos más un umbral; 5) una *función de activación* o *transferencia*, Act_i , que combina las señales de entrada con el estado de activación inicial para producir un nuevo valor de activación; 6) una *función de salida*, $\varphi(\bullet)$, que transforma el estado final de activación de la neurona en las señales de salida, las cuales son enviadas a los nodos conectados a ella; 7) una *señal de salida* o *output*, Y ; 8) una *regla* o *algoritmo* de aprendizaje.

² Términos que se emplearán de manera intercambiable a lo largo del documento.

Gráfica 1.2
Esquema básico de un Perceptrón Simple



Puede apreciarse que una unidad de procesamiento funciona como un sistema input-output, ya que usa los valores de los nodos de entrada para calcular el valor de la neurona de salida. Si definimos las señales de entrada, X_1, \dots, X_n , y la señal de salida, Y , como vectores m -dimensionales; es decir, $X_n = [x_{n1} \dots x_{nm}]$ y $Y = [y_1 \dots y_m]$,³ podemos especificar matricialmente un PS a través de las siguientes expresiones:

$$Net = \mathbf{W}\mathbf{X} + \alpha \quad (1.1)$$

$$Act_t = \gamma(Act_{t-1}, Net_t) \quad (1.2)$$

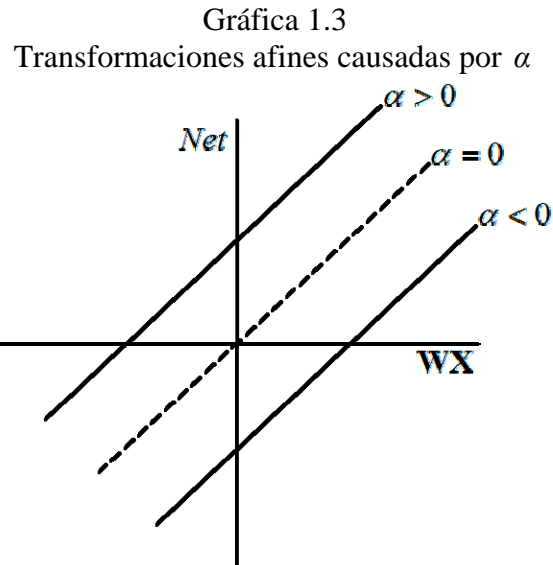
$$\mathbf{Y} = \varphi(Act_t) \quad (1.3)$$

Donde: \mathbf{X} y \mathbf{W} son matrices $n \times 1$ y $1 \times n$, respectivamente; $\mathbf{Y} = [Y]$.

La ecuación (1.1) representa la forma en que se ponderarán las señales de entrada y los pesos sinápticos; en este caso, se trata de una combinación lineal entre los inputs y sus respectivas intensidades de conexión más un umbral. En esta expresión si w_n es positivo la señal de entrada tendrá un efecto *excitador*, es decir, la neurona se activará; por el contrario, si w_n es negativo la sinapsis será *inhibidora*, es decir, la neurona se desactivará; por último, si w_n es igual a cero, significa que no existe conexión entre los nodos. Además, es importante advertir que el efecto de un cambio en el nodo X_n sobre Net será siempre $w_n \Delta X_n$, independientemente de que ΔX_n sea positivo o negativo; en otras palabras, los efectos son simétricos. Ahora bien, el uso de α tiene el propósito de buscar una

³ Donde m es el número de elementos u observaciones. Es importante mencionar que no es necesario que la señal de salida contenga los mismos elementos que las señales de entrada.

transformación afín para \mathbf{WX} porque dependiendo si éste es positivo o negativo, la relación Net es modificada (gráfica 1.3).⁴ Nótese que si $\mathbf{WX} = 0$, entonces $Net = \alpha$.



Por otro lado, la ecuación (1.2) combina el potencial post-sináptico, Net , con el estado inicial de la neurona, Act_{t-1} , para producir un nuevo estado de activación acorde con la información recibida;⁵ sin embargo, en la mayoría de las arquitecturas de RNA's aplicadas se considera que el estado actual de la neurona no depende de su estado previo. De esta forma, podemos expresar la función de transferencia como:

$$Act_t = \gamma(Net_t) \quad (1.4)$$

Además, si consideramos que $\gamma(\bullet)$ tiene carácter determinista, es decir, que se trata de una función identidad, entonces podemos simplificar la expresión anterior de la siguiente forma:

$$Act = Net \quad (1.5)$$

La ecuación (1.3) es la señal de salida; es decir, la transformación de (1.2) o (1.5) a través de la función $\phi(\bullet)$, que permite capturar las posibles relaciones no lineales entre Net y el output (véase sección 1.2.1).

⁴ El parámetro α puede pensarse como un peso w_0 asociado a una señal de entrada $X_0 = 1$. Haykin, 1999.

⁵ Toda neurona presenta cierto valor o *estado* inicial, de reposo o excitación que puede ser continuo, discreto, limitado o ilimitado. Freeman y Skapura (*op. cit.*)

Finalmente, la regla de aprendizaje es la manera en que se calcula y actualiza la matriz \mathbf{W} , con el objetivo de minimizar el error entre la señal de salida calculada y la señal de salida observada (véase sección 1.2.2).

Al simplificar el sistema de ecuaciones (1.1)-(1.3) podemos reescribirlo como:

$$Net = \mathbf{W}\mathbf{X} + \alpha \quad (1.6)$$

$$\mathbf{Y} = \varphi(Net) \quad (1.7)$$

Resulta interesante notar que la ecuación (1.6) es idéntica al *modelo de regresión lineal*: los nodos de entrada y salida son equivalentes a las variables independientes y dependiente, respectivamente; mientras que los pesos sinápticos son semejantes a los coeficientes estimados y el sesgo, al intercepto. Si el sistema (1.6)-(1.7) tuviera más de un nodo de salida, $\mathbf{Y}^T = [Y_{1,t} \ \dots \ Y_{k,t}]$, y el conjunto de inputs fueran los p rezagos de las señales de salida, $\mathbf{X}^T = [Y_{1,t-p} \ \dots \ Y_{k,t-p}]$, obtendríamos una RNA similar a un modelo de *Vectores Autorregresivos* (González, 2000). Más aún: al introducir una conexión entre los outputs Y_1, \dots, Y_k definiríamos un *sistema de ecuaciones simultáneas*.

1.2.1. La función de activación-salida

La función de activación-salida transforma el estado final de activación de la neurona, Net , en las señales de salida. Por lo general, $\varphi(\bullet)$ es de tipo *sigmoidea* como la logística o la tangente hiperbólica —ecuación (1.8) y (1.9), respectivamente—. Este tipo de funciones, no lineales, crecientes y restringidas en un intervalo monotónico con límites superiores e inferiores, se caracterizan por presentar una derivada definida en todos sus puntos.

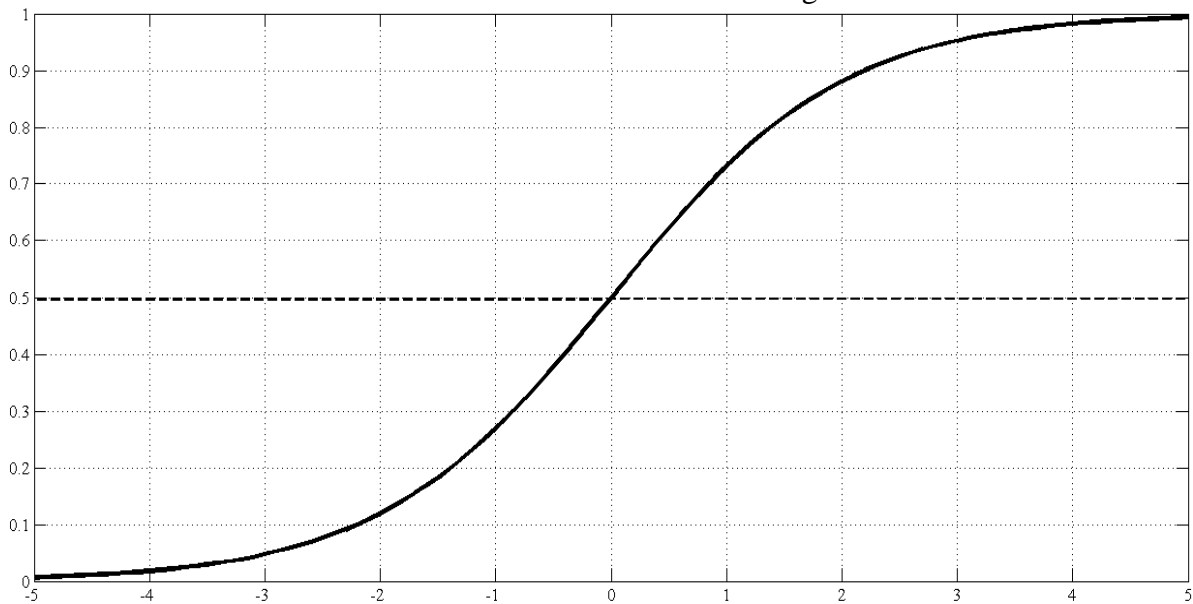
$$\varphi(Net) = \frac{1}{1 + e^{-Net}} \quad (1.8)$$

$$\varphi(Net) = \frac{e^{Net} - e^{-Net}}{e^{Net} + e^{-Net}} \quad (1.9)$$

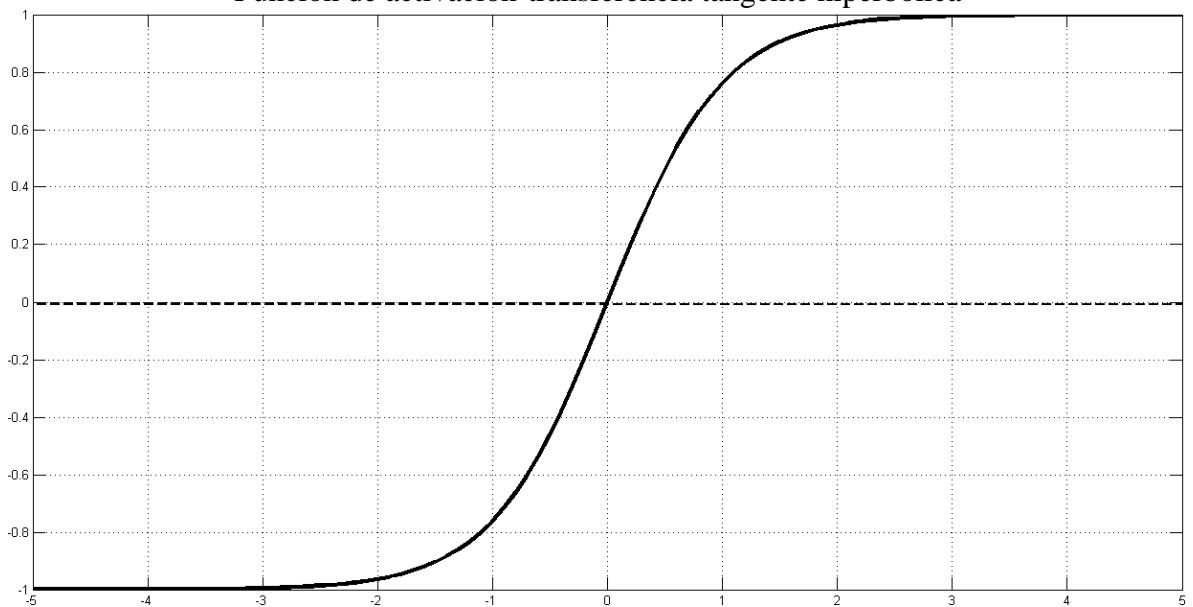
La función logística tiene un rango de 0 a 1 (gráfica 1.4); en consecuencia, cuando $Net = 0$ la salida del nodo es 0.5, es decir, la neurona tiene cierta actividad aún en ausencia de estimulación; es importante notar que al aumentar Net la unidad aumenta su activación y al disminuir, se inhibe. Por otro lado, la función tangente hiperbólica tiene un rango de -1 a 1 (gráfica 1.5); de esta forma, si $Net = 0$ la salida es 0, lo cual permite resolver problemas

continuos acotados o no. El objetivo esencial para utilizar estas funciones es capturar las posibles relaciones no lineales entre *Net* y el output; además, su empleo es apropiado en la modelación económica, debido a que las formas funcionales *a priori* entre variables son, frecuentemente, desconocidas.⁶

Gráfica 1.4
Función de activación-transferencia logística



Gráfica 1.5
Función de activación-transferencia tangente hiperbólica



⁶ Una forma funcional incorrecta lleva a una deficiente estimación de los parámetros.

1.2.2. La regla de aprendizaje

El *conocimiento* de una neurona se encuentra representado en los pesos sinápticos entre los nodos; así, el proceso de aprendizaje o *entrenamiento* implica cierto número de cambios en la intensidad de las conexiones; es decir, el Perceptrón *aprende* al modificar de forma óptima los valores de la matriz \mathbf{W} .

Básicamente, existen dos clases de algoritmos para la actualización de los pesos: 1) supervisados y 2) no supervisados.⁷ Dentro del conjunto de reglas de aprendizaje supervisado pueden distinguirse tres tipos: 1) por corrección de error, 2) por refuerzo y 3) estocásticas. En la práctica, los *algoritmos por corrección de error* son los más utilizados. Este tipo de reglas se caracterizan por establecer un *agente* que compara la señal de salida calculada con la señal de salida deseada y, si éstas son diferentes, ajusta la intensidad de las conexiones a través de minimizar el error cometido hasta que la señal de salida sea igual a la deseada; en otras palabras, el aprendizaje por corrección de error confronta el output estimado, $\hat{\mathbf{Y}}$, con el output observado, \mathbf{Y} , y ajusta \mathbf{W} iterativamente hasta que $\hat{\mathbf{Y}} = \mathbf{Y}$.

A continuación se presenta el algoritmo básico de aprendizaje del PS:

1. Inicio. Se establecen los pesos sinápticos de \mathbf{W} con números aleatorios.⁸
2. Iteración i -ésima. Se calcula:

$$Net_i = \mathbf{W}_i \mathbf{X} + \alpha_i$$

$$\hat{\mathbf{Y}}_i = \varphi(Net_i)$$

Donde: i indica la iteración tal que $i = 1, \dots$

3. Se calcula el error de salida:

$$E_i = \mathbf{Y} - \hat{\mathbf{Y}}_i \tag{1.10}$$

Donde: $\hat{\mathbf{Y}}$ es la señal de salida calculada.

4. Corrección de los pesos sinápticos. Si $\hat{\mathbf{Y}}_i \neq \mathbf{Y}$, entonces \mathbf{W} y α se modifican a través de las siguientes expresiones:

$$\left. \begin{aligned} \mathbf{W}_{i+1} &= \mathbf{W}_i + \eta E_i \mathbf{X}^T \\ \alpha_{i+1} &= \alpha_i + \eta E_i \end{aligned} \right\} \tag{1.11}$$

⁷ Su implementación dependerá de la *arquitectura* neuronal utilizada. Para mayor detalle, consúltese Flórez y Fernández (2008).

⁸ Los cuales pueden estar en los intervalos (0,1) o (-1,1), dependiendo si se utiliza la función (1.8) o la (1.9).

Donde: η es una constante positiva (conocida como *tasa de aprendizaje*) tal que $0 < \eta < 1$; si ésta es pequeña, la neurona aprenderá con lentitud y, si es grande, lo hará con mayor rapidez (comúnmente, $\eta = 0.5$); T indica transposición.

5. Parada. Si no se han modificado las intensidades de las conexiones en las últimas r iteraciones —ecuación (1.12)— el algoritmo se detiene, ya que el PS se ha estabilizado. En otro caso, se va al paso dos con $i = i + 1$.

$$\left. \begin{array}{l} \mathbf{W}_s = \mathbf{W}_i \\ \alpha_s = \alpha_i \end{array} \right\} \quad (1.12)$$

Donde: $s = i + 1, \dots, i + r$.

Aunque eficiente, el algoritmo anterior presenta una limitación importante: si las señales de entrada no son linealmente separables,⁹ el aprendizaje del PS no converge después de un número finito de iteraciones (Minsky y Papert, 1969). Por esta razón, Werbos (1974) y, a la postre, Rumelhart *et al.* (*op. cit.*), Levenberg (*op. cit.*) y Marquardt (*op. cit.*) plantean, desarrollan y perfeccionan —respectivamente— el algoritmo *Backpropagation*, una regla de aprendizaje más poderosa y eficiente.

1.3. El Perceptrón Multicapa

La palabra *arquitectura* está reservada para la descripción formal de los elementos topológicos de una RNA: 1) número de capas, 2) número de nodos por capa, 3) tipo de conexión entre nodos y 4) algoritmo de aprendizaje. Las arquitecturas neuronales pueden clasificarse según la estructura de las capas, el flujo de datos, el tipo de respuesta de la red, etc.;¹⁰ sin embargo, debido a la enorme riqueza de arquitecturas y de reglas de aprendizaje existentes, en esta sección se desarrollarán únicamente las características fundamentales del *Perceptrón Multicapa* (PM) y del algoritmo *Backpropagation* (BP).

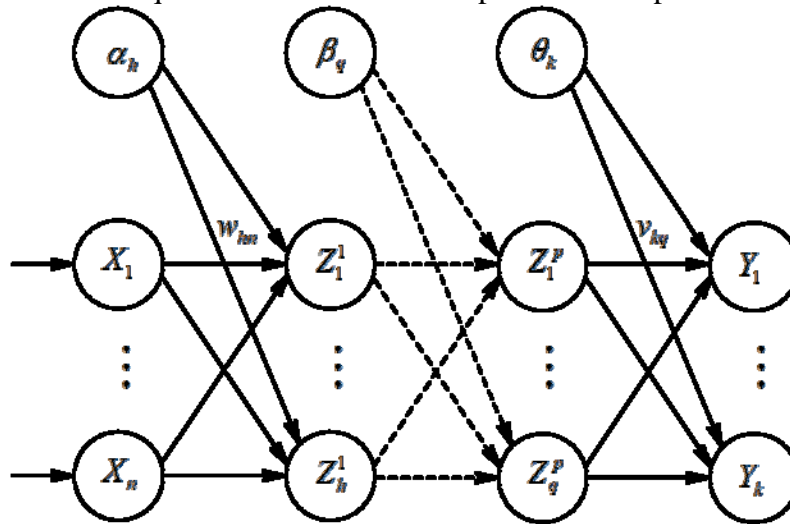
⁹ Por ejemplo: se dice que dos conjuntos de puntos A y B son linealmente separables en un espacio n -dimensional si existen $n + 1$ números reales, w_1, \dots, w_n, α , tal que cada punto $(X_1, \dots, X_n) \in A$ satisface $\mathbf{W}_i \mathbf{X} \geq \alpha$ y que cada punto $(X_1, \dots, X_n) \in B$ satisface $\mathbf{W}_i \mathbf{X} < \alpha$. Véase gráfica 1.7.

¹⁰ Véase Flórez y Fernández (*op. cit.*).

En esencia, el PM es una extensión y generalización del PS.¹¹ Se trata de una arquitectura formada por diversas capas de neuronas interconectadas *hacia adelante* (figura 1.6),¹² las cuales pueden clasificarse en tres tipos:

1. Capa de entrada: es la que recibe y procesa directamente los inputs, X_1, \dots, X_n .
2. Capas ocultas: son internas a la RNA y su objetivo es capturar todas las relaciones no directas entre las señales de entrada y de salida, Z_h^1, \dots, Z_q^p .
3. Capa de salida: es la que transfiere la información procesada al exterior de la red, Y_1, \dots, Y_k .

Gráfica 1.6
Esquema básico de un Perceptrón Multicapa

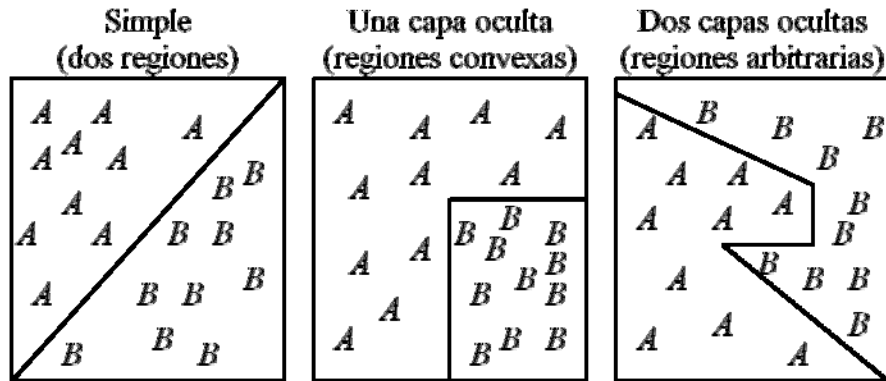


Si bien es posible construir arquitecturas con un número indeterminado de capas ocultas, Funahashi (1989) demostró que un PM con una capa oculta es un aproximador universal de funciones y puede representar adecuadamente regiones polinomiales convexas, tanto cerradas como abiertas; por otro lado, Lapedes y Faber (1987) comprobaron que un PM con dos capas ocultas puede representar regiones arbitrariamente complejas (gráfica 1.7). En general, una arquitectura con al menos una capa oculta y con suficientes nodos puede aprender y reproducir cualquier tipo de función o relación continua entre un grupo de señales de entrada y de salida.

¹¹ El PS es considerado una arquitectura monocapa tipo *feedforward*.

¹² Es decir, una capa recibe señales de entrada desde una capa anterior (la cual está más cerca de la capa de entrada) y envía señales de salida a una capa posterior (la cual está más próxima a la capa de salida). Este tipo de RNA's se conocen como arquitecturas *feedforward*.

Gráfica 1.7
 Tipo de Perceptrón y de región clasificable¹³



Si definimos las señales de entrada, X_1, \dots, X_n , y de salida, Y_1, \dots, Y_k , como vectores m -dimensionales; es decir, $X_n = [x_{n1} \dots x_{nm}]$ y $Y_k = [y_{k1} \dots y_{km}]$,¹⁴ podemos especificar matricialmente un PM $[n - h - k]$ ¹⁵ a través de las siguientes expresiones:

$$Net = \mathbf{W}\mathbf{X} + \boldsymbol{\alpha} \quad (1.13)$$

$$\mathbf{Z} = \varphi(Net) \quad (1.14)$$

$$\mathbf{Y} = \delta(\mathbf{V}\mathbf{Z} + \boldsymbol{\beta}) \quad (1.15)$$

Donde: \mathbf{X} , \mathbf{Z} y \mathbf{Y} son matrices de señales $n \times 1$, $h \times 1$ y $k \times 1$, respectivamente; \mathbf{W} es una matriz $h \times n$ con los pesos sinápticos de la capa de entrada; \mathbf{V} es una matriz $k \times h$ con los pesos sinápticos de la capa oculta; $\boldsymbol{\alpha}$ es una matriz $h \times 1$ con los umbrales de la capa de entrada y $\boldsymbol{\beta}$ es una matriz $k \times 1$ con los umbrales de la capa oculta.

Al igual que la ecuación (1.6), la expresión (1.13) establece una combinación lineal entre los inputs y sus respectivas intensidades de conexión más un umbral de activación.

La ecuación (1.14) transforma el potencial post-sináptico, Net , a través de la función $\varphi(\bullet)$ y envía la nueva señal a la capa oculta, \mathbf{Z} , con el objetivo de capturar patrones complejos entre los inputs y el output. Como señala González (*op. cit.*), es importante aclarar que los h nodos de la capa oculta no representan ningún concepto real;

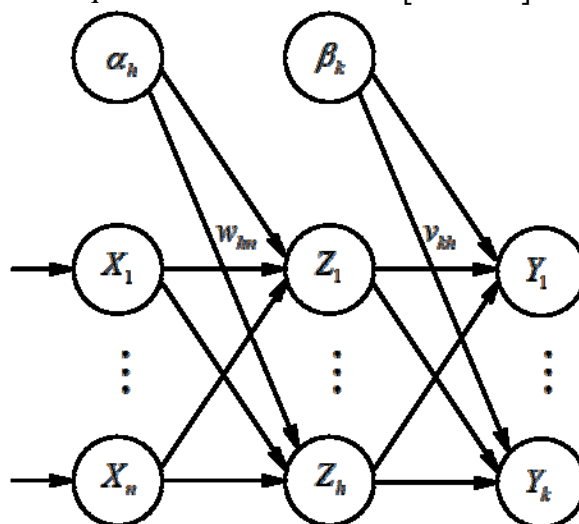
¹³ Cuando la función de activación-salida es no lineal —como la ecuación (1.8) o (1.9)—, entonces las regiones no están limitadas por rectas, sino por curvas.

¹⁴ Donde m es el número de elementos u observaciones.

¹⁵ Donde n , h y k hacen referencia al número de nodos en la capa de entrada, oculta y de salida, respectivamente. Véase gráfica 1.8.

es decir, no tienen interpretación o significado económico directo.¹⁶ Sin embargo, esto no representa una limitación; por el contrario, como se indicó en párrafos anteriores, una arquitectura con una capa oculta y suficientes neuronas en ella puede aproximar cualquier función lineal o no lineal, lo cual sugiere que una RNA correctamente especificada puede utilizarse como una poderosa herramienta en la identificación y reproducción de Procesos Generadores de Información no conocidos *a priori*.

Gráfica 1.8
Esquema básico de un PM $[n - h - k]$



Por otro lado, aunque las funciones $\varphi(\bullet)$ y $\delta(\bullet)$ pueden ser distintas para cada una de las capas de la red —inclusive para cada una de las neuronas—, por lo general, en la modelación de variables económicas $\varphi(\bullet)$ puede ser cualquiera de las ecuaciones (1.8) o (1.9) mientras que $\delta(\bullet)$ se determina como lineal. La ecuación (1.15), al igual que la ecuación (1.7), es la señal de salida calculada. Puede observarse que si el conjunto de inputs del sistema (1.13)-(1.15) fueran los p rezagos de las señales de salida, $\mathbf{X}^T = [Y_{1,t-p} \ \dots \ Y_{k,t-p}]$, obtendríamos una RNA similar a un modelo de *Vectores Autorregresivos*. Como dato adicional para futuras investigaciones, resulta interesante conocer y comparar el costo de formulación (en términos de parámetros) de estas

¹⁶ McNelis (2005) interpreta las neuronas de la capa oculta como variables *latentes*; es decir, expectativas subjetivas o fuerzas conductoras del mercado: los "animal spirits" de Keynes.

metodologías:¹⁷ $q = (2p + 1) \times l$ para un VAR irrestricto y $g = [(n + 1) \times h] + [(h + 1) \times k]$ para una RNA Multicapa; donde q y g son los grados de libertad de un VAR irrestricto y una RNA Multicapa, respectivamente; p es el número de variables; l es el número de rezagos; n , h y k son el número de nodos en la capa de entrada, oculta y de salida, respectivamente.

Finalmente, para calcular y actualizar las matrices de pesos sinápticos, \mathbf{W} y \mathbf{V} , se utiliza el algoritmo BP, con el objetivo de minimizar de forma óptima el error de estimación.

1.3.1. El Algoritmo Backpropagation

Como se indicó en la sección 1.1.2, se dice que una red aprende cuando ésta modifica sus pesos sinápticos una y otra vez hasta encontrar un conjunto que minimice el error de salida cometido. En el caso del PM, el aprendizaje se lleva a cabo a través del algoritmo *Backpropagation* (BP), el cual fue propuesto por Werbos (*op. cit.*) y, posteriormente, desarrollado por Rumelhart *et al.* (*op. cit.*).

El objetivo del algoritmo BP es estimar las matrices \mathbf{W} y \mathbf{V} minimizando una función de costo —ecuación (1.21)— a través de su gradiente; así, una vez que las señales de entrada se aplican a la primera capa de la red y éstas se transforman y propagan a través de la capa oculta hasta generar un conjunto de outpus estimados,¹⁸ se calcula una señal de error para cada unidad de salida —ecuación (1.16)—.

En seguida, las señales de error se transmiten hacia atrás, partiendo de la capa de salida, hacia todos los nodos de la capa oculta —ecuaciones (1.17)-(1.18)—.¹⁹ Es importante notar que las unidades ocultas sólo reciben una fracción de la señal total del error, basándose aproximadamente en la contribución relativa que haya aportado la unidad a la salida original. Este proceso se repite hasta que todos los nodos de la red hayan recibido una señal de error que describa su contribución relativa del error total; de esta forma, con base en la señal de error percibida, se actualizan los pesos sinápticos de cada

¹⁷ En la actualidad, los Vectores Autoregresivos son muy populares para el modelado de variables económicas, siempre y cuando las relaciones subyacentes sean lineales.

¹⁸ Proceso conocido como *Propagación hacia delante*.

¹⁹ Proceso conocido como *Propagación hacia atrás*.

unidad para hacer que la red converja asintóticamente hacia una solución estable — ecuaciones (1.19)-(1.20)—.

A continuación se presenta el algoritmo básico BP:

1. Inicio. Se establecen los pesos sinápticos de \mathbf{W} y \mathbf{V} con números aleatorios.²⁰
2. Iteración i -ésima. Se calcula:

$$Net_i = \mathbf{W}_i \mathbf{X} + \mathbf{a}_i$$

$$\mathbf{Z}_i = \varphi(Net_i)$$

$$\hat{\mathbf{Y}}_i = \delta(\mathbf{V}_i \mathbf{Z}_i + \mathbf{\beta}_i)$$

Donde: i indica la iteración tal que $i = 1, \dots$

3. Se calcula el error para los nodos de la capa de salida:

$$\mathbf{E}_i = \mathbf{Y} - \hat{\mathbf{Y}}_i \quad (1.16)$$

Donde: \mathbf{E} es una matriz $k \times 1$; $\hat{\mathbf{Y}}$ es la señal de salida calculada.

4. Si $\hat{\mathbf{Y}}_i \neq \mathbf{Y}$, entonces se propaga el error hacia los nodos de la capa oculta y de entrada:

$$\boldsymbol{\zeta}_i = \boldsymbol{\Lambda}_i \mathbf{E}_i \quad (1.17)$$

$$\boldsymbol{\xi}_i = \boldsymbol{\Omega}_i \mathbf{V}_i^T \boldsymbol{\zeta}_i \quad (1.18)$$

Donde: $\boldsymbol{\Lambda} = \text{diag}(\delta'(\bullet)_1, \dots, \delta'(\bullet)_k)$; $\boldsymbol{\Omega} = \text{diag}(\varphi'(\bullet)_1, \dots, \varphi'(\bullet)_h)$; $\boldsymbol{\zeta}$ y $\boldsymbol{\xi}$ son matrices $k \times 1$ y $h \times 1$, respectivamente.

5. Se actualizan los pesos de la capa de entrada y de la capa oculta:

$$\left. \begin{aligned} \mathbf{W}_{i+1} &= \mathbf{W}_i - \eta \boldsymbol{\xi}_i \mathbf{X}^T \\ \mathbf{a}_{i+1} &= \mathbf{a}_i - \eta \boldsymbol{\xi}_i \end{aligned} \right\} \quad (1.19)$$

$$\left. \begin{aligned} \mathbf{V}_{i+1} &= \mathbf{V}_i - \eta \boldsymbol{\zeta}_i \mathbf{Z}_i^T \\ \mathbf{\beta}_{i+1} &= \mathbf{\beta}_i - \eta \boldsymbol{\zeta}_i \end{aligned} \right\} \quad (1.20)$$

Donde: η es una constante positiva (conocida como *tasa de aprendizaje*) tal que $0 < \eta < 1$; si ésta es pequeña, la neurona aprenderá con lentitud y, si es grande, lo hará con mayor rapidez (comúnmente, $\eta = 0.5$); T indica transposición.

²⁰ Los cuales pueden estar en los intervalos (0,1) o (-1,1), dependiendo si se utiliza la función (1.8) o la (1.9).

6. Parada. Se busca minimizar el índice de rendimiento de la red —ecuación (1.21)—; en este caso, el Error Cuadrático Medio. Si no se cumple con el error objetivo, se va al paso dos con $i = i + 1$.

$$\varepsilon = \frac{1}{M} \mathbf{E}^T \mathbf{E} \quad (1.21)$$

Donde M es el número de patrones de entrenamiento.

La virtud de este proceso consiste en que, a medida que se entrena la red, los nodos de la capa oculta se organizan a sí mismos de tal modo que las distintas neuronas aprenden a reconocer distintas características de las señales de entrada. Así, cuando se les presente un conjunto arbitrario de señales de entrada que contenga ruido o que esté incompleta, las unidades de la capa oculta responderán con una salida activa si las nuevas entradas contienen información semejante a aquella característica que las unidades individuales hayan aprendido a reconocer durante su entrenamiento; y a la inversa, las unidades de las capas ocultas tienen una tendencia a inhibir sus salidas si la información de entrada no contiene las características apropiadas.

En la actualidad, existen diferentes variantes del algoritmo BP como el *Levenberg-Marquardt* (LM), el cuál consiste en calcular y actualizar los pesos de una matriz Ψ que contiene los pesos de toda la red,²¹ a través de la siguiente expresión:²²

$$\Psi_{i+1} = \Psi_i - \left((\mathbf{J}_i^T \mathbf{J}_i + \lambda_i \mathbf{I})^{-1} \mathbf{J}_i^T \mathbf{E}_i \right) \quad (1.22)$$

Donde: $\mathbf{E}_i = \mathbf{Y} - \hat{\mathbf{Y}}_i$; $\hat{\mathbf{Y}}$ es la señal de salida calculada; \mathbf{J} es la matriz Jacobiana; λ es un parámetro de ajuste que se modifica en cada iteración usando la tasa de aprendizaje η tal que $0 < \eta < 1$; T indica transposición.

Como señalan Abolfazl *et al.* (*op. cit.*), el algoritmo BP-LM tiene un excelente desempeño en el entrenamiento de RNA's donde el rendimiento de la arquitectura esté determinado por el Error Cuadrático Medio, ya que esta regla de aprendizaje converge a una solución estable (mínimo local o global) en menos iteraciones; sin embargo, a diferencia del algoritmo BP, el BP-LM requiere más cálculos por iteración.

²¹ Es decir, $\Psi = \{\mathbf{W}, \mathbf{V}\}$.

²² Para consultar un desarrollo más extenso y elegante, véase Abolfazl *et al.* (2005) y Hagan y Menhaj (1994).

1.4. Estadísticos de bondad de ajuste y correcta especificación

Como señalan McNelis (*op. cit.*) y McMenamin (1996), el complejo sistema no lineal que representa una RNA necesita que se tomen en cuenta los siguientes criterios para su correcta especificación:

1.4.1. Número de neuronas en la capa oculta

Determinar la dimensión de la capa oculta, Z_h , es la parte más complicada en la construcción de una arquitectura neuronal; gran parte de la eficiencia de una RNA recaerá en la elección del número de neuronas ocultas. Aunque se trata (en esencia) de un proceso *ensayo-error*, existen diversas propuestas que pueden ayudar a fijarlas:²³

$$h = \sqrt{n \times k} \quad (1.23)$$

$$h = 2k \pm 1 \quad (1.24)$$

$$h = \frac{M/10}{n(k+1)} - 1 \quad (1.25)$$

Donde: n , h y k son el número de nodos en la capa de entrada, oculta y de salida, respectivamente; M es el número de patrones de entrenamiento.

Cabe mencionar que un número muy grande de nodos ocultos hará que el sistema se ajuste demasiado a las singularidades de los inputs durante el entrenamiento, aprendiendo incluso el ruido presente en ellos, por lo que perderá capacidad de generalización.

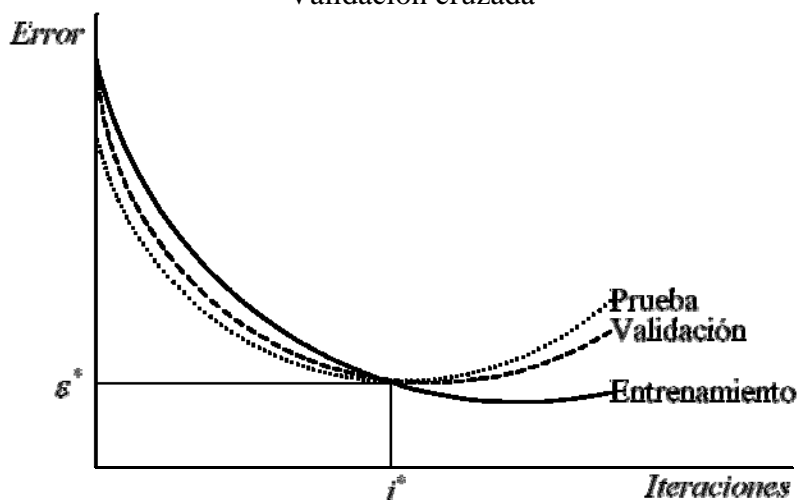
1.4.2. Validación cruzada

El excesivo ajuste de la red neuronal a las particularidades de los patrones empleados en la fase de entrenamiento es un fenómeno conocido como *sobreaprendizaje*. Una RNA con sobreaprendizaje o *sobreajuste* no puede generalizar satisfactoriamente las relaciones subyacentes en los datos; por esta razón, el algoritmo BP necesita que las señales de entrada se dividan en tres conjuntos: 1) de entrenamiento, 2) de validación y 3) de prueba. De esta forma, cuando la red actualice los pesos de la capa de entrada y de la capa oculta —ecuaciones (1.19)-(1.20)— se podrá observar el proceso de cómo el error de entrenamiento y el de validación disminuyen (a medida que aumenta el número de iteraciones) hasta alcanzar un mínimo en la superficie del error; mientras que, al mismo tiempo, los datos de

²³ Véase Flórez y Fernández (*op. cit.*).

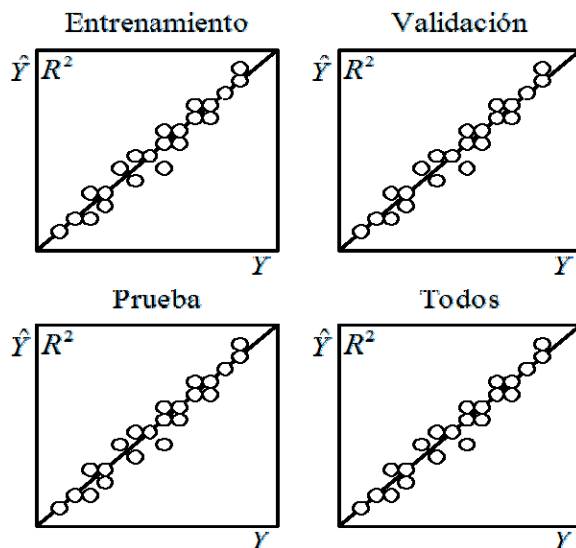
prueba proporcionan una estimación insesgada del error de generalización. Este proceso se conoce como *validación cruzada* (gráfica 1.9) y sirve para detectar el punto óptimo de aprendizaje.

Gráfica 1.9
Validación cruzada



Además de lo anterior, podemos realizar un análisis de regresión lineal entre las señales de salida observadas y las calculadas a través de los conjuntos de datos de entrenamiento, de validación y de prueba (gráfica 1.10.), para conocer la bondad de ajuste de la arquitectura neuronal.

Gráfica 1.10.
Análisis de regresión²⁴



²⁴ Las R^2 se calculan a partir de la ecuación (1.26).

1.4.3. Criterios dentro de la muestra²⁵

Si se desea realizar un análisis de inferencia a través de una RNA, entonces es necesario un apropiado *ajuste* entre la señal de salida calculada y la observada, así como verificar que se satisfagan las propiedades estadísticas de los errores de estimación, esto es: normalidad, homoscedasticidad y ausencia de autocorrelación serial. A continuación se presentan los estadísticos básicos dentro de la muestra:

1. Bondad de ajuste. Indica la proximidad existente entre las señales de salida observadas y las calculadas a través de los conjuntos de datos de entrenamiento, de validación y de prueba.

$$R^2 = 1 - \frac{\mathbf{E}^T \mathbf{E}}{(\mathbf{Y} - \bar{\mathbf{Y}})^T (\mathbf{Y} - \bar{\mathbf{Y}})} \quad (1.26)$$

$$\bar{R}^2 = 1 - (1 - R^2) \frac{M - 1}{M - g} \quad (1.27)$$

Donde: $g = (n + 1)h + (h + 1)k$ es el número de parámetros estimados y M el número total de observaciones.

2. Error Estándar. Es una medida basada en la varianza estimada de los residuales.

$$s = \sqrt{\frac{\mathbf{E}^T \mathbf{E}}{M - g}} \quad (1.28)$$

3. Normalidad. El estadístico Jarque-Bera permite contrastar la hipótesis de normalidad en los residuos:

$$JB = \frac{M}{6} \left(S^2 + \frac{(K - 3)^2}{4} \right) \quad (1.29)$$

Donde: K es la kurtosis y S , el sesgo.

4. Homoscedasticidad. El estadístico de Engle sirve para comprobar la ausencia de Heteroscedasticidad Condicional Autorregresiva; en una arquitectura bien especificada la varianza debe ser constante:

$$ARCH = \mu_0 + \sum_{\tau=1}^g \mu_{\tau} \mathbf{E}_{t-\tau}^2 + \nu_t \quad (1.30)$$

Donde: $\nu \sim N(0, \sigma^2)$.

²⁵ Los datos dentro de la muestra se conforman por los conjuntos de señales de entrenamiento y de validación.

5. No autocorrelación serial. Para evidenciar autocorrelación serial de cualquier orden se utiliza el estadístico Ljung-Box —ecuación (1.31)—, el cual examina la significancia conjunta de las primeras l autocorrelaciones con una distribución χ^2 con L grados de libertad:

$$Q(L) = (M)(M + 2) \sum_{l=1}^L \frac{\rho_l^2(\mathbf{E})}{(M - l)} \quad (1.31)$$

Donde: $\rho_l(\mathbf{E})$ es la función de autocorrelación parcial.

6. Criterios de información. Los estadísticos de Akaike y de Schwarz —ecuaciones (1.32) y (1.33), respectivamente— resultan herramientas muy útiles cuando se determina el número óptimo de nodos en la capa oculta o para comparar modelos:

$$AIC = -\frac{2LV}{M} + \frac{2g}{M} \quad (1.32)$$

$$SC = -\frac{2LV}{M} + \frac{g \log M}{M} \quad (1.33)$$

$$\text{Donde: } LV = \frac{M}{2} (1 + \log(2\pi)) + \log \left(\frac{\mathbf{E}^T \mathbf{E}}{M} \right)$$

1.4.4. Criterios fuera de la muestra²⁶

Finalmente, si se desea construir una RNA únicamente para fines de pronóstico, entonces se debe contrastar su desempeño predictivo o de generalización a través de dos métodos complementarios: 1) el gráfico y 2) el estadístico. El primero consiste en analizar visualmente la simulación histórica de $\hat{\mathbf{Y}}$; el segundo, en evaluar los siguientes criterios:²⁷

1. Error Medio.

$$EM = \frac{1}{M} \sum_{m=1}^M \mathbf{E}_m \quad (1.34)$$

2. Error Medio Absoluto:

$$EMA = \frac{1}{M} \sum_{m=1}^M |\mathbf{E}_m| \quad (1.35)$$

²⁶ Los datos fuera de la muestra se conforman por el conjunto de señales de prueba.

²⁷ Véase Pindyck *et al.* (1997).

3. Error Cuadrático Medio:

$$ECM = \frac{1}{M} \mathbf{E}^T \mathbf{E} \quad (1.36)$$

4. Raíz del Error Cuadrático Medio:

$$RECM = \sqrt{\frac{1}{M} \mathbf{E}^T \mathbf{E}} \quad (1.37)$$

5. Coeficiente de *Theil - U* :²⁸

$$Theil - U = \frac{RECM}{\sqrt{\frac{1}{M} \mathbf{E}^{*T} \mathbf{E}^*}} \quad (1.38)$$

Donde: \mathbf{E}^* es el error de pronóstico *ingenuo*.

²⁸ Véase Bliemel (1973).

II. REDES NEURONALES ARTIFICIALES PARA EL PRONÓSTICO DEL PIB: MÉXICO, 1980.1-2008.3

*Me lo contaron y lo olvidé;
lo vi y lo entendí;
lo hice y lo aprendí.
Confucio*

En el análisis económico aplicado es fundamental contar con instrumentos matemáticos y estadísticos que proporcionen información relevante a los agentes económicos para la toma de decisiones, por ello es necesario definir modelos robustos que den cuenta de la(s) variable(s) de nuestro interés. En el presente capítulo se estiman dos modelos de Redes Neuronales Artificiales para modelar el Producto Interno Bruto de México, 1980.1-2007.4; el primero univariado y el segundo, multivariado.¹

2.1. El vector de información

El siguiente vector contiene un conjunto de información general que puede dar cuenta de la economía mexicana en un contexto de economía abierta y pequeña:

$$\mathbf{X}^T = [Y_t \quad I_t \quad M_t \quad Q_t \quad Y_t^*] \quad (2.1)$$

Donde Y_t = Producto Interno Bruto de México;² I_t = Inversión Total; M_t = Oferta Monetaria;³ Q_t = Tipo de Cambio (México-Estados Unidos);⁴ Y_t^* = Producto Interno Bruto de Estados Unidos.

Un primer punto que debe tomarse en cuenta para la aplicación de cualquier metodología de modelación para el pronóstico, es la transformación adecuada de cada una de las series (Juselius, 2006), ya que la inferencia estadística está basada en los supuestos de medias, varianzas y covarianzas fijas. Por tanto, antes de trabajar con el vector (2.1) debemos asegurar que los datos sean estacionarios; es decir, que tanto el primer y segundo momentos sean constantes a través del tiempo.

¹ Estimaciones a través del software MATLAB 2009.

² A precios de 1993.

³ M = Billetes y monedas en poder del público + Cuentas de cheques m.n en bancos residentes + Cuentas de cheques m.e. en bancos residentes + Depósitos en cuenta corriente en bancos residentes + Depósitos a la vista de las Sociedades de Ahorro y Préstamo + Activos financieros internos en poder de residentes + Activos financieros internos en poder de no residentes + Captación de sucursales y agencias de bancos mexicanos en el exterior.

⁴ $Q = \text{NER} * (\text{P}^{\text{us}}/\text{P}^{\text{mex}})$, NER = Tipo de Cambio Nominal (pesos por dólar), P^{us} y P^{mex} = Índice de Precios del Consumidor de México y de Estados Unidos, respectivamente.

Las gráficas 2.1 y 2.2 presentan las series del vector de información en logaritmos y primera diferencia, respectivamente. La inspección visual revela que los supuestos de media y varianza constantes no se cumplen para las series en niveles, pero se satisfacen al diferenciarlas una vez respecto al tiempo, a pesar de que existe evidencia de periodos de alta volatilidad durante el periodo de estudio.⁵ Una de las ventajas de trabajar con la metodología RNA es la tolerancia a información incompleta o no lineal en los datos; de esta forma, aunque la serie i no tenga la misma longitud periódica que las demás variables, no es necesario reducir el rango de estudio. Es decir, no hay pérdida de información.

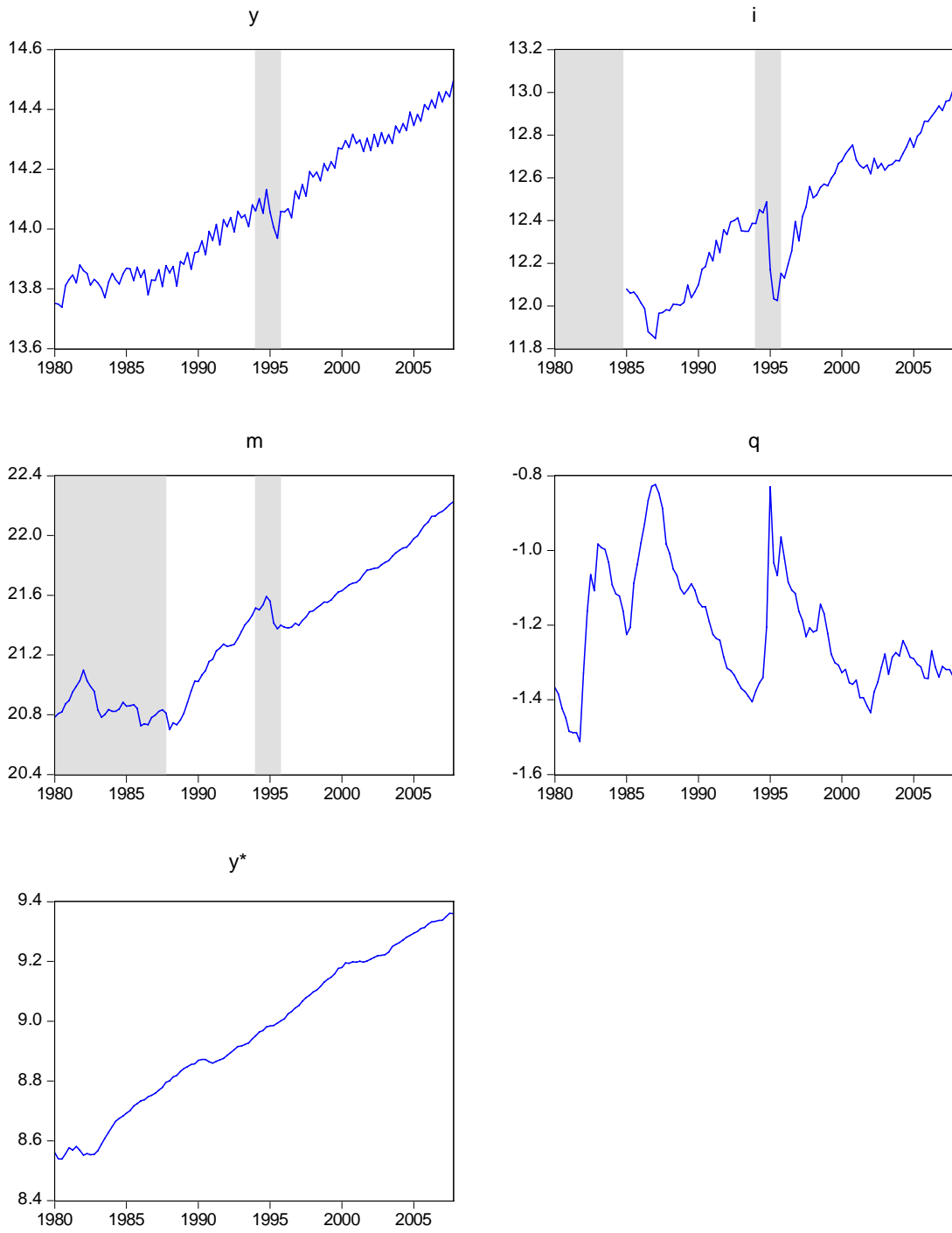
Como sugiere Juselius (*op. cit.*), es una excelente idea contrastar los periodos de alta volatilidad de las series con los eventos económicos y, de esta forma, justificarlos a través de hallar intervenciones o reformas económicas cruciales. Sin embargo, es importante tener en mente que no siempre es posible conocer cuándo una intervención resulta suficientemente fuerte para producir un efecto “extraordinario”.⁶

Los hechos estilizados permiten identificar los siguientes sucesos que provocaron profundos cambios estructurales para la economía nacional: durante el periodo de 1982-1988 se cambia el modelo económico; a partir de estos años el crecimiento se descansa en el mercado internacional, ya que éste se ve como única salida al agotamiento del modelo anterior, representado por estancamiento de la actividad productiva desde los años setenta; por tanto, se inicia una etapa de privatización de las empresas y una política económica que se apega a las reglas del libre mercado. México se adhiere al GATT en 1986 (Acuerdo General Sobre Aranceles y Comercio), las devaluaciones cambiarias se usan como mecanismos compensadores de la liberación externa. La inflación sube en promedio de casi 100% cada año, el desempleo crece a una tasa del 15% entre 1983 y 1985 junto con drásticas caídas en la producción. El número de paraestatales baja de 1155 en 1982 a 412 en 1988. Ante la severa crisis económica, se establecen los Pactos de Crecimiento Económico con los diversos sectores sociales, mediante los cuales el gobierno coordina y acuerda precios relativos (factores y bienes), y ancla el tipo de cambio nominal.

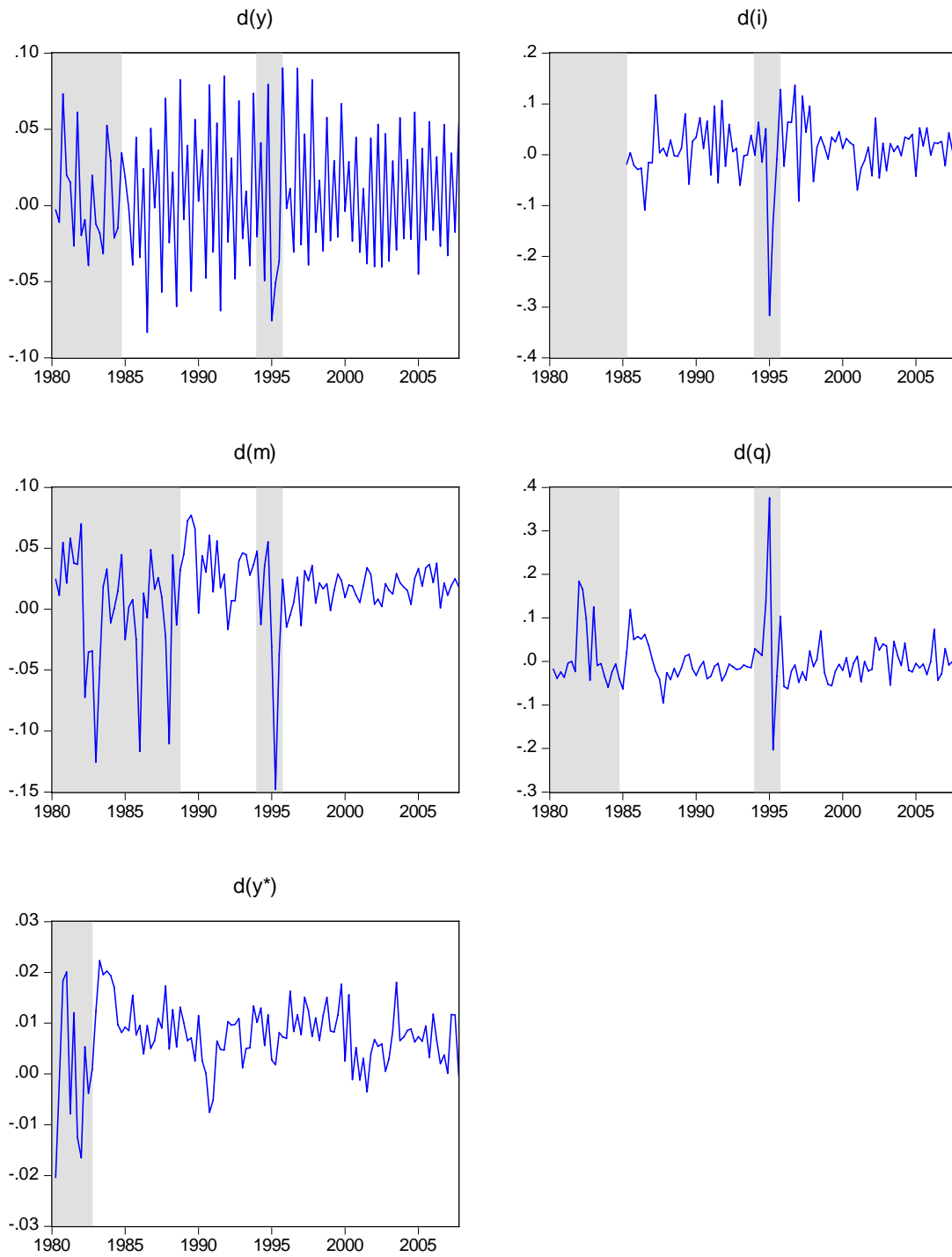
⁵ Para visualizar con mayor facilidad la alta volatilidad de las variables empleadas, se han sombreado los periodos más representativos.

⁶ En esencia, cada mes, trimestre o año es sujeto a alguna clase de intervención política; sin embargo, la mayoría de éstas tienen poco impacto sobre los datos y los modelos. Juselius (*op. cit.*).

Gráfica 2.1
Series en logaritmos, 1980.1-2007.4



Gráfica 2.2
Series en primera diferencia, 1980.2-2007.4



A partir de 1988, entre los eventos económicos más importantes destacan: en febrero de 1991, George Bush, Carlos Salinas de Gortari y Brian Mulroney (presidentes de EE.UU., México y Canadá respectivamente) anuncian formalmente que sus gobiernos se preparan para negociar un histórico acuerdo comercial de libre comercio, el cual será un catalizador del crecimiento económico del hemisferio, a través del aumento en la inversión, el comercio y el empleo; las negociaciones concluyen formalmente el 11 de agosto de 1992 con la elaboración del documento final, pero el TLCAN entra en operación hasta el primero de enero de 1994. La privatización bancaria, llevada a cabo sin contar con un marco regulatorio adecuado, desemboca en la crisis de 1995, cuya consecuencia es la quiebra masiva de bancos y su posterior rescate bajo el polémico esquema del Fobaproa. El alto déficit en cuenta corriente y una baja capacidad para hacer frente a los compromisos de la deuda, aunado a una serie de aumentos en las tasas de interés estadounidenses, obligan a devaluar el peso, creando una reacción en cadena en toda América Latina caracterizada por la fuga de capitales (efecto “Tequila”).

Con base en el análisis gráfico anterior procedemos a determinar el orden de integración de las series a través de las pruebas de raíces unitarias convencionales en la literatura, con el fin de conocer cuántas veces debemos diferenciar cada serie para que ésta se vuelva estacionaria. La información estadística del cuadro 2.1 nos permite concluir que las series en niveles no rechazan la hipótesis nula de raíz unitaria y, por tanto, su orden de integración es $I(1)$; mientras que, por otro lado, sus primeras diferencias sí rechazan la hipótesis nula; es decir, su orden de integración es $I(0)$. Estos resultados justifican la redefinición del vector (2.1) como:

$$\mathbf{X}^T = [\Delta y_t \quad \Delta i_t \quad \Delta m_t \quad \Delta q_t \quad \Delta y_t^*] \quad (2.2)$$

Es importante notar que la información contenida en este nuevo vector posee las características necesarias para realizar inferencia estadística, esto es: medias, varianzas y covarianzas constantes a través del tiempo. Ahora bien, la literatura sugiere que, además del análisis anterior, se realicen las pruebas de *causalidad en el sentido de Granger* (Granger, 1969) para determinar si un grupo de series se explican conjuntamente, tanto en niveles como en primeras diferencias, a lo largo del tiempo. Los resultados que se obtengan nos ayudarán a establecer en las siguientes secciones los modelos RNA más apropiados para el vector (2.2).

Cuadro 2.1
Estadísticas básicas y pruebas de raíces unitarias, 1980.1-2007.4

	y	$d(y)$	i	$d(i)$	m	$d(m)$	q	$d(q)$	y^*	$d(y^*)$
Mean	14.06	0.00	12.41	0.01	21.34	0.01	-1.21	0.00	8.95	0.00
Median	14.04	-0.00	12.41	0.01	21.38	0.01	-1.24	-0.01	8.94	0.00
Maximum	14.49	0.09	13.01	0.13	22.22	0.07	-0.82	0.37	9.36	0.02
Minimum	13.73	-0.08	11.84	-0.31	20.70	-0.14	-1.51	-0.20	8.53	-0.02
Std. Dev.	0.21	0.04	0.31	0.06	0.44	0.03	0.16	0.06	0.25	0.00
Skewness	0.27	0.15	-0.04	-1.72	0.21	-1.93	0.55	2.33	-0.08	-0.90
Kurtosis	1.72	1.94	1.83	11.57	1.86	8.48	2.64	15.14	1.81	5.29
Jarque-Bera Probability	8.92 0.01	5.61 0.06	5.25 0.07	323.70 0.00	6.95 0.03	208.54 0.00	6.24 0.04	783.12 0.00	6.72 0.03	39.52 0.00
Sum Sum.Sq.Dev.	1575.55 5.15	0.74 0.20	1142.60 9.01	0.93 0.32	2390.93 22.27	1.44 0.15	-136.15 2.87	0.02 0.42	1003.34 6.97	0.80 0.00
Observations	112	111	92	91	112	111	112	111	112	111
ADF	-3.40 (8) ³	-3.31 (7) ²	-1.84 (5) ¹	-4.22 (4) ¹	-2.05 (1) ³	-7.38 (0) ²	-2.09 (0) ²	-9.17 (0) ¹	-2.53 (2) ³	-5.77 (1) ²
DF-GLS	-2.34 (8) ³	-2.79 (7) ²	0.51 (5) ²	-3.35 (3) ²	-1.71 (1) ³	-7.17 (0) ²	-1.90 (0) ³	-8.85 (0) ²	-2.63 (2) ³	-1.29 (1) ²
PP	-0.29 (8) ²	-11.22 (7) ¹	-0.35 (5) ²	-10.00 (4) ¹	2.62 (1) ¹	-6.82 (0) ¹	-2.09 (0) ²	-9.17 (0) ¹	-2.90 (2) ³	-8.12 (1) ²
KPSS	2.44 (8) ³	0.36 (7) ²	19.19 (5) ²	0.01 (4) ¹	7.02 (1) ³	0.32 (0) ²	3.65 (0) ³	0.10 (0) ²	1.88 (2) ³	0.19 (1) ²

Notas: ¹ Sin constante ni tendencia; ² constante; ³ constante y tendencia; número de rezagos entre paréntesis; todas las pruebas son válidas al 95% de confianza.

Por un lado, la información vertida en el cuadro 2.2 revela que, para las datos en niveles, rechazamos la hipótesis nula de *no causalidad en el sentido de Granger* para las series y , i y m hasta con cuatro rezagos; es decir, existe evidencia estadística de que i y m explican el comportamiento de y hasta por un año; es importante notar que la serie q determina a y sólo a partir del segundo rezago, mientras que y^* lo hace únicamente durante los dos primeros trimestres. Por otro lado, el cuadro 2.3 muestra que, con excepción de y^* , todas las series diferenciadas rechazan la hipótesis nula hasta por cuatro rezagos, lo cual evidencia una fuerte relación entre los datos en el corto plazo. Los diagramas de dispersión no sólo confirman visualmente estas conclusiones, sino también dejan ver que entre y y q existen ciertas no-linealidades en niveles (gráfica 2.3),⁷ las cuales son más evidentes entre y , m y y^* en primeras diferencias (gráfica 2.4).

⁷ Una gran dispersión entre dos variables refleja una débil o compleja correlación; es decir, los puntos no están suficientemente agrupados como para asegurar la existencia de una relación (causal) lineal.

Cuadro 2.2

Análisis de causalidad en el sentido de Granger: series en logaritmos, 1980.1-2007.4

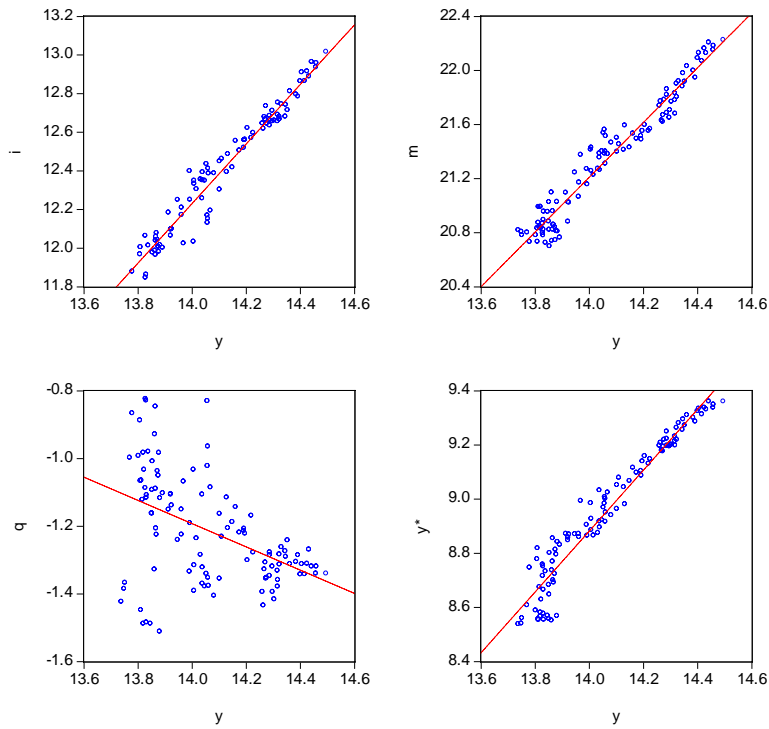
Rezago		<i>i</i> does not Granger Cause <i>y</i>	<i>y</i> does not Granger Cause <i>i</i>	<i>m</i> does not Granger Cause <i>y</i>	<i>y</i> does not Granger Cause <i>m</i>	<i>q</i> does not Granger Cause <i>y</i>	<i>y</i> does not Granger Cause <i>q</i>	<i>y</i> * does not Granger Cause <i>y</i>	<i>y</i> does not Granger Cause <i>y</i> *
1	F-Statistic	7.62	0.10	16.03	1.30	1.03	1.42	20.42	3.57
	Prob.	0.00	0.74	0.00	0.25	0.31	0.23	0.00	0.06
	Obs	91		111		111		111	
2	F-Statistic	11.14	23.85	5.41	6.01	6.40	2.01	3.47	0.98
	Prob.	0.00	0.00	0.00	0.00	0.00	0.13	0.03	0.37
	Obs	90		110		110		110	
3	F-Statistic	11.19	15.59	3.79	3.91	9.35	2.22	2.39	0.49
	Prob.	0.00	0.00	0.01	0.01	0.00	0.09	0.07	0.68
	Obs	89		109		109		109	
4	F-Statistic	31.72	8.87	6.50	3.34	15.93	2.18	2.28	0.54
	Prob.	0.00	0.00	0.00	0.01	0.00	0.07	0.06	0.70
	Obs	88		108		108		108	

Cuadro 2.3

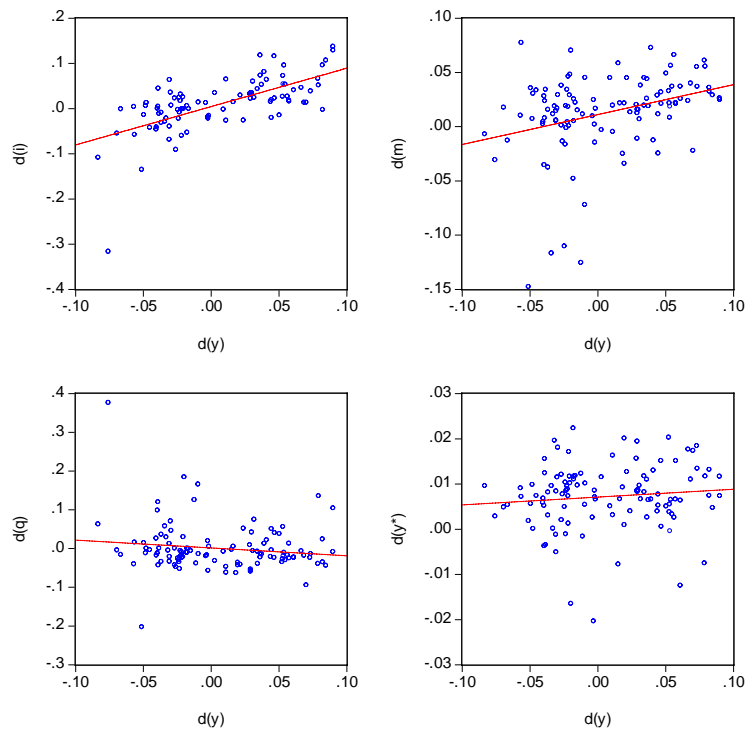
Análisis de causalidad en el sentido de Granger: series en primera diferencia, 1980.2-2007.4

Rezago		<i>d(i)</i> does not Granger Cause <i>d(y)</i>	<i>d(y)</i> does not Granger Cause <i>d(i)</i>	<i>d(m)</i> does not Granger Cause <i>d(y)</i>	<i>d(y)</i> does not Granger Cause <i>d(m)</i>	<i>d(q)</i> does not Granger Cause <i>d(y)</i>	<i>d(y)</i> does not Granger Cause <i>d(q)</i>	<i>d(y</i> *) does not Granger Cause <i>d(y)</i>	<i>d(y)</i> does not Granger Cause <i>d(y</i> *)
1	F-Statistic	22.12	36.85	9.89	5.98	10.57	1.94	2.23	0.79
	Prob.	0.00	0.00	0.00	0.01	0.00	0.16	0.13	0.37
	Obs	90		110		110		110	
2	F-Statistic	17.11	18.06	4.87	3.30	13.83	1.39	0.79	0.42
	Prob.	0.00	0.00	0.00	0.04	0.00	0.25	0.45	0.65
	Obs	89		109		109		109	
3	F-Statistic	42.92	9.37	9.21	2.62	19.79	1.30	1.72	0.66
	Prob.	0.00	0.00	0.00	0.05	0.00	0.27	0.16	0.57
	Obs	88		108		108		108	
4	F-Statistic	15.52	4.99	3.16	2.19	12.22	1.21	2.40	3.68
	Prob.	0.00	0.00	0.01	0.07	0.00	0.30	0.05	0.00
	Obs	87		107		107		107	

Gráfica 2.3
 Diagrama de dispersión, series en logaritmos, 1980.1-2007.4



Gráfica 2.4
 Diagrama de dispersión, series en primera diferencia, 1980.2-2007.4



2.2. Dos modelos de Redes Neuronales Artificiales

Básicamente, existen dos grandes grupos de metodologías para el pronóstico: el univariado y el multivariado. El primer grupo busca identificar patrones (formas funcionales) en los datos a través de modelos *caja negra* que utilizarán sólo las observaciones pasadas de la variable de estudio para explicar su comportamiento futuro; el segundo grupo asume que existen relaciones de *causa y efecto* entre la variable de interés y otras variables. Aunque Fildes (1989) señala la dificultad de encontrar reglas sencillas mediante las cuales el pronosticador puede seleccionar un método específico,⁸ los modelos RNA permiten trabajar bajo ambos enfoques permitiendo reconocer tanto formas funcionales lineales como no lineales en los datos.

2.2.1. RNA Univariada

Como señalan Makridakis *et al* (1998), existen dos razones para utilizar modelos univariados o de *caja negra*: 1) cuando el sistema no puede ser entendido o, si éste puede serlo, es difícil medir las relaciones que rigen su comportamiento; 2) cuando el objetivo es únicamente pronosticar lo que sucederá y no explicar lo que ya ocurrió.

Bajo la lógica de que las observaciones pasadas revelan más información acerca del comportamiento del PIB en el futuro inmediato, a continuación se presenta la RNA Univariada adoptada en este trabajo:

$$\begin{aligned} Net &= \mathbf{W}\mathbf{X} + \boldsymbol{\alpha} \\ \mathbf{Z} &= \varphi(Net) \\ \mathbf{Y} &= \delta(\mathbf{V}\mathbf{Z} + \boldsymbol{\beta}) \end{aligned} \tag{2.3}$$

Donde: $\mathbf{X}^T = [\Delta y_{t-1} \quad \Delta y_{t-2} \quad \Delta y_{t-3} \quad \Delta y_{t-4} \quad \Delta y_{t-5} \quad \Delta y_{t-6}]$; \mathbf{W} es una matriz 3×6 con los pesos sinápticos de la capa de entrada; \mathbf{Z} es una matriz de señales 3×1 ; \mathbf{V} es una matriz 1×3 con los pesos sinápticos de la capa oculta;⁹ $\boldsymbol{\alpha}$ es una matriz 3×1 con los umbrales de la capa de entrada; $\boldsymbol{\beta}$ es una matriz 1×1 con los umbrales de la capa oculta; $\mathbf{Y} = [\Delta y_t]$; $\varphi(\bullet)$ y $\delta(\bullet)$ son funciones tangente hiperbólica y lineal, respectivamente;

⁸ Se tienen muchas y grandes ventajas al considerar más de un método de pronóstico. Entre más difieran éstos, mejor. Armstrong (1978).

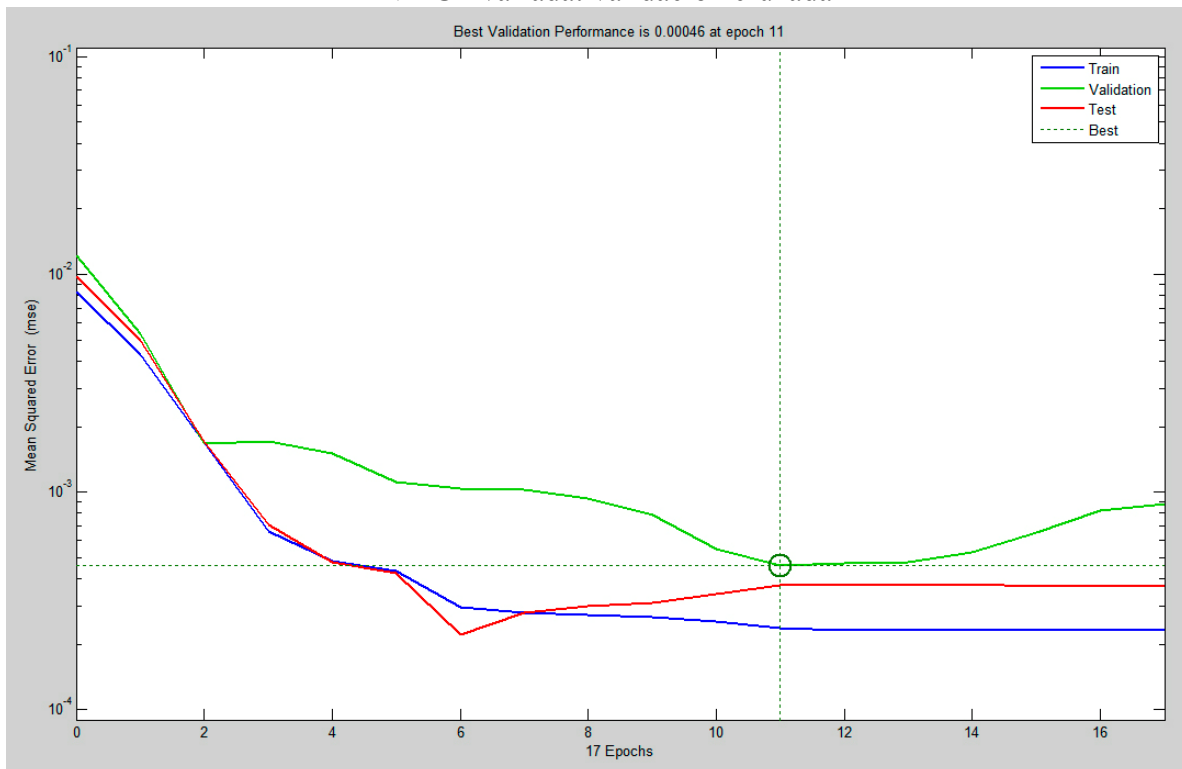
⁹ El número de nodos de la capa oculta se determinó a través de la expresión $h = \sqrt{n \times k}$; donde: n , h y k son el número de nodos en la capa de entrada, oculta y de salida, respectivamente.

cálculo y actualización de las matrices W y V a través del algoritmo BP *Levenberg-Marquardt*.

Como se señalaba en la sección 1.4.2 un excesivo ajuste de la red neuronal a las particularidades de los patrones empleados en la fase de entrenamiento es un fenómeno conocido como *sobreaprendizaje*; una RNA con sobreaprendizaje o *sobre-ajuste* no puede generalizar satisfactoriamente las relaciones subyacentes en los datos. Para evitar el sobreaprendizaje se utiliza la validación cruzada, la cual consiste en dividir aleatoriamente las señales de entrada en tres conjuntos para detectar el punto óptimo donde se minimiza el error de generalización. Bajo esta premisa, la información contenida en la matriz X se dividió de la siguiente manera: 1) conjunto de entrenamiento, 80%; 2) conjunto de validación, 10%; y 3) conjunto de prueba, 10%.

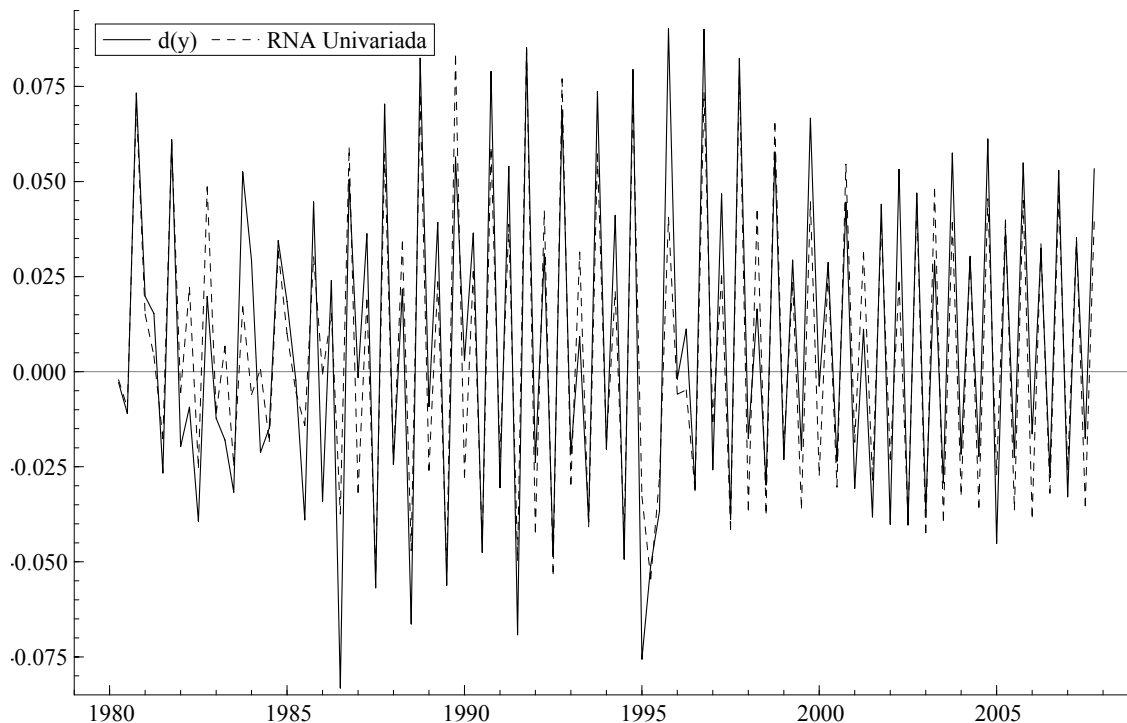
A través de la gráfica 2.5 se puede observar el proceso de cómo el error de entrenamiento y el de validación disminuyen –a medida que aumenta el número de iteraciones– hasta alcanzar un mínimo en la superficie del error ($\varepsilon=0.000046$); mientras que, al mismo tiempo, los datos de prueba proporcionan una estimación insesgada del error de generalización. Puede concluirse que no existe un sobre-ajuste de la RNA a los datos.

Gráfica 2.5
RNA Univariada: validación cruzada



La gráfica 2.6 presenta la simulación histórica de \hat{Y} . Como puede observarse, la bondad de ajuste es satisfactoria (91.42%);¹⁰ sin embargo, Flowers (1989) señala que debe tenerse en mente que no existe garantía de que un buen ajuste histórico proporcione pronósticos precisos en el futuro.

Gráfica 2.6
RNA Univariada: simulación histórica, 1980.1-2007.4

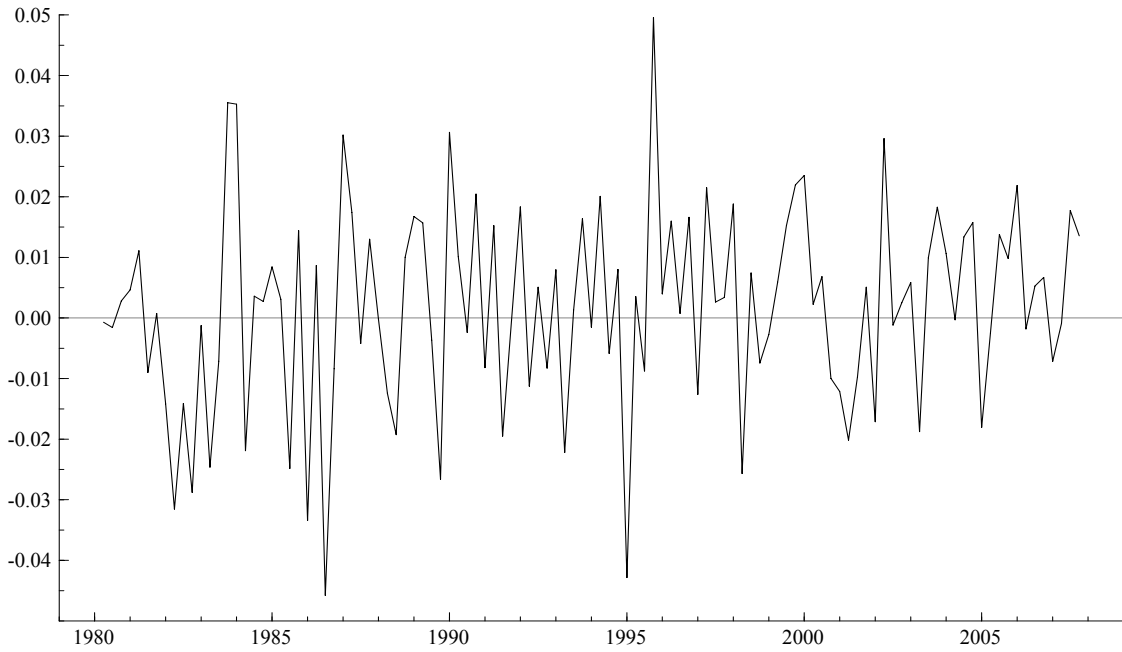


Por otro lado, el gráfico 2.7 muestra el error de estimación: puede observarse que no existen patrones sistemáticos, puntos aberrantes o cambios estructurales importantes que violen los supuestos básicos de un ruido Gaussiano. La información vertida en los cuadros 2.4 y 2.5 confirman que la serie es normal,¹¹ no está autocorrelacionada y es homoscedástica; además, las diferentes pruebas de raíces unitarias aplicadas corroboran que ε es un proceso estacionario. Por tanto, podemos concluir que la RNA Univariada está correctamente especificada.

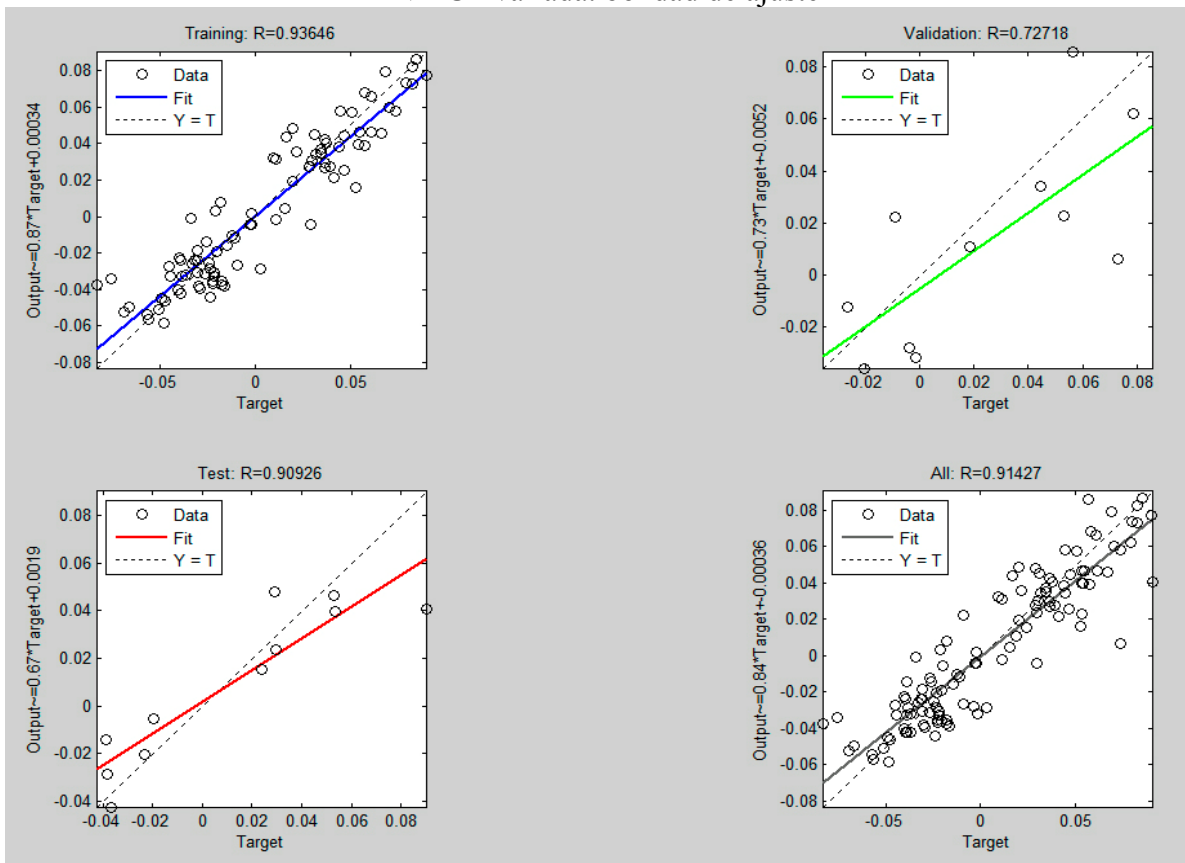
¹⁰ La gráfica 2.8 presenta un análisis de regresión lineal entre las señales de salida observadas y las calculadas a través de los conjuntos de datos de entrenamiento, de validación y de prueba.

¹¹ El estadístico Jarque-Bera (1.73) confirma que el proceso se distribuye normalmente y la prueba de hipótesis H_0 evidencia que su media es estadísticamente igual a cero.

Gráfica 2.7
RNA Univariada: error de estimación, 1980.1-2007.4



Gráfica 2.8
RNA Univariada: bondad de ajuste



Cuadro 2.4
Error de estimación: estadísticas básicas
y pruebas de raíces unitarias, 1980.1-2007.4

ε	
Mean	0.00
Median	0.00
Maximum	0.04
Minimum	-0.04
Std. Dev.	0.01
Skewness	-0.23
Kurtosis	3.39
Jarque-Bera	1.73
Probability	0.42
H ₀ : Mean=0	0.99
Probability	0.32
Sum	0.17
Sum Sq. Dev.	0.03
Observations	111
ADF	-10.82 (0) ¹
DF-GLS	-11.22 (0) ²
PP	-10.82 (0) ¹
KPSS	0.02 (0) ²

Notas: ¹ Sin constante ni tendencia; ² constante y tendencia; número de rezagos entre paréntesis; todas las pruebas son válidas al 99% de confianza.

Cuadro 2.5
Error de estimación: estadísticos de correcta especificación

Rezago	Q-stats ¹	p-values	Arch-stats ²	p-values
1	0.25	0.61	0.56	0.45
2	1.56	0.45	2.06	0.35
3	3.13	0.37	3.25	0.35
4	3.36	0.49	3.13	0.53
5	3.84	0.57	3.85	0.56
6	7.32	0.29	3.82	0.70
7	7.63	0.36	3.78	0.80
8	8.86	0.35	4.51	0.80
9	9.10	0.42	4.69	0.86
10	11.58	0.31	5.54	0.85

Notas: ¹ H₀: No autocorrelación, ² H₀: Homoscedasticidad

2.2.2. RNA Multivariada

Los modelos multivariados o *causales* representan fenómenos reales de forma simplificada, para ello se determina una o varias variables endógenas y un conjunto de variables explicativas. Debe tomarse en cuenta que para decidir qué tan complejo debe ser el modelo deben tomarse en cuenta los usos que se le darán (Aykaç y Borges, 1989).¹² Por tanto, si el único objetivo es obtener pronósticos exactos, entonces el criterio de cuáles variables usar como dependientes reside en su poder causal sobre la(s) variable(s) dependiente(s).¹³ De esta forma, con base en la información estadística del cuadro 2.3, a continuación se presenta la RNA Multivariada adoptada en este trabajo:

$$\begin{aligned} Net &= \mathbf{W}\mathbf{X} + \boldsymbol{\alpha} \\ \mathbf{Z} &= \varphi(Net) \\ \mathbf{Y} &= \delta(\mathbf{V}\mathbf{Z} + \boldsymbol{\beta}) \end{aligned} \quad (2.4)$$

Donde: $\mathbf{X}^T = [\Delta y_{t-1} \quad \Delta y_{t-2} \quad \Delta i_t \quad \Delta i_{t-1} \quad \Delta m_t \quad \Delta m_{t-1} \quad \Delta q_t \quad \Delta q_{t-1} \quad \Delta y_t^* \quad \Delta y_{t-1}^*]$; \mathbf{W} es una matriz 3×10 con los pesos sinápticos de la capa de entrada; \mathbf{Z} es una matriz de señales 3×1 ; \mathbf{V} es una matriz 1×3 con los pesos sinápticos de la capa oculta;¹⁴ $\boldsymbol{\alpha}$ es una matriz 3×1 con los umbrales de la capa de entrada; $\boldsymbol{\beta}$ es una matriz 1×1 con los umbrales de la capa oculta; $\mathbf{Y} = [\Delta y_t]$; $\varphi(\bullet)$ y $\delta(\bullet)$ son funciones tangente hiperbólica y lineal, respectivamente; cálculo y actualización de las matrices \mathbf{W} y \mathbf{V} a través del algoritmo BP *Levenberg-Marquardt*.

Como se señaló en el apartado anterior, para evitar el sobreaprendizaje del sistema (2.4) se dividieron aleatoriamente las señales de entrada en tres conjuntos para realizar la validación cruzada: 1) un conjunto de entrenamiento, 80%; 2) un conjunto de validación, 10%; y 3) un conjunto de prueba, 10%. El proceso de cómo el error de entrenamiento y el de validación disminuyen –a medida que aumenta el número de iteraciones– hasta alcanzar un mínimo en la superficie del error ($\varepsilon=0.00024$); mientras que, al mismo tiempo, los datos de prueba proporcionan una estimación insesgada del error de generalización, puede

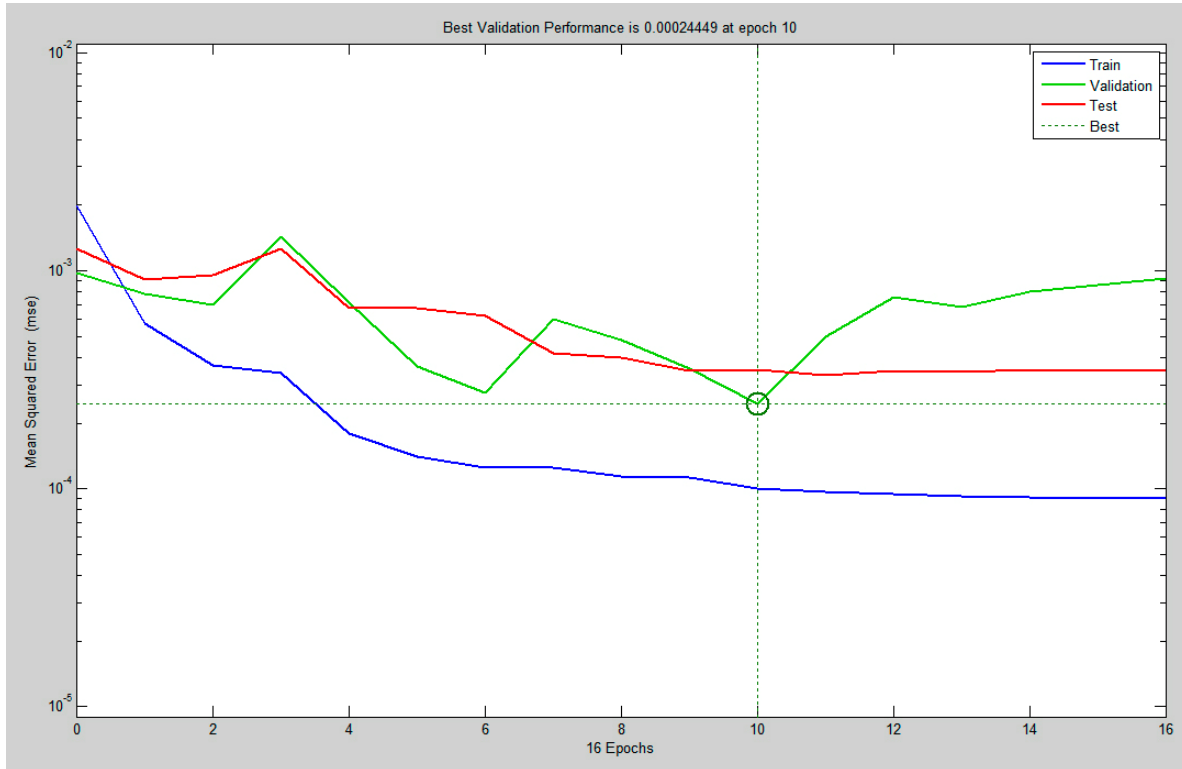
¹² Los modelos de series de tiempo son más apropiados para los pronósticos a corto plazo, mientras que para la planeación a mediano plazo y para el análisis estructural, es más conveniente usar datos de corte transversal. Aykaç y Borges (*op. cit.*).

¹³ La construcción de un modelo macroeconómico formal satisfactorio es una tarea extremadamente desafiante y costosa: el pronosticador enfrenta problemas teóricos y estadísticos de forma simultánea.

¹⁴ El número de nodos de la capa oculta se determinó a través de la expresión $h = \sqrt{n \times k}$; donde: n , h y k son el número de nodos en la capa de entrada, oculta y de salida, respectivamente.

observarse a través de la gráfica 2.9. Puede concluirse que no existe un sobre-ajuste de la RNA a los datos.

Gráfica 2.9
RNA Multivariada: validación cruzada



La gráfica 2.10 presenta la simulación histórica de \hat{Y} . Como puede observarse, la bondad de ajuste es satisfactoria (94.89%).¹⁵

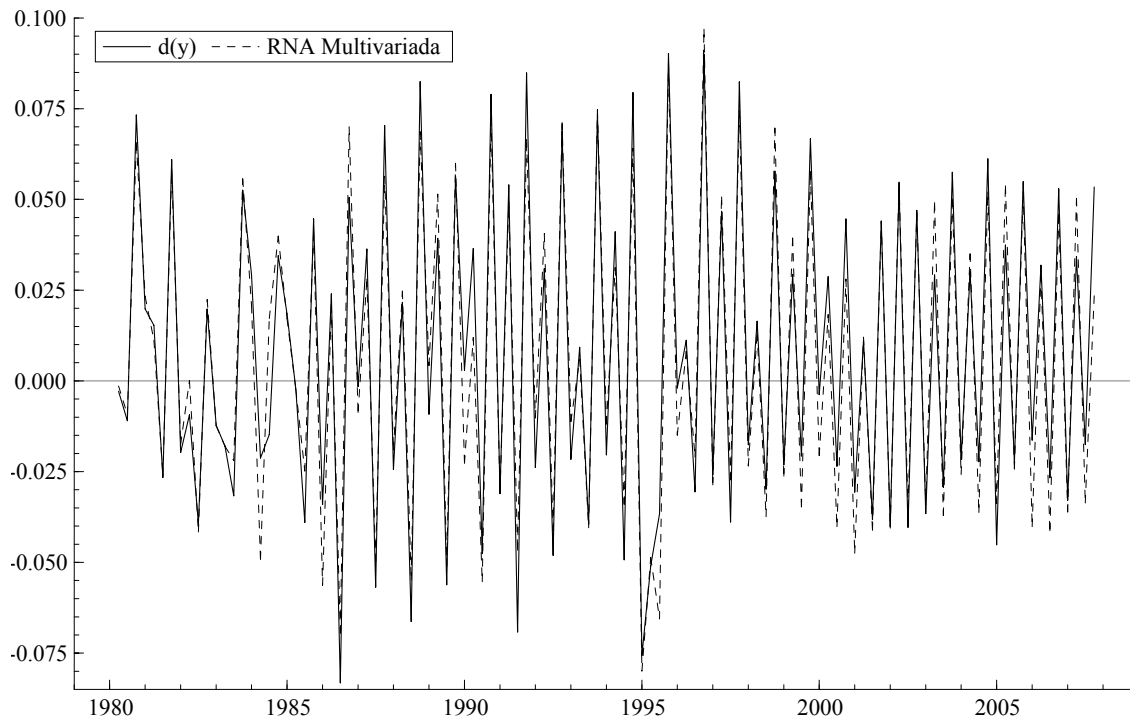
Por otro lado, el gráfico 2.11 muestra el error de estimación: puede observarse que no existen patrones sistemáticos, puntos aberrantes o cambios estructurales importantes que violen los supuestos básicos de un ruido Gaussiano. La información vertida en el cuadro 2.6 confirma que la serie es normal,¹⁶ no está autocorrelacionada y es homoscedástica (cuadro 2.7); además, las diferentes pruebas de raíces unitarias aplicadas corroboran que ε es un proceso estacionario. Por tanto, podemos concluir que la RNA Multivariada está correctamente especificada. Es importante rescatar el hecho de que, aunque Δi no tiene la

¹⁵ La gráfica 2.12 presenta un análisis de regresión lineal entre las señales de salida observadas y las calculadas a través de los conjuntos de datos de entrenamiento, de validación y de prueba.

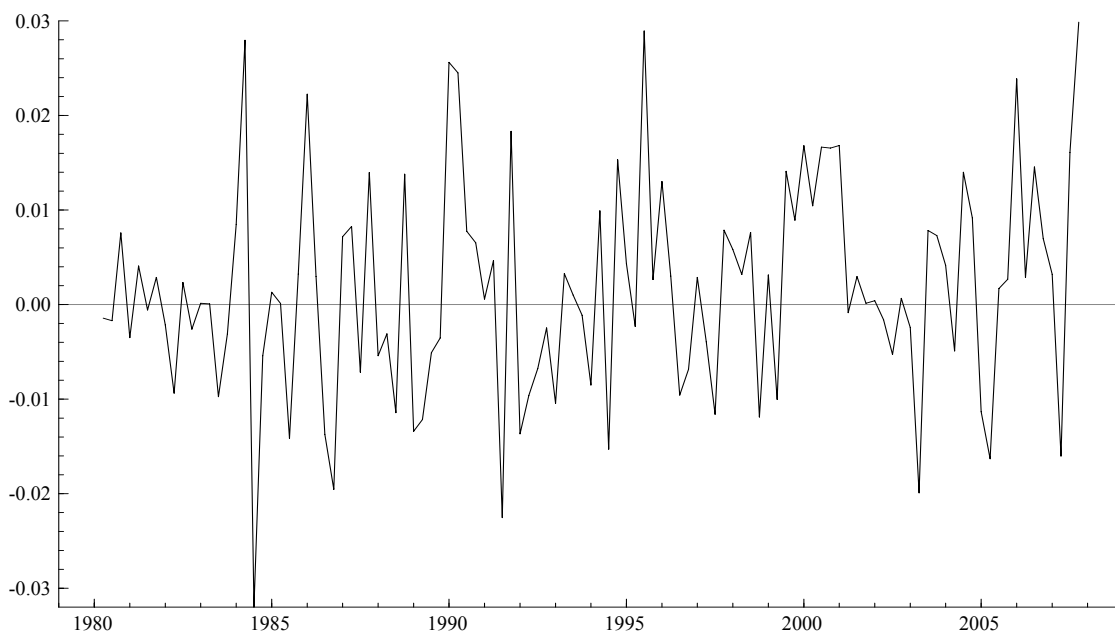
¹⁶ El estadístico Jarque-Bera (0.58) confirma que el proceso se distribuye normalmente y la prueba de hipótesis H_0 evidencia que su media es estadísticamente igual a cero.

misma longitud periódica que las demás variables, no fue necesario reducir el rango de estudio. Es decir, no hubo pérdida de información al estimar \hat{Y} .

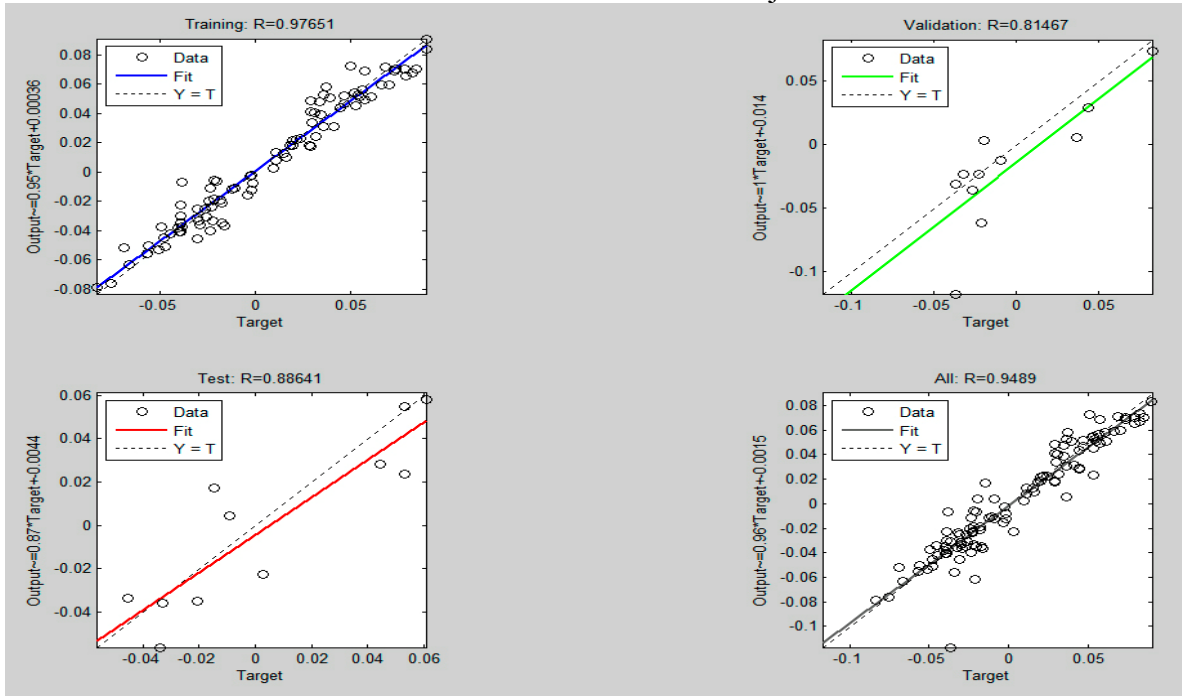
Gráfica 2.10
RNA Multivariada: simulación histórica, 1980.1-2007.4



Gráfica 2.11
RNA Multivariada: error de estimación, 1980.1-2007.4



Gráfica 2.12
RNA Multivariada: bondad de ajuste



Cuadro 2.6
Error de estimación: estadísticas básicas
y pruebas de raíces unitarias, 1980.1-2007.4

	ε
Mean	0.00
Median	0.00
Maximum	0.02
Minimum	-0.03
Std. Dev.	0.01
Skewness	0.13
Kurtosis	3.23
Jarque-Bera Probability	0.58 0.74
H_0 : Mean=0 Probability	1.40 0.16
Sum Sum Sq. Dev.	0.16 0.01
Observations	111
ADF	-9.57 (0) ¹
DF-GLS	-9.70 (0) ²
PP	-9.57 (0) ¹
KPSS	0.02 (0) ³

Notas: ¹ Sin constante ni tendencia; ² constante; ³ constante y tendencia; número de rezagos entre paréntesis; todas las pruebas son válidas al 99% de confianza.

Cuadro 2.7

Error de estimación: estadísticos de correcta especificación

Rezago	Q-stats ¹	p-values	Arch-stats ²	p-values
1	0.57	0.44	2.78	0.09
2	0.63	0.72	3.32	0.19
3	2.02	0.56	3.62	0.30
4	2.33	0.67	3.65	0.45
5	2.82	0.72	3.95	0.55
6	3.76	0.70	3.99	0.67
7	3.92	0.78	3.83	0.79
8	4.00	0.85	4.93	0.76
9	4.06	0.90	4.90	0.84
10	5.96	0.81	5.33	0.86

Notas: ¹ H₀: No autocorrelación, ² H₀: Homoscedasticidad

III. RESULTADOS

Seis honrados servidores me enseñaron cuanto sé; sus nombres son: cómo, cuándo, dónde, qué, quién y por qué.

Rudyard Kipling

Las diversas pruebas gráficas y estadísticas realizadas en el capítulo anterior (secciones 2.2.1 y 2.2.2), nos permitieron concluir que los errores de estimación tanto de la RNA Univariada como de la RNA Multivariada son normales, no están autocorrelacionados y son homoscedásticos; es decir, no existen patrones sistemáticos, puntos aberrantes o cambios estructurales importantes que violen los supuestos básicos de un ruido Gaussiano; por tanto, concluimos que los sistemas (2.3) y (2.4) están correctamente especificados. Con base en lo anterior, el siguiente paso es evaluar y comparar las arquitecturas neuronales.

3.1. Evaluación y comparación de modelos

Para evaluar y contrastar el desempeño de generalización de las arquitecturas neuronales se utilizan dos métodos complementarios: 1) el gráfico y 2) el estadístico. El primero consiste en analizar visualmente la simulación histórica de la señal de salida calculada y el segundo, en evaluar los diversos estadísticos descritos en la sección 1.4.4.

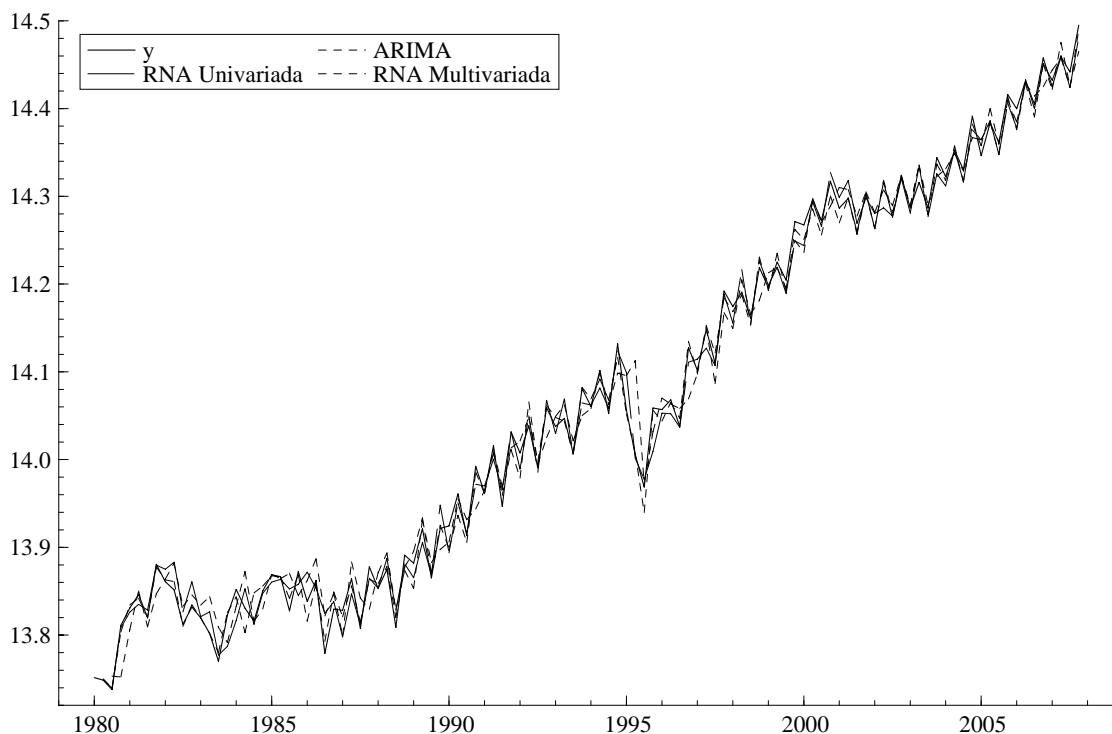
3.1.1. Pronósticos dentro de la muestra

Para valorar los pronósticos de un modelo, Pindyck *et al.* (1997) sugieren que, como primer paso, se evalúe gráficamente la simulación histórica. La gráfica 3.1 permite confirmar el óptimo desempeño de nuestras estimaciones; sin embargo, la información estadística vertida en el cuadro 3.1 permite realizar conclusiones mucho más robustas que el simple análisis visual.

En términos generales, los pronósticos dentro de la muestra son mucho más precisos con la RNA Multivariada que los de la RNA Univariada e incluso mejores que los arrojados por un modelo ARIMA:¹ el estadístico de Theil (en sus dos versiones) confirman esta aseveración. No obstante, en términos generales, ambas arquitecturas son igual de confiables.

¹ Por su simplicidad, los modelos propuestos por Box y Jenkins (1976) se consideran muy útiles para contrastar metodologías más sofisticadas. Los estadísticos de bondad de ajuste y de correcta especificación pueden ser consultados en el anexo.

Gráfica 3.1
 Pronósticos dentro de la muestra: simulación histórica, 1980.2-2007:4



Cuadro 3.1
 Pronósticos dentro de la muestra: series en niveles, 1980.2-2007:4

	ARIMA	RNA Univariada	RNA Multivariada
Mean Error	0.0004	0.0015	0.0015
Mean Absolute Error	0.0189	0.0129	0.0088
Root Mean Square Error	0.0246	0.0166	0.0115
Mean Square Error	0.0006	0.0002	0.0001
Theil's U	0.5613	0.3812	0.2636
Mean Pct Error	0.0000	0.0001	0.0001
Mean Abs Pct Error	0.0013	0.0009	0.0006
Root Mean Square Pct Error	0.0017	0.0011	0.0008
Theil's Relative U	0.5633	0.3821	0.2629

3.1.2. Pronósticos fuera de la muestra

La precisión es la prueba de los pronósticos; por tanto, las estimaciones fuera de la muestra nos permiten evaluar la capacidad de generalización de cualquier modelo, tanto matemático como econométrico. La información estadística vertida en el cuadro 3.2, nos permite concluir que los pronósticos fuera de la muestra de la RNA Univariada son más precisos que los de la RNA Multivariada o los del modelo ARIMA; sin embargo, este último tiene un desempeño relativamente mejor que la RNA Multivariada. Aykac y Borges (*op. cit.*) nos

explican: "en la medida en que las series de tiempo cambian lentamente y si no se espera que las otras variables que afectan el comportamiento de la variable que se está pronosticando cambien significativamente, es difícil que en el corto plazo un modelo multivariado supere en utilidad a los modelos univariados".

Sin embargo, debe tomarse en cuenta que una de las ventajas de la arquitectura neuronal multivariada sobre la univariada es que esta última no puede ser usada para realizar escenarios hipotéticos ("qué pasa si") y, por tanto, no pueden describir adecuadamente resultados alternativos; es decir, aun cuando es más exacto, el contenido de información de un pronóstico univariado es limitado, ya que no abarca las influencias futuras de otras variables; de esta forma, los pronósticos derivados de los modelos multivariados tienen mayor precisión a mediano y largo plazos.

Cuadro 3.2
Pronósticos fuera de la muestra: series en niveles, 2008.1-2008.3

	ARIMA	RNA Univariada	RNA Multivariada
Mean Error	-0.0091	-0.0078	-0.0141
Mean Absolute Error	0.0117	0.0078	0.0151
Root Mean Square Error	0.0133	0.0084	0.0188
Mean Square Error	0.0001	0.0000	0.0003
Theil's U	0.3333	0.2107	0.4727
Mean Pct Error	-0.0006	-0.0005	-0.0009
Mean Abs Pct Error	0.0008	0.0005	0.0010
Root Mean Square Pct Error	0.0009	0.0005	0.0013
Theil's Relative U	0.3331	0.2106	0.4728

3.2. Pronósticos del PIB, 2008.1-2008.3

Finalmente, el cuadro 3.3 presenta los pronósticos del PIB para el periodo 2008.1-2008.3, éstos deben entenderse como un ejercicio empírico de los fundamentos teóricos desarrollados a lo largo del primer y segundo capítulos; sin embargo, evidencian las posibilidades de implementar la metodología RNA para elaborar escenarios congruentes de pronóstico:

Cuadro 3.3
Pronósticos del PIB, 2008.1-2008.3

	Trimestre	ARIMA	RNA Univariada	RNA Multivariada	Observado
Valor	2008.1	1,910,494.04	1,910,189.67	1,884,301.79	1,887,229.71
	2008.2	1,948,998.22	1,967,597.51	1,985,899.89	1,956,873.59
	2008.3	1,919,096.49	1,894,159.90	1,938,471.18	1,882,761.24
Tasa de crecimiento trimestral	2008.1	-2.76	-3.24	-4.61	-4.45
	2008.2	2.00	4.17	5.10	3.62
	2008.3	-1.55	-3.26	-0.94	-3.86
Tasa de crecimiento anual	2008.1	1.89	2.06	2.73	2.57
	2008.2	2.52	2.22	2.63	2.67
	2008.3	4.32	1.03	1.18	0.65

CONCLUSIONES

En el presente trabajo se modeló y pronosticó el Producto Interno Bruto, 1980.1-2008.3, a través de la metodología de Redes Neuronales Artificiales. Para ello se definieron los conceptos básicos y la técnica fundamental de una RNA Multicapa Supervisada, así como las características principales de la *regla de aprendizaje* para la construcción de este tipo de arquitecturas: el *Algoritmo Backpropagation*, con el objetivo de proporcionar una guía teórica y empírica que pudiese ser entendida por quien cuente con una formación matemática y económica básica.

A través del modelado del PIB de México, bajo el consejo de Kennedy (*op. cit*) de "anteponer la nitidez de los datos a la divinidad econométrica", se pudo constatar las ventajas de una RNA (tanto univariada como multivariada): no fue necesario elaborar modelos *a priori*, hubo generalización satisfactoria del PGI, así como tolerancia a información incompleta o no lineal. De esta forma, como señala Zhang *et al.* (*op. cit*), las RNA's resultaron una eficaz herramienta para detectar y reproducir relaciones lineales y no lineales de nuestro vector de información. Es decir, bajo el hecho de que la realidad económica es compleja y caótica y no simple y ordenada, el uso de la metodología RNA para el modelado y pronóstico nos permitió relajar el supuesto de que los agentes conocen el modelo real de la economía: los factores económicos y extraeconómicos que gobiernan el comportamiento del PIB.

Si bien es cierto que la compleja forma funcional de un sistema *caja negra* como las RNA's hace difícil (pero no imposible) interpretar económicamente los parámetros estimados, su potencia no consiste necesariamente en la *elegancia* de una solución simple, sino en la *generalización* de una respuesta concreta a partir de información de la que se desconoce un modelo *a priori*. El objetivo del economista no es simplificar hasta sus últimas consecuencias la realidad, sino proporcionar información de mayor calidad, así como reducir la incertidumbre de los tomadores de decisiones, especialmente en el área de los pronósticos. Aunque el modelado y pronóstico de las arquitecturas desarrolladas a lo largo del segundo capítulo son satisfactorios, es importante tomar en cuenta las palabras de Jain y Mao (1996): "no hay evidencia conclusiva de su superioridad sobre las técnicas convencionales". Por tanto, el uso de esta técnica reside en el criterio del economista y de la naturaleza de la aplicación. De esta forma, como señala González (*op. cit.*): "la metodología

RNA debe ser vista como un complemento natural al análisis de regresión lineal". Debido a la creciente complejidad de los problemas a los que se enfrentan los investigadores, es preciso enriquecer los métodos y técnicas de análisis para el procesamiento de información y, de esta forma, obtener soluciones más rápidas y eficaces.

ANEXO

Cuadro A.1
Modelo ARIMA: especificación

Dependent Variable: D(Y)				
Method: Least Squares				
Sample (adjusted): 1981Q2 2007Q4				
Included observations: 107 after adjustments				
Convergence achieved after 24 iterations				
MA Backcast: 1980Q2 1981Q1				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D95	-0.087596	0.015488	-5.655665	0.0000
AR(1)	-0.036023	0.011266	-3.197571	0.0018
AR(4)	0.959332	0.010544	90.98772	0.0000
MA(2)	0.290548	0.072509	4.007060	0.0001
MA(4)	-0.674943	0.072060	-9.366406	0.0000
R-squared	0.861928	Mean dependent var		0.006208
Adjusted R-squared	0.856513	S.D. dependent var		0.043618
S.E. of regression	0.016523	Akaike info criterion		-5.322580
Sum squared resid	0.027845	Schwarz criterion		-5.197681
Log likelihood	289.7580	Hannan-Quinn criter.		-5.271948
Durbin-Watson stat	2.240677			
Inverted AR Roots	.98	-.01+.99i	-.01-.99i	-1.00
Inverted MA Roots	.83			

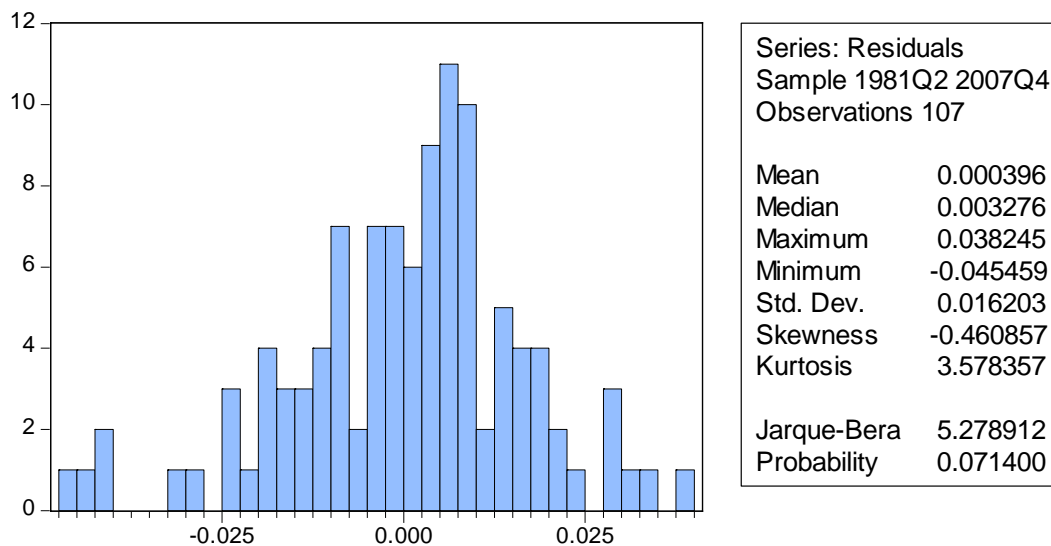
Cuadro A.2
Modelo ARIMA: raíces del sistema

Inverse Roots of AR/MA Polynomial(s)		
Specification: D(Y) AR(1) AR(4) MA(2) MA(4) D95		
Sample: 1980Q1 2007Q4		
Included observations: 107		
AR Root(s)	Modulus	Cycle
-0.998804	0.998804	
-0.009004 ± 0.989551i	0.989592	3.976963
0.980790	0.980790	
No root lies outside the unit circle. ARMA model is stationary.		
MA Root(s)	Modulus	Cycle
5.22e-18 ± 0.989731i	0.989731	4.000000
0.830073	0.830073	
-0.830073	0.830073	
No root lies outside the unit circle. ARMA model is invertible.		

Cuadro A.3
Modelo ARIMA: correlograma

Actual and ARMA Model Correlogram							
Specification: D(Y) AR(1) AR(4) MA(2) MA(4) D95							
Sample: 1980Q1 2007Q4							
Included observations: 107							
<u>Autocorrelation</u>			<u>Partial Autocorrelation</u>				
	Actual	Model	Difference	Actual	Model	Difference	
0	1.000	1.000	0.000	0	1.000	1.000	0.000
1	-0.756	-0.837	0.080	1	-0.756	-0.837	0.080
2	0.625	0.936	-0.311	2	0.124	0.785	-0.661
3	-0.739	-0.834	0.095	3	-0.549	-0.080	-0.469
4	0.866	0.932	-0.066	4	0.548	0.409	0.139
5	-0.711	-0.836	0.126	5	0.085	-0.056	0.142
6	0.565	0.928	-0.362	6	-0.249	0.239	-0.488
7	-0.695	-0.834	0.139	7	-0.138	-0.006	-0.133
8	0.842	0.924	-0.082	8	0.372	0.169	0.204
9	-0.696	-0.836	0.139	9	-0.026	-0.030	0.004
10	0.561	0.920	-0.359	10	-0.090	0.102	-0.192

Gráfica A.1
Error de estimación: estadísticas básicas



Cuadro A.4
Error de estimación: estadísticos de correcta especificación

Rezago	Q-stats ¹	p-values	Arch-stats ²	p-values
1	1.79	0.18	0.81	0.36
2	2.41	0.94	0.95	0.38
3	2.02	0.11	0.66	0.57
4	1.75	0.14	1.75	0.14
5	1.47	0.20	0.78	0.56

Notas: ¹ H₀: No autocorrelación, ² H₀: Homoscedasticidad

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*, Vo. 6.
2. Anderson, J. A., J. Silverstein, S. Ritz y R. Jones (1977). "Distinctive Features, Categorical Perception and Probability Learning: Some Applications on a Neural Model", *Psychological Review*, 84.
3. Armstrong, J. S. *Long-Range Forecasting: From Crystal Ball to Computer*. John Wiley, New York.
4. Aykac, A. y A. Borges (1989). "Métodos econométricos para aplicaciones gerenciales", en Makridakis, S. y S. Wheelwright, *Manual de técnicas de pronóstico*. Editorial Limusa.
5. Bliemel, F. (1973). "Theil's Forecast Accuracy Coefficient: A Clarification", *Journal of Marketing Research*, Vol. 10, No. 4.
6. Box, G. E. P. y G. M. Jenkins (1976.), *Time series analysis: Forecasting and control*. Holden Day, San Francisco.
7. Brunak, S. y B. Lautrup (1990). *Neural Networks*, World Scientific, Singapore.
8. Carpenter, G. y S. Grossberg (1986). "Absolutely Stable Learning of Recognition Codes by a Self-Organizing Neural Network", *American Institute of Physics Conference Proceedings 151: Neural Networks for Computing*.
9. Fildes, R. (1989). "Elaboración de pronósticos: principios", en Makridakis, S. y S. Wheelwright, *Manual de técnicas de pronóstico*. Editorial Limusa.
10. Flórez, R. y J. M. Fernández (2008). *Las Redes Neuronales Artificiales. Fundamentos teóricos y aplicaciones prácticas*. Serie Metodología y Análisis de Datos en Ciencias Sociales, Netbiblo.
11. Flowers, A. D. (1989). "Métodos de atenuación para la planeación y el control a corto plazo", en Makridakis, S. y S. Wheelwright, *Manual de técnicas de pronóstico*. Editorial Limusa.
12. Freeman, J. A. y D. M. Skapura (1993). *Redes neuronales. Algoritmos, aplicaciones y técnicas de programación*, Addison-Wesley/Diaz de Santos.
13. Funahashi, K. I. (1989). "On the approximate realization of continuous mappings by neural Networks", *Neural Networks*, 2.
14. Fukushima, K. (1980). "Neocognitron: a Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position", *Biological Cybernetics*, 36.
15. Granger, C. W. J. (1969). "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods", *Econometrica*, 37.
16. Granger, C. W. J. (1991). "Developments in the Nonlinear Analysis of Economic Series", *Scandinavian Journal of Economics*, 93, 2.
17. González, S. (2000). "Neural Networks for Macroeconomic Forecasting: A Complementary Approach to Linear Regression Models", *Finance Canada Working Paper*.
18. Hagan, M. T. y M. B. Menhaj (1994). "Training Feedforward Networks with Marquardt Algorithm", *IEEE Transactions on Neural Networks*, Vol. 5, No. 6.
19. Haykin, S. (1999). *Neural Networks. A comprehensive foundation*. Second Edition, Prentice Hall.

20. Hebb, D. O. (1949). *The Organization of Behavior*, John Wiley & Sons, New York.
21. Hinton, G. E., D. H. Ackley y J. Sejnowski (1984). *Boltzmann Machines: Constraint Satisfaction Networks that Learn*, Technical Report CMU-CS-84-119, Carnegie-Mellon University, Department of Computer Science.
22. Hopfield, J. J. (1982). "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", *Proceedings of the National Academy of Sciences*, 79.
23. Jain, A. K. y J. Mao (1996). "Artificial Neural Networks: A Tutorial", *IEEE*.
24. Juselius, K. (2006). *The Cointegrated VAR Model. Methodology and Applications*", Advance Texts in Econometrics, Oxford University Press.
25. Kimball, A. W. (1957). "Errors of the third kind in statistical consulting", *Journal of the American Statistical Association*, 52.
26. Kennedy, P. E. (2002), "Sinning in the Basement: What are the Rules? The Ten Commandments of Applied Econometrics," *Journal of Economic Surveys*, Vol. 16
27. Kohonen, T. (1990). "The Self-Organizing Map", *Proceedings of the IEEE*, 78:9.
28. Kohonen, T. (1997). *Self Organizing Maps*, Springer Verlag, Berlín.
29. Lapedes, A. S. y R. M. Faber (1987). "Nonlinear signal processing using neural networks: Prediction and system modeling", *Technical Report LA-UR-87-2662*.
30. Levenberg, K. (1944). "A Method for the Solution of Certain Non-linear Problems in Least Squares". *Quarterly of Applied Mathematics*, 2(2):164–168, Jul.
31. Lucas, R. E. (1976). "Econometric Policy Evaluation: A Critique", *Carnegie Rochester Conference Series on Public Policy*, 1.
32. Loría, E. (2007). *Econometría con aplicaciones*, Editorial Pearson Educación, México.
33. Makridakis, S., S. Wheelwright y R. J. Hyndman (1998). *Forecasting: Methods and Applications*. John Wiley, New York.
34. Marquardt, D.W. (1963). "An Algorithm for the Least-Squares Estimation of Nonlinear Parameters". *SIAM Journal of Applied Mathematics*, 11(2):431–441, Jun.
35. McCulloch, W. S. y W. Pitts (1943). "A Logical Calculus of Ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biophysics*, Vol. 5.
36. McMenamin, J. S. (1996). "Why Not Pi? A Primer on Neural Networks for Forecasting", *Technical White Papers*, Itron.
37. McNelis, P. D. (2005). *Neural Networks in Finance. Gaining Predictive Edge in the Market*. Academiv Press Advanced Finance Series, Elsevier Academic Press.
38. Menger, K. (1960). "A counterpart of Occam's razor in pure and applied mathematics ontological uses", *Synthese*, Volume 12, Number 4
39. Minsky, M., (1954). *Neural Nets and the Brain-Model Problem*, PhD Dissertation, University of Princeton.
40. Minsky, M. y S. Papert (1969). *Perceptrons: An Introduction to Computational Geometry*, MIT Press, Cambridge.
41. Pyndick, R. S., D. L. Rubinfeld, B. H. Hall y S. L. Schmukler (1997). *Handbook to Accompany Econometric Models and Economic Forecasts*. McGraw-Hill.
42. Rodríguez-Fernández, J. L. (1999). "Ockham's Razor", *Endeavour*, 23.
43. Rosenblatt, R (1958). "The Perceptron. A Probabilistic Model for Information Storage and Organization in the Brain", *Psychological Review*, 65.
44. Rumelhart, D. E., G. E. Hinton y R. K. Williams (1986). "Learning Representations by Backpropagation Errors", *Nature*, 323.

45. Sims, C. A. (1980). "Macroeconomics and reality", *Econometrica*, Vol. 48, 1.
46. Trippi, R. R. y E. Turban (1993). *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real-World Performance*, Probus.
47. Tukey, J. W. (1962). "Statistical methods in psychology journals", *American Psychologist*, 54.
48. Werbos, P. (1974). "Beyond Regression: New Tools for Prediction and Analysis in the Behavior Sciences", PhD thesis, Dept. of Applied Mathematics, Harvard University, Cambridge.
49. Widrow, B. y M. E. Hoff (1960). *Adaptive Switching Circuits*, IRE WESCON Convention Record, 4.
50. Zhang, G., B. E. Patuwo y M. Y. Hu (1997). "Forecasting with Artificial Neural Networks: the State of the Art", *International Journal of Forecasting*, 14.