

## PROCESOS GAUSSIANOS EN LA PREDICCIÓN DE LAS FLUCTUACIONES DE LA ECONOMÍA MEXICANA\*

*Irene García, Loren Trigo,  
Sabatino Costanzo  
y Enrique ter Horst\*\**

### RESUMEN

La capacidad de algunas redes neuronales para predecir la dirección de la economía de México —representada por el LEI— cuyos insumos son las versiones simultáneas (suavizante y predictiva) de un Proceso Gaussiano alimentado por un índice de acciones y uno de bonos —ambos representativos del mercado mexicano—, es comparada favorablemente (por medio del método de Anatolyev y Gerko para evaluar la precisión de un predictor), con la capacidad predictiva de redes desarrolladas para el mismo fin por dos de los autores de este artículo en uno artículo anterior, cuyos insumos son rezagos de dichos índices.

\* Irene García, departamento de Cómputo Científico y Estadística Universidad Simón Bolívar, Venezuela (correo electrónico: irene@cesma.usb.ve). L. Trigo, ProAlea International Consultants Inc. La Florida, Caracas (correo electrónico: loren.trigo@gmail.com). S. Costanzo, Vicepresidencia de Estrategia e Investigación, Grupo de Empresas Econoinvest, Instituto de Estudios Superiores de Administración, Caracas (correo electrónico: scostanzo@econoinvest.com). Enrique ter Horst, profesor asociado, Euomed Management, Marsella, e Instituto de Estudios Superiores de Administración, Caracas (correo electrónico: enriquetorhorst@gmail.com).

\*\* *Palabras clave:* redes neuronales, predicción, fluctuaciones, economía, mercados, mercados de capitales, procesos gaussianos. *Clasificación JEL:* G. Artículo recibido el 19 de noviembre de 2008 y aceptado el 25 de noviembre de 2009. Este trabajo fue realizado con el apoyo económico de la Academia de Ciencias Cívicas, Matemáticas y Naturales de Venezuela por medio de la beca para realizar estudios de doctorado otorgada a I. García.

## ABSTRACT

The ability of some neural nets to predict the direction of the Mexican economy — represented by its LEI — when taking as inputs the simultaneous versions (smoothing and predictive) of a Gaussian Process fed with a Stock Index and a Bonds Index representing the Mexican market, is favorably compared — through the Anatolyev and Gerko predictive accuracy test — with the predictive ability of nets developed for a similar purpose by the authors in a previous paper, and whose inputs are the lagged indexes of the Mexican capital markets and some of their moving averages.

## INTRODUCCIÓN

Este artículo analiza la capacidad de las redes neuronales para predecir el signo de las variaciones mensuales de la economía de México, usando como insumos tanto el suavizamiento conjunto como la proyección predictiva de dicho suavizamiento realizados con un proceso gaussiano bidimensional que, al integrar simultáneamente — tanto en el suavizamiento como en la proyección — a un índice de bonos soberanos con uno de acciones — ambos representativos del mercado de capitales —, proporciona a la red de los dos insumos (suavizamiento y proyección) que ésta requiere para predecir la tendencia futura (al alza o la baja) del índice compuesto (LEI) de México, considerado como un índice coincidente y no adelantado de la economía. La comparación por medio del método de Anatolyev y Gerko — desarrollado para evaluar la precisión de un predictor — de los resultados obtenidos en este trabajo, con los de redes análogas desarrolladas con el mismo fin por dos de los autores del presente artículo en un trabajo anterior (Trigo y Costanzo, 2007), cuyos insumos son rezagos y promedios móviles de dichos índices, mostró que en este caso la metodología que incorpora procesos gaussianos en el preprocesamiento de los datos es mejor que la basada en rezagos. Para la aplicación tanto del presente artículo como del citado líneas arriba se utilizaron las mismas redes neuronales de alimentación frontal con algoritmo de aprendizaje de percolación reversa (una optimización del algoritmo de aprendizaje de propagación reversa). La evaluación de ambas metodologías se hizo sobre la base de conjuntos extramuestrales compuestos por un porcentaje fijo (10%) los datos totales. El desempeño relativo de las redes tanto en el artículo anterior como en el presente fue medido en términos de la suma de la magnitud de los movimientos del índice LEI,

LEI, cuyos signos fueron predichos correctamente por éstas, suponiendo implícitamente para ello que el índice de actividad económica era “comerciable” y evaluando el resultado con la prueba estadística de precisión en el acierto direccional y magnitudinal de Anatolyev y Gerko (2005). Los resultados obtenidos mostraron que la capacidad predictiva de la red neuronal alimentada con datos provenientes del mercado de capitales preprocesados por un método gaussiano bidimensional es superior a la capacidad predictiva de redes similares alimentadas con los mismos datos preprocesados con rezagos y promedios móviles. En suma, el uso de procesos gaussianos para preprocesar los datos que reflejan el movimiento de los mercados de capitales, hace a la red que utiliza este insumo más apta para predecir el desempeño de la economía que las que utilicen los mismos datos preprocesados con métodos tradicionales.

El método más conocido para predecir los ciclos de negocios es el índice LEI o Composite Index of Leading Economic Indicators, índice de índole lineal que ha predicho cada una de las ocho últimas recesiones en los Estados Unidos. Este índice, publicado por el Conference Board para los Estados Unidos, y México entre otros, se compone de 10 índices económicos: horas promedio trabajadas a la semana por obreros de producción en industrias manufactureras; número promedio de reclamaciones de seguro de desempleo; nuevas órdenes de los manufactureros para bienes de consumo; velocidad relativa en que vendedores pueden entregar pedidos a compañías industriales, según el índice del National Association of Purchasing Managers; nuevas órdenes recibidas por manufactureros en bienes de capitales no relacionados con la defensa; número de permisos concedidos para edificios residenciales; cambio en el mercado de acciones; provisión monetaria  $M2$  (ajustada por inflación); la dispersión del rendimiento, y el índice de expectativas del consumidor.

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 puede ser 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 pregunta 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 es-

tadísticas 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 recurre a dos argumentos. El primero es que los mercados de capitales (*equity & debt*) son la principal fuente de financiación de la economía y en esta capacidad actúan como estimulantes. El segundo recurre a una metáfora de las finanzas o 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 e inversión (Trigo y Costanzo, 2007).

Aun cuando ha habido cierto desacoplamiento entre el mercado de acciones y el de bonos soberanos debido al inusualmente largo mercado alcista de los años noventa en los Estados Unidos, tanto el 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 (Ruggiero, 2001).

## 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 de las utilizadas por nosotros, así como de su entrenamiento. Centramos nuestro análisis en el empleo de las redes neuronales multicapas con aprendizaje supervisado, ya que 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 (por ejemplo, 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.

## II. METODOLOGÍA Y DATOS

La metodología utilizada por los autores de este artículo para construir la red y optimar su entrenamiento comprende 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.

Ya que este problema no ha recibido solución analítica, debemos recurrir a la sistematización de un método de prueba y error. Por esta razón optimamos 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 las redes candidatas obtenidas con *i*).

El programa *Braicel*, utilizado para generar todas nuestras redes neuronales es un *Add-in* de EXCEL producido por Promised Land Technologies cuyo algoritmo de búsqueda requiere que el cuerpo de datos se divida en tres conjuntos: de entrenamiento, de entrenamiento con prueba y de prueba o validación — aquí llamado conjunto extramuestral—. El primer conjunto (entrenamiento) proporciona 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 tercer conjunto (extramuestral) se compone de datos completamente nuevos para la red que permiten evaluar su desempeño. En general se recomienda utilizar 60% de los datos como conjunto de entrenamiento, 30% de los datos como conjunto de entrenamiento con prueba y 10% de los datos como conjunto extramuestral.

En este estudio los conjuntos de datos tienen las siguientes fechas: 5 de noviembre de 1995-11 de noviembre de 1998, entrenamiento y entrenamiento con prueba México; 12 de noviembre de 1998-11 de noviembre de 2001, extramuestral México. Dado el número de observaciones disponibles para este estudio decidimos incluir como entradas a las redes dos de los 10 índices componentes del LEI, esto es, un índice del mercado accionario y un índice del mercado de bonos soberanos. Los índices utilizados como entradas fueron: México, *i*) *MSCI Mexico Former Dead-Price Index* (~MP), y *ii*) *JPM Elmi Mexico (L) Dead-Return Ind.* (OFCL).

En principio podemos usar cualquier conjunto crudo de datos para entrenar nuestras redes. Sin embargo, en la práctica suele ser de gran importancia realizar un preprocesamiento apropiado de los datos antes de alimen-

tar con ella a la red. Dicho preprocesamiento, de ser adecuado, ayuda a la red a aprender mejor. En el caso del presente artículo, cada variable de entrada de la red fue optimada (preprocesada) antes de alimentar a la red con ella.

### *Preprocesamiento de los datos*

En la práctica, realizar un preprocesamiento apropiado de los datos antes de alimentar con ellos a la red neuronal ayuda a que ésta aprenda mejor. En este artículo se utilizó un modelo no paramétrico basado en una regresión no lineal de procesos gaussianos que es fácil de manejar e interpretar y que efectivamente ayuda a la red a mejorar la predicción.

## III. LOS PROCESOS GAUSSIANOS

Para comprender el modelo utilizado se debe comenzar por entender lo que es un proceso gaussiano. Se puede pensar en un proceso gaussiano (PG) como una generalización de una distribución gaussiana aplicada a un vector de infinitas variables (función), así que la inferencia toma lugar en el espacio de las funciones, de allí que el modelo es no paramétrico, es decir, no se fija de antemano el número y la naturaleza de los parámetros.

En la comunidad estadística existen muchas teorías para modelos con procesos gaussianos en series de tiempo, pero aunque han sido conocidos por largo tiempo, su uso puede quizá ser establecido desde el final del siglo XIX y su aplicación a problemas reales está aún en la fase inicial. En términos formales, un proceso gaussiano es una sucesión de variables aleatorias indizadas en la que cualquier subconjunto resulta tener distribución conjunta gaussiana.

En este trabajo,  $y$   $f(x)$  es el índice LEI (variables aleatorias) y el proceso está indizado respecto a  $x$ , el conjunto de posibles variables explicativas de (el índice de acciones y el índice de bonos). El proceso se especifica completamente con una media  $m(x)$  y una función de covarianza  $k(x, x')$  y se denota  $f(x) \sim \mathcal{GP}(m(x), k(x, x'))$ .

### 1. Modelo

El modelo utilizado en el preprocesamiento (Rasmussen, 1996, Rasmussen y Williams 2005), está definido en términos de una combinación lineal de funciones de base fijas dadas por los elementos del vector  $\phi(x)$  de modo que:

$$y = f(x) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

$$f(x) = w^t \phi(x)$$

en que  $x$  es el vector de entrada y  $w$  representa el vector de pesos. En general, los modelos lineales empleados hasta ahora en la bibliografía para predecir ciclos económicos presentan baja capacidad predictiva, sin embargo, éste modelo lineal utilizado para preprocesar los datos no tiene el mismo inconveniente como muestra líneas abajo.

Para hacer la estimación de los parámetros del modelo se utiliza un enfoque bayesiano, por lo cual imponemos una distribución *a priori* para  $w$  de la forma:  $p(w) = \mathcal{N}(w|0, \lambda^{-1}I)$ , en la que  $\lambda$  representa la precisión de la distribución. Nótese que cada  $w$  induce una función de probabilidad en  $f(x)$ . Con la estructura especificada, el modelo es un proceso gaussiano. En efecto, si queremos evaluar  $f(x)$  para valores específicos del vector  $x$ , digamos  $x_1, \dots, x_N$ , estamos interesados en la distribución conjunta de  $f(x_1), \dots, f(x_N)$ . Al denotar  $f$  como un vector de componentes  $f_n = f(x_n), n = 1, \dots, N$  se tiene que  $f = \Phi w$ . Entonces, por ser  $f(x)$  una combinación lineal de variables aleatorias gaussianas independientes, resulta también gaussiana y el proceso está definido por:

$$E[f] = E[w] = 0$$

$$cov[f] = E[ff^t] = E[w\Phi^t] = \Phi^{-1} \Phi^t = K$$

en que  $K$  donde tiene elementos  $K_{ij} = k(x_i, x_j) = \lambda^{-1} \phi^t(x_i) \phi(x_j)$ .

De esta manera, el modelo presentado es un proceso de ruido con distribución gaussiana de la forma  $p(y|f, \lambda^{-1}I_N)$ , en el que  $y = (y_1, \dots, y_N)^t$  y  $f = (f_1, \dots, f_N)^t$ , y, de la definición de proceso gaussiano, resulta que la distribución marginal de  $p(f)$  es  $\mathcal{N}(f|0, K)$ . Así, la marginal es  $p(y) = \mathcal{N}(y|0, C)$ , con  $C$  matriz de covarianza con elementos:

$$C(x_i, x_j) = k(x_i, x_j) = \lambda^{-1} \phi^t(x_i) \phi(x_j)$$

## 2. Predicción

Dado un conjunto de datos de entrenamiento  $y = (y_1, \dots, y_N)^t$  correspondientes a los vectores  $x_1, \dots, x_N$ , la idea es predecir el valor  $y_{N+1}$ , lo cual re-

quiere evaluar la distribución predictiva  $p(y_{N+1}|y)$ . Para encontrar  $p(y_{N+1}|y)$ , se coloca  $p(y_{N+1}) = N(y_{N+1}|0, C_{N+1})$ , en que  $C_{N+1}$  es una matriz de covarianza que se particiona de la siguiente forma:

$$C_{N+1} = \begin{bmatrix} C_N & k \\ k^t & c \end{bmatrix}$$

con  $C_N$  de dimensión  $N \times N$  para  $i, j = 1, \dots, N$ , el vector  $k$  con elementos para  $i = 1, \dots, N$  y el escalar  $c = k(x_{N+1}, x_{N+1})^{-1}$ . Luego, la distribución predictiva  $p(y_{N+1}|y)$  es gaussiana con media y covarianzas dadas por:

$$m(x_{N+1}) = k^t C_N^{-1} y$$

$$^2(x_{N+1}) = c - k^t C_N^{-1} k$$

### 3. Selección del kernel

Como la distribución predictiva depende del valor de  $x_{N+1}$  sólo por medio de la función  $k(\dots)$ , su selección es decisiva para la predicción. La función de covarianza define cercanía o similitud: entradas que son cercanas probablemente tendrán valores de respuesta  $y$  cercanos, además la covarianza debe llevar codificada las suposiciones que se tienen acerca de la función que deseamos estimar  $f$ .

En este artículo se usó la familia paramétrica exponencial cuadrada como *kernel*, esta es (véase apéndice 2):

$$k(x_i, x_j) = \frac{1}{f} \exp \left[ -\frac{1}{2} (x_i - x_j)^t M (x_i - x_j) \right] \frac{1}{n} \delta_{ij}$$

Los hiperparámetros del *kernel* (la precisión del ruido y los de  $k(\dots)$ ) se infieren a partir de los datos, maximizando la verosimilitud  $p(y | \dots)$  respecto a  $\dots$ . La matriz

$$M = \begin{bmatrix} l_1 & 0 \\ 0 & l_2 \end{bmatrix}$$

en la que  $l_1$  y  $l_2$  desempeñan el papel de parámetros de longitud-escala, es decir, cuán lejos se debe mover (a lo largo del eje de una cierta entrada) para que los valores de la función no estén correlacionados.



#### 4. Datos o instrumentación

Los datos utilizados corresponden al país México, son mensuales y la serie va desde el 15 de enero de 1994 hasta el 15 de noviembre de 2001, es decir, el tamaño muestral es 95, en el que:  $y$  es el vector que deseamos estimar y corresponde al índice LEI en México;  $x_1$  representa un vector de la primera variable explicativa que en este caso corresponde al índice de bonos soberanos *JPM Elmi Mexico Dead-Return Ind*;  $x_2$  denota un vector de la segunda variable explicativa que corresponde al índice del mercado accionario: *MSCI Mexico Former Dead-Price Index*.

El método descrito se usó como preprocesamiento de los datos en dos versiones: *i*) como predictor y *ii*) como suavizador de la serie. En ambos casos los resultados que se muestran se obtuvieron con los siguientes valores iniciales de los hiperparámetros:  $l_1 = l_2 = 1$ ,  $\frac{\sigma_f^2}{\sigma_n^2} = 1$  y  $\frac{\sigma_n^2}{\sigma_n^2} = 0.1$ , con los datos previamente estandarizados, pues el proceso está centrado en 0. Estos valores provienen de la estimación de una regresión lineal multivariante como base para instrumentar dichos parámetros, así como la varianza del ruido (Rasmussen y Williams, 2005, Wolpert, Ickstadt y Hansen, 2003). Para hacer la predicción se tomó una ventana móvil de tamaño 5 (datos de entrenamiento) y como prueba el dato siguiente. Para el suavizado la ventana también es móvil pero de tamaño 70.

#### 5. Capacidad predictiva del modelo basado en PG

La manera más sencilla de evaluar la capacidad predictiva del modelo es graficar la economía observada *versus* la estimada por el modelo propuesto, lo cual corresponde a la gráfica 1 y en la que la línea continua indica la media predictiva obtenida con el modelo y las barras grises las bandas de confianza asociadas a cada predicción a 95% de confianza. Las cruces representan los valores observados. Es evidente que los valores observados y los estimados siguen la misma tendencia y que el error en la predicción es pequeño. Resalta la capacidad del modelo para estimar la caída del LEI en los índices del tiempo entre 10 y 20.

Con el fin de comparar los resultados de este modelo con los de la red neural publicada en un trabajo previo a éste (Trigo y Costanzo, 2007), se mide el desempeño del modelo por el número de predicciones correctas del signo de la variación del índice y suponiendo además para ello que el índice

de actividad económica es “comerciable” a fin de aplicar la prueba estadística de precisión en el acierto direccional y magnitudinal (Anatolyev y Gerko, 2005).

Dado que el modelo basado en PG muestra un indicador de Anatolyev máximo, se pasa a calcular algo más refinado para determinar la rentabilidad, esto es, la curva de patrimonio acumulado inducido por las decisiones de mercado tomadas con base en las recomendaciones del modelo de PG.

En contraposición a la metodología descrita líneas arriba — en la que los insumos de la red son preprocesados mediante de un proceso gaussiano —, en un trabajo anterior (Trigo y Costanzo, 2007), se utiliza un método de optimización consistente en utilizar cada variable de entrada para generar otras 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 utilizaba aisladamente para proyectar el LEI correspondiente mediante el siguiente procedimiento, que coincide con la estratégica de compraventa descrita por Anatolyev y Gerko (2005).

[...] Si se asume que el LEI es comerciable con precio igual al valor numérico del índice y 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 “preciot” y el “preciot-1” del LEI, de lo contrario se resta esa misma cantidad como “pérdida”; esto es, se construyeron 24 sistemas de compra-venta del LEI, cada uno con una entrada rezagada distinta y 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). Los sistemas de compraventa se jerarquizaron en base a los resultados obtenidos y se escogió en consecuencia al rezago 20 como óptimo [...].

Los resultados de la optimización de los rezagos llevaron a la siguiente selección: *Mexico, MSCI Mexico former dead-Price Index (~MP)* rezago 20 meses, y *JPM Elmi Mexico (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 tres valores

distintos 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 tienen 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 (Trigo y Costanzo, 2007).

#### IV. EVALUACIÓN DEL DESEMPEÑO DE LAS REDES: DESCRIPCIÓN DE LA PRUEBA ANATOLYEV Y GERKO

La evaluación del desempeño de las redes relativa a los datos del conjunto extramuestral fue hecha con base en el número —y a la 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)—, en el supuesto de que el índice de actividad económica es “comerciable”. Los resultados mostraron: *i*) que las dos redes neuronales (la red con preprocesamiento gaussiano y la red con preprocesamiento simple) poseen capacidad predictiva —es decir, que en ambos casos tratados, la red predice al LEI con un margen de certeza de Anatolyev de más de 90%—; *ii*) que la capacidad predictiva de la red con preprocesamiento gaussiano de los insumos es superior que la capacidad predictiva de la red con insumos preprocesados por medio de rezagos y promedios móviles, y *iii*) que la capacidad predictiva de ambas redes es superior a la de una regresión múltiple equivalente.

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, la cual 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 cobre 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 (Anatolyev y Gerko, 2005).

Comparación de la capacidad predictiva de las redes neuronales (alimen-

tadas con datos preprocesados respectivamente con metodología gaussiana y con metodología simple), con la de una regresión múltiple equivalente. La capacidad de los dos 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 cada uno de los modelos—, es contrastada a continuación con la “rentabilidad” que éste habría obtenido de haber seguido las recomendaciones de compra y de venta señaladas por una regresión múltiple comparable.

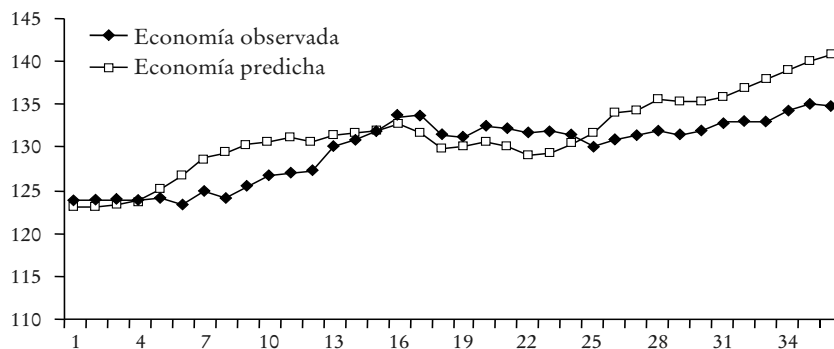
## V. ANÁLISIS DE LOS RESULTADOS

### *Capacidad predictiva de las redes neuronales*

A continuación se resume la capacidad predictiva de las redes dentro del conjunto extramuestral. Siguen las gráficas de las curvas de la economía observada (en azul marino) y estimada (en fucsia) por la red neuronal de preprocesamiento simple para México (gráfica 1).

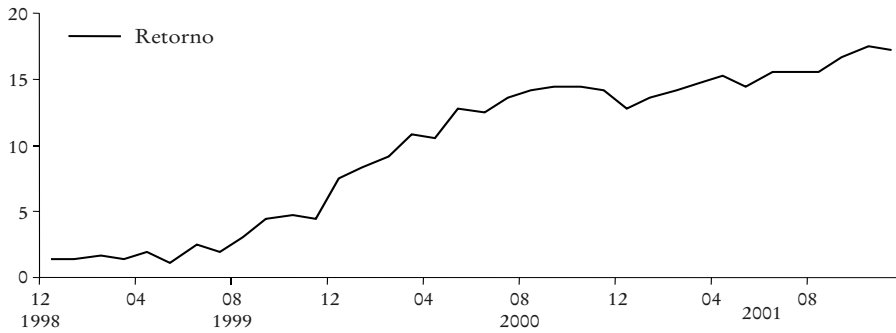
Adviértase a continuación la “adquisición” consistente de “ganancias” y la baja volatilidad de la curva de patrimonio acumulado correspondiente a la gráfica anterior —que en realidad representa el “éxito predictivo acumulado” de la estrategia utilizada por la red con PP simple para predecir el LEI de México. Las mismas observaciones son pertinentes para el caso de pre procesamiento gaussiano de los datos (PP gaussiano) presentadas en la gráfica 3.

GRÁFICA 1. *LEI observado vs LEI predicho por la red con preprocesamiento gaussiano de los datos<sup>a</sup>*

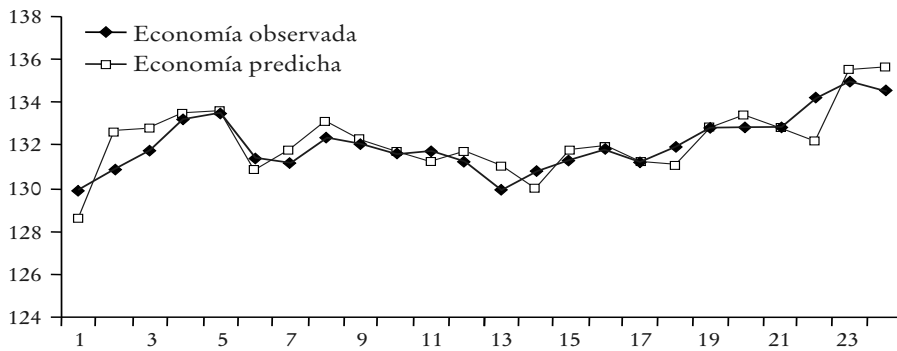


<sup>a</sup> *Excess Profit* de la red neuronal con PP gaussiano de los datos, 2.158; *Prob.* predictiva de la red neuronal de PP gaussiano de los datos, 0.969.

GRÁFICA 2. *Curva de patrimonio acumulado de la estrategia de compraventa de la red neuronal con PP simple*

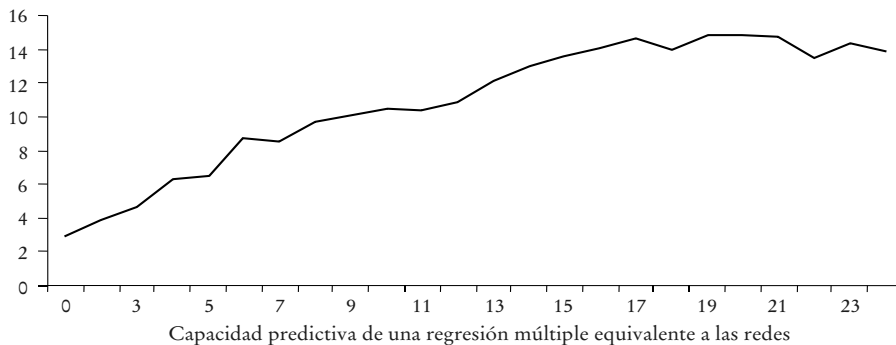


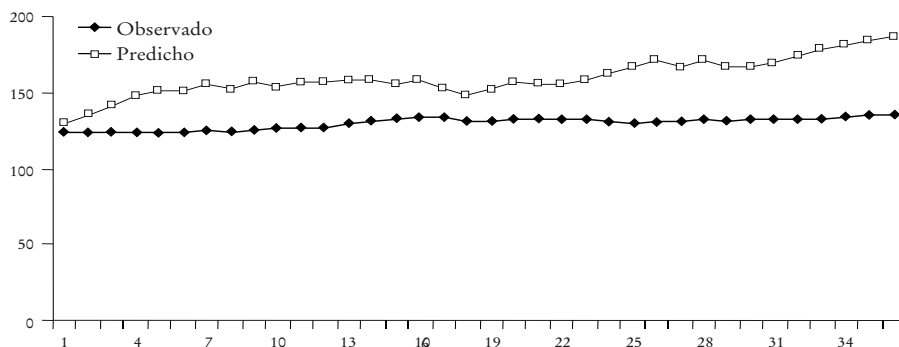
GRÁFICA 3. *LEI observado vs LEI predicho por la red con preprocesamiento gaussiano de los datos<sup>a</sup>*



<sup>a</sup> *Excess Profit* de la red neuronal con PP gaussiano de los datos, 2.158; *Prob.* predictiva de la red neuronal de PP gaussiano de los datos, 0.969.

GRÁFICA 4. *Curva de patrimonio acumulado de la red con PP gaussiano*



GRÁFICA 5. *Regresión múltiple en la economía observada vs la predicha*

<sup>a</sup> *Excess Profit* de la regresión, 1.93; *Prob. predictiva* de la regresión, 0.897.

Nótese en la gráfica 4 la “adquisición” consistente de “ganancias” y la baja volatilidad de la curva de patrimonio acumulado correspondiente a la gráfica 3 —que en realidad representa el *éxito predictivo acumulado* de la estrategia utilizada por la red con PP Gaussiano para predecir el LEI de México.

La regresión múltiple tuvo las mismas entradas (*inputs*) que las redes y fue efectuada con base en primeras diferencias de Durbin calculado en dos pasos (Trigo y Costanzo, 2007).

## CONCLUSIONES

En este artículo hemos constatado que la capacidad de las redes neuronales para predecir la dirección de la economía de México —representada por el LEI— depende no sólo de las variables independientes utilizadas, sino también de la calidad de esta última así como sus versiones simultáneas (suavizante y predictiva) que se puede hacer mediante del preprocesamiento de los datos con el uso de un proceso gaussiano alimentado por un índice de acciones y uno de bonos (ambos representativos del mercado mexicano). Dicho proceso gaussiano no sólo es capaz de suavizar simultáneamente el índice de acciones y el de bonos sin eliminar la dependencia (posiblemente no lineal) que podría existir entre las dos variables, sino que puede hacer lo mismo para cualquier número de variables independientes que se hubiesen utilizado en el análisis.

El poder predictivo de la red alimentada con datos suavizados por medio del proceso gaussiano es mayor que el de otros métodos, tal como lo muestra el cuadro 1 de este artículo.

CUADRO 1. Resumen de los resultados comparativos de las redes neuronales con PPS PPG & regresión para México<sup>a</sup>

<i>Red neuronal con PP aussiano vs red neuronal con PP simple vs regresión (según Anatolyev)</i>	<i>Excess Profit de la regresión</i>	<i>Prob. predictiva de la regresión</i>	<i>Excess profit de lar red neuronal de preprocesa- miento simple de la data</i>	<i>Prob. predictiva de la red neuronal de preprocesa- miento simple de la data</i>	<i>Excess profit de la red neuronal de preprocesa- miento gaussiano de los datos</i>	<i>Prob. predictiva de la red neuronal de prerocesa- miento gaussiano de los datos</i>
México	1.93	0.897	2.06	0.960	2.158	0.969

a La regresión múltiple tuvo las mismas entradas, que las dos redes y fue efectuada con base en primera diferencias de Durbin calculado en dos pasos (Trigo y Costanzo, 2007).

Efectivamente, no sólo la probabilidad predictiva de la red neuronal sino también el *Excess Profit* son mejores (0.969 y 2.158, respectivamente) cuando se preprocesa los datos utilizando el proceso gaussiano que cuando se preprocesa utilizando rezagos o promedios móviles, o cuando se usa simplemente una regresión lineal. Dichos resultados generalizan y mejoran significativamente los hallados en un trabajo que precede al presente de Trigo y Costanzo (2007). Es de hacer notar que no hemos encontrado en la bibliografía académica trabajos similares que utilicen métodos de suavizamiento no paramétricos (de tipo gaussiano) de los datos para entrenar redes neuronales. Dicha idea ha resultado ser útil y fácil de aplicar para mejorar la calidad de los pronósticos hechos por medio de estgos instrumentos.

### APÉNDICE 1

Este apéndice contiene dos ejemplos particulares: los pares de ETF [\$HUI, GLD] y [\$XOI, USO]. Los símbolos \$HUI y \$XLE representan los ETF (*Exchange Traded Funds*) de acciones de empresas relacionadas con la prospección, la exploración, la extracción y la comercialización del oro y de los hidrocarburos, respectivamente. Los símbolos GLD y USO representan los ETF del oro y del crudo en los Estados Unidos, respectivamente.

En primer lugar se evalúa cuán eficiente son los datos del fondo \$HUI preprocesados con promedios móviles para predecir el movimiento del fondo GLD utilizando redes neuronales; luego, cuán eficiente son los datos del fondo \$HUI preprocesados con métodos gaussianos para predecir el comportamiento del mismo fondo GLD, también utilizando redes neuronales, y por último se compararán ambos resultados para poner en evidencia la ventaja de utilizar un proceso gausiano en vez de un promedio móvil como preprocesador de datos en este caso particular.

Este mismo ejercicio se realiza con el par [ $\$XLE$ ,  $USO$ ] para mostrar un resultado análogo.

Aunque el “demostrar” formalmente que “el suavizamiento de los datos con procesos gaussianos es superior al suavizamiento con promedios móviles cuando se trata de mejorar las predicciones de las redes neuronales” va mucho más allá del alcance de éste artículo, los ejemplos que mostramos a continuación son prueba adicional de que, como en el caso del  $LEI$  de México, el resultado de alimentar la red neuronal con datos preprocesados utilizando procesos gaussianos es mejor —en términos predictivos— que el resultado obtenido cuando se alimenta la red con los mismos datos preprocesados aún con el “mejor” promedio móvil que se pueda hallar en cada caso.

Hemos observado que esto ocurre también en otros casos similares en los que el precio del índice de las acciones que dependen de un producto (*commodity*) dado es utilizado como entrada (*input*) para predecir el precio de ese producto —de hecho, de la observación de varios ejemplos similares a esos dos provino originalmente la idea planteada en el trabajo original.

*Ejemplo 1:* caso [ $\$HUI$ ,  $GLD$ ]. *i*) evaluación de la predicción del fondo  $GLD$  usando el fondo  $\$HUI$  preprocesado con proceso gaussiano: prueba de Anatolyev; EP  $PROB$  2.745491;  $PROB$ , 0.993958.

*ii*) evaluación de la predicción del fondo  $GLD$  usando el fondo  $\$HUI$  preprocesado con promedio móvil: Anatolyev; EP, 2.745491;  $PROB$ , 0.960803.

*Ejemplo 2:* caso [ $\$XLE$ ,  $USO$ ]. *i*) evaluación de la predicción del fondo  $USO$  usando el fondo  $\$XLE$  preprocesado con proceso gaussiano: test de EP, 2.148366,  $PROB$ , 0.968315. *ii*) Evaluación de la predicción del fondo  $USO$  usando el fondo  $\$XLE$  preprocesado con promedio móvil: prueba de Anatolyev; EP, 1.103543;  $PROB$ , 0.730208.

## APÉNDICE II

En este artículo se usó la familia paramétrica exponencial cuadrada como kernel. Esta es:

$$k(x_i, x_j) = \frac{2}{f} \exp \left( -\frac{1}{2} (x_i - x_j)^t M (x_i - x_j) \right) \frac{2}{n} ij$$

Aunque existe una variedad de funciones de covarianzas que pueden ser utilizadas, seleccionar la exponencial cuadrada tiene como principal atractivo garantizar que las predicciones son invariantes con cambios de origen en el espacio de las entradas (Williams y Rasmussen, 1996 y Rasmussen, 1996).



## REFERENCIAS BIBLIOGRÁFICAS

- Anatolyev, Stanislav (2005), "A Trading Approach to Testing for Predictability", *Journal of Business and Economic Statistics*, Vol. 23, No. 4, pp. 455-461.
- Jagric, Timotej (2003), "A Nonlinear Approach to Forecasting with Leading Economic Indicators", *Studies in Nonlinear Dynamics & Econometrics* 7, núm. 2.
- Jurik, Mark (1994), "Back Percolation, Assigning Local Error in Feed-forward Perception Methods", *Braincel*, New Haven, Promised Land Technologies, Inc.
- Lahiri, Kajal, y Geoffrey Moore (1991), *Leading Economic Indicators: New Approaches and Forecasting Records*, Nueva York, Cambridge University Press.
- Parisi F., Antonino, Franco Parisi F. y José Luis Guerrero C. (2003), "Modelos predictivos de redes neuronales en índices bursátiles", *EL TRIMESTRE ECONÓMICO*, vol. LXX (4), 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.
- Rasmussen, C. E. (1996), "Evaluation of Gaussian Processes and Other Methods for Non-linear Regression", tesis doctoral, University of Toronto, disponible en (<http://www.kyb.mpg.de/publications/pss/ps2304.ps>) y ([www.es.utoronto.ca/~carl/](http://www.es.utoronto.ca/~carl/)).
- , y Christopher K. I. Williams (2005), *Gaussian Processes for Machine Learning*. Cambridge, Massachusetts y Londres, Inglaterra, MIT Press.
- Ruggiero, Murray (1996), "Using Correlation Analysis to Predict Trends", *Futures Magazine*, febrero, pp. 46-49.
- (1997), *Cybernetic Trading Strategies*, John Wiley & Sons, Inc.
- (1998), "Intermarket Analysis is Fundamentally Sound", *Futures Magazine*, abril, pp. 58-62.
- (2001), "Intermarket Analysis & Economic Forecasting", *Futures Magazine*, marzo, pp. 60-62.
- (2001), "The Economy & Markets: Trading the Relationship", *Futures Magazine*, abril, pp. 60-62.
- Tkacz, G. (2000), "Neuronal Network Forecasting of Canadian GDP Growth", *The International Journal of Forecasting*.
- Trigo, Lorenza, y Sabatino Costanzo (2007), "Redes neuronales en la predicción de las fluctuaciones de la economía a partir del movimiento de los mercados de capitales", *EL TRIMESTRE ECONÓMICO*, vol. LXXIV (2), núm. 294, pp. 415-440.
- Walczak, 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.

- Cerpa, Narciso, y Steven Walczak (1999), "Heuristic Principles for the Design of Artificial Neural Networks", *Information and Software Technology*, vol. 41, núm. 2, pp. 109-119.
- , 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.
- Williams, C. K., y C. E. Rasmussen (1996), *Gaussian processes for regression*. D. S. Touretzky, M. C. Mozer, y M. E. Hasselmo (comps.), *Advances in Neural Information Processing Systems 8*, MIT Press.
- Wolberg, John R. (2000), *Expert Trading Systems*, Nueva York, John Wiley & Sons, Inc.
- Wolpert, L. Robert, Katja Ickstadt y Martin B. Hansen (2003), "A Nonparametric Bayesian Approach to Inverse Problems", J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith y M. West (comps.), *Bayesian Statistics*, Oxford University Press, 2003.