

# REDES NEURONALES EN LA PREDICCIÓN DE LAS FLUCTUACIONES DE LA ECONOMÍA A PARTIR DEL MOVIMIENTO DE LOS MERCADOS DE CAPITALES\*

*Loren Trigo y Sabatino Costanzo\*\**

## RESUMEN

Este estudio analiza la capacidad de las redes neuronales para predecir la dirección de las economías de los Estados Unidos y México, con los índices rezagados de los mercados de capitales de cada país como insumos y el índice compuesto de indicadores adelantados de cada país (LEI, aquí tratado como índice coincidente) como salida resultante. La capacidad predictiva estable y significativa de las redes neuronales utilizadas fue establecida y su superioridad predictiva respecto a la de una regresión múltiple comparable fue medida con el método estadístico de medición de la precisión predictiva elaborado por Anatolyev y Gerko.

## ABSTRACT

This study analyzes the ability of neuronal nets to predict the direction of the U. S. and Mexican economies, with the lagged indexes of the capital markets of each country as inputs and the compound leading economic indicator index (LEI, here treated as a coincident indicator) as output. Neuronal net significant and stable predictive ability was established and the predictive superiority of the net with respect to a comparable multiple regression was measured using the predictive accuracy test designed by Anatolyev and Gerko.

## INTRODUCCIÓN

La drástica reducción de tiempo de computación requerido para generar y procesar modelos no lineales basados en inteligencia artificial ha estimulado su aplicación a temas macroeconómicos que antes

\* *Palabras clave:* redes neuronales, predicción, fluctuaciones, economía, mercados de capitales. *Clasificación JEL:* G. Artículo recibido el 7 de marzo de 2005 y aceptado el 29 de junio de 2006.

\*\* L. Trigo, Centro de Producción e Innovación Tecnológica del Instituto de Estudios Superiores de Administración (IESA), Caracas (correo electrónico: loren.trigo@iesa.edu.ve). S. Costanzo, Centro de Finanzas del Instituto de Estudios Superiores de Administración (IESA), Caracas (correo electrónico: sabatino.costanzo@iesa.edu.ve).

eran considerados complejos. Parte de este impulso se debe a que hoy en día se duda que los modelos lineales puedan predecir los ciclos de negocios. Varios estudios (Kaiser y Maraval, 1999; Verbrugge, 1997, y Kim *et al*, 1996, entre otros) han establecido que los ciclos de negocio son asimétricos. Otros estudios recientes (Granger y Terasvirta, 1993; Diebold y Rudebusch, 1993; Swansen y White, 1995, y Jaditz *et al*, 1998) han demostrado que los modelos de predicción macroeconómica mejoran cuando incluyen componentes no lineales. Con base en el ejemplo de Natter *et al* (1994), Moody *et al* (1993), White (1996), Tkacz (2000) y Jagric (2003), quienes han creado modelos de predicción macroeconómica basados en redes neuronales que se desempeñan mejor que los modelos lineales, el objetivo de este estudio es reconfirmar la validez de los modelos no lineales con base en redes neuronales para la proyección macroeconómica, mas de tal manera que sean de utilidad para la comunidad de inversionistas potenciales. En este estudio utilizamos la medida estadística de la precisión en el acierto direccional y magnitudinal establecida por Anatolyev y Gerko (2005).

El método más conocido para predecir los ciclos de negocios es el índice LEI (*Leading Economic Indicators*) o índice compuesto —de índole lineal—, que ha predicho cada una de las pasadas ocho recesiones en los Estados Unidos. Este índice, publicado mensualmente por la Conference Board para los Estados Unidos y México entre otros, se compone de 10 índices económicos: *i*) horas promedio trabajadas a la semana por obreros de producción en industrias manufactureras (.1946); *ii*) número promedio de reclamos de seguro de desempleo (.0268); *iii*) nuevas órdenes de los manufactureros para bienes de consumo (.0504); *iv*) velocidad relativa en que vendedores pueden entregar pedidos a compañías industriales, según el índice de la National Association of Purchasing Managers (.0296); *v*) nuevas órdenes recibidas por manufactureros en bienes capitales no relacionados con la defensa (.0139); *vi*) número de permisos concedidos para edificios residenciales; (.0205); *vii*) cambio en el mercado de acciones, según el precio de las acciones en el índice Standard & Poor's 500 (.0309); *viii*) provisión monetaria M2 (ajustada por inflación) (.2775); *ix*) la dispersión del rendimiento, esto es, la diferencia entre la tasa de interés a largo plazo (bonos soberanos a diez

años) y la de corto plazo (el interés libre de riesgo o notas del tesoro) (.3364), y  $x$ ) el índice de expectativas del consumidor (.0193).

Aunque la Conference Board le asigna al cambio en el mercado de acciones un factor de estandarización de sólo .0309, según el masivo compendio publicado en 1961 por Geoffrey Moore del National Bureau of Economic Research, los precios accionarios fueron clasificados como indicadores adelantados 31 veces, como coincidentes 14 veces y como rezagados sólo cinco veces. Renshaw (1995) muestra que aun cuando el mercado de acciones no predice bien las recesiones económicas, sí es un excelente predictor de las recuperaciones. Respecto al *yield spread* que recibe de la Conference Board el segundo factor más alto de estandarización, Estrella y Mishkin (1997, 1998) y Dotsey (1998), entre otros, documentan la efectividad de la dispersión de los bonos soberanos como predictor de los ciclos de negocios, en particular de las recesiones.

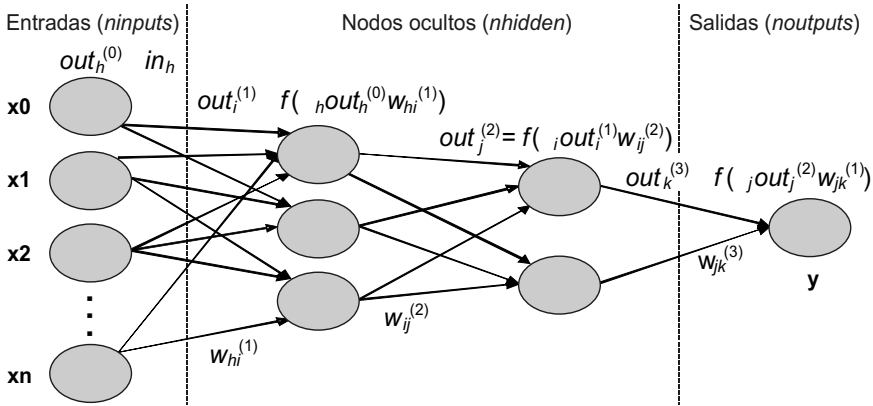
Es muy conocido que, según la percepción de muchos operadores financieros exitosos, el rendimiento histórico de los mercados de acciones y bonos puede ser utilizado para predecir el rendimiento futuro de los mismos. Se puede ir aún más allá al afirmar que dicho rendimiento histórico es también utilizado para predecir el movimiento futuro de otros índices económicos relevantes que acompañan a estos dos mercados en la composición del LEI. Esto induce a preguntarse por qué y en qué medida en los Estados Unidos esos cambios históricos en los mercados de acciones y de bonos soberanos se corresponden con —y permiten predecir— los cambios en las estadísticas que influyen directamente en el crecimiento o la contracción económicas, como son los índices de ventas, sueldos, ingresos por impuestos y gastos de capital. Para explicar este efecto se puede recurrir a dos argumentos: *i*) que los mercados de capitales (*equity & debt*) son la principal fuente de financiación de la economía y en esta capacidad actúan como estimulantes; *ii*) recurre a una metáfora de las finanzas comportamentalistas según la cual en las economías con mercados de capitales maduros un alza del mercado de valores hace “sentirse próspero” al inversionista, estimulándolo a aumentar su consumo y su inversión. Ruggiero (2001) arguye que aun cuando ha habido cierto desfase entre el mercado de acciones y el de bonos soberanos debido al inusualmente largo mercado alzista

de los años noventa en los Estados Unidos, tanto uno como el otro están positivamente correlacionados con el LEI estadounidense a largo plazo, y documenta una correlación de 0.6089 entre este LEI y un simple oscilador del índice S&P500 durante el intervalo 1963-2001. Situemos pues nuestro estudio dentro de esta polémica, del lado de los que arguyen que las expectativas de la economía no estimulan los mercados de capitales sino que la relación apunta más bien hacia la dirección opuesta. Además, a fin de comprobar si la metáfora comportamentalista tiene valor general, hemos escogido proyectar el LEI de dos países cuyos mercados de capitales tienen distintos grados de madurez y participación: México y los Estados Unidos.

### I. REDES NEURONALES

Antes de abordar el tema de la metodología y los datos utilizados en este estudio, cabe aquí dar una idea general de lo que son las redes neuronales y, en particular, de las características utilizadas en este estudio así como de su entrenamiento. Centramos nuestro análisis en el empleo de las redes neuronales multicapas con aprendizaje supervisado ya que (como veremos líneas abajo) son capaces de ajustarse a cualquier función y se desempeñan particularmente bien con las series de tiempo.

Al igual que otros modelos de proyección, como las regresiones y las extrapolaciones (promedios móviles, suavizamientos), los modelos basados en redes neuronales utilizan entradas para generar un resultado que es una proyección. Sin embargo, lo que distingue a los modelos neuronales de otros modelos es su capacidad de aprender y adaptarse al entorno. Las redes neuronales se componen de tres elementos fundamentales: *i*) unidades de procesamiento o nodos capaces de operar de manera paralela; *ii*) funciones de transferencia (o funciones de activación) que transforman la información en los nodos, y *iii*) pesos de conexión que determinan la importancia relativa entre nodos. Estos tres elementos se utilizan para construir una red que puede ser de una o más capas. En el caso de las redes multicapas de alimentación frontal (MLP o *Multi Layer Perceptron*), las salidas de una capa constituyen la entrada a la próxima capa, tal y como se muestra en la gráfica 1a. Al hablar de una red de  $N$  capas, la capa

GRÁFICA 1a. *Red de tres capas*

de entrada es la capa 0 y la de salida es la capa  $N$ . La(s) capa(s) intermedia(s) se llama(n) —por consenso— capa(s) oculta(s).

Las redes neuronales aprenden y se adaptan modificando los pesos asociados a las conexiones entre los nodos. Los factores comprendidos en el aprendizaje realizado por redes multicapas son los siguientes:

- i)* El propósito del entrenamiento o aprendizaje por la red neuronal es minimizar los errores de salida en un conjunto particular de datos de entrenamiento haciendo una serie de actualizaciones de los pesos conectores  $w_{ij}$ . Las reglas que rigen la actualización de los pesos constituyen el algoritmo de aprendizaje de la red.
- ii)* Definimos una función error  $E(w_{ij})$  que “mide” cuán lejos está la red actual de la red deseada (correctamente entrenada).
- iii)* Derivadas parciales de la función error  $E(w_{ij})/w_{ij}$  nos dicen en qué dirección necesitamos movernos en el espacio de pesos para reducir el error. Este componente del algoritmo de aprendizaje se llama gradiente descendiente.
- iv)* La tasa de aprendizaje especifica los tamaños de los pasos que tomamos en el espacio de pesos para cada iteración de la ecuación de gradiente descendiente que actualiza los pesos.
- v)* Sólo las salidas de la capa final aparecen en la función error. Sin embargo, este error dependerá de todas las capas de pesos ante-

riores y un componente del algoritmo de aprendizaje llamado propagación en reversa los ajustará todos. La propagación en reversa automáticamente ajusta las salidas de las capas ocultas anteriores de tal manera que las capas formen representaciones intermedias (ocultas) apropiadas.

- vi) Continuamos paso a paso a través del espacio de pesos hasta que los errores son “suficientemente pequeños”.
- vii) Si escogemos funciones de activación (transferencia) neuronales con derivadas que asumen formas particularmente sencillas, podemos hacer que los cálculos para la actualización de los pesos sean muy eficientes.

## II. ENTRENAMIENTO DE UNA RED MULTICAPAS (MLP)

i) Tomamos el conjunto de pautas de entrenamiento que deseamos que la red aprenda  $\{in_i^p, out_j^p: i = 1 \dots ninputs, j = 1 \dots noutputs, p = 1 \dots npatterns\}$ . ii) Configuramos la red con unidades de entrada (*ninputs*) completamente conectados a unidades ocultas (*nhidden*) a través de pesos conectores  $w_{ij}$ , que a su vez están conectados completamente a unidades de salida (*noutputs*) a través de conexiones con pesos  $w_{jk}$ . iii) Generamos pesos iniciales aleatorios, por ejemplo del rango  $[-smwt, smwt]$ . iv) Seleccionamos una función error  $E(w_{jk})$  y tasa de aprendizaje apropiadas. v) Aplicamos la actualización de pesos  $w_{ij} = w_{ij} - \eta \frac{\partial E}{\partial w_{ij}}$  para cada peso  $w_{ij}$  para cada pauta de entrenamiento  $p$ . Un conjunto de actualizaciones de todos los pesos para todas las pautas de entrenamiento se llama una época de entrenamiento. vi) Repetimos el paso v) hasta que la función error de la red es “suficientemente pequeña”.

Entonces tenemos una red neuronal entrenada. Como se deduce de esto, el entrenamiento de una red implica la definición de un número significativo de parámetros lo cual hace posible la sobreparametrización del modelo o sobreajuste del modelo a la data. Existe una técnica para optimar el entrenamiento de las redes neuronales evitando el sobreajuste que presentamos líneas abajo, pero para comprenderla hay que entender cómo surge este sobreajuste en el caso de las redes neuronales. Esto es lo que intentamos esbozar a continuación.

### III. PODER COMPUTACIONAL DE LAS REDES MULTICAPAS

El teorema de la aproximación universal para un MLP probado independientemente por Cybenko (1989) y Hornik *et al* (1989) dice que cualquier función continua capaz de mapear intervalos de números reales a algún intervalo-salida (*output*) de números reales puede ser aproximada arbitrariamente cercana por un MLP con sencillamente una capa oculta. Este resultado es cierto sólo para clases restringidas de funciones de activación o transferencia, como lo son las funciones sigmoideas o de distribución logística. En otras palabras, sea  $(\times)$  una función monotónica-incrementante continua no constante y acotada. Entonces para cualquier función continua  $f(x)$  con  $x = \{x_i \in [0,1]: i = 1, \dots, m\}$  y  $0 < b_j < 1$ , existe un entero  $M$  y constantes reales  $\{w_{jk}, b_j, w_{jk}: j = 1, \dots, M, k = 1, \dots, m\}$  tal que

$$F(x_1, \dots, x_m) = \sum_{j=1}^M \sum_{k=1}^m w_{jk} x_k + b_j$$

es una realización aproximada de  $f(\times)$ , esto es,  $|F(x_1, \dots, x_m) - f(x_1, \dots, x_m)| < \epsilon$  para todo  $x$  que yace en el espacio de entrada. Claramente esto aplica a un MLP con  $M$  unidades ocultas ya que  $(\times)$  puede ser un sigmoide,  $w_{jk}, b_j$  pueden ser capas ocultas y sesgos, y los  $w_{jk}$  pueden ser pesos de salida. De aquí sigue que, dado un número suficiente de unidades ocultas, un MLP de dos capas puede aproximar cualquier función continua. Para la interpretación estadística de las redes neuronales referimos al lector a Bishop (1995).

### IV. METODOLOGÍA Y DATOS

Explicado en qué consiste una red neuronal y lo que es capaz de aprender, queda por analizar la metodología utilizada por los autores de este trabajo para construirla y optimar su entrenamiento. Hemos visto que la construcción y optimación de una red neuronal implica definir un número considerable de parámetros, como número de nodos, número de capas, tipo de función de transferencia, algoritmo de aprendizaje, tasa de aprendizaje, manejo del error, pesos iniciales, etc. Cada combinación posible de estos parámetros determina una red distinta, lo cual nos plantea el problema combinatorio de cómo hallar la red óptima.

Puesto que este problema no ha recibido solución analítica, debemos recurrir a la sistematización de un método de prueba y error. Por esta razón decidimos optimar las redes por medio de dos programas complementarios: *i*) un algoritmo de búsqueda del programa de redes neuronales (Braincel) capaz de generar y evaluar diversas arquitecturas de redes, con base en las diferentes combinaciones de parámetros posibles, y *ii*) un programa elaborado por los autores capaz de optimar el entrenamiento de las redes candidatas obtenidas con *i*).

El algoritmo de búsqueda de Braincel requiere que el cuerpo de datos se divida en tres conjuntos: *i*) de entrenamiento, *ii*) de entrenamiento con prueba y *iii*) de prueba o validación aquí llamado conjunto extramuestral. El primer conjunto (entrenamiento) aporta los datos para que la red “aprenda” o se entrene en la búsqueda de pautas comparando sus predicciones con las históricas. El segundo conjunto (entrenamiento y prueba) sirve para supervisar el entrenamiento de la red con el objetivo de eliminar el sobreajuste. El método utilizado es la interrupción temprana del entrenamiento y el suavizamiento de los datos, como se explica líneas abajo.

Para los algoritmos iterativos de gradiente descendente (como propagación reversa en lotes y gradientes conjugados), el error de la red disminuirá según vaya aumentando el número de épocas de entrenamiento. Por esto, al inicio, el error en la data de entrenamiento y prueba disminuirá según se vaya reduciendo el subajuste, pero finalmente, este error comenzará a aumentar de nuevo según ocurre el sobreajuste. La solución natural para obtener la mejor generalización, es decir el error más bajo en el conjunto de entrenamiento y prueba, es la utilización del método de interrupción temprana del entrenamiento. De manera sencilla, una red se debe entrenar hasta que el error asociado a la data de entrenamiento y prueba comience a subir de nuevo, para entonces interrumpir el entrenamiento. En este punto se espera que el error de generalización también comience a subir. El tercer conjunto (extramuestral) se compone de datos completamente nuevos para la red que permiten evaluar su desempeño. En general se recomienda utilizar 60% de la data como conjunto de entrenamiento, 30% como conjunto de entrenamiento con prueba y 10% como conjunto extramuestral.



En este estudio los conjuntos de datos tienen las siguientes fechas: 7 de octubre de 1986-7 de octubre de 2001 (entrenamiento y entrenamiento con prueba Estados Unidos); 8 de octubre de 2001-1 de octubre de 2004 (extramuestral Estados Unidos); 5 de noviembre de 1995-11 de noviembre de 1998 (entrenamiento y entrenamiento con prueba México), y 12 de noviembre de 1998-11 de noviembre de 2001 (extramuestral México).

Cabe señalar que las redes construidas con Braincel utilizan como algoritmo de aprendizaje de red la llamada percolación reversa (*Back-Percolation*) —que es una modificación del algoritmo de propagación reversa—, establecida por Mark Jurik en 1984. El algoritmo de propagación reversa, ampliamente utilizado, tiene la desventaja respecto al de percolación reversa de que en el primero el error que posee el nodo de salida no se puede comunicar a todas las capas precedentes. Esto causa que la precisión de la red disminuya según se incrementa el número de nodos y capas. En contraste, la percolación reversa permite que el error se propague del nodo de salida hacia todos los demás nodos y capas precedentes y que el número de nodos y capas pueda aumentar sin empeorar el desempeño de la totalidad (Jurik, 1994, p. 85). Es notorio que ambos algoritmos de aprendizaje son equivalentes en el caso de redes cuya arquitectura consiste de una única capa oculta.

Es importante destacar que en el caso de este estudio el algoritmo de búsqueda de Braincel arrojó congruentemente que las redes asociadas a los mejores resultados son redes con una sola capa oculta, una capa de entrada y una capa (consistente en un nodo) de salida. Las características de las redes usadas para proyectar se presentan en el cuadro 1. Si tomamos como entradas de la red a varios cortes temporales ( $k$  trozos) de los datos de la serie  $x(t)$ , es decir,  $t, t-1, t-2, \dots, t-k+1$ , y como salida la predicción para  $x(t+1)$ , entonces podemos reformular nuestro problema predictivo como el de hallar el mapa de entradas-salidas (*input-output*) óptimo (Weigend y Gershenfeld, 1994). Sin embargo, las redes desarrolladas en este estudio, aunque similares en la utilización de cortes temporales, son diferentes en otros aspectos.

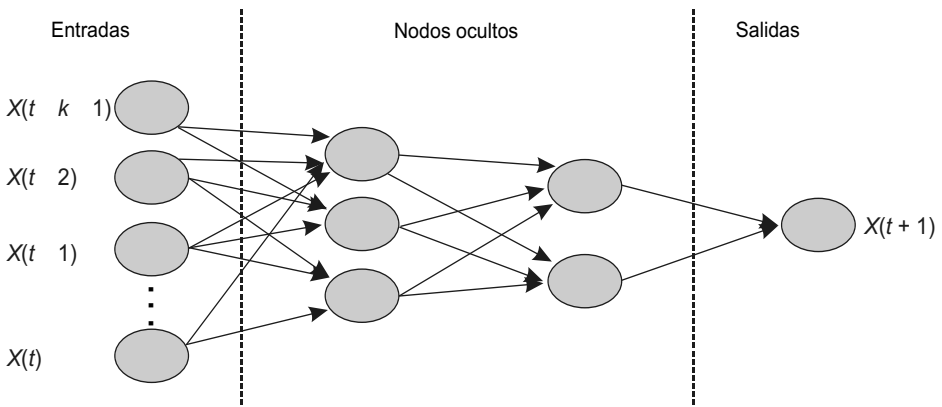
La variable de salida de las redes que presentamos aquí está dada por la variación del índice adelantado macroeconómico de un país

CUADRO 1. *Arquitectura de las redes neuronales utilizadas para proyectar las variaciones del LEI*

<i>Salida</i>	<i>Capa 0</i>	<i>Capa 1 (oculta)</i>	<i>Capa 2</i>
<b>Funciones de transferencia</b>			
LEI Estados Unidos	Logística	Logística	Lineal
LEI México	Logística	Logística	Lineal
	<i>Capa 0 + bias</i>	<i>Capa 1 (oculta) + bias</i>	<i>Capa 2</i>
<b>Número de nodos</b>			
LEI Estados Unidos	8	4	1
LEI México	8	4	1
	<i>Aprendizaje</i>	<i>Intervalo de pesos iniciales</i>	<i>Error de entrenamiento</i>
<b>Parámetros de aprendizaje</b>			
LEI Estados Unidos	0.076	0.4	4.75
LEI México	0.187	0.4	4.82

(el LEI en los Estados Unidos y el LEI en México, ambos publicados por la Conference Board) correspondiente al periodo  $t$ . En contraste con el PIB (producto interno bruto), el cual suele publicarse anualmente, los índices adelantados que intentamos proyectar se publican mensualmente y proporcionan un número suficiente de observaciones para el entrenamiento y prueba de las redes. El LEI recibió en este estudio tratamiento de indicador coincidente —y no adelantado— de la economía. Para esto fue necesario rezagar las entradas respecto al LEI como se describe líneas abajo.

GRÁFICA 1b.



Desde el punto de vista heurístico, para que una red neuronal sea entrenada óptimamente conviene tener alrededor de 10 veces más observaciones que el número de pesos conectores que tenga la red. Como el número de entradas determina el número de pesos conectores, se debe reducir el número de entradas al permitido por el número de observaciones disponible. Por ejemplo, si una red tiene dos entradas y una salida, requerirá óptimamente  $(2 + 1) * 10 * 10 = 300$  observaciones, que dividido entre 12 meses es igual a 25 años de data. Dado el número de observaciones disponibles para este estudio decidimos incluir como entradas a las redes dos de los diez índices componentes del LEI, esto es, para cada país un índice del mercado accionario y un índice del mercado de bonos soberanos correspondientes. Los índices utilizados como entradas fueron para Estados Unidos: *i*) S&P 500 Composite-Price Index (dólares) y *ii*) JPM United States Govt. Bond (dólares)-Price Index; para México: *i*) MSCI Mexico Former Dead-Price Index (pesos) y *ii*) JPM ELMÍ México (L) Dead-Return Ind (OFCL).

En principio podemos usar cualquier conjunto simple de datos para entrenar nuestras redes. Sin embargo, en la práctica suele ser de gran importancia realizar un preprocesamiento apropiado de la data antes de alimentar con ella a la red. Dicho preprocesamiento, de ser adecuado, ayuda a la red a aprender mejor. En el caso del presente trabajo, cada variable de entrada de la red fue optimada (preprocesada) antes de alimentar a la red con ella. El método de optimación consistió en que cada variable de entrada fue utilizada para generar 24 variables de entrada consistentes en rezagos sucesivos de la variable de entrada original, cada una más rezagada que la próxima desde ( $X - 1$  mes), ( $X - 2$  meses), ( $X - 3$  meses)... hasta ( $X - 24$  meses). Cada variable de entrada rezagada se utilizó aisladamente para proyectar el LEI correspondiente mediante el siguiente procedimiento, que coincide con la estrategia de compraventa descrita por Anatolyev y Gerko (2005) (véase líneas abajo): *i*) se supone que el LEI es comerciable con precio igual al valor numérico del índice; *ii*) si la entrada y el LEI coinciden en la dirección del cambio (si ambos suben o ambos bajan), se suma al “patrimonio” o *equity* del agente una “ganancia” igual al valor absoluto de la diferencia entre el “precio<sub>*t*</sub>” y el “precio<sub>*t-1*</sub>” del LEI, de lo contrario se resta esa

CUADRO 2. *Valores del índice de EP*

<i>Rezago</i>	<i>EP Anatolyev</i>	<i>Rezago</i>	<i>EP Anatolyev</i>
Lag_2	0.71339569	Lag_14	0.95337281
Lag_3	0.28314134	Lag_15	2.74800468
Lag_4	1.53075007	Lag_16	0.46402437
Lag_5	1.80635256	Lag_17	0.09574475
Lag_6	2.13690593	Lag_18	0.43578441
Lag_7	1.3699712	Lag_19	0.01960145
Lag_8	0.66849162	Lag_20	2.25037464
Lag_9	1.35038554	Lag_21	2.04505793
Lag_10	0.66499122	Lag_22	1.2619383
Lag_11	0.42872635	Lag_23	0.90530506
Lag_12	0.68200151	Lag_24	0.39314648
Lag_13	1.38125953		

misma cantidad como “pérdida”; *iii*) esto es, se construyeron 24 sistemas de compraventa del LEI, cada uno con una entrada rezagada distinta; *iv*) se calculó el cociente estadístico EP de estos 24 sistemas de compraventa, esto es, “Excess Profitability” o rentabilidad excedente (y la probabilidad de predicción asociada) de Anatolyev y Gerko (2005) (véase líneas abajo); *v*) los sistemas de compraventa se jerarquizaron con base en los resultados obtenidos; por ejemplo, para escoger el rezago apropiado del índice de acciones de México como predictor de la economía, se obtuvo una tabla de valores del índice de EP, gracias a la cual se escogió el rezago 20 como óptimo (véase cuadro 2).

Los resultados de la optimización de los rezagos llevaron a la siguiente selección: para los Estados Unidos, *i*) S&P 500 Composite-Price Index (dólares), rezago 7 meses y *ii*) JPM United States Govt. Bond (dólares)-Price Index, rezago 4 meses; para México: *i*) MSCI México Former Dead -Price Index (pesos), rezago 20 meses y *ii*) JPM ELMI México (L) Dead -Return Ind (OFCL), rezago 15 meses.

Cada una de las entradas de rezago seleccionadas se utilizó para construir tres promedios móviles que también fueron usados como entradas de la red. Estos tres promedios móviles se construyeron introduciendo otros tres valores de  $\alpha$  (0.1, 0.3 y 0.5) en la siguiente fórmula de suavizamiento: promedio móvil,  $M_t = \alpha x_t + (1 - \alpha) * M_{t-1}$ . Los valores escogidos para  $\alpha$  permiten considerar los tres casos posibles: en el cuerpo de datos, *i*) las últimas observaciones tie-

nen importancia preponderante, *ii*) la última fracción considerable tiene importancia preponderante y *iii*) casi todos los datos son de importancia primordial. En total, cada red fue alimentada con ocho variables de entrada: seis suavizadas y dos sin suavizar.

## V. EVALUACIÓN DEL DESEMPEÑO DE LAS REDES

La evaluación del desempeño de las redes relativa a los datos del conjunto extramuestral fue hecha con base en el número —y magnitud— de los aciertos en la predicción de la variación del índice por medio de la prueba estadística de precisión en el acierto direccional y magnitudinal —*Excess Profitability* o EP de Anatolyev y Gerko (2005)—, suponiendo para ello que el índice de actividad económica es “comerciable”. Los resultados mostraron: *i*) que la red neuronal posee capacidad predictiva —es decir, que en ambos casos tratados, la red predice al LEI con un margen de certeza de más de 90%—; *ii*) que esta capacidad predictiva es congruente —es decir, que es estable durante el periodo de prueba—, y *iii*) que la capacidad predictiva de la red es superior a la de una regresión múltiple equivalente. Dado que el instrumento usado tanto para evaluar la capacidad predictiva de la red y demostrar su congruencia como para compararla con la de una regresión es la prueba de predictibilidad de Anatolyev y Gerko, ofrecemos a continuación una descripción de dicha prueba estadística que mide la precisión en el acierto direccional y magnitudinal de un predictor y los sintetiza en dos parámetros llamados respectivamente *Excess Profitability* o EP y *Probability of Prediction* o *Prob.*, la cual es la probabilidad de que una variable normal centrada en 0 y con desviación estándar igual a 1 asuma un valor ubicado dentro del intervalo que va de  $-EP$  a  $EP$ , es decir, la probabilidad de no independencia de las predicciones y las observaciones.

Según Anatolyev y Gerko (2005), “si  $y_t$  es el rendimiento de un activo financiero o índice y  $\hat{y}_t \geq 0$  una predicción o proyección continuamente distribuida de  $y_t$ , la predicción  $\hat{y}_t$  depende sólo de la data del conjunto de información  $I_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$  o, de manera más general, del conjunto de información extendido  $I_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$ , el cual puede incluir otras variables históricas. Considere ahora la siguiente regla de compraventa basada en  $\hat{y}_t$ :

Compre acciones que valgan el patrimonio actual si  $\hat{y}_t \geq 0$ , (1)

Venda acciones que valgan el patrimonio actual si no

Esto es, un inversionista va largo si la predicción para el próximo periodo es positiva, y va corto si no lo es. Llamemos esta regla la estrategia de compraventa. Equipado con esta estrategia, el inversionista modifica su posición cada periodo de compraventa y cierra la posición al final de cada periodo. Entonces, el rendimiento durante un periodo de la estrategia de compraventa es:

$$r_t = \text{sign}(\hat{y}_t)y_t \quad (2)$$

en el que  $\text{sign}(\cdot)$  toma el valor de  $-1$  cuando su argumento es negativo y el valor de  $1$  cuando su argumento es no negativo. Implícitamente suponemos que la distribución de  $\text{sign}(\hat{y}_t)$  es no degenerada.

La estrategia de compraventa (1) describe el comportamiento de un “analista técnico artificial” neutral al riesgo, si seguimos la terminología de Skouras (2001). La rentabilidad de la estrategia de compraventa (1) fue evaluada por Gençay (1998) como si las predicciones tuvieran valor económico en la práctica. Usando dos decenios y medio de DJIA Gençay (1998) encuentra que esta estrategia de compraventa es capaz de proporcionar ganancias perceptibles en comparación con las de una estrategia de *buy and hold*. Aquí se usa esta estrategia de compraventa para construir una prueba formal de predictabilidad promedio de los rendimientos, la cual se basa en la rentabilidad extramuestral de la estrategia de compraventa. El lector debe tener presente que el proceso de compraventa es tan sólo un experimento cerebral, y que no hace diferencia alguna el que haya límites de mercado (como costos de comercio y restricciones a la venta corta) que impidan o permitan la ejecución de la estrategia.

De manera formal, la hipótesis nula es la de que hay independencia de media condicional,  $H_0: E[y_t | I_{t-1}] = \text{const.}$  Técnicamente se requiere que según la hipótesis nula se dé una propiedad más fuerte:  $\hat{y}_t$  sea independiente de  $y_t$  para todos los rezagos y avances. El rendimiento esperado de un periodo de la estrategia de compraventa (1) es  $E[r_t]$ , el cual es estimable congruentemente con la hipótesis nula por los siguientes dos estimadores:

$$A_T = \frac{1}{T} \sum_{t=1}^T r_t \quad (3)$$

y

$$B_T = \frac{1}{T} \sum_{t=1}^T \text{sign}(\hat{y}_t) \frac{1}{T} \sum_{t=1}^T y_t \quad (4)$$

Ciertamente, con la hipótesis nula  $A_T \stackrel{P}{\rightarrow} E[r_t]$  y  $B_T \stackrel{P}{\rightarrow} E[\text{sign}(\hat{y}_t)] E[y_t]$   $E[\text{sign}(\hat{y}_t) E[y_t]]^{H_0} = E[\text{sign}(\hat{y}_t) E[y_t | I_{t-1}]] = E[\text{sign}(\hat{y}_t) y_t] = E[r_t]$ . Mientras que  $A_T$  es el rendimiento promedio que resulta del uso de la estrategia de compraventa,  $B_T$  es (la estimación del) rendimiento promedio de una estrategia de referencia (*benchmark*) que genera señales de compraventa aleatoriamente con probabilidades que corresponden a las proporciones de “compras” y “ventas” implicadas *ex post* por la estrategia de compraventa. Cuando  $y_t$  es predecible, la predicción en tiempo real y la inversión siguiendo la estrategia de compraventa generarán un rendimiento más alto que el de la estrategia de referencia, la diferencia entre  $A_T$  y  $B_T$  será considerable y la prueba tendrá poder.

Para completar la construcción de la prueba, queda por computar la varianza de  $A_T$  y  $B_T$  con la hipótesis nula. Sea  $p_{\hat{y}} = \Pr\{\text{sign}(\hat{y}_t) = 1\}$ , entonces

$$\text{var}(A_T - B_T) = 4 \frac{T}{T^2} \frac{1}{T} p_{\hat{y}}(1 - p_{\hat{y}}) \text{var}(y_t) \quad (5)$$

La manera más simple de estimar esta varianza es

$$\hat{V}_{EP} = \frac{4}{T^2} \hat{p}_{\hat{y}}(1 - \hat{p}_{\hat{y}}) \frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^2$$

en la que hemos corregido por los grados de libertad al estimar la varianza de  $y_t$ , y en la que

$$\hat{p}_{\hat{y}} = \frac{1}{2} \left( 1 + \frac{1}{T} \sum_{t=1}^T \text{sign}(\hat{y}_t) \right)$$

es un estimador congruente de  $p_{\hat{y}}$ . El estimado  $\hat{V}_{EP}$  es positivo por construcción a menos de que por casualidad todas las predicciones tengan el mismo signo. La estadística de nuestra prueba, de tipo Hausman y que llamamos estadística de rentabilidad excedente (EP), y su distribución asintótica son:

$$EP = \frac{A_T - B_T}{\sqrt{\hat{V}_{EP}}} \sim N(0, 1) \quad (6)$$

según Anatolyev y Gerko (2005). Adviértase que la fórmula (6) no es más que un cociente de rendimientos/volatilidad, indicándonos la relación intrínseca entre EP y el cociente clásico de Sharpe que también calculamos líneas abajo.

### 1. *Capacidad predictiva de la red neuronal*

El cociente EP analiza la capacidad predictiva de la red neuronal en términos del tamaño del intervalo que va de  $EP_{t-1}$  a  $EP_t$ , en el que EP es una variable normal estándar centrada en 0 y con desviación estándar igual a 1. Se refuta la hipótesis nula de que hay independencia de media condicional  $H_0: E[y_t | I_{t-1}] = const$  si el intervalo es suficientemente grande dando un área probabilística con la curva normal de, por ejemplo, .90 o más (en el caso de  $EP \in [-1.69, 1.69]$ ) o .95 o más (en el caso de  $EP \in [-1.965, 1.965]$ ), según el nivel de confianza requerido.

### 2. *Congruencia histórica de la capacidad predictiva de la red*

Para analizar la estabilidad de la capacidad predictiva de cada red se calculó la desviación estándar de la frecuencia acumulativa con que la red predijo correctamente —según el cociente EP—, la variación del LEI a lo largo de la duración del conjunto extramuestral.

### 3. *Comparación de la capacidad predictiva de la red neuronal con la de una regresión múltiple equivalente*

Por último, la capacidad de los modelos de redes neuronales para predecir la variación del LEI —junto con la rentabilidad que un inversionista habría obtenido de haber seguido las recomendaciones de compra y de venta arrojadas por el modelo— fue contrastada con la rentabilidad que éste habría obtenido de haber seguido las recomendaciones de compra y de venta hechas por una regresión múltiple comparable.



## VI. ANÁLISIS DE LOS RESULTADOS

1. *Capacidad predictiva de la red neuronal*

El cuadro 3 resume la capacidad predictiva de las redes dentro del conjunto extramuestral. Además del EP se calculó una serie de medidas secundarias, de las cuales la única que pudiera no ser de uso general es el *drawdown*. El *drawdown* en cualquier punto del tiempo es el decrecimiento en el valor neto de la cuenta del patrimonio desde el valor histórico máximo de la misma (Wolberg, 2000, p. 39):

Cociente del nivel del patrimonio = nivel del patrimonio/nivel máximo del patrimonio

*Drawdown* = 1 - cociente del nivel del patrimonio  
(*Equity Ratio* = *Equity/Max Equity*. *Drawdown* = 1-*Equity Ratio*)

También se calculó la proporción Sharpe, cociente conocido en el mundo financiero y cuyo comportamiento se asemeja al del EP:

Proporción Sharpe  $\frac{ROI}{\text{desviación estándar del ROI}}$

$ROI = \frac{\text{utilidad [final]} - \text{utilidad [inicial]}}{\text{desviación estándar del ROI}}$

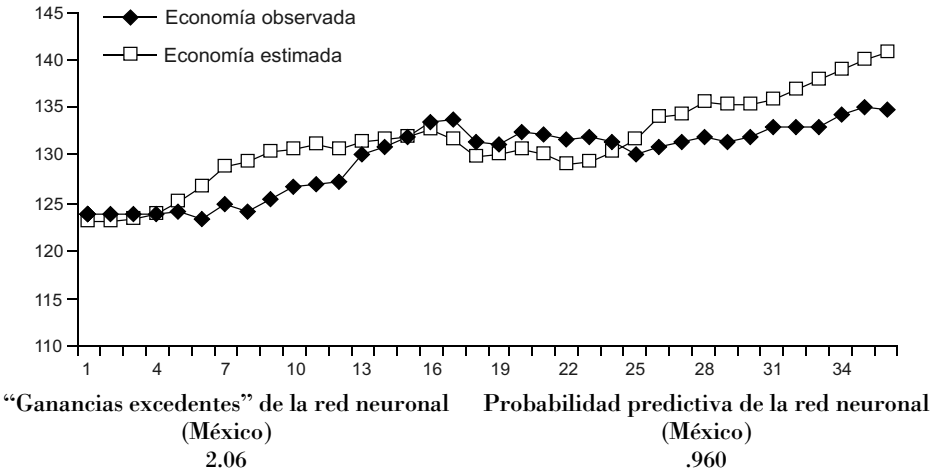
desviación estándar del ROI

CUADRO 3. *Capacidad predictiva de las redes*

<i>Estrategia</i>	<i>EP</i>	<i>Profit</i>	<i>ROI</i>	<i>Desviación estándar de ROI</i>	<i>Proporción Sharpe</i>	<i>Promedio Drawdown</i>	<i>Maximum Drawdown</i>
Red-LEI- Estados Unidos	1.81	5.50	0.05	0.01	3.79	0.18	1.11
Red-LEI- México	2.06	17.75	0.15	0.05	3.09	0.07	0.29
	<i>Porcentaje derecho</i>	<i>Promedio porcentual derecho</i>	<i>Desviación estándar del porcentaje derecho</i>	<i>Pérdida total</i>	<i>Porcentaje de pérdida</i>	<i>95% del intervalo confidencial del porcentaje promedio derecho</i>	
Red-LEI- Estados Unidos	0.67	0.67	0.10	3.30	0.33	0.63	0.70
Red-LEI- México	0.67	0.64	0.08	5.13	0.43	0.62	0.67

GRÁFICA 2. *México: Economía estimada vs observada*

(Diciembre de 1998-septiembre de 2001)



Además, para ambas redes la prueba EP muestra que dentro de un nivel de más de 90% de confianza podemos rechazar la hipótesis nula de independencia (falta de capacidad predictiva) y por tanto aceptar la hipótesis de la capacidad de las redes utilizadas para anticipar las fluctuaciones de las economías de los Estados Unidos y México. Como evidencia complementaria obsérvese que la media del porcentaje de aciertos está por encima de 50% y su estrecho intervalo de confianza, también está bien por encima de 50%. La proporción Sharpe mayor que 1 indica que los vaivenes de la ganancia o pérdida (asociados al riesgo) son menores que el ROI (la ganancia).

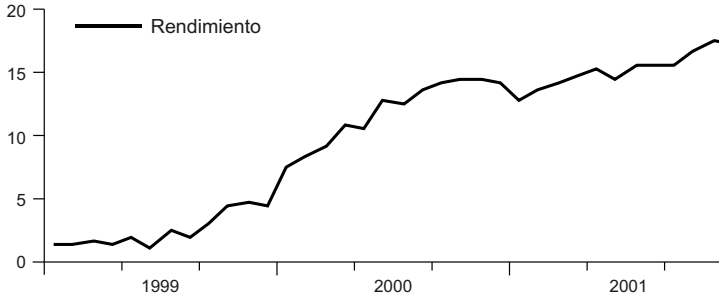
En la gráfica 3 mostramos las curvas de la economía observada y estimada por la red neuronal para México. Adviértase a continuación la adquisición congruente de rendimientos y la baja volatilidad de la curva patrimonio acumulado correspondiente a la gráfica 2 —que en realidad representa el “éxito predictivo acumulado” de la estrategia utilizada por la red para predecir el LEI de México—. Las mismas observaciones son pertinentes para el caso de los Estados Unidos.

## 2. *Congruencia histórica de la capacidad predictiva de la red*

Las gráficas 6 y 7 muestran que la capacidad predictiva se mantiene a lo largo del tiempo dentro del periodo extramuestral. El porcen-

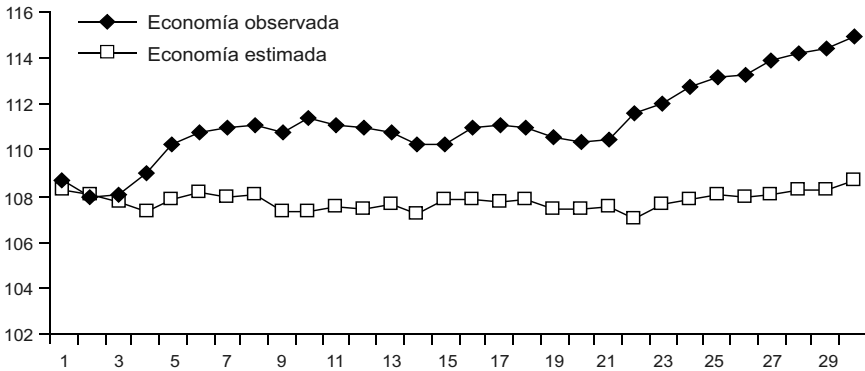
**GRÁFICA 3. México: Curva de patrimonio acumulado de la estrategia de compraventa de la red neuronal**

(Diciembre de 1998-agosto de 2001)



**GRÁFICA 4. Estados Unidos: Economía estimada vs observada**

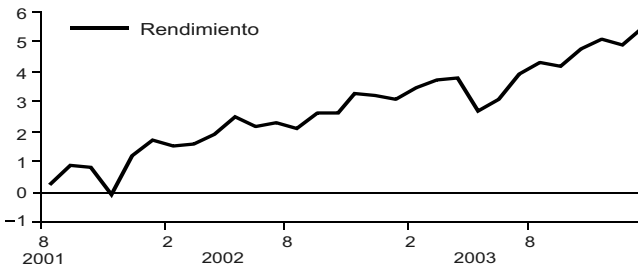
(Agosto de 2001-septiembre de 2003)

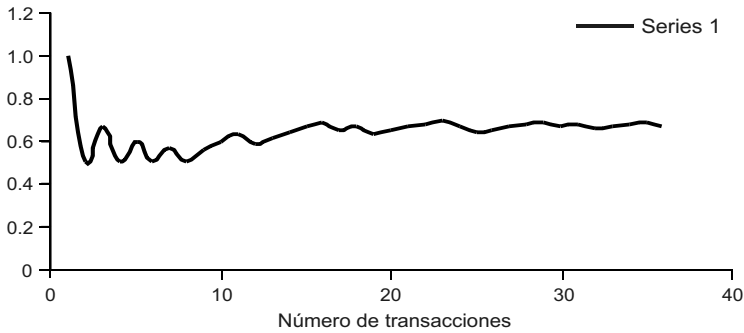
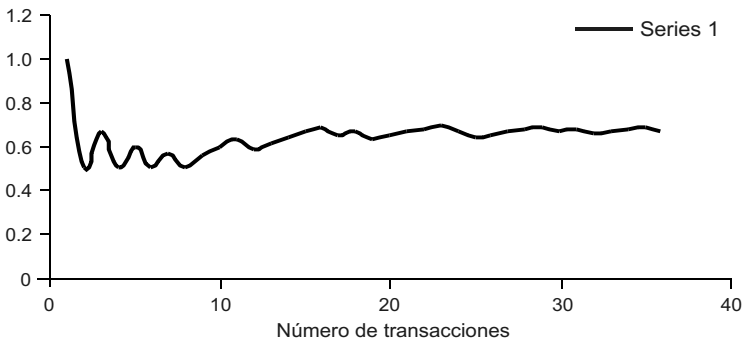


“Ganancias excedentes” de la red neuronal (Estados Unidos) 1.81      Probabilidad predictiva de la red neuronal (Estados Unidos) .929

**GRÁFICA 5. Estados Unidos: Curva de patrimonio acumulado de la estrategia de compraventa de la red neuronal**

(Agosto de 2001-agosto de 2003)



GRÁFICA 6. *Estados Unidos: Porcentaje de aciertos de la red*GRÁFICA 7. *México: Porcentaje de aciertos de la red*

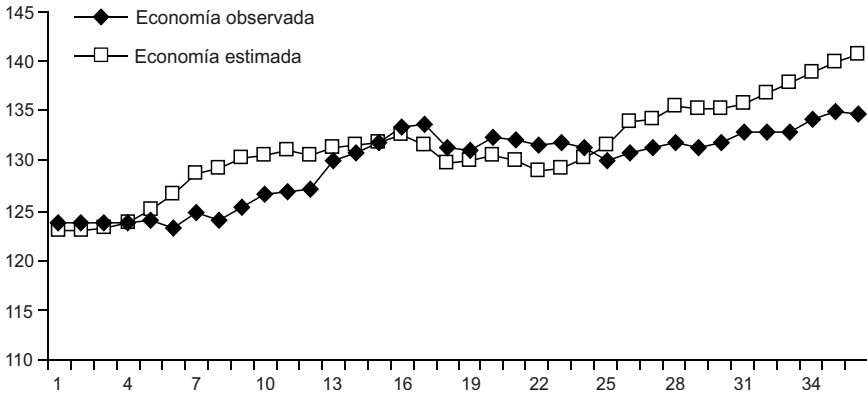
taje de aciertos mantiene una pendiente cercana a 0, lo cual sugiere que las redes mantienen a lo largo del tiempo una precisión constante en la proyección de la dirección del LEI.

### 3. *Comparación de la capacidad predictiva de la red neuronal con la de una regresión múltiple equivalente*

Las regresiones múltiples tuvieron las mismas entradas (*inputs*) que las redes (esto es, los mismos rezagos del índice de bonos y de acciones) y fueron efectuadas con base en primeras diferencias con de autocorrelación de Durbin calculado en dos pasos (véase gráficas 8-11 y cuadro 4). Adviértase que la red es mejor predictora que la regresión en ambos casos: tanto para México como para los Estados Unidos.

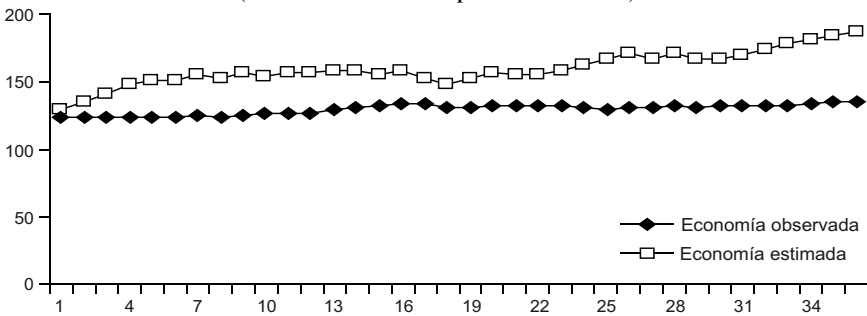
**GRÁFICA 8. México: Red neuronal**

(Diciembre de 1998-septiembre de 2001)



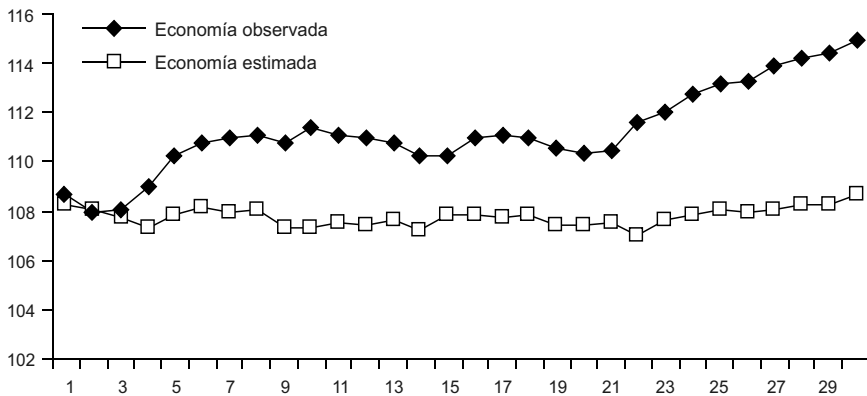
**GRÁFICA 9. México: Regresión múltiple**

(Diciembre de 1998-septiembre de 2001)



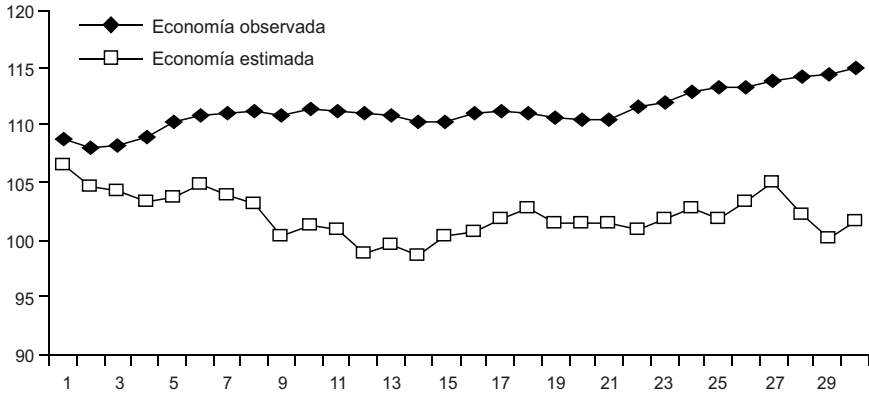
**GRÁFICA 10. Estados Unidos: Red neuronal**

(Agosto de 2001-septiembre de 2003)



GRÁFICA 11. *Estados Unidos: Regresión múltiple*

(Agosto de 2001-septiembre de 2003)



CUADRO 4. *Resumen de los resultados comparativos de la red neuronal vs regresión*

<i>Red neuronal vs regresión (según Anatolyev)</i>	<i>Ganancias excedentes de la regresión</i>	<i>Probabilidad predictiva de la regresión</i>	<i>Ganancias excedentes de la red neuronal</i>	<i>Probabilidad predictiva de la red neuronal</i>
México	1.93	.897	2.06	.960
Estados Unidos	1.67	.907	1.81	.929

## CONCLUSIONES

Las redes neuronales tienen capacidad predictiva significativa y esa capacidad predictiva es estable durante un intervalo que queda por determinar. Esto significa que las redes neuronales debidamente entrenadas son una fuente valiosa de información al momento de invertir en una empresa ubicada en una economía foránea o local. Debido a que las entradas se rezagaron un mínimo de cuatro meses para la economía de los Estados Unidos y un mínimo de 15 meses para la mexicana, las redes proyectan con bastante anticipación, lo cual es conveniente para algunas estrategias de inversión.

La hipótesis comportamentalista de la relación entre los mercados de capitales maduros (el sentimiento de prosperidad) y la propensión al gasto —relación que serviría para explicar el crecimiento de la economía— obviamente no se aplica a países con mercados de ca-

pitales incipientes que afectan y son afectados de manera directa sólo por un grupo de inversionistas relativamente pequeño. En este trabajo los mercados de capitales mexicano y estadounidense sirven de entrada a redes que proyectan con éxito la economía de esos países. Este estudio confirma la pertinencia de los mercados de capitales como predictores del comportamiento macroeconómico y del de los ciclos de negocios (Wheelwright y Makridakis, 1985; Stock y Watson, 1989).

#### REFERENCIAS BIBLIOGRÁFICAS

- Anatolyev, Stanislav (2005), "A Trading Approach to Testing for Predictability", *Journal of Business and Economic Statistics*, vol. 23, núm. 4, páginas 455-461.
- , y Alexander Gerko (2005), "A Trading Approach to Testing for Predictability", *Journal of Business and Economic Statistics*, vol. 23, núm. 4, páginas 455-461.
- Bishop, Christopher (1995), *Neuronal Networks for Pattern Recognition*, Oxford, Clarendon Press.
- Cerpa, Narciso, y Steven Walczak (1999), "Heuristic Principles for the Design of Artificial Neural Networks", *Information and Software Technology*, volumen 41, núm. 2, pp. 109-119.
- Conference Board (2004), "Definición del LEI", <http://www.tchindicators.org/methodology/component-description.cfm>.
- Cybenko, G. (1989), "Approximation by Superpositions of a Sigmoidal Function", *Math. Control Signals Systems*, vol. 2, pp. 303-314.
- Diebold, F. X., y G. D. Rudebusch (1993), "Further Evidence on Business-Cycle Duration Dependence", J. H. Stock y M. Watson (comps.), *Business Cycles, Indicators and Forecasting*, Chicago, The University of Chicago Press.
- Dotsey, Michael (1998), "The Predictive Content of the Interest Rate Term Spread for Future Economic Growth", *Economic Quarterly* (verano).
- Ducker, Michael J. (1997), "Strengthening the Case for the Yield Curve as a Predictor of U.S. Recessions", *Federal Reserve Bank of St. Louis Review*, vol. 79, marzo-abril, pp. 41-51.
- Estrella, Arturo, y Gikas A. Hardouvelis (1991), "The Term Structure as a Predictor of Real Economic Activity", *Journal of Finance*, vol. 46, junio, páginas 555-576.
- Estrella, Arturo, y Frederic S. Mishkin (1997), "The Predictive Power of the Term Structure of Interest Rates in Europe and the United States: Implications for the European Central Bank", *European Economic Review*, vol. 41, julio, pp. 1375-1401.

- Gençay, R. (1998), "Optimization of Technical Trading Strategies and the Profitability in Security Markets", *Economics Letters* 59, pp. 249-254.
- Getely, E. (1996), *Neuronal Networks for Financial Forecasting*, Nueva York, John Wiley & Sons, Inc.
- Granger, C. W. J., y T. Terasvirta (1993), *Modelling Nonlinear Economic Relationships*, Oxford, Oxford University Press.
- Haykin, Simon (1999), *Neuronal Networks, A Comprehensive Foundation*, Prentice Hall.
- Hornik, K., M. Stinchcombe y H. White (1989), "Multilayer Feed-Forward Networks Are Universal Approximators", *Neural Networks*, vol. 2, pp. 359-366.
- Jaditz, T., L. A. Riddick y C. L. Sayers (1998), "Multivariate Nonlinear Forecasting: Using Financial Information to Forecast the Real Sector", *Macroeconomics Dynamics* 2, pp. 369-382.
- Jadric, Timotej (2003), "A Nonlinear Approach to Forecasting with Leading Economic Indicators", *Studies in Nonlinear Dynamics & Econometrics* 7, núm. 2, artículo 4, Berkeley Electronic Press.
- Jurik, Mark (1994), "Back Percolation, Assigning Local Error in Feed-forward Perception Methods", *Braincel*, New Haven, Promised Land Technologies, Inc.
- Kaiser, R., y A. Maraval (1999), "Estimation of the Business Cycle—A Modified Hodrick Prescott Filter", *Spanish Economic Review* 1, pp. 175-206.
- Kim *et al* (1996), "Detecting Asymmetries in Observed Linear Time Series and Unobserved Disturbances", *Studies in Nonlinear Dynamics and Econometrics* 1, pp. 131-143.
- Lahiri, Kajal, y Geoffrey Moore (1991), *Leading Economic Indicators: New Approaches and Forecasting Records*, Nueva York, Cambridge University Press.
- Mitchel, W. C., y A. F. Burns (1983), "Statistical Indicators of Cyclical Revivals", *Business Cycles, Inflation, and Forecasting*, Cambridge, National Bureau of Economic Research, segunda edición.
- Moody, J., A. Levin y S. Rehfuss (1993), "Predicting the U.S. Index of Industrial Production", *Neuronal Network World* 3-6, pp. 791-794.
- Moore, Geoffrey (1961), *Business Cycle Indicators*, Princeton, Princeton University Press, vol. 1, tabla 3.2, p. 56.
- Murphy, John J. (1991), *Intermarket Technical Analysis*, Nueva York, John Wiley & Sons, Inc.
- (2004), *Intermarket Analysis: Profiting from Global Market Relationships*, Nueva York, John Wiley & Sons, Inc.
- Natter, M., C. Haefke, T. Soni y H. Otruba (1994), "Macroeconomic Forecasting Using Neuronal Networks", *Second International Workshop on Neuronal Networks in the Capital Markets* (NNCM), Pasadena.
- Parisi, Antonino, Franco Parisi y José Luis Guerrero C. (2003), "Modelos predictivos de redes neuronales en índices bursátiles", *EL TRIMESTRE ECONÓMICO*, vol. LXX, núm. 280, pp. 721-744.



- Pesaran, M. H., y A. Timmermann (1992), "A Simple Nonparametric Test of Predictive Performance", *Journal of Business and Economic Statistics* 10, pp. 461-465.
- Refenes, A. P. (1995), *Neuronal Networks in the Capital Markets*, Nueva York, John Wiley & Sons, Inc.
- Renshaw, Edward F. (1995), "Essays on the Stock Market", Department of Economics, New York State University at Albany, Albany, <http://www.albany.edu/~renshaw/>
- Ruggiero, Murray (1996), "Using Correlation Analysis to Predict Trends", *Futures Magazine*, febrero, pp. 46-49.
- (1997), *Cybernetic Trading Strategies*, Nueva York, John Wiley & Sons, Inc.
- (1998), "Intermarket Analysis is Fundamentally Sound", *Futures Magazine*, abril, pp. 58-62.
- (2001a), "Intermarket Analysis & Economic Forecasting", *Futures Magazine*, marzo, pp. 60-62.
- (2001b), "The Economy & Markets: Trading the Relationship", *Futures Magazine*, abril, pp. 60-62.
- Schwager, J. (1995), *Fundamental Analysis*, Nueva York, John Wiley & Sons, Inc.
- Skouras, Spyros P. (2001), "Financial Returns and Efficiency as Seen by an Artificial Technical Analyst", *Journal of Economic Dynamics and Control* (25)1-2, pp. 213-244.
- Stock, J. H., y M. Watson (1989), "New Indexes of Leading and Coincident Economic Indicators", *NBER Macroeconomics Annual*, pp. 351-394.
- , y — (comps.) (1993a), *Business Cycles, Indicators and Forecasting*, Chicago, The University of Chicago Press.
- , y — (1993b), "A Procedure for Predicting Recessions with Leading Indicators: Economic Issues and Recent Experience", J. H. Stock y M. Watson (comps.), *Business Cycles, Indicators and Forecasting*, Chicago, The University of Chicago Press.
- Swansen, N., y H. White (1995), "A Model Selection Approach to Real-time Macroeconomic Forecasting Using Linear Models and Artificial Neuronal Networks", Discussion Papers, Department of Economics, Pennsylvania State University.
- Tkacz, G. (2000), "Neuronal Network Forecasting of Canadian GDP Growth", proxima edición en *International Journal of Forecasting*.
- Trippi, Robert R., y Efraim Turban (1996), *Neuronal Networks in Finance and Investing*, Chicago, Irwin Professional Publishing.
- Verbrugge, R. (1997), "Investigating Cyclical Asymmetries", *Studies in Non-linear Dynamics and Econometrics* 2, pp. 15-22.
- Waczak, Steven (2001), "An Empirical Analysis of Data Requirements for Financial Forecasting with Neural Networks", *Journal of Management Information Systems*, vol. 17, núm. 4, pp. 203-222.

- Weigend, A. S., y N. A. Gershenfeld (1994), *Time Series Prediction: Forecasting the Future and Understanding the Past*, Addison-Wesley.
- Wheelwright, Steven C., y Spyros Makridakis (1985), *Forecasting Methods for Management*, Nueva York, John Wiley & Sons, Inc.
- White, Halbert (1996), "Economic Prediction Using Neuronal Networks: The Case of IBM Daily Stock Prices", Robert R. Trippi y Efraim Turban, *Neuronal Networks in Finance and Investing*, Chicago, Irwin Professional Publishing.
- Wolberg, John R. (2000), *Expert Trading Systems*, Nueva York, John Wiley & Sons, Inc.