

Análisis y pronósticos de bonos y acciones de Argentina

Un enfoque no lineal usando redes neuronales

Marcelo R. Ferreyra

Resumen: En este documento se detallan los diferentes pasos que se siguieron para el análisis y el desarrollo de un sistema orientado a la compra y venta de bonos y acciones del mercado Argentino. El estudio se realizó principalmente sobre los bonos Brady PAR y FRB y las acciones TEAR, YPF, PEREZ, GALICIA y el índice Merval.

Introducción

Existen varias estrategias para operar en el mercado, aunque una de las más usadas es una combinación entre el análisis técnico, las noticias y los fundamentos. Estos últimos dan una base de confianza o desconfianza que puede suavizar o acentuar una noticia inesperada, aunque frente a la ausencia de noticias, el análisis técnico es una referencia muy fuerte para seguir.

El enfoque de esta investigación está orientado al desarrollo de modelos validados estadísticamente y luego seguidos en tiempo real operando de acuerdo a la señal producida. Normalmente se les da el nombre de sistemas mecánicos para significar que la única forma de operar con estadísticas de apoyo tales como máximas pérdidas esperadas, máximo número de pérdidas seguidas esperadas, etc., es haber medido correctamente el modelo en el pasado y luego apegarse a él hasta que su desempeño, contrastado con su pasado, caiga en un determinado nivel.

Medir correctamente el modelo no es una tarea trivial, fundamentalmente debido a la poca cantidad de datos disponibles y relevantes para el desarrollo del mismo. Es así que deben usarse diferentes técnicas de validación para obtener estadísticas confiables (Pardo 1992).

Un modelo puede estar basado sobre diferentes técnicas, tales como modelos AR y sus derivados, análisis técnico, regresión lineal, redes neuronales, algoritmos genéticos, etc. Este documento enfoca las redes neuronales, debido a su capacidad de encontrar relaciones no lineales en datos con ruido.

La primera parte de este informe trata acerca del análisis del mercado de capitales en cuanto a su autocorrelación o independencia de sus datos. Después de dar una breve reseña sobre la Hipótesis del Mercado Eficiente, se verán otras técnicas que intentan encontrar la presencia de relaciones no lineales, tales como el Análisis Reescalado (Rescaled Range) y la prueba BDS. Para un mejor entendimiento de estos métodos se repasan conceptos estadísticos como la distribución normal y la correlación.

En la segunda parte se presenta lo básico sobre redes neuronales y sus aplicaciones financieras. Un análisis un poco más detallado se realiza sobre la aplicación a series de tiempo, ya que este es el objetivo final.

Por último se detallan los resultados del modelo aplicado tanto a los bonos como a las acciones.

Enfoques tradicionales

Un buen punto para comenzar cuando se quiere llegar a un objetivo, es saber donde está uno parado. En nuestro caso esto significa conocer cuáles son las prácticas más usadas para llevar adelante una cartera de activos y fundamentalmente cuál es el paradigma aceptado por la comunidad que integra el mercado de capitales internacionales.

El primero, y quizás más grande dilema que surge y parece quedar sin respuesta es la predictibilidad del mercado. ¿Los precios se mueven al azar?. Depende el punto de vista. Un académico dirá que sí, un trader dirá que no. ¿Quién tiene la razón?

No existe una respuesta simple a esta cuestión, ya que muchos estudios referentes a correlaciones lineales y no lineales son cuestionados debido a los métodos usados. Además, aunque existan personas en el mundo que venzan el mercado una y otra vez, como es el caso de Warren Buffet o Peter Lynch, está dentro de las probabilidades el que lo hagan por azar. Al menos eso dicen los defensores del mercado eficiente, aunque no sepan cómo medir estas probabilidades.

En 1959 Osborne afirmó que la diferencia de precios del mercado sigue un camino al azar (Random Walk) y en 1965 Eugene Fama formalizó la Hipótesis del Mercado Eficiente (EMH). Un mercado eficiente se define como un gran número de participantes racionales que compiten tratando de predecir el mercado de forma tal de maximizar sus ganancias, y en donde la información importante está a disposición de todos. En un mercado eficiente el precio de un activo será una buena estimación de su valor intrínseco (Fama 1965).

En esencia, la EMH dice que el mercado está compuesto por demasiada cantidad de personas para estar errado. Cualquier noticia se descuenta en los precios inmediatamente, de manera tal que el activo adquiere el precio justo, sin ventajas para nadie.

La EMH y los caminos al azar justifican el uso de las herramientas estadísticas tradicionales para el estudio de los cambios de precios que supuestamente se dan al azar y con distribución log-normal.

En estos últimos 30 años ha habido una gran cantidad de informes que apoyan la eficiencia del mercado, pero también han surgido muchos investigadores cuyos trabajos sugieren lo contrario. Véase Vaga (1994), Peters (1991), Peters (1994), Brock (1991), Tsibouris (1995). Por supuesto, las controversias continúan hoy en día, y es probable que uno de los motivos sea que el ruido que contienen las series de precios esconde los patrones que puedan existir.

La distribución de los retornos

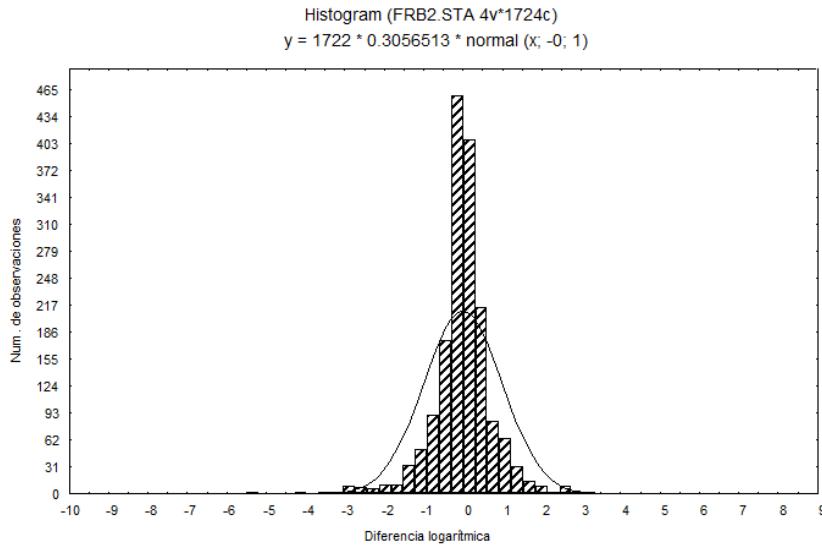
Si tenemos una serie histórica de precios, podemos analizar la distribución de sus diferencias de precios diarios de la siguiente manera:

Para evitar que las diferencias de precios al principio de la serie sean menores que al final, debido a un incremento de los precios (ej. El Dow Jones a principios del 87 era de unos 2.000 puntos), tomamos diferencias porcentuales o diferencias de los logaritmos de los precios (de ahí distribución log-normal).

Intuitivamente sabemos que habrá mayor cantidad de cambios de precios “pequeños” que “grandes” y que estas diferencias serán tanto positivas como negativas. Si construimos un histograma con estas diferencias obtendremos una aproximación a una curva de distribución llamada normal o de gauss.

El siguiente gráfico muestra la distribución de las diferencias logarítmicas del FRB en columnas que indican la cantidad de observaciones que se encontraron en cada rango. La curva superpuesta es la distribución normal.

Las diferencias están normalizadas, a cada valor se le resta la media y se lo divide por la desviación estándar. Para comenzar nuestro análisis centrémonos en la curva ideal de distribución, en este caso la curva normal. Notamos que la curva es simétrica y el máximo valor corresponde a las diferencias de precios cercanas a cero (en realidad el valor cero da el máximo), y a medida que nos acercamos hacia los extremos su valor va disminuyendo. A pesar de que no se puede ver en el gráfico, los extremos de la curva jamás tocan la línea de cero, aunque más allá de ± 3 su valor es ínfimo.



Si suponemos que el área total que encierra la curva es 1, entonces para calcular cuál es la probabilidad de encontrar una diferencia de precios en un rango de valores, sólo hay que calcular el área correspondiente. Por ejemplo la probabilidad de que ocurran cambios de precios no más allá de ± 2 desviaciones estándar es de 95.44% y no más allá de ± 3 de 99.74%.

Si ahora comparamos la distribución de nuestra serie de precios del FRB con respecto a la normal, encontramos que es mucho más alta en el centro y que los extremos se extienden con frecuencias observadas mucho más allá de 2 ó 3 desviaciones estándar. De hecho, de acuerdo a la distribución normal, en las 1722 diferencias de precios que corresponden al período analizado del FRB (Agosto 92 – Mayo 99) deberíamos haber encontrado unas 5 cuyo valor absoluto fuese mayor que 3 desviaciones estándar, pero encontramos 31.

Existen algunas pruebas estadísticas que comparan cuantitativamente una distribución con respecto a la normal. Así, la kurtosis mide cuánto más alto es el pico en cero y cuánto más anchas son las colas. Un valor de 0 indica una distribución normal. El FRB tiene una kurtosis de 24, muy lejos de la normal.

La prueba de Shapiro-Wilk mide la correlación del número de observaciones ordenadas por desviaciones estándar de la distribución normal y de la distribución a analizar. El resultado W para este caso es de 0.782, valor demasiado bajo (W va de 0 a 1) que sugiere una distribución que se aparta en forma significativa de la normal. Para más información véase Frees (1996).

En resumen, podemos ver que la distribución de los retornos porcentuales del FRB no puede ser considerada normal. Lo mismo pasa con otros bonos y con las acciones o el índice Merval.

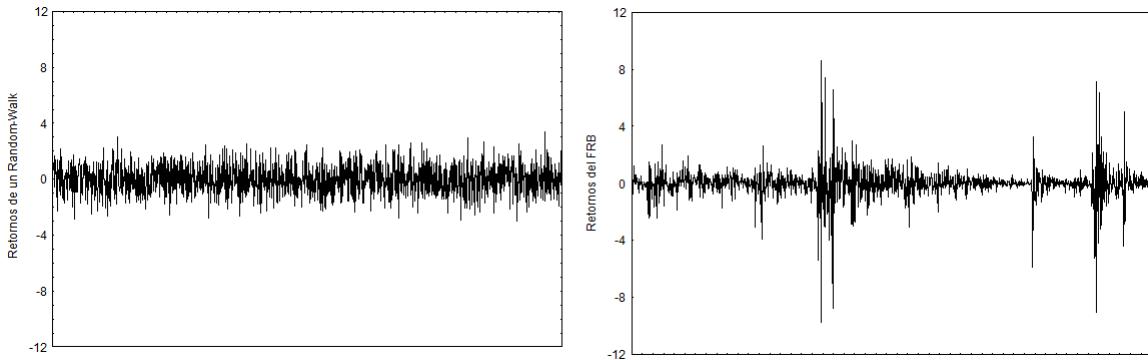
La volatilidad

Otro aspecto interesante de los retornos es su dispersión o volatilidad, que se mide con la desviación estándar. En el caso de una serie de precios que siguiera un camino al azar (random walk), podríamos medir la volatilidad de sus retornos en diferentes puntos de la misma y encontraríamos que siempre es la misma.

Cuando se miden en diferentes puntos varios estadísticos, por ejemplo la media y la desviación estándar, y se encuentran los mismos valores, se dice que la serie es estacionaria, y su análisis es más simple que en el caso de una serie de precios no estacionaria. De hecho normalmente se aplica algún procedimiento para convertir una serie no estacionaria en estacionaria, si es que es el caso. Por ejemplo, la media de un índice de precios como el Merval o el Dow Jones, será distinta según el período que se mida. Para evitar este comportamiento se la diferencia y se trabaja con estas diferencias o retornos logarítmicos o porcentuales como ya se explicó.

Si en vez de medir la media, medimos la desviación estándar o volatilidad en la serie de retornos de un bono, una acción o un índice encontraremos cambios abruptos en diferentes períodos. Es precisamente este cambio de las volatilidades lo que hace tan complejo modelar un activo financiero.

Los dos gráficos que siguen muestran, el de la izquierda los retornos de un camino al azar y el de la derecha los retornos del FRB. Se pueden observar períodos de alta volatilidad y períodos de tranquilidad. El último corresponde a enero del 99, respondiendo a la devaluación de Brasil.



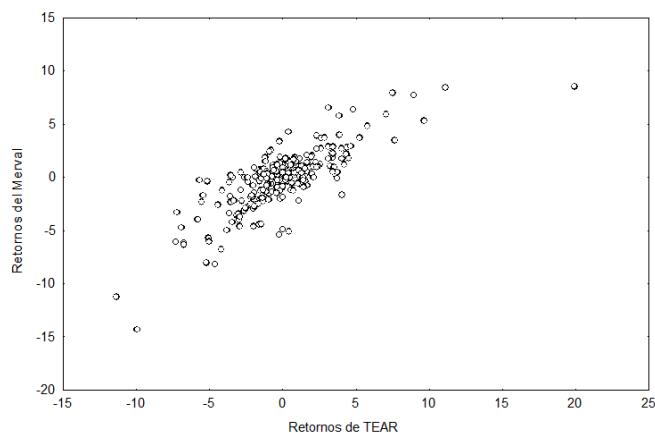
En estos gráficos también se puede apreciar que un precio real no puede responder a un camino al azar. Para más información véase Mandelbrot (1999).

Si contamos con una serie de retornos lo suficientemente larga, digamos unas 2.000 observaciones, y vamos midiendo la volatilidad desde el principio y tomando cada vez mayor cantidad de datos podemos verificar si hay algún valor al que se converge. En el caso de un camino al azar, rápidamente se converge a la volatilidad con la que fue generado, pero en el caso de una serie de precios reales no encontraremos ningún valor de convergencia. Cuando sucede esto se dice que los datos tienen volatilidad infinita. Véase el capítulo 14, "Fractal Statistics" en Peters (1994).

Las correlaciones

El término correlación sugiere dos variables (como mínimo) que se encuentran relacionadas de alguna manera. Por ejemplo podríamos suponer que los retornos de una acción que integra un índice están correlacionados con los retornos de este índice. La relación que queremos encontrar o que suponemos que existe es: el retorno de la acción es positivo, entonces el retorno del índice también lo es, y viceversa.

Una manera de comprobarlo es dibujando un gráfico de dispersión (scatter plot), como el siguiente, que corresponde a los retornos diarios de TEAR y Merval durante el año 1998.

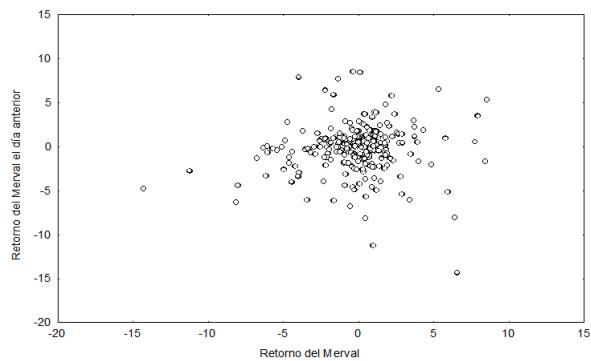


Con cada par de valores se coloca un punto en el gráfico. Aquí vemos que, cuando los retornos de TEAR tienden a ser mayores, también sucede lo mismo con los del Merval. El estadístico más usado para medir la correlación entre dos variables es el de Pearson, que mide las relaciones lineales entre las variables y en este caso es 0.798. El coeficiente de correlación r puede tomar valores entre -1 y 1 . Cuando es 1 existe una perfecta correlación lineal, si es -1 significa que cuando una variable sube la otra baja y si es 0 no existe ninguna correlación lineal. Cabe aclarar que puede ser que r sea 0 aunque exista una correlación no lineal perfecta como es el siguiente caso:

i	1	2	3	4	5
xi	-2	-1	0	1	2
yi	4	1	0	1	4

El coeficiente de correlación entre los 5 valores que toman las variables x e y en este ejemplo es 0 aunque existe una perfecta relación entre ellas, $y = x^2$.

Para saber si los retornos de los precios de hoy influirán sobre los de mañana, hallamos la autocorrelación, que no es más que una correlación entre una serie de números y la misma serie desplazada una posición. Por ejemplo el siguiente gráfico muestra la autocorrelación de los retornos del Merval en el año 1998.



Se puede observar que no es tan evidente una relación lineal, de hecho el coeficiente r es de 0.083. ¿Cómo sabemos que este coeficiente no es significante? Por otro lado, ¿cómo sabemos que el r de 0.798 hallado entre los retornos de TEAR y Merval no es casual?. Si queremos estar seguros en un 95% entonces calculamos la correlación máxima que tendrían dos series al azar por medio de la siguiente fórmula:

$$r_{max} = 2/\sqrt{n}$$

En donde n es la cantidad de datos. Esto significa que si tuviéramos 100 pares de series aleatorias no sería extraño encontrar que 5 de ellas tienen un coeficiente de correlación mayor que r_{max} sólo por azar. El valor 0.083 se calculó basándose en 247 valores, por lo que para aceptar un coeficiente estadísticamente significativo (a un 95% de confianza), éste debería ser mayor de 0.127. Como no lo es, decimos que el r encontrado no es significante (Gottman 1981).

Podríamos probar desplazamientos mayores a un día (¿el precio de ayer influye sobre el de hoy?), pero la siguiente tabla muestra que no existen correlaciones significativas.

	1 día	2 días	3 días	4 días	5 días	6 días
r	0.083	-0.087	0.006	0.088	0.11	-0.106

Los fractales

Por medio de la geometría fractal se pueden describir las formas naturales de una manera más precisa que a través de la geometría euclíadiana. “Un fractal es un objeto en el cual las partes están, de alguna manera, relacionadas con el todo” (Peters 1991). “Un fractal es una forma geométrica que consiste en un motivo que se repite a sí mismo en cualquier escala a la que se observe” (Sametband 1994, p 65). Pensemos en un árbol. Cada rama contiene a su vez ramas más pequeñas que contienen a su vez ramas más pequeñas hasta llegar a la hoja. Si nos dan a observar dos gráficos del mercado, uno diario y otro semanal, pero sin indicar la escala, no podremos identificarlos. A diferentes escalas, el mercado se ve similar; o dicho de una manera más precisa, tiene características estadísticas similares. La mayoría de las formas naturales y las series de tiempo se describen muy bien con fractales.

Una de los aspectos de los objetos fractales es que su dimensión normalmente no es un número entero. Pensemos en una hoja de papel. Tiene largo y ancho, por lo que su dimensión es sin lugar a dudas 2. Ahora hagamos un bollo con ella. ¿Cuál es su dimensión?. Ya no es más 2, porque sale del plano, pero tampoco es 3 porque no es un sólido. Su dimensión es un número que está entre 2 y 3.

Una serie de datos aleatorios con distribución normal tiene la capacidad de llenar todo el plano en el que se dibuja, por lo que su dimensión es 2. En cambio, si en la serie, los datos no son totalmente independientes, sino que existe alguna autocorrelación, entonces esta serie no podrá ocupar todos los puntos del plano porque su valor futuro está condicionado por el pasado. En este caso la dimensión será un valor entre 1 y 2.

Una de las maneras de medir la dimensión de una serie de datos es a través del exponente de Hurst que es el tema de la siguiente sección.

El exponente de Hurst

Alrededor de 1907 y por unos 40 años, el hidrólogo Harold E. Hurst trabajó en el proyecto de una represa en el río Nilo en Egipto. La política de apertura y cierre de las compuertas es algo bastante complejo, debido a que durante épocas lluviosas no deben generarse inundaciones y en las sequías no debe dejarse el río sin agua. Hurst pensó en un principio que sería lógico suponer que los factores que determinan la cantidad de agua al año (las lluvias) siguen un camino al azar. Cuando él decidió probar esta hipótesis, nos dió una nueva herramienta estadística: el exponente de Hurst (Peters 1991). Éste puede distinguir series al azar de aquellas que no lo son.

Básicamente el exponente de Hurst, H de aquí en más, mide cómo varía la diferencia entre la máxima y la mínima desviación acumulada respecto de la media. Si la serie corresponde a un camino aleatorio, entonces esta variación deberá ser función de la raíz cuadrada del tiempo. Esto ya había sido notado por Einstein cuando estudió el movimiento browniano que describe el movimiento errático que realiza una partícula suspendida en un fluido. Einstein encontró que la distancia que cubre la partícula es proporcional a la raíz cuadrada del tiempo en que se mide (Lavenda).

$$R/S = (a * N)^H$$

R es el rango o la diferencia entre la máxima y la mínima desviación acumulada respecto de la media. S es la desviación estándar, por lo que a fin de comparar diferentes series de datos se calcula el rango “reescalado” (rescaled range) R/S , que será proporcional a la cantidad de observaciones elevado a la H .

Si H es 0.5 entonces estamos frente a una serie de datos al azar. Caso contrario la serie de datos contiene correlaciones lineales o no lineales, de corto y/o largo alcance.

Existe una diferencia muy grande entre H y el coeficiente de correlación r . H mide relaciones tanto lineales como no lineales y además no se limita a relaciones de datos cercanos sino que puede detectar correlaciones de largo alcance. Cuando hablamos de relaciones no lineales queremos significar que los cambios en una variable no son proporcionales a los cambios de la otra.

Otros dos beneficios de H son que se puede usar con series cuya distribución no es normal y además es capaz de detectar ciclos no periódicos.

Si H es mayor que 0.5 decimos que la serie es persistente y contiene tendencias. Si, por el contrario, H es menor que 0.5 decimos que es antipersistentes ya que cambia más a menudo que una serie aleatoria. Hurst encontró un valor de 0.90 en sus estudios del Nilo e intrigado por este valor tan lejos del esperado estudió otros fenómenos naturales encontrando en todos ellos valores de H de alrededor de 0.70.

Estos valores de H indican una dependencia entre el pasado y el futuro. Cada observación lleva una memoria de todos los sucesos ocurridos en el pasado, al menos teóricamente.

Peters (1991, 1994) midió el H de varias series de datos de EEUU y algunos países de Europa y encontró valores que difieren de 0.5 en forma significativa, aunque hay algunos estudios sobre el posible sesgo que pueden ser introducidos en los cálculos del mismo (Moody 1995). Peters también detectó un ciclo no periódico (que no se repite con la misma frecuencia) de alrededor de 4 años sobre el Dow Jones.

Los estudios realizados sobre el Merval, el FRB, y algunas de las acciones más líquidas muestran valores de H que si bien se apartan de 0.5 no llegan a ser significativos. Uno de los motivos puede ser la poca cantidad de datos que el cálculo de H requiere. En el apéndice se encuentran detalles sobre el algoritmo de cálculo y los valores encontrados.

Mandelbrot demostró que la inversa de H es la dimensión fractal. Una serie aleatoria tiene un H de 0.5 por lo que su dimensión es $1/0.5=2$. Si H es mayor que 0.5, la dimensión de la serie de datos será un número menor que 2.

Los procesos caóticos

Cuando decimos que un sistema se comporta de modo aleatorio, en general estamos asumiendo que existen tantos factores o grados de libertad a considerar que es imposible predecir el resultado. Dicho de otra manera, es aleatorio porque es muy complejo. Desde este punto de vista un sistema puede ser modelado como totalmente predecible, en base por ejemplo a las leyes de Newton, o como totalmente al azar como cuando se modela un gas compuesto por un número muy grande de partículas usando la mecánica estadística. Pero hacia 1975¹ gran cantidad de científicos alrededor del mundo comenzaron a notar un nuevo tipo de movimiento que hoy llamamos Caos (Alligood 1997). Este movimiento o evolución de un sistema parece aleatorio pero no lo es. Es más, pasará como aleatorio todas las pruebas tradicionales de estadística. Tampoco son necesarios muchos grados de libertad para generarlo. Sistemas muy simples pueden generar caos en que la total predictibilidad está disfrazada de aleatoriedad. La clave está en la sensibilidad a las condiciones iniciales. Esto significa que si dos sistemas regidos por las mismas leyes hacen su punto de partida en puntos ligeramente diferentes, al cabo de cierto tiempo los estados en que se encuentran serán completamente distintos. No interesa qué tan cercanos estén los puntos de partida, esta diferencia aunque infinitesimal, se amplificará y en cierto tiempo los sistemas diferirán completamente.

Consideremos las implicaciones que tendría la sensibilidad a las condiciones iniciales si el mercado se comportara en forma caótica. Cualquier modelo que realicemos tendrá en cuenta precios anteriores del instrumento a predecir y/o de otros mercados. Estos precios contienen ruido que viene de valores mal tomados, valores faltantes y proyectados con los anteriores, etc., por lo que para comprobar la robustez del modelo se deberían probar pequeñas variaciones de los datos. Si los resultados son similares, suponemos que el modelo es robusto, pero bajo la hipótesis de un mercado caótico y sensible a las condiciones iniciales, los distintos pronósticos a mediano o largo plazo diferirán completamente.

¹ Los fenómenos no lineales con alta sensibilidad a las condiciones iniciales no eran desconocidos por los grandes matemáticos y físicos del siglo pasado, pero al no contar con computadoras, su análisis era muy engorroso y se hacía impracticable. Véase Sametband (1994) para una excelente introducción a la complejidad.

Lorenz (1995) fue uno de los primeros en aplicar las leyes del caos en la predicción de las condiciones meteorológicas. Su frase “el aleteo de una mariposa en Londres puede generar un huracán en Centro América”, ejemplifica la sensibilidad a las condiciones iniciales. Predecir si lloverá un determinado día dentro de 20 días es y será imposible, independiente de la tecnología usada, porque jamás se podrán medir las condiciones iniciales con la suficiente exactitud requerida.

En septiembre de 1987 se reunieron en el Santa Fe Institute unas 20 personas de áreas tan variadas como economía, física, biología y computación para hablar de “la economía como un sistema complejo que evoluciona” y a partir de aquí fueron aumentando los trabajos en dinámica no lineal, caos y complejidad como modelos económicos y aplicados al análisis de datos.

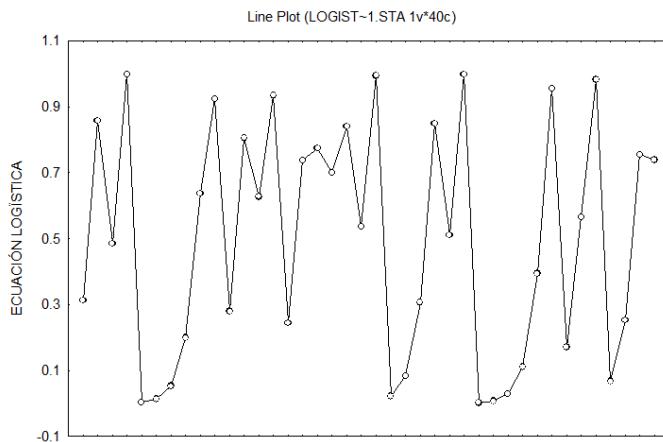
Consideremos un modelo simple del mercado de capitales en el que precios bajos atraen la atención de compradores, mientras que altos precios atraen a los vendedores (Baestaens 1994). Supongamos que los precios aumentan con la demanda a una tasa a de la siguiente manera:

$$P_{t+1} = a * P_t$$

Supongamos también que mientras el precio sube los vendedores comienzan a presionar reduciendo el precio a una tasa proporcional al cuadrado del precio anterior. Podemos expresar este modelo poco realista pero instructivo como sigue (Peters 1991):

$$P_{t+1} = a * P_t - a * P_t^2 = a * P_t * (1 - P_t) \quad (1)$$

Si a vale 2 y el precio inicial 0.3, iterando (1), al cabo de algunos ciclos alcanzaremos el valor 0.5. Es más, no interesa con qué precio comencemos (mayor que 0 y menor que 1), tarde o temprano terminaremos en 0.5. Es como si la ecuación (1), al menos con $a=2$, tuviese un punto de atracción en 0.5. Si ahora a pasa a valer 3 en vez de 2, veremos que ya no hay un solo punto al cual converge, sino dos. A medida que aumenta el parámetro a notaremos que los valores de convergencia se irán duplicando, hasta que $a=4$ en el que no existe ningún valor de convergencia, sino que cada nueva iteración parece dar como resultado un número al azar.



Ecuación logística con $a=4$ (símil aleatorio)

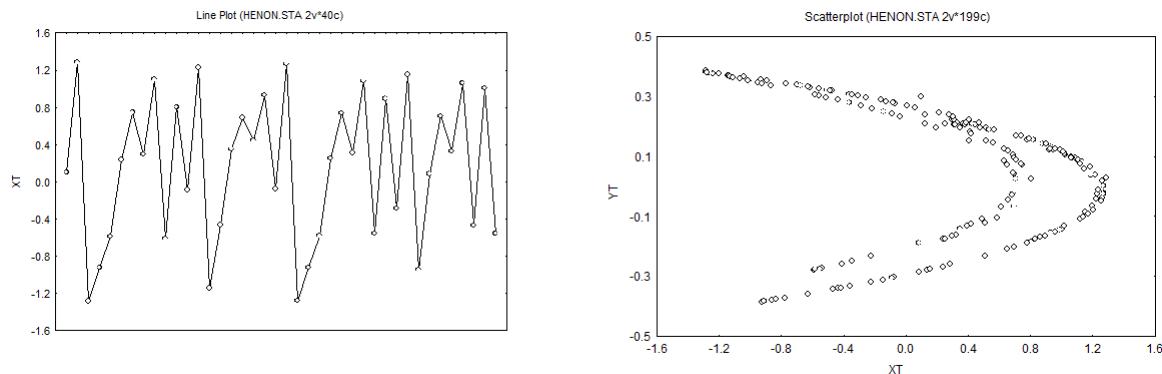
En este modelo los precios no se mueven por factores externos (noticias inesperadas) sino debido a la no linealidad misma de la serie. A esta ecuación se la conoce como Logística y es una de las más simples ya que su valor depende solamente del valor anterior. No todas las ecuaciones dependen de un solo valor. Un modelo un poco más real del mercado seguramente dependerá de varias variables.

Para estudiar el comportamiento de una serie caótica se utiliza un espacio de fases en el que cada variable que contribuye al sistema se usa como eje de un gráfico de dispersión. Por ejemplo la ecuación de Henon se genera sobre la base de dos variables:

$$x_{t+1} = 1 + y_t - a * x_t^2$$

$$y_{t+1} = b * x_t$$

Si graficamos la variable x cuando $a=1.4$ y $b=0.3$ vemos que el comportamiento parece aleatorio aunque su valor depende del valor anterior de x y del valor anterior de y . Cuando estudiamos el espacio de fases a través de un gráfico de dispersión de dos dimensiones con las variables x e y , vemos un objeto al que llamamos *atractor extraño*, en particular éste es el atractor de Henon. Los valores de la serie dependerán del valor inicial de x e y , pero el atractor siempre se verá igual, no importa que valores iniciales elijamos.



Si prestamos atención a los detalles del atractor veremos que es un fractal cuya dimensión es 1.26. Es más que una línea y menos que un plano, tal como una serie de retornos del mercado de capitales (Peters 1991).

Por supuesto, no siempre contamos con las variables del sistema que genera una determinada serie de datos, como es el caso del mercado de capitales. Cuando este es el caso, se pueden usar como variables para la reconstrucción del espacio de fases, la serie a analizar y sucesivos valores anteriores de la misma (lags). De esta manera se podría, por ejemplo, reconstruir el espacio de fases de la ecuación de Henon usando $x(t)$ y $x(t-1)$ en vez de $x(y)$ e $y(t)$.

Si en el espacio de fases de una serie de datos encontramos un atractor, entonces significa que cada punto de la serie no tiene la total libertad de ubicarse en cualquier región de este espacio, por lo que concluimos que los datos no son independientes.

El BDS test

En enero de 1983 la revista Physical Review Letters publicó un artículo en el que P. Grassberger y I. Procaccia comentan su creación de una nueva manera de medir la dimensión de un atractor a través de la correlación integral que indica la probabilidad de que un par de puntos de un atractor estén dentro de una distancia R uno de otro. Este método está basado en que el atractor que se forme crecerá hasta que la dimensión del espacio de fases sea mayor que la de éste. Una vez alcanzada esta dimensión, ésta representará la cantidad mínima de variables que se necesitarán para modelar la serie de datos.

Brock, Hsieh y LeBaron (1991) utilizaron la correlación integral como base para un test al que llamaron BDS test y que mide cuánto se aparta una determinada serie de datos de una independiente e idénticamente distribuida (IID). Para estimar intervalos de confianza, realizaron pruebas de Montecarlo sobre series IID de diferentes longitudes y publicaron las tablas correspondientes.

Ellos midieron varias series pertenecientes a mercados de capitales, entre ellas el S&P500 y encontraron períodos en que estas series se apartaban en forma significativa de una serie IID.

Si realizamos el BDS test sobre el Merval o los bonos, también encontraremos períodos significativamente diferentes de una serie IID pero, antes de sacar conclusiones, analicemos estos resultados con detenimiento.

En Brock (1991) página 99, se analizaron dos períodos del S&P500, 1928-39 (*sp1*) y 1950-62 (*sp2*), encontrándose que *sp1* es significativamente distinto de una serie IID, pero *sp2* no lo es. Si sobre estos mismos períodos analizamos la varianza notaremos que en *sp1* varía entre 1 y 3, pero en *sp2* varía de 0.32 a 0.85. Estos cambios en la varianza (volatilidad) en *sp1*, pero no en *sp2*, nos sugieren que lo que realmente nos dice el BDS test es que los retornos son claramente no IID sólo en *sp1*.

En conclusión, el BDS test aplicado a mercados financieros con cambios bruscos de volatilidades sólo nos indica si la serie es o no IID, pero no nos dice nada acerca de la independencia de sus datos, como lo sugieren algunos autores (Lin 1997).

Resumiendo los análisis

Podemos resumir todo lo visto hasta ahora en los siguientes puntos:

1. La distribución de los retornos logarítmicos está muy lejos de ser normal. Los cambios abruptos de precios se dan mucho más a menudo de lo esperado según la distribución normal. En el caso del FRB vimos que cambios de precios más allá de 3 desviaciones estándar son 5 veces superiores a lo esperado.
2. Los retornos diarios no muestran un coeficiente de correlación lineal significativo. Parece que no existiera ninguna relación lineal en los datos.
3. El análisis R/S no muestra un valor de H significativamente diferente de 0.5, por lo que a primera vista pareciese que no existen correlaciones lineales o no lineales de corto o largo alcance. Una explicación de esto podría ser que la cantidad de datos es insuficiente con respecto a la cantidad normalmente necesaria para este test.
4. El BDS test muestra que los datos contienen períodos que no son IID, aunque esto no nos dice nada acerca de la independencia de los mismos.

En estos cuatro puntos no encontramos demasiada evidencia para afirmar o negar independencia de algún tipo, por lo que el siguiente paso será modelar las series de datos y analizar su desempeño frente a períodos posteriores a los usados para parametrizar los modelos.

Considerando la posibilidad de que el mercado contenga relaciones no lineales una de las herramientas que se ajustan son las redes neuronales, no sólo por tener la capacidad de encontrar este tipo de relaciones, sino por su robustez frente a datos con demasiado ruido.

Si podemos validar en forma confiable un modelo de redes neuronales aplicados a los mercados ya sea de bonos como de acciones, podremos concluir que los precios no se comportan de forma independiente.

Para esto primero veremos algunas nociones de las redes neuronales para luego pasar directamente a la problemática de construir modelos y medir su desempeño.

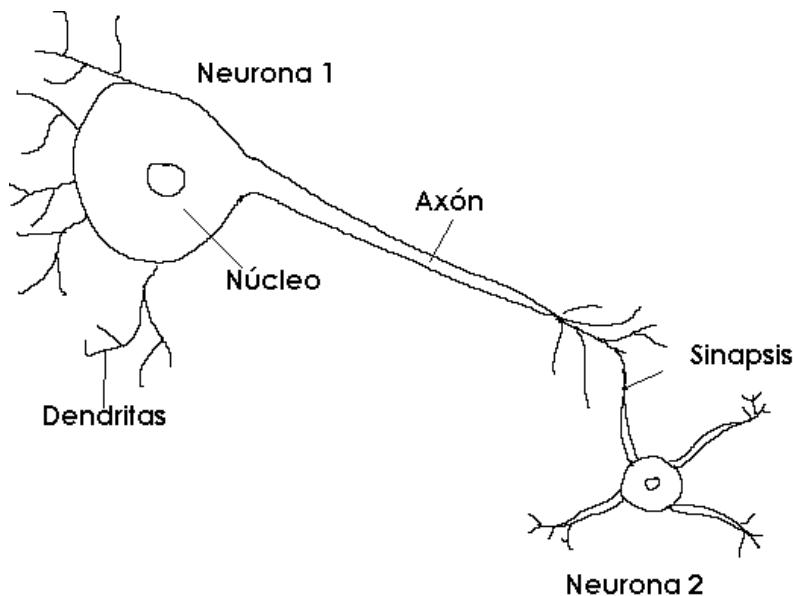
Redes neuronales

Desde los comienzos de las redes neuronales con la publicación del paper de McCulloch y Pitts, pasando por el perceptrón de Rosenblatt en 1958, ha habido un enorme avance no solo en lo que se refiere a potencia de computación sino también en los diferentes tipos de redes y algoritmos que fueron surgiendo. Nosotros nos concentraremos en un determinado tipo de redes a las que se las denomina “multilayer feed forward neural networks”.

Hay dos maneras de ver una red neuronal, desde un punto de vista físico, en el cual se hace una analogía con las redes neuronales biológicas y desde un punto de vista matemático, haciendo un paralelo con los modelos

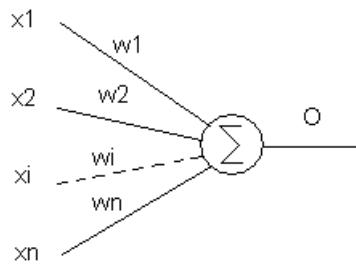
estadísticos de regresión. Comenzaremos viendo una descripción muy simplificada del funcionamiento del cerebro a nivel neuronal para luego describir cómo se simula por medio de una computadora. Finalmente compararemos una red neuronal con un modelo regresivo.

Un cerebro humano está formado por unos 10 mil millones de *neuronas* que se conectan entre sí, mediante 1 billón de conexiones aproximadamente. Cada neurona se compone de un *núcleo*, que es donde se procesa la información en forma de impulsos electroquímicos, provenientes de otras neuronas y recibidos a través de las *dendritas*, y de un haz de conducción, llamado *axón*, por medio del cual se conecta a otras neuronas para enviar su respuesta a los estímulos.



La unión de una rama del axón de una neurona con una dentrita de otra recibe el nombre de *sinapsis* y regula la fuerza de las conexiones en diferentes grados que van desde conexiones excitatorias hasta inhibitorias. Cada vez que se aprende algo, se modifican las sinapsis, o lo que es lo mismo, cambia el grado de conexión entre dos neuronas. Nuestra memoria reside en las sinapsis y es por ello que las células del cerebro o neuronas, nunca mueren como las demás células. Si lo hicieran no tendríamos memoria de largo plazo.

Tomando las redes neuronales biológicas como modelo, nace a fines de los cincuenta, principios de los sesenta, una nueva tecnología de procesamiento de la información: las redes neuronales artificiales (RNA). Básicamente una RNA consiste de elementos de procesamiento de información llamados neuronodos ó simplemente neuronas que reciben datos provenientes de otras neuronas, los procesan y los envían a su vez, a otras neuronas. Cada neurona realiza una suma ponderada por el grado de conexión de los datos que recibe, simulando la sinapsis y finalmente aplica una función a esta suma para obtener un dato de salida. Normalmente la función que aplica es no lineal y se la denomina función de transferencia, y se pueden citar como ejemplos una tangente hiperbólica o una sigmoidal.



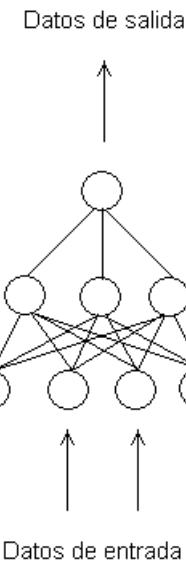
En donde O se calcula como función de Y:

$$O = f(Y)$$

Y el valor Y es:

$$Y = \sum_{i=1}^{i=n} x_i * w_i$$

Las neuronas se agrupan en capas. La que recibe los datos se llama capa de entrada (input layer). Ésta, a su vez se conecta con una o más capas intermedias (hidden layers), las que finalmente se conectan a la capa de salida (output layer). Un gráfico aclara el funcionamiento:



En este gráfico la capa de entrada consta de 4 neuronas que se conectan con cada una de las 3 neuronas de la capa intermedia. Por último, las 3 neuronas de la capa intermedia se conectan con la única neurona de la capa de salida.

Un cerebro tiene una gran estructura y la habilidad de construir sus propias reglas por medio de lo que generalmente se conoce como “experiencia” (Haykin 1994). Para que una RNA pueda resolver un problema, se utiliza el mismo procedimiento: *que aprenda de la experiencia*. En vez de darle las reglas en forma de “si pasa A entonces haga B”, como un programa convencional, se utilizan ejemplos pertenecientes al dominio del problema dado con sus correspondientes respuestas. Una vez procesados todos los ejemplos, incluso más de una vez cada uno si es necesario, la RNA habrá “aprendido” y estará lista para responder a problemas similares que se le planteen. Se dice que la red está entrenada. Podemos ver una RNA como un programa

capaz de encontrar estructuras (patterns) en los datos y luego de entrenada puede reconocer estas estructuras en nuevos datos que no se usaron para su entrenamiento. Además, y esto es muy importante, las estructuras no tienen que ser idénticas, sino solo similares.

Veamos ahora una RNA desde un punto de vista matemático. La respuesta frente a una determinada configuración de los datos de entrada será una combinación de los procesamientos individuales de cada neurona. Por ejemplo, si designamos con x cada entrada que recibe la capa intermedia con n neuronas, cada neurona i dará como resultado:

$$H_i = f(a_{i0} + a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n)$$

En donde los coeficientes a representan los ponderadores antes mencionados. A su vez, si suponemos que sólo tenemos una capa intermedia que se conecta a una sola neurona de salida, ésta última recibirá como entrada los H_i y su resultado será:

$$Y = f(b_0 + b_1H_1 + b_2H_2 + \dots + b_nH_n)$$

Entrenar una red significa encontrar el conjunto de ponderadores (a,b) que minimicen los errores dados por

$$E = \frac{1}{n} \sum_{i=1}^n (R_i - Y_i)^2$$

en donde R es la respuesta buscada en cada uno de los n ejemplos e Y es la respuesta dada por la RNA. Recordemos que la función de transferencia f introduce la no linealidad en la respuesta. Encontrar los ponderadores es equivalente a encontrar los coeficientes en una regresión lineal, en donde los datos de entrenamientos representan las variables independientes y cada respuesta la variable dependiente.

Los ponderadores que minimizan el error E dado más arriba en una RNA se encuentran mediante un algoritmo, y el más conocido es “Backpropagation”. Éste consta de dos fases, la primera evalúa el error cometido por la red a la que se le presenta uno o más ejemplos. La segunda usa este error para modificar los ponderadores. Todo este proceso, llamado entrenamiento, se repite una y otra vez con todos los ejemplos disponibles hasta que el error haya alcanzado un determinado nivel prefijado.

El entrenamiento no garantiza obtener el menor error posible, por lo que siempre es aconsejable reentrenar una RNA varias veces comenzando siempre con un conjunto de ponderadores distintos y al azar.

Hasta ahora vimos que el objetivo del entrenamiento en una RNA es minimizar los errores cometidos en la evaluación de un conjunto de ejemplos, al que llamamos conjunto de entrenamiento (training set). ¿Pero qué pasa si ahora fijamos los ponderadores de una red ya entrenada y le presentamos un nuevo conjunto de ejemplos diferentes a todos los del entrenamiento? La respuesta debiera ser: los errores cometidos son similares a los cometidos en el conjunto de entrenamiento. Pero no siempre es así. Depende entre otros factores de qué tan similares son los nuevos datos y qué tan bueno fue el entrenamiento.

La importancia de la similitud de los datos es muy intuitiva, si los ejemplos del entrenamiento contienen solamente estructuras que corresponden a mercados en alza, probablemente un mercado en baja sea difícil de reconocer. Por qué un entrenamiento puede resultar regular o malo, es un poco más difícil de explicar.

Los datos, o ejemplos que usamos para entrenar una RNA son el resultado de las transacciones que se realizan en el mercado de capitales a través del tiempo, por ello se los llama series de tiempo, y se las puede registrar por cada transacción (tic), o cada determinado tiempo, por ejemplo cada hora, día, mes o año.

Si deseamos un sistema que nos permita predecir que sucederá dentro de los próximos 5 días, probablemente no tenga sentido utilizar datos registrados con cada tic (datos intradiarios), sino que lo ideal será usar datos diarios. Pero los datos diarios registran el precio perteneciente a la última transacción del día que no tiene porque ser un precio representativo del día. A esto hay que agregar que existen mercados, como el de los bonos Brady, en los que se dificulta tener un precio de cierre y en los que sólo se cuenta con un precio de oferta y de compra (bid y ask).

Cuando trabajamos con series históricas debemos estar preparados a encontrarnos con datos mal tomados, irreales (por ejemplo el promedio entre el precio de oferta y de compra), poco representativos del día, ausencia de datos, etc. Todos estos factores agregan ruido o errores a nuestra serie de datos.

Otro problema es la cantidad disponible. Normalmente no contamos con demasiados datos, y si lo son, cabe preguntarse si las relaciones o estructuras que queremos descubrir con una RNA se mantienen en el tiempo. Por ejemplo, ¿las relaciones entre los mercados antes y después del Tequila serán las mismas?

Vimos que entrenar una red es buscar los ponderadores que minimizan los errores cometidos por ésta en todos los datos de entrenamiento. Pero estos datos contienen ruido, y si no somos cautelosos es probable que la red en vez de encontrar estructuras solamente memorice estos datos con ruido. Ser cauteloso significa no sobredimensionar una RNA y no tratar de obtener el menor error posible, y una manera de lograrlo es suspender el entrenamiento de tanto en tanto y verificar la red con otros datos diferentes, llamado datos de validación. En el momento que el error sobre los datos de validación deje de disminuir, debemos dejar de entrenar.

Caso Real

La siguiente es una descripción simplificada de un modelo basado en redes neuronales que sirve para operar tanto sea en el mercado de bonos como en el de acciones de Argentina.

Como datos de entrada al modelo se tomaron algunos bonos Brady y el índice Merval. Lo que se trata de pronosticar es cómo será la tendencia de las próximas 2 semanas del bono FRB. Si la tendencia pronosticada indica una baja del bono se debe vender, caso contrario se debe comprar.

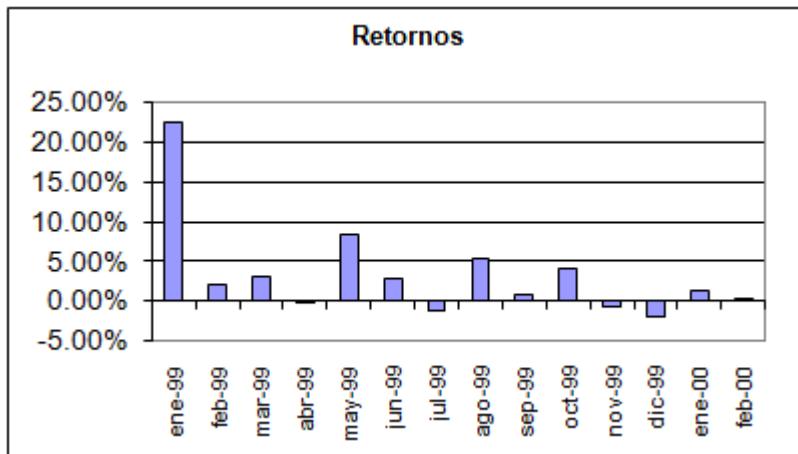
En el modelo ya desarrollado y probado se aprecian algunos aspectos que lo hacen muy valioso:

- El cambio de tendencia pronosticado se detecta con dos días de anticipación.
- La cantidad de cambios de tendencia, si bien depende de la volatilidad del mercado, es de unas 2 veces al mes en promedio, lo que minimiza el impacto de los costos de entrada y salida al mercado.
- Debido al horizonte pronosticado (dos semanas) y considerando la alta correlación entre los bonos, el Merval y las acciones que lo componen, se puede usar con cualquiera de ellos.

El método usado para la elección del modelo, más allá de cuestiones técnicas, tiene que ver con la confiabilidad del mismo. Por ejemplo, cómo asegurar que el desempeño logrado con datos históricos se mantendrá en el futuro. Después de todo, es muy fácil justificar y “pronosticar” un determinado acontecimiento después de que sucedió.

Una primera estrategia para robustecer la elección de un modelo es dejar de lado los últimos meses y usar el resto de datos para el desarrollo del mismo. Luego de finalizado el modelo se debería probar con los datos apartados. Si los resultados obtenidos fuesen similares a los anteriores, es probable que la red neuronal haya aprendido las relaciones. Cuanto mayor sea el período de pruebas, mayor será la confiabilidad. Pero debido a la poca cantidad de datos (recuérdese que a partir del Tequila probablemente cambió la estructura interna del mercado) con los que se cuenta, debe existir un compromiso entre cantidad de datos para entrenar una red y cantidad de datos para probarla.

Por supuesto, la mejor prueba es la que se realiza en tiempo real (exante) y el modelo que aquí se presenta fue medido de las dos maneras. Entre enero y julio del año 1999 se realizó una prueba con datos históricos (expost) pero no utilizados por la red en su entrenamiento. Desde agosto de 1999 hasta febrero del 2000 la prueba es en tiempo real (exante), esto significa que el pronóstico es real y con dos días de anticipación a los hechos.



Los mejores rendimientos del modelo en la primera mitad del año 1999 respecto del la segunda mitad, se explican con la diferente volatilidad que tuvieron esos períodos. Durante el primer semestre la volatilidad del bono FRB fue de 1.4% mientras que el segundo tuvo una volatilidad menor a la mitad, un 0.59%.

Las experiencias con modelos sobre otros mercados y durante otros períodos muestran que a mayor volatilidad mayor rendimiento. Es más, la mayoría de los modelos desarrollados tuvieron excelentes rendimientos durante las crisis.

Modelar el mercado de capitales no es una tarea fácil y no existe consenso sobre si tiene sentido hacerlo o no, ya que muchos suponen que los precios siguen un camino al azar. Probablemente haya períodos aleatorios y otros no tanto. Las redes neuronales muestran que esto podría ser verdad y quizás relacionado con los cambios de volatilidades.

Los resultados obtenidos hasta ahora han sido lo suficientemente buenos como para seguir investigando en redes neuronales y sus aplicaciones en mercados de capitales.

BIBLIOGRAFIA

Alligood (1997). Chaos. An Introduction to Dynamical Systems – K. Alligood, T. Sauer and J. Yorke. Springer.

Brock (1991). Nonlinear Dynamics, Chaos, and Instability – W. Brock, D. Hsieh, and B. LeBaron. The MIT Press.

Baestaens (1994). Neural Network Solutions for Trading in Financial Markets – Dirk-Emma Baestaens, W.M. Van Den Bergh, and D. Wood. Pitman Publishing.

Deboeck (1994). Trading on the Edge – Guido Deboeck. John Wiley & Sons, Inc.

Fama (1965). The Behavior of Stock Market Prices – Eugene Fama. Journal of Business 38, 1965.

Frees (1996). Data Analysis using regression models – Checking Normality pág. 92. Edward W. Frees. Prentice Hall.

- Gottman (1981). Time-series analysis: A comprehensive introduction for social scientists – John M. Gottman. Cambridge University Press.
- Grassberger (1983). Characterization of Strange Attractors – P. Grassberger and I. Procaccia. Physical Review Letters. January 1983.
- Haykin (1994). Neural Networks. A Comprehensive Foundation. – Simon Haykin. Prentice-Hall, Inc.
- Lavenda (1985). El movimiento browniano – Bernard H. Lavenda. Ciencia y Técnica. Abril 1985.
- Lin (1997). The ABC's of BDS – Kenneth Lin. Journal of Computational Intelligence in Finance, vol 5/num 4
- Lorenz (1995). La esencia del caos – Edward N. Lorenz. Debate S.A. Madrid 1995.
- Mandelbrot (1999). Panorama of grid-bound self-affine variability. Multifractals and 1/f Noise – Benoit B. Mandelbrot. Springer
- Moody (1995). Improved Estimates for the Rescaled Range and Hurst Exponents – John Moody and Lizhong Wu. Third International Conference (London, October 1995).
- Pardo (1992). Design, Testing, and Optimization of Trading Systems – Robert Pardo. John Wiley & Sons, Inc.
- Peters (1991). Chaos and Order in the Capital Markets – Edgard E. Peters. John Wiley & Sons, Inc.
- Peters (1994). Fractal Market Analisys – Edgard E. Peters. John Wiley & Sons, Inc.
- Refenes (1995). Neural Networks in the Capital Markets – Apostolos-Paul Refenes (Editor). John Wiley & Sons, Inc.
- Sametband (1994). Entre el orden y el caos: la complejidad – Moisés José Sametband. Fondo de Cultura Económica de Argentina.
- Trippi (1993). Neural Networks in Finance and Investing – R. Trippi and E. Turban (Editors). Probus Publishing Company.
- Vaga (1994). Profiting from Chaos – Tonis Vaga. McGraw-Hill, Inc.