



UNIVERSIDAD DE CHILE
FACULTAD DE ECONOMÍA Y NEGOCIOS

ALGORITMOS GENÉTICOS VERSUS FILTRO DE KALMAN
EN LA PREDICCIÓN DE ACCIONES NORTEAMERICANAS:
GE, GM, IBM, UTX Y VZ

Seminario para optar al título de Ingeniero Comercial
Mención Economía

Profesor guía: Antonino Parisi Fernández
Ph.D Finanzas y Economía

Autor: Felipe Asenjo Wilkins

Santiago de Chile
Septiembre de 2006

*A Oscar Asenjo, mi Padre, por inculcarme el Amor
y la Pasión por la Universidad de Chile
y por ser el más grande de mis ídolos*

*A Antonino Parisi, por su confianza,
sus consejos y orientación*

*A David Díaz, por su disposición,
ayuda y apoyo en todo momento*

ALGORITMOS GENÉTICOS VERSUS FILTRO DE KALMAN EN LA PREDICCIÓN DE ACCIONES NORTEAMERICANAS: GE, GM, IBM, UTX Y VZ

Felipe Asenjo Wilkins

Profesor Guía: Antonino Parisi Fernández, Ph.D

Resumen

Utilizando valores de cierres semanales, correspondientes al período comprendido entre el 20 de Agosto de 2001 al 14 de Agosto de 2006, se analiza la eficiencia de modelos multivariantes dinámicos, optimizados por *algoritmos genéticos* y *filtro de kalman*, para predecir el signo de las variaciones semanales en la cotización bursátil de GE, GM, IBM, UTX y VZ. Los resultados fueron comparados con los de un modelo AR(1) y de un modelo multivariable ARIMAX(2,2,2). Los mejores modelos producidos por el *algoritmo genético* arrojaron un porcentaje de predicción de signo (PPS), para un conjunto extramuestral de 52 datos semanales, de un 77%, 71%, 81%, 75% y 75%, para las acciones GE, GM, IBM, UTX y VZ, respectivamente. La capacidad predictiva resultó significativa en cada una de las acciones, de acuerdo al *test de acierto direccional* de Pesaran & Timmerman (1992). Al analizar el PPS de los modelos de *filtro de kalman*, se encontró que estos fueron menores, resultando significativos en el caso de GE, GM, IBM y UTX. Por otro lado, el PPS de los modelos AR(1), se encontró que estos fueron no significativos para todas las acciones en estudio. Los modelos multivariantes ARIMAX(2,2,2) registraron un PPS más alto que los de *filtro de kalman* para el caso de UTX, siendo no significativo. Además, los modelos construidos por el *algoritmo genético* generaron en promedio el mayor retorno acumulado, excepto en el caso de GM, donde la rentabilidad más alta fue registrada por el modelo de *filtro de kalman*. Ampliando el tamaño del conjunto extramuestral hasta cinco años, se llega a concluir que el PPS de todos los modelos analizados es decreciente, llegando a observar que, independientemente de la significancia de la capacidad predictiva de los modelos de proyección, estos superaron en rentabilidad a la estrategia de inversión pasiva en todas las acciones analizadas, por lo que prácticamente siempre fue mejor gestionar la cartera indexada en función de alguno de ellos.

Palabras Claves: Filtro de Kalman, Algoritmos Genéticos, Modelo Multivariable Dinámico, Porcentaje de Predicción de Signo, Test de Acierto Direccional, Capacidad Predictiva.

ÍNDICE

1.	Introducción.....	5
2.	Metodología y Datos.....	7
	A. Filtro de Kalman.....	7
	A.1. Especificación del Filtro de Kalman.....	9
	A.2. Estimación del Filtro de Kalman.....	10
	B. Algoritmos Genéticos.....	13
	B.1. Especificación.....	13
	B.2. Codificación, población inicial y generaciones.....	14
	C. Evaluación de la capacidad predictiva y significancia estadística de los modelos.....	17
3.	Resultados.....	20
4.	Conclusiones.....	22
	Bibliografía.....	23
	Anexos.....	25

1. Introducción

Este estudio analiza la eficiencia de modelos multivariantes dinámicos construidos a partir de metodologías de *algoritmo genético* y *filtro de kalman* para predecir el signo de la rentabilidad o variación semanal de GE¹, GM², IBM³, UTX⁴ y VZ⁵. La contribución de este estudio radica en mostrar la utilidad de las metodologías de *algoritmo genético* y *filtro de kalman* en la determinación de un modelo multivariable dinámico, a fin de medir su capacidad de predecir el movimiento en el precio (sobre la base del número de veces en que la dirección pronosticada es la correcta) y así determinar cuál metodología es la más apropiada. Se entiende que la predicción de la dirección del movimiento de la acción permite desarrollar estrategias de transacción efectivas, las cuales pueden arrojar mejores resultados que aquellas basadas en la proyección del valor de la variable observada (Leung, Daouk & Chen, 2000).

La hipótesis de mercados eficientes (Fama, 1970) establece que los retornos de los activos financieros no son predecibles y que no es posible generar beneficios anormales, debido a que el mercado incorpora toda la información relevante y que por tanto, la valoración de diversos activos refleja dicha información. Como el surgimiento de nueva información es de carácter aleatorio, los cambios en precios también lo serán. En este contexto, la evidencia empírica ha modelado los precios accionarios como un *random walk* o camino aleatorio. No obstante, existe evidencia significativa de que los precios accionarios no siguen un camino aleatorio y que en cierta forma son predecibles. Más, trabajos realizados por Lo & MacKinlay (1988), Conrad & Kaul (1988) y Gavridis (1998), sostienen que el mercado bursátil puede ser predecible sin dejar de ser eficiente.

De este modo, con el pasar de los años se ha demostrado que el concepto estocástico que esta detrás de las fluctuaciones de precios accionarios es errónea. En consecuencia, se asume que las regularidades que puedan contener las series históricas de variaciones de

¹ General Electric Company.

² General Motors Corporation.

³ International Business Machines Corporation.

⁴ United Technologies Corporation.

⁵ Verizon Communications Inc.

precios accionarios, pueden ser capturadas por los modelos de *series de tiempo*, de *redes neuronales* y de *algoritmos genéticos*.

No obstante, el grado de predictibilidad de los retornos es por lo general considerado como económicamente no significativo, por lo que Hodgson y Nicholls (1991) evaluaron la significancia económica de predecir la dirección de los cambios en los precios de los activos. En esta materia, Leung, Daouk & Chen (2000) compararon la capacidad predictiva de modelos de clasificación (*linear discriminant análisis, logit model, probit model* y *probabilistic neural network*) con los de estimación de nivel (*adaptive exponencial smoothing, vector autoregression with Kalman Filter, multivariate transfer function* y *multilayered feedforward neural network*) y concluyeron que los primeros (los cuales son propuestos para predecir el signo o la dirección del retorno de los índices bursátiles S&P 500, FTSE 100 y Nikkei 225) se desempeñan mejor que los segundos en términos de su tasa de acierto, medida como el número de aciertos en que la dirección pronosticada es la correcta y, además, son capaces de generar beneficios más altos.

Basado en lo anterior es que este estudio analiza la eficiencia de modelos multivariados dinámicos contruidos a partir de *algoritmos genéticos* y *filtro de kalman* para predecir el signo de las variaciones semanales en los precios de GE, GM, IBM, UTX y VZ, las cuáles fueron escogidas en forma aleatoria⁶. Los mejores modelos producidos por el *algoritmo genético* arrojaron un porcentaje de predicción de signo, para un conjunto extramuestral de 52 datos semanales, de un 77%, 71%, 81%, 75% y 75%, para GE, GM, IBM, UTX y VZ, respectivamente. La capacidad predictiva resultó significativa, de acuerdo al *test de acierto direccional* de Pesaran & Timmermann (1992). Al analizar el porcentaje de predicción de signo de los modelos de *filtro de kalman*, se encontró que estos fueron menores (69%, 69%, 69%, 60% y 67%) y resultaron ser significativos, a excepción de VZ. La capacidad predictiva de los modelos AR(1) resultó ser no significativa para todas las acciones, con la menor rentabilidad y PPS. En cuanto a los modelos multivariados dinámicos tipo ARIMAX(2,2,2), se registró una capacidad predictiva de 63%, 67%, 63%,

⁶ Se escogió en forma aleatoria 5 acciones que pertenecieran al índice Dow Jones (DJI). Se entiende que el proceso de selección sea aleatorio, con el fin de probar la eficacia de los modelos y abstraerse de posibles particularidades de cada acción.

62% y 67%, siendo significativa en tres de cinco acciones: GM, IBM y VZ. Por último los modelos construidos por el *algoritmo genético* generaron el mayor retorno acumulado promedio para GE, IBM, UTX y VZ. Posteriormente se procedió a incrementar al tamaño del conjunto extramuestral, formando segmentos de dos, tres, cuatro y hasta cinco años, llegando a determinar que el PPS es decreciente para todos los modelos en estudio. No obstante, las conclusiones son similares a las obtenidas en el caso de un conjunto extramuestral de 52 datos semanales.

Este trabajo se divide en cuatro secciones: la sección dos explica la metodología empleada en la investigación, la sección tres muestra el análisis de los resultados, y finalmente, la sección cuatro presenta las conclusiones del estudio.

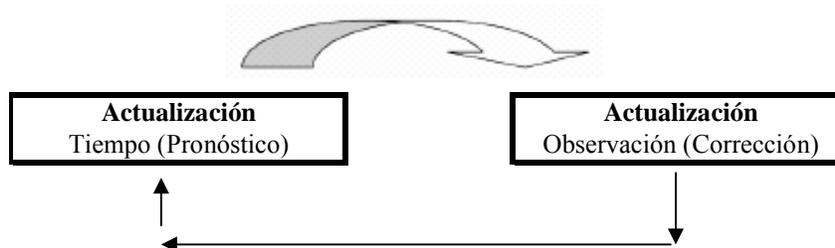
2. Metodología y Datos

Se utilizaron 311 valores de cierre semanales de las variables incluidas en los modelos, correspondientes al período entre el 04 de agosto de 1997 y el 09 de diciembre de 2002. El Anexo 3 presenta las estadísticas de las series estudiadas.

A. Filtro de Kalman

El *filtro de kalman*, introducido por Kalman (1960), consiste en un algoritmo recursivo, que provee una solución eficiente al método de mínimos cuadrados ordinarios. Esta solución permite obtener un estimador lineal, insesgado y óptimo del estado de un proceso en cada momento del tiempo utilizando la información disponible en el momento (t-1) y, posteriormente, actualizando dicho estimador con la información adicional disponible al momento t. Este proceso se realiza mediante un algoritmo de predicción y uno de corrección, con los cuales se van ajustando los parámetros del sistema, que posteriormente son optimizados para encontrar aquellos que mejor explican el fenómeno de estudio. La figura 1 muestra una descripción del ciclo.

Figura 1. El Ciclo del Filtro de Kalman



Éste permite combinar modelos de corte transversal y de series de tiempo, ya sea para la estimación de modelos autorregresivos (Blake, 2002), la estimación de modelos cuyos parámetros cambian en el tiempo y la estimación de modelos con componentes no observables (Schwartz, 1997). Además, es el principal algoritmo para estimar modelos dinámicos, especificados a través de la forma estado – espacio. Con respecto a este último, puede ser usado en muestras finitas para *forecast for Gaussian ARMA models, multivariate ARMA models, MIMIC (multiple indicators and multiple causes), Markov switching models, y time varying (random) coefficient models*.

La intuición que está detrás del *filtro de kalman* consiste en estimar e inferir los parámetros relevantes de un sistema, a partir de observaciones imprecisas e inciertas. En base a las observaciones de ciertas variables, las cuales pueden contener errores, es posible estimar otras variables de interés que no se observan. Así, al ser un proceso recursivo, a medida que se incorporan nuevas observaciones en la estimación, es posible obtener la información deseada. Su ventaja radica en que permite la existencia de errores en el proceso de obtención de datos, debido a que es un proceso que se adapta en el tiempo, pues a medida que incorpora nueva información, cambia la ponderación que tienen las nuevas observaciones y a las estimaciones históricas que realiza. Además, permite incorporar explícitamente los errores de predicción, sin restringirlos.

Esta metodología tiene una serie de aplicaciones financieras. Penacho (1991) utiliza el *filtro de kalman* para determinar tasas de interés reales e inflación. Schwartz (1997) muestra una aplicación de *filtro de kalman* en la predicción de precios de commodities, para uno, dos y hasta tres factores de riesgo. Díaz & Schwartz (2000), estiman modelos de

hasta dos factores de riesgo para el precio del petróleo, con el fin de encontrar una estrategia de cobertura óptima. Naranjo (2002) utilizó *filtros de kalman* para modelos lognormales del precio del cobre y petróleo, siendo el primer trabajo con paneles dinámicos incompletos. Cortázar, Schwartz & Naranjo (2003) utilizando un panel incompleto de datos de PRC y PRBC, muestran un modelo de Vacicek con tres factores de riesgo, a través de *filtro de kalman*.

A.1. Especificación del Filtro de Kalman

El espacio de estados es una representación matricial del sistema observado, el cual esta descrito por las siguientes dos ecuaciones:

$$y_t = c_t + Z_t x_t + e_t \quad (1)$$

$$x_t = d_t + T_t x_{t-1} + v_t \quad (2)$$

La primera es la *ecuación de medición* o *corrección*, que muestra la relación entre las variables observables y las no observables. En ella se explica la evolución de k variables no observables, mediante n variables observables; con y_t , un vector de dimensión n que contiene variables observables⁷; x_t , un vector de dimensión k de variables no observadas⁸; Z_t , una matriz de dimensión $n \times k$ que relaciona linealmente las variables observadas y no observadas; y c_t , un vector de dimensión n constantes. El vector e_t contiene los errores de medición, los que tienen una distribución normal:

$$e_t \sim N(0, H(\psi)) \quad (3)$$

, donde $H(\psi)$ es la matriz de varianzas - covarianzas de los errores de predicción con dimensión $n \times n$, la que es función de ψ , la matriz de parámetros del modelo.

⁷ En este trabajo, se utiliza como variable observable a las variaciones del índice S&P 500. Este índice fue escogido debido a que arrojó los mejores resultados para el *filtro de kalman*.

⁸ Las variables de interés que se intentan predecir son las respectivas cotizaciones bursátiles de las acciones.

La segunda ecuación es la *ecuación de transición o predicción*, que muestra la evolución de las variables no observables del sistema. En ella tenemos que x_t es un vector de k variables no observables en el tiempo t , d_t es un vector de k constantes, T_t es una matriz de dimensión $k \times k$ que relaciona los estados de la variable no observable. Nuevamente el vector v_t contiene los errores, los que tienen una distribución normal:

$$v_t \sim N(0, Q_t) \quad (4)$$

, donde Q_t es la matriz de varianzas - covarianzas de los errores de la ecuación de transición con dimensión $k \times k$, que no depende de ψ , la matriz de parámetros del modelo.

Los errores e_t y v_t se asumen independientes entre sí, con la siguiente estructura de varianza contemporánea:

$$\Omega_t = \text{Var} \begin{bmatrix} e_t \\ v_t \end{bmatrix} = \begin{bmatrix} H_t & G_t \\ G_t' & Q_t \end{bmatrix} \quad (5)$$

De acuerdo a lo anterior, es posible generalizar esta metodología, permitiendo que el sistema de matrices y vectores $\Xi \equiv \{\alpha_t, \gamma_t, \beta_t, \eta_t, H_t, Q_t, G_t\}$ dependa de variables observadas X_t y de parámetros no observados θ .

A.2. Estimación del Filtro de Kalman

El objetivo del *filtro de kalman* es encontrar un modelo multivariable dinámico que contenga estimadores insesgados y lineales para la predicción de signo de las variaciones semanales en los precios de GE, GM, IBM, UTX y VZ. Mediante un proceso recursivo se busca estimar el valor de \hat{x} (variable observable o explicativa) de acuerdo a la información histórica, añadiendo una corrección proporcional al término de error de predicción, de modo de minimizar este último.

Sea $f_t = (y_1, y_2, \dots, y_t)$, el conjunto de información ocurrida hasta el instante t . Previo a este instante t , en la etapa de predicción, se proyecta el estado actual de las variables no observables (x_t) y de la matriz de varianza – covarianza del error. Luego, esta información es utilizada para predecir en el siguiente instante. La ecuación de predicción se muestra a continuación:

$$\hat{x}_{t/t-1} \equiv E(x_t / f_{t-1}) = \gamma_t + \eta_t x_{t-1} \quad (6)$$

A partir de la ecuación anterior, se proyecta el valor de x al instante t , utilizando la información disponible al instante anterior ($t-1$). La mejor proyección que se puede obtener, consiste en usar el último valor conocido de la variable x . Además, esta proyección tiene una matriz de varianza – covarianza dada por:

$$P_{t/t-1} = E[(x_t - \hat{x}_{t/t-1})(x_t - \hat{x}_{t/t-1})' / f_{t-1}] = \eta_t P_{t-1} \eta_t' + V_t \quad (7)$$

A partir de la estimación de la variable no observada es posible predecir la variable observada de nuestro interés, lo que está definido a continuación:

$$\hat{y}_{t/t-1} = \alpha_t + \beta_t \hat{x}_{t/t-1} \quad (8)$$

De este modo, al estimar la variable de interés y luego observar su valor, es posible actualizar las variables y mejorar la estimación. Para realizar esto se utiliza el vector de errores de predicción:

$$e_t = y_t - \hat{y}_{t/t-1} \quad (9)$$

, con matriz de varianza – covarianza:

$$F_t = E[e_t e_t' / f_t] = \beta_t P_{t/t-1} \beta_t' + H_t \quad (10)$$

La actualización del vector de variables no observables queda definido por:

$$\hat{x}_t = E(x_t / f_t) = \hat{x}_{t/t-1} + P_{t/t-1} \beta_t' F_t^{-1} e_t \quad (11)$$

, con matriz de varianza – covarianza dada por:

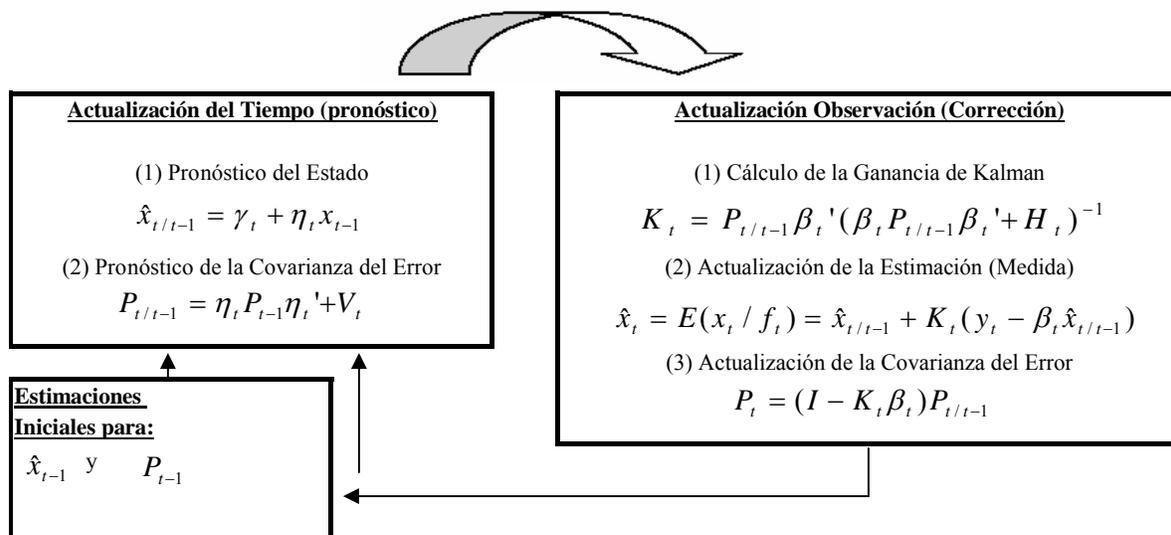
$$P_t = E[(x_t - \hat{x}_t)(x_t - \hat{x}_t)' / f_t] = P_{t/t-1} - P_{t/t-1} \beta_t' F_t^{-1} \beta_t P_{t/t-1} \quad (12)$$

, al término K_t se le conoce como ganancia de Kalman. K_t es una matriz de $k \times n$, que minimiza el error cuadrático medio.

$$K_t = P_{t/t-1} \beta_t' F_t^{-1} = \frac{P_{t/t-1} \beta_t'}{\beta_t P_{t/t-1} \beta_t' + H_t} \quad (13)$$

Finalmente, la figura 2 muestra un cuadro completo de la operación del filtro. Como se ve en la figura, posterior a cada par de actualizaciones, el proceso es repetido tomando como punto inicial las nuevas estimaciones del estado y las respectivas covarianzas del error. Esta naturaleza recursiva es la característica esencial del *filtro de kalman*.

Figura 2. Visión completa del Filtro de Kalman⁹



⁹ Para mayor descripción, véase, Solera, A., “El Filtro de Kalman”. (2003)

Cabe destacar que este proceso es por lo general realizado por máxima verosimilitud. No obstante, en este trabajo el proceso de optimización del *filtro de kalman* es realizado a partir de *brute force*, que busca los parámetros que maximizan el porcentaje de predicción de signo (PPS). Para ello, se utilizan 100.000 iteraciones para cada una de las acciones a ser consideradas en este estudio.

B. Algoritmos Genéticos

El *Algoritmo genético* es una función matemática que simula el proceso evolutivo de las especies, cuyo objetivo es encontrar una solución óptima para problemas específicos de maximización o minimización. De este modo, recibe como input generaciones de soluciones para un problema específico, arrojando los especímenes más aptos como output. Estas soluciones mejoradas se reproducen y generan descendientes, dando origen a nuevas generaciones con características más aptas que generaciones anteriores.

Los *algoritmos genéticos* utilizan códigos (cromosomas) para representar cada una de las soluciones del problema, por lo tanto, es necesario codificar todo el rango de soluciones posibles. Una de las codificaciones más utilizadas para la representación de soluciones es a través de cadenas binarias, es decir, por medio de un conjunto de ceros y unos (Davis, 1994).

Esta metodología puede ser utilizada en aplicaciones financieras (Bauer, 1994). La evidencia empírica muestra la aplicación de *algoritmo genéticos* en la calificación de créditos bancarios, en la predicción de la probabilidad de bancarrota de la firmas y desarrollar técnicas de transacción que permitieran asignar montos de inversión.

B.1. Especificación

El objetivo del *algoritmo genético* es encontrar un modelo multivariable dinámico que maximice el porcentaje de predicción de signo de las variaciones semanales en los precios de GE, GM, IBM, UTX y VZ.¹⁰ Los datos utilizados corresponden a los precios de

¹⁰ Parisi, Parisi y Cornejo (2003), utilizan *algoritmos genéticos* para encontrar modelos multivariables dinámicos que maximicen el porcentaje de predicción de signo de las variaciones semanales de los índices

cierre semanales para el período comprendido entre el 20 de Agosto de 2001 y el 14 de Agosto de 2006.

Los modelos multivariantes dinámicos utilizados para predecir la variación semanal en los precios de GE, GM, IBM, UTX y VZ ($\Delta GE_t, \Delta GM_t, \Delta IBM_t, \Delta UTX_t$ y ΔVZ_t), se presentan a continuación:

$$\begin{aligned}
 \Delta GE_t &= \alpha_1 \cdot \Delta GE_{t-1} + \dots + \alpha_{AR} \cdot \Delta GE_{t-AR} + \phi_1 \cdot \Delta DJI_{t-1} + \dots + \phi_X \cdot \Delta DJI_{t-X} + \delta_1 \cdot \varepsilon_{t-1} + \dots + \delta_{MA} \cdot \varepsilon_{t-MA} \\
 \Delta GM_t &= \alpha_1 \cdot \Delta GM_{t-1} + \dots + \alpha_{AR} \cdot \Delta GM_{t-AR} + \phi_1 \cdot \Delta DJI_{t-1} + \dots + \phi_X \cdot \Delta DJI_{t-X} + \delta_1 \cdot \varepsilon_{t-1} + \dots + \delta_{MA} \cdot \varepsilon_{t-MA} \\
 \Delta IBM_t &= \alpha_1 \cdot \Delta IBM_{t-1} + \dots + \alpha_{AR} \cdot \Delta IBM_{t-AR} + \phi_1 \cdot \Delta DJI_{t-1} + \dots + \phi_X \cdot \Delta DJI_{t-X} + \delta_1 \cdot \varepsilon_{t-1} + \dots + \delta_{MA} \cdot \varepsilon_{t-MA} \\
 \Delta UTX_t &= \alpha_1 \cdot \Delta UTX_{t-1} + \dots + \alpha_{AR} \cdot \Delta UTX_{t-AR} + \phi_1 \cdot \Delta DJI_{t-1} + \dots + \phi_X \cdot \Delta DJI_{t-X} + \delta_1 \cdot \varepsilon_{t-1} + \dots + \delta_{MA} \cdot \varepsilon_{t-MA} \\
 \Delta VZ_t &= \alpha_1 \cdot \Delta VZ_{t-1} + \dots + \alpha_{AR} \cdot \Delta VZ_{t-AR} + \phi_1 \cdot \Delta DJI_{t-1} + \dots + \phi_X \cdot \Delta DJI_{t-X} + \delta_1 \cdot \varepsilon_{t-1} + \dots + \delta_{MA} \cdot \varepsilon_{t-MA}
 \end{aligned}
 \tag{14}$$

, donde ε_t es el término de error del modelo. Los subíndices AR, MA y X, corresponden al máximo número de rezagos de las variables independientes. Notar que el modelo multivariable dinámico utilizado es un modelo de series de tiempo, cuya variable dependiente esta expresada en función de rezagos de la propia variable dependiente, de rezagos de variables exógenas como el índice DJI¹¹ y de rezagos del término de error.

B.2. Codificación, población inicial y generaciones

El máximo número de rezagos para los términos AR, MA y X fue 6, por lo que el máximo número de variables para los modelos fue de 18. Cada variable es representada como un *bit* y, por lo tanto, el largo del código binario (cromosoma) es igual al número máximo de variables del modelo¹². De acuerdo a las matemáticas binarias, una cadena de

bursátiles IPC, TSE, DJI y Nasdaq. Éste trabajo utiliza esta metodología con el fin de compararla con la de *filtro de kalman*, para la cotización bursátil de GE, GM, IBM, UTX y VZ.

¹¹ Se escogió el índice Dow Jones, debido a que otorgó los mejores resultados bajo la metodología de *algoritmo genético*.

¹² El primer bit de la cadena representará al precio de la acción, en primera diferencia y rezagado un período; el bit 2 representará al rezago 2; el bit 7 representará al residuo del modelo rezagado un periodo; el bit 12 representará el sexto rezago del residuo; el siguiente bit representará el primer rezago del índice DJI en primera diferencia; y, finalmente, el último bit representará al sexto rezago. Cada bit puede tomar el valor de

largo “L”, permite representar “ 2^L ” soluciones posibles. En consecuencia, se cuenta con un total de 262.144 modelos para predecir la dirección de los cambios en precio de GE, GM, IBM, UTX y VZ, entre los cuales, el algoritmo genético realizará una búsqueda de la mejor solución.

Para iniciar el proceso de optimización genética¹³ el algoritmo generó 100 modelos aleatorios de 18 bits de largo. De acuerdo a Bauer (1994), es importante que la población esté compuesta por un número no tan pequeño de potenciales soluciones, normalmente entre 100 y 200 individuos, de manera de garantizar que sea seleccionado un grupo amplio de individuos representativos de las mejores soluciones. A continuación, el algoritmo procedió a estimar cada uno de los 100 modelos pertenecientes a la primera generación, utilizando *brute force*¹⁴.

El proceso de *brute force* es un proceso aleatorio que prueba todas las combinaciones posibles de parámetros a fin de encontrar aquellos que mejor explican el fenómeno de estudio. En este trabajo, el proceso de *brute force* comienza tomando las primeras 52 semanas de observaciones históricas, y realiza una proyección para la semana 53 en adelante. De esta forma, se utilizan 52 semanas como datos intramuestrales y se proyecta la variación del precio de la acción para el conjunto extramuestral de 260 datos. Por lo tanto, se cuenta con 260 proyecciones extramuestrales hechas por cada uno de los 100 modelos, los cuales son evaluados en función del PPS encontrado.

Luego los modelos se ordenan de manera descendente, por PPS. A continuación, se seleccionan los 10 primeros modelos, correspondientes a los 10 PPS más altos, a los cuales se les llama padres. De estos 10 padres, el 100% pasa directamente a la siguiente generación, de manera de asegurar que las mejores soluciones alcanzadas hasta ese momento sobrevivan a través de las generaciones. A partir de la combinación de los

0 o 1: cuando tome el valor de 0, la variable que representa no será incluida en el modelo, mientras que cuando tome el valor de 1, la variable representada sí formará parte del modelo. Una vez realizada la codificación de los modelos, se comenzó a trabajar con el *algoritmo genético*. La primera generación de modelos (o población inicial), de un tamaño de 100, se obtuvo aleatoriamente desde el rango de interés.

¹³ Véase Cuadro 3 (Anexo 1)

¹⁴ El número de iteraciones utilizadas fue de 20.000 por modelo.

cromosomas de los 10 padres se generan 80 hijos por el método de cruce doble¹⁵. Además, se generan 10 hijos en forma aleatoria para cada generación, con lo cuál la siguiente generación queda formada por 100 modelos (10 padres con mejor PPS de la generación anterior, 80 hijos de esos padres y 10 hijos creados en forma aleatoria).

Se permitió una mutación con una probabilidad de 8.33%, es decir, de manera aleatoria, algunos genes correspondientes a los 80 hijos cambian su posición binaria de 1 a 0 o viceversa. Con esto se permite que el proceso de optimización tenga acceso a potenciales soluciones no cubiertas en la combinación de los padres, pero se hace de manera que su incidencia no sea un factor preponderante, ya que de hacerlo originaría un proceso de evolución aleatorio. Por otra parte, la clonación está implícita en el proceso ya que no se restringió el cruce entre dos modelos iguales. Una vez terminada la nueva generación de modelos, se calculan nuevamente los PPS, continuando con el proceso de selección y generando nuevas descendencias hasta concluir con la generación número 15.

Con respecto a la probabilidad de mutación y al número de generaciones, debemos precisar que en términos generales existe una relación inversa entre estas dos variables, es decir, a medida que aumentamos la probabilidad de mutación disminuye la cantidad de generaciones necesarias para alcanzar el óptimo. Esto porque el aumentar la probabilidad de mutación repercute directamente en la muestra de posibles soluciones a ser probadas por el algoritmo, aumentando también la probabilidad de encontrar el óptimo en menos generaciones. En caso de ocupar una tasa de mutación inferior, sería de esperar que la convergencia al mejor modelo fuera más lenta, por lo que serían necesarias más generaciones. Sin embargo, como se mencionó anteriormente, tasas muy altas de mutación podrían provocar que el proceso de selección se volviera aleatorio, por lo que la tasa usada debe garantizar un proceso evolutivo que mejore las especies a través de las generaciones.

¹⁵ Véase Cuadro 1 (Anexo 1)

Creemos que los parámetros escogidos¹⁶ cumplen con lo anterior, dejando abierto el debate con respecto a la combinación óptima de número de generaciones y probabilidad de mutación, así como también el análisis en profundidad de la sensibilidad de la optimización a dichos parámetros. El Gráfico 1 muestra que, por lo general, el algoritmo convergió al modelo óptimo antes de la generación 15 (se encontró siempre antes de la generación 8). Al mismo tiempo, dada la enorme combinatoria y capacidad de cálculo requerida para analizarla, se ha mantenido constante el tamaño inicial del proceso de *brute force* en 52 datos, centrando la búsqueda sólo en la combinatoria de variables explicativas. Un enfoque más amplio de optimización genética debería tratar de optimizar la búsqueda de variables explicativas y al mismo tiempo, el tamaño de observaciones óptimas necesarias para un proceso de *brute force*, siendo éste último un desafío propuesto para nuevas investigaciones.

C. Evaluación de la capacidad predictiva y significancia estadística de los modelos

Tanto los modelos generados por *algoritmos genéticos* como el modelo de *filtro de kalman* fueron estimados a partir de los datos del conjunto intramuestral. Este proceso fue seguido por una evaluación empírica sobre la base de los datos del conjunto extramuestral, el cual se compone de 260 observaciones semanales. Como se mencionó anteriormente, el desempeño relativo de los modelos fue medido por el número de predicciones correctas del signo de la variación del precio. Para ello se comparó el signo de la proyección con el signo de la variación observada en cada i -ésima semana, donde $i=1, 2, \dots, s$, siendo s el número de observaciones del conjunto extramuestral. Si los signos entre la proyección y el observado coinciden, entonces se anota un *acierto*, aumentando la efectividad del modelo analizado, en caso contrario, el modelo disminuye su capacidad predictiva. El porcentaje de predicción de signo (PPS) de cada modelo se calculó de la siguiente forma:

$$\text{Porcentaje} = \frac{\sum_{t=n+1}^{n+m} P_i}{m}, \text{ con } P_i = \begin{cases} 1, & \text{si } \Delta \text{ Precio}_t * \hat{\Delta} \text{ Precio}_t > 0 \\ 0, & \text{otro caso} \end{cases} \quad (15)$$

¹⁶ Población inicial de 100 modelos, 15 generaciones, y tasa de mutación de 8.33%.

, donde Δ representa la variación observada y $\hat{\Delta}$ la variación estimada. Además, para medir la significancia estadística de la capacidad predictiva de cada uno de los modelos, se aplicó el test de acierto direccional de Pesaran & Timmermann (1992). Cabe destacar que, se procedió a calcular el PPS para uno, dos, tres, cuatro y hasta cinco años, con el fin de verificar si la capacidad predictiva de los modelos es robusta en el tiempo. A continuación se muestra una tabla que resume los PPS y RPA calculados.

Período	PPS	Rentabilidad Promedio Acumulada
21-Agosto-2000 al 13-Agosto-2001	PPS Intramuestral de 1 Año	RPA Intramuestral de 1 Año
14-Agosto-2001 al 19-Agosto-2002	PPS Extramuestral de 1 Año	RPA Extramuestral de 1 Año
14-Agosto-2001 al 18-Agosto-2003	PPS Extramuestral de 2 Años	RPA Extramuestral de 2 Años
14-Agosto-2001 al 16-Agosto-2004	PPS Extramuestral de 3 Años	RPA Extramuestral de 3 Años
14-Agosto-2001 al 15-Agosto-2005	PPS Extramuestral de 4 Años	RPA Extramuestral de 4 Años
14-Agosto-2001 al 14-Agosto-2006	PPS Extramuestral de 5 Años	RPA Extramuestral de 5 Años

Adicionalmente, los resultados del mejor modelo multivariado dinámico y del *filtro de kalman* de cada una de las acciones analizadas fueron comparados con los de un modelo ingenuo o AR(1) y un modelo multivariable ARIMAX(2,2,2). La estructura del modelo AR(1) se presenta en la siguiente ecuación:

$$\Delta \text{Precio}_{it} = \alpha_0 + \alpha_1 \cdot \Delta \text{Precio}_{it-1} + \varepsilon_t \quad (16)$$

, donde Δ indica primera diferencia, ε_t representa el término de error en el momento “t”, y α_0 y α_1 son los coeficientes del modelo¹⁷. Por otro lado, la estructura del modelo ARIMAX(2,2,2) se presenta a continuación:

$$\Delta \text{Precio}_{it} = \alpha_0 \cdot \Delta \text{Precio}_{it-1} + \alpha_1 \cdot \Delta \text{Precio}_{it-2} + \alpha_2 \cdot \Delta DJI_{it-1} + \alpha_3 \cdot \Delta DJI_{it-2} + \alpha_4 \cdot \varepsilon_{it-1} + \alpha_5 \cdot \varepsilon_{it-2} \quad (17)$$

, donde Δ indica primera diferencia, ε_t representa el término de error en el momento “t”, y α_0 , α_1 , α_2 , α_3 , α_4 y α_5 son los coeficientes del modelo. Esta comparación adicional se realizó

¹⁷ Las estadísticas de los modelos AR(1) se encuentran en el Anexo 5.

para probar empíricamente la hipótesis planteada en la introducción, permitiéndonos medir la eficacia de una estrategia *activa*, la cual es representada por las técnicas de *algoritmo genético* y *filtro de kalman*, y la eficacia de una estrategia *pasiva*, representada por el modelo ingenuo y la estrategia de *buy & hold*.

Sin embargo, y de acuerdo a Sharda (1994), es necesario ser cuidadosos al realizar tal comparación ya que, por definición, un modelo ingenuo no utiliza ningún tipo de recomposición de la muestra utilizada en las estimaciones, al contrario del proceso de *brute force* o a la utilización de sólo un dato, como lo requiere el *filtro de kalman*. Así, dicha comparación está sujeta a una muestra sesgada, ya que el modelo ingenuo utilizará la totalidad de la muestra de tamaño “*n*” para medir su capacidad predictiva, mientras que el proceso de *brute force* utilizará “*s*” observaciones para hacerlo, donde “*s*” es igual a “*n - k*” donde “*k*” es el número de observaciones perteneciente al subconjunto intramuestral o de entrenamiento. Para solucionar dicho problema el modelo ingenuo utilizará las primeras 52 observaciones para estimar los parámetros, y las restantes 260 para medir su capacidad predictiva, donde cada predicción es actualizada al transcurrir una semana¹⁸.

También se calculó la rentabilidad de seguir las recomendaciones de *compra* (cuando el valor proyectado para el futuro cercano indicaba un alza en el precio) y *venta* (en el caso contrario) de los modelos analizados. Para el cálculo de las rentabilidades de cada una de las técnicas se consideró un monto de inversión inicial de US\$ 100.000, valorándose la estrategia como el valor total de la cartera, es decir, la suma de la posición en la acción, más el dinero en efectivo con el que se contaba. Las señales de compra fueron ejecutadas en la medida que la cartera contaba con dinero en efectivo al momento de la señal, y las recomendaciones de venta fueron ejecutadas en la medida que se contaba con unidades de la acción respectiva. No se consideraron *ventas cortas*, costos de transacción ni de almacenaje.

¹⁸ Una completa descripción de la metodología utilizada en la estimación del modelo ingenuo es mostrado en el Anexo 4.

3. Resultados

Los mejores modelos multivariantes dinámicos, obtenidos a través del *algoritmo genético* (AG) para cada uno de las acciones en estudio, presentan la siguiente forma funcional:

$$\begin{aligned}
 \Delta GE_t &= \alpha_2 \cdot \Delta GE_{t-2} + \alpha_3 \cdot \Delta GE_{t-3} + \alpha_4 \cdot \Delta GE_{t-4} + \alpha_6 \cdot \Delta GE_{t-6} + \phi_2 \cdot \Delta DJI_{t-2} + \phi_3 \cdot \Delta DJI_{t-3} + \phi_5 \cdot \Delta DJI_{t-5} + \delta_2 \cdot \varepsilon_{t-2} + \delta_6 \cdot \varepsilon_{t-6} \\
 \Delta GM_t &= \alpha_2 \cdot \Delta GM_{t-2} + \alpha_4 \cdot \Delta GM_{t-4} + \alpha_5 \cdot \Delta GM_{t-5} + \phi_1 \cdot \Delta DJI_{t-1} + \phi_4 \cdot \Delta DJI_{t-4} + \phi_6 \cdot \Delta DJI_{t-6} + \delta_5 \cdot \varepsilon_{t-5} + \delta_6 \cdot \varepsilon_{t-6} \\
 \Delta IBM_t &= \alpha_3 \cdot \Delta IBM_{t-3} + \alpha_4 \cdot \Delta IBM_{t-4} + \alpha_5 \cdot \Delta IBM_{t-5} + \phi_1 \cdot \Delta DJI_{t-1} + \phi_3 \cdot \Delta DJI_{t-3} + \phi_4 \cdot \Delta DJI_{t-4} + \phi_5 \cdot \Delta DJI_{t-5} + \delta_4 \cdot \varepsilon_{t-4} + \delta_5 \cdot \varepsilon_{t-5} \\
 \Delta UTX_t &= \alpha_3 \cdot \Delta UTX_{t-3} + \alpha_4 \cdot \Delta UTX_{t-4} + \phi_1 \cdot \Delta DJI_{t-1} + \delta_1 \cdot \varepsilon_{t-1} + \delta_2 \cdot \varepsilon_{t-2} + \delta_4 \cdot \varepsilon_{t-4} + \delta_6 \cdot \varepsilon_{t-6} \\
 \Delta VZ_t &= \alpha_1 \cdot \Delta VZ_{t-1} + \alpha_5 \cdot \Delta VZ_{t-5} + \alpha_6 \cdot \Delta VZ_{t-6} + \phi_1 \cdot \Delta DJI_{t-1} + \phi_2 \cdot \Delta DJI_{t-2} + \phi_3 \cdot \Delta DJI_{t-3} + \phi_6 \cdot \Delta DJI_{t-6} + \delta_5 \cdot \varepsilon_{t-5} + \delta_6 \cdot \varepsilon_{t-6}
 \end{aligned}
 \tag{18}$$

El Cuadro 4 muestra que los mejores modelos producidos por el *algoritmo genético* arrojaron un PPS de 77%, 71%, 81%, 75%, y 75%, para las acciones, GE, GM, IBM, UTX y VZ, respectivamente. Esta capacidad predictiva, estimada sobre un conjunto extramuestral de 52 datos semanales (1 año), resultó ser significativa al 5% en cada uno de las acciones, de acuerdo al test de acierto direccional. Comparativamente, la capacidad predictiva del *filtro de kalman* resultó menor y significativa en cuatro de las cinco acciones analizadas: GE, GM, IBM y UTX. Por su parte, los modelos multivariantes ARIMAX(2,2,2), resultaron ser no significativos a excepción de GM, IBM y VZ.

Para un conjunto extramuestral de 104 datos semanales (2 años), los modelos producidos por el *algoritmo genético* arrojaron un PPS de 74%, 73%, 77%, 68%, y 71%, para las acciones, GE, GM, IBM, UTX y VZ, respectivamente. Esta capacidad predictiva, resultó significativa al 5% en cada una de las acciones, de acuerdo al test de acierto direccional. La capacidad predictiva del *filtro de kalman* resultó ser significativa sólo para cuatro de las cinco acciones analizadas: GE, GM, IBM y VZ, con un PPS menor que el de *algoritmo genético*. En comparación, los modelos multivariantes ARIMAX(2,2,2), resultaron no significativos para todas las acciones en estudio.

Por otro lado, para un conjunto extramuestral de 156 datos semanales (3 años), los modelos producidos por el *algoritmo genético* arrojaron un PPS de 69%, 67%, 70%, 63%, y 67%, para las acciones, GE, GM, IBM, UTX y VZ, respectivamente. Esta capacidad

predictiva, resultó significativa al 5% en cada una de las acciones, de acuerdo al test de acierto direccional. La capacidad predictiva del *filtro de kalman* resultó ser significativa sólo para cuatro de las cinco acciones analizadas: GE, GM, IBM y VZ, con un PPS menor que el de *algoritmo genético*. Con respecto a los modelos multivariantes ARIMAX(2,2,2), éstos resultaron significativos sólo para el caso de VZ. Para el caso de GE y VZ, el PPS de dichos modelos fue superior que el de *filtro de kalman*. En resumen, de acuerdo a la tabla 4, la capacidad predictiva de los modelos es decreciente a medida que aumenta el conjunto extramuestral.

A continuación se calculó la rentabilidad promedio anual (RPA) que hubiera logrado un inversionista de haber seguido las recomendaciones de compra/venta de estos modelos, asumiendo una inversión inicial de US\$100.000. Los resultados muestran que para GE, la rentabilidad de todos los modelos fue negativa, para un horizonte extramuestral de un año. No obstante, a partir del segundo año en adelante, la mejor rentabilidad es alcanzada por los modelos de *algoritmos genéticos*. La peor rentabilidad fue representada por el modelo de *filtro de kalman*, la cual fue negativa en todos los años. Para el caso de GM, la mejor rentabilidad es alcanzada con el modelo de *filtro de kalman*, seguido por el modelo de *algoritmos genéticos*. Con respecto a IBM, UTX y VZ, los modelos de *algoritmos genéticos* lograron en promedio una rentabilidad mayor que modelos de *filtro de kalman*, multivariantes ARIMAX(2,2,2) y el modelo ingenuo AR(1). Sin embargo, para el caso de IBM y UTX, los modelos de *filtro de kalman* alcanzaron la segunda mejor rentabilidad. Sólo en el caso de VZ, los modelos multivariantes ARIMAX(2,2,2) superaron a los modelos de *filtro de kalman*, pero con una rentabilidad menor que los de *algoritmos genéticos*.

Las estrategias pasivas e ingenuas, para tres de las cinco acciones en estudio: GM, IBM y VZ, fueron las que obtuvieron los peores desempeños, siendo superadas por los *algoritmos genéticos* y el *filtro de kalman*, tanto en PPS, como en RPA. Los resultados apuntan a que los modelos multivariados obtenidos con los *algoritmos genéticos* son más robustos y permitirían obtener mejores rentabilidades que los modelos de *filtro de kalman*.

4. Conclusiones

Los resultados muestran que los modelos de *algoritmos genéticos* obtuvieron la mejor RPA en cuatro de las cinco acciones analizadas. Además, dichos modelos fueron superiores a los de *filtro de kalman* en términos de la capacidad de predicción del signo de las variaciones en el precio de las acciones analizadas.

Los resultados apuntan a la conveniencia de realizar estrategias activas de transacción en las acciones estudiadas que modelen el cambio de signo esperado de los activos, ya sea con técnicas como *algoritmos genéticos* o *filtro de kalman*, ya que estrategias alternativas de inversión, como la modelación ingenua o la estrategia pasiva, no fueron capaces de generar buenos resultados, tanto en términos estadísticos como económicos. Al mismo tiempo, los resultados confirman la importancia de medir no sólo la capacidad predictiva y la rentabilidad acumulada para los activos en el período estudiado, sino que también la posibilidad de equivocarse al determinar las preferencias entre una y otra técnica, dadas las características puntuales de los datos. Es por esta razón que se espera en el futuro próximo desarrollar procesos de robustez de los resultados como el *bootstrapping*.

Bibliografía

Conrad J. & Kaul G. (1988). Time-variation in expected returns. *Journal of Business* 61, 409-425.

Conrad J. & Kaul G. (1989). Mean reversion in short-horizon expected returns. *Review of Financial Studies* 2, 225-240.

Cortázar, G., Schwartz, E. & Naranjo, L. (2003). Term structure estimation in low-transaction frequency markets: A Kalman filter approach with incomplete panel data. Working Paper, Anderson School of Management at UCLA.

Díaz, A. (2000). Modelo de precios de commodities, filtro de Kalman y su aplicación a políticas óptimas de hedging en mercados de futuros. Tesis de Magister en Ciencias de la Ingeniería, Pontificia Universidad Católica de Chile.

Fama, Eugene (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25, p.p. 383-417.

Fama, E. & French K. R. (1988). Permanent and temporary components of stock prices. *Journal of Political Economy* 98, 247-273.

Hodgson, A. & Nicholls, D. (1991). The impact of index futures markets on Australian share market volatility. *Journal of Business Finance and Accounting* 18, 267-280.

Kalman, E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* D(82): 34-45.

Leung Mark T., Daouk Hazem, Chen An-Sing (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting* (16)2, pp.173-190.

Lo, A. & MacKinley A. C. (1988). Stock market price do not follow random walk: Evidence from a simple specification test. *Review of Financial Studies* 1, 41-66.

Naranjo, L. (2002). Modelos lognormales de precios de commodities y calibración mediante el filtro de Kalman utilizando paneles de datos incompletos de futuros de cobre y petróleo. Tesis de Magister en Ciencias de la Ingeniería, PUC.

Parisi, A., Parisi, F. & Guerrero, J. L. (2003). Modelos Predictivos de Redes Neuronales en Índices Bursátiles. *El Trimestre Económico*, México. (Forthcoming).

Parisi, A., Parisi, F. & Díaz, D. (2006). Modelos de Algoritmos Genéticos y Redes Neuronales en la Predicción de Índices Bursátiles Asiáticos.

Parisi, A., Parisi, F. & Cornejo, E. (2003). Algoritmos Genéticos Recursivos en la Predicción de Índices Bursátiles de América del Norte: IPC, TSE, NASDAQ y DJI.

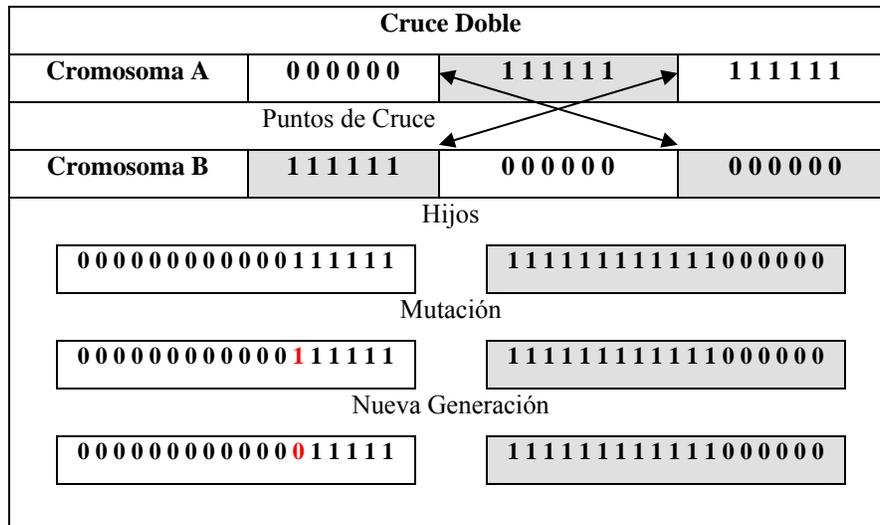
Pennacchi, G. (1991). Identifying the dynamics of real interest rates and inflation: Evidence using survey data. *The Review of Financial Studies* 4(1): 53-86.

Pesaran, M.H. & Timmermann A. (1992). A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics* 10, pp. 461-465.

Schwartz, E. (1997). The stochastic behaviour of commodity prices: Implications for valuations and hedging. *Journal of Finance* 52 (3): 923-973.

Anexos 1

CUADRO N° 1
Esquema del Operador de Cruce



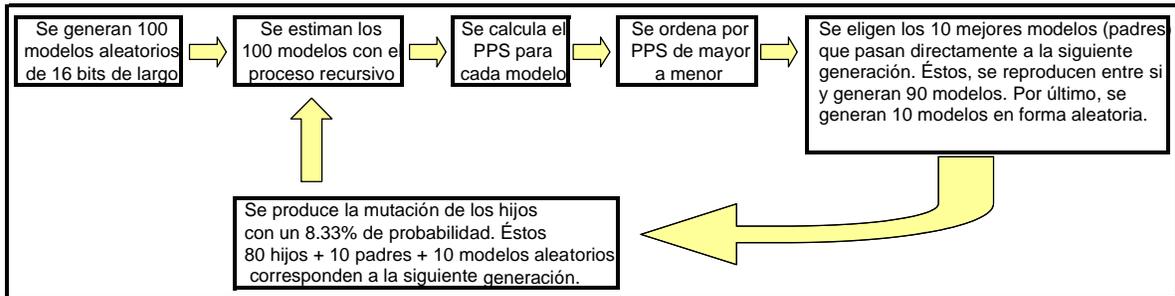
En este cuadro el Cromosoma A representa a un padre, individuo o modelo, cuyas variables explicativas corresponden a 6 rezagos del error, y a 6 rezagos del índice DJI. El Cromosoma B representa un modelo cuyas variables explicativas corresponden a 6 rezagos de la variable dependiente (variación en el precio de la acción). Cada hijo toma una porción de los genes de sus padres, de acuerdo al operador de cruce doble. Con una probabilidad del 8.33% algunos genes de los hijos mutarán para dar paso a potenciales soluciones no cubiertas en la selección aleatoria de la primera generación y sus sucesivas descendencias.

CUADRO N° 2
Terminología

Términos	Definiciones
Entorno o Medio Ambiente	Función Objetivo
Individuo Ejemplar	Una de las posibles soluciones
Población	Conjunto de soluciones en un momento determinado
Generación	Nombre que identifica a la población, en un momento determinado
Cromosoma	Código Binario
Genes	Los bits que conforman el código binario

CUADRO N° 3

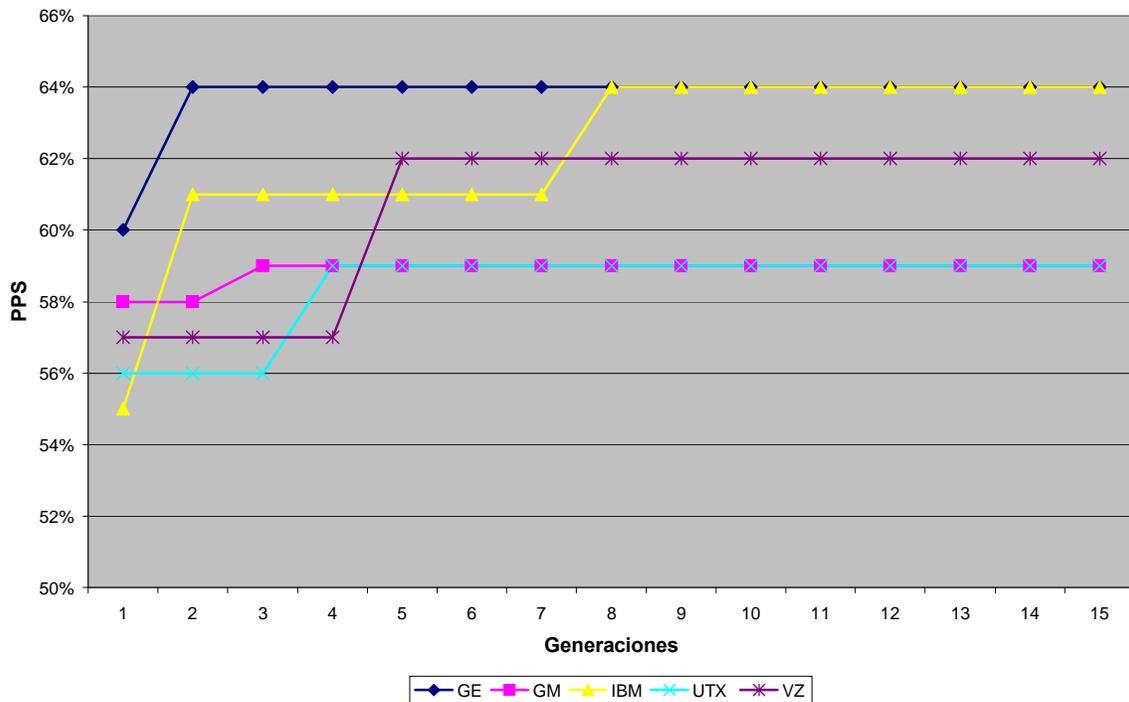
Esquema del Algoritmo genético de Selección de Modelos



Fuente: Elaboración propia. En este cuadro se pueden apreciar las diferentes etapas del proceso realizado por el algoritmo genético.

GRÁFICO N° 1

Evolución de PPS a 5 Años



Fuente: Elaboración propia. En este gráfico podemos apreciar que, en términos generales, el algoritmo converge a una solución óptima con anterioridad a la generación 9.

Anexos 2

CUADRO N° 4:
Resumen de los Resultados

Modelo	PPS 1 Año	PPS 2 Años	PPS 3 Años	PPS 4 Años	PPS 5 Años	RPA 1 Año	RPA 2 Años	RPA 3 Años	RPA 4 Años	RPA 5 Años
GE - General Electric Company										
Filtro de Kalman	69.23% (2,97)*	61.54% (2,63)*	54.49% (2,06)*	54.33% (2,13)*	53.28% (2,08)*	-28.76%	-31.87%	-30.86%	-30.86%	-30.86%
Algoritmo Genético	76.92% (3,99)*	74.04% (4,96)*	69.23% (4,83)*	65.38% (4,45)*	64.86% (4,79)*	-14.52%	8.82%	15.18%	10.97%	6.87%
ARIMAX (2,2,2)	63.46% (1,95)	59.62% (1,98)*	56.41% (1,61)	54.33% (1,25)	51.74% (0,56)	-8.19%	-16.03%	13.45%	9.57%	3.93%
AR(1)	53.85% (1,29)	53.85% (0,46)	50.00% (0,08)	50.48% (0,14)	49.81% (0,34)	-7.01%	-7.01%	-9.54%	-8.89%	-6.54%
Estrategia Pasiva	-	-	-	-	-	-24.81%	-24.08%	-13.76%	-9.99%	-8.93%
GM - General Motors Corporation										
Filtro de Kalman	69.23% (2,96)*	63.46% (3,10)*	60.90% (3,16)*	58.65% (2,73)*	55.60% (2,44)*	58.13%	58.46%	74.70%	64.38%	23.28%
Algoritmo Genético	71.15% (3,16)*	73.08% (4,73)*	66.67% (4,16)*	61.06% (3,16)*	59.07% (2,93)*	21.80%	40.80%	33.10%	13.20%	-30.30%
ARIMAX (2,2,2)	67.31% (2,60)*	56.73% (1,41)	56.41% (1,63)	58.65% (2,51)*	55.21% (1,68)	-17.39%	-8.30%	-20.91%	-20.09%	-10.27%
AR(1)	55.77% (1,31)	52.88% (0,40)	48.08% (0,97)	49.04% (0,54)	49.42% (0,12)	-8.95%	-10.89%	-13.67%	-28.98%	-24.51%
Estrategia Pasiva	-	-	-	-	-	-9%	-17%	-12.15	-23.95%	-29.44%

* En el test DA el valor de Z crítico es de 1,96, para un nivel de significancia del 5%.

CUADRO N° 4:
Resumen de los Resultados
(Continuación)

Modelo	PPS 1 Año	PPS 2 Años	PPS 3 Años	PPS 4 Años	PPS 5 Años	RPA 1 Año	RPA 2 Años	RPA 3 Años	RPA 4 Años	RPA 5 Años
IBM - International Business Machines Corporation										
Filtro de Kalman	69.23% (2,94)*	64.42% (3,02)*	60.26% (2,7)*	60.10% (2,89)*	59.46% (2,95)*	6.13%	4.87%	3.97%	7.90%	0.99%
Algoritmo Genético	80.77% (4,53)*	76.92% (5,65)*	69.87% (4,97)*	67.31% (4,96)*	64.09% (4,50)*	9.19%	42.68%	46.38%	52.67%	53.86%
ARIMAX (2,2,2)	63.46% (1,96)*	59.62% (2,01)*	54.49% (1,13)	55.29% (1,49)	53.28% (1,03)	-9.42%	-0.13%	8.56%	-7.46%	-10.90%
AR(1)	46.15% (0,00)	50.00% (1,10)	50.00% (1,07)	51.92% (1,13)	52.51% (1,14)	-10.37%	-10.37%	-10.37%	-10.37%	-10.37%
Estrategia Pasiva	-	-	-	-	-	-24.07%	-16.76%	-95.41%	-65.19%	-117.78%
UTX - United Technologies Corporation										
Filtro de Kalman	59.62% (1,97)*	58.65% (1,42)	59.62% (1,70)	55.29% (0,47)	54.44% (0,0)	4.78%	23.94%	49.04%	67.93%	93.43%
Algoritmo Genético	75.00% (3,63)*	68.27% (3,64)*	63.46% (3,21)*	61.06% (3,13)*	59.07% (2,88)*	37.57%	99.23%	103.28%	107.32%	151.45%
ARIMAX (2,2,2)	61.54% (1,67)	53.85% (0,79)	51.28% (0,35)	50.96% (0,32)	50.19% (0,13)	17.76%	32.16%	51.74%	53.35%	79.03%
AR(1) [†]	28.85% (0,24)	37.50% (0,28)	42.95% (0,59)	42.79% (1,28)	45.56% (0,61)	-29.49%	-12.01%	1.36%	15.72%	44.10%
Estrategia Pasiva	-	-	-	-	-	-11.86%	20.91%	42.99%	56.50%	90.48%

En el test DA el valor de Z crítico es de 1,96, para un nivel de significancia del 5%.

[†] En el caso de la acción UTX, el modelo AR(1) no puede ser invertido debido a que el PPS intramuestral fue de 61.54%.

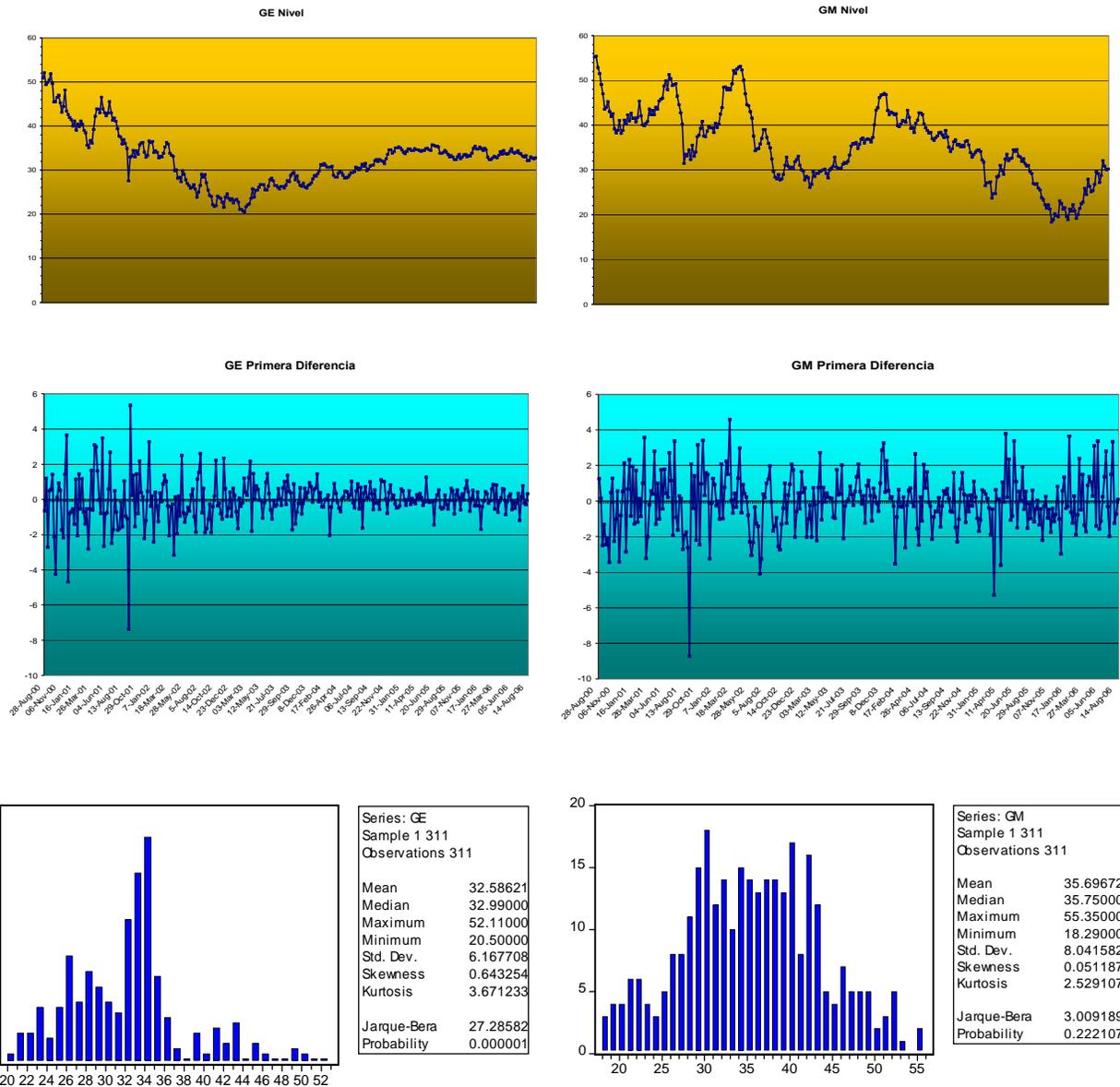
CUADRO N° 4:
Resumen de los Resultados
(Continuación)

Modelo	PPS 1 Año	PPS 2 Años	PPS 3 Años	PPS 4 Años	PPS 5 Años	RPA 1 Año	RPA 2 Años	RPA 3 Años	RPA 4 Años	RPA 5 Años
VZ - Verizon Communications Inc.										
Filtro de Kalman	67.31% (0,54)	61.54% (1,8)*	57.69% (1,68)*	56.73% (1,63)	55.60% (1,55)	-3.87%	-22.61%	-21.45%	-24.81%	-30.93%
Algoritmo Genético	75.00% (3,68)*	71.15% (4,33)*	67.31% (4,32)*	64.42% (4,14)*	61.78% (3,77)*	11.28%	27.09%	41.32%	55.75%	60.80%
ARIMAX (2,2,2)	67.31% (2,12)*	59.62% (1,88)	58.33% (2,08)*	58.17% (2,36)*	56.37% (2,05)*	-22.81%	7.62%	25.41%	11.22%	17.60%
AR(1)	50.00% (0,99)	51.92% (0,08)	50.00% (0,07)	48.56% (0,37)	49.03% (0,37)	-4.76%	-22.41%	-27.16%	-43.22%	-46.53%
Estrategia Pasiva	-	-	-	-	-	-35.91%	-23.87%	-12.33%	-23.50%	-16.35%

* En el test DA el valor de Z crítico es de 1,96, para un nivel de significancia del 5%.

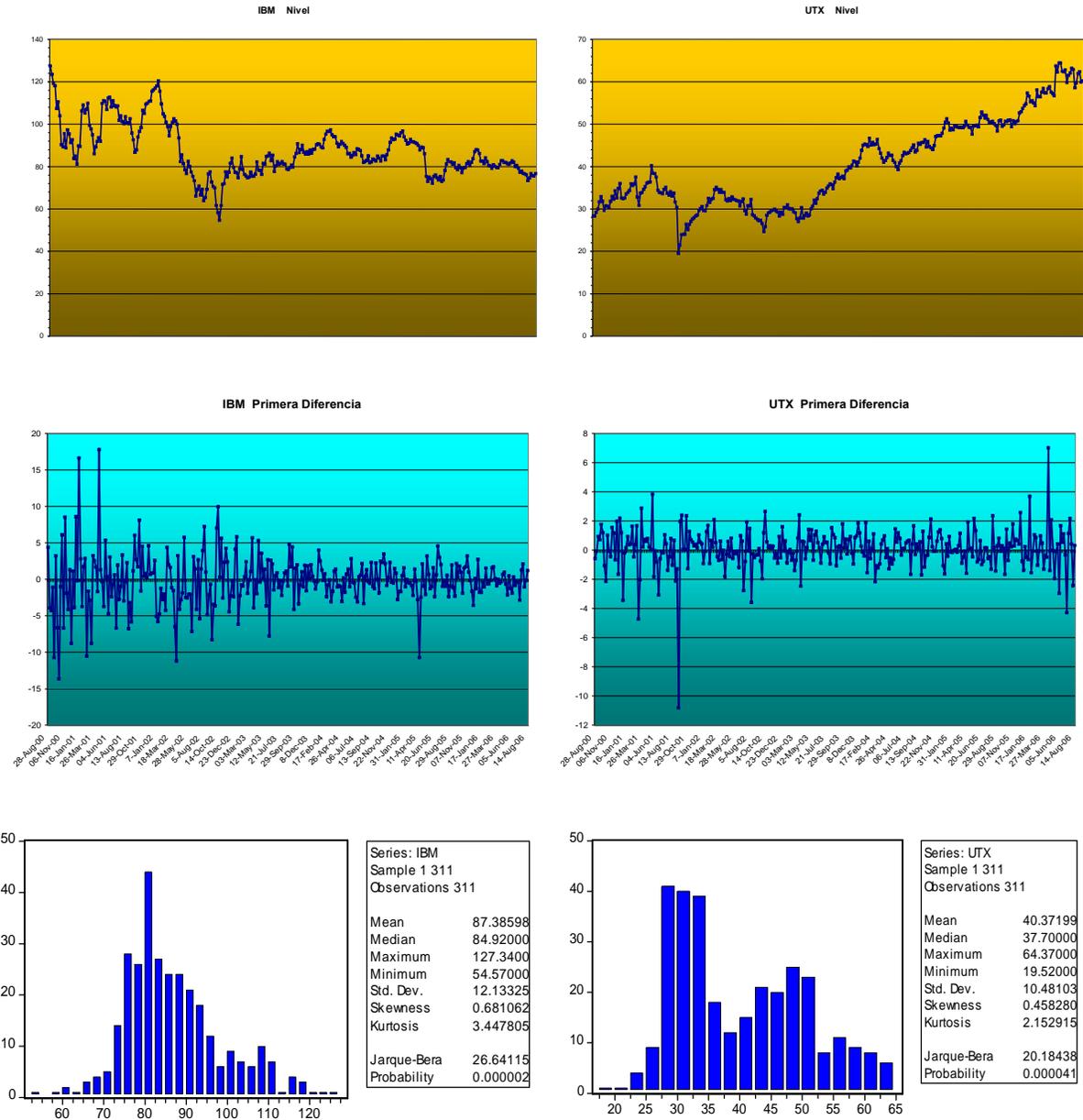
Anexos 3

Características de las series General Electric Company (GE) y General Motors Corporation (GM)



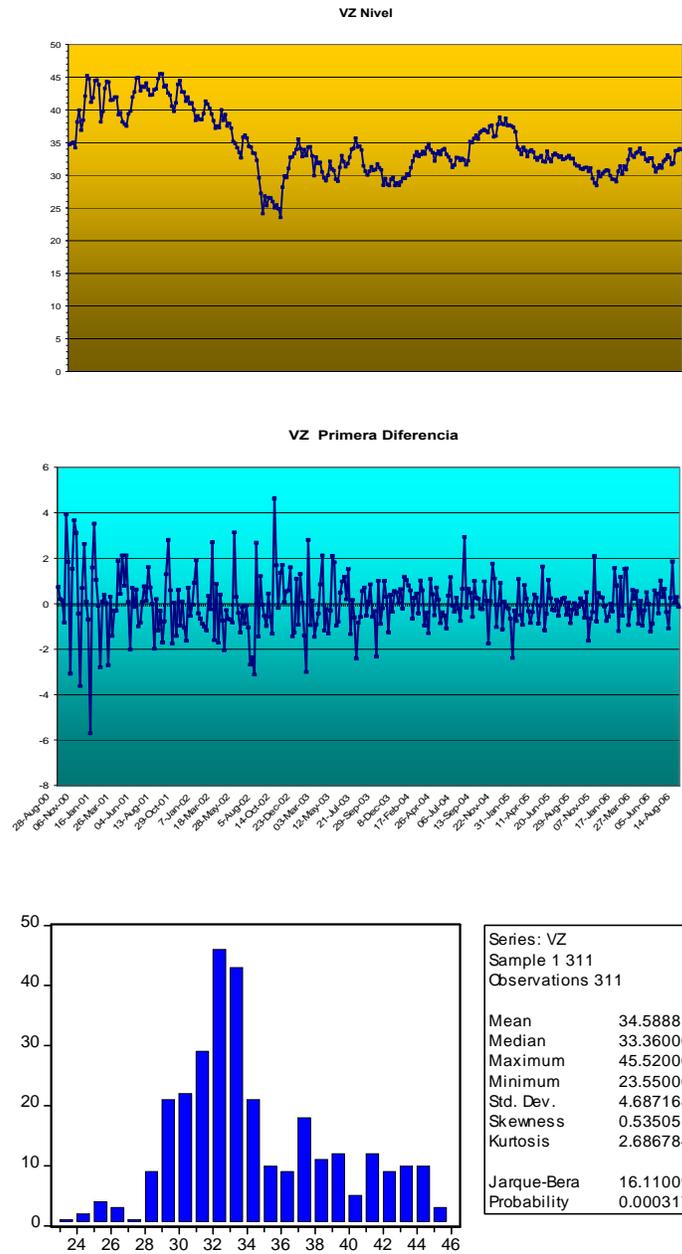
Fuente: Elaboración propia.

Características de las series International Business Machines Corporation (IBM) y United Technologies Corporation (UTX)



Fuente: Elaboración propia.

Características de las series Verizon Communications Inc. (VZ)



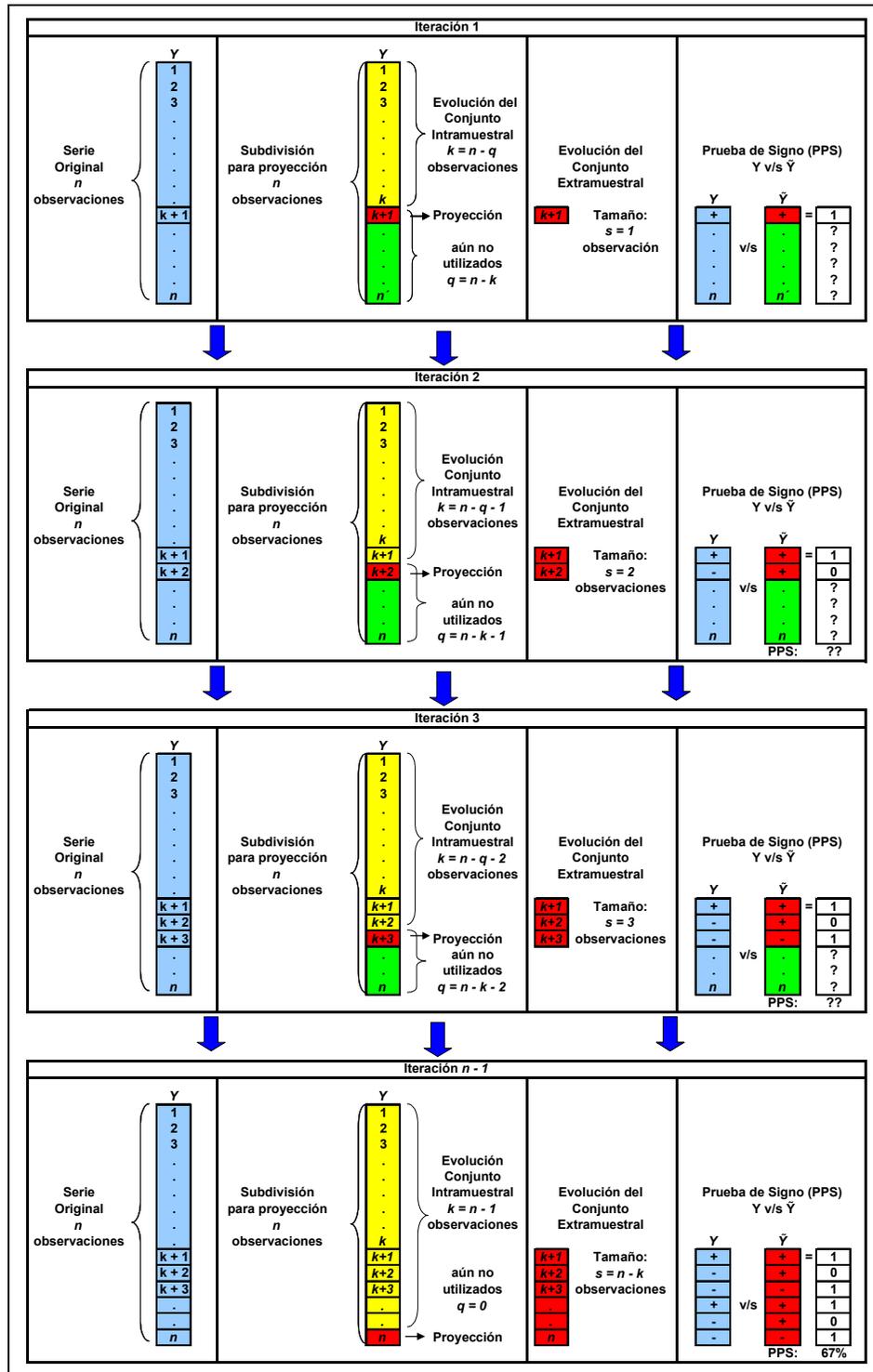
Fuente: Elaboración propia.

Anexos 4

Funcionamiento Recursivo en Modelos de Predicción

Esta metodología ha sido empleada anteriormente para medir el desempeño de modelos de redes neuronales que buscan predecir períodos de recesión en los Estados Unidos (Qi, 2001; Estrella y Mishkin, 1998) y para proyectar el signo de la variación de índices bursátiles internacionales (Parisi, Parisi y Guerrero, 2003). El funcionamiento recursivo consiste en agregar, mediante un algoritmo iterativo, nueva información a la ya estudiada por el modelo de predicción. Desde el punto de vista financiero, esto nos permite reflejar que es necesario que los agentes ajusten sus expectativas frente a la nueva información obtenida, pero sin dejar de considerar la totalidad de la información pasada, pues si razonan utilizando toda la información disponible sus predicciones serán más completas y por lo tanto, probablemente más acertadas. De esta forma, si consideramos que inicialmente el conjunto extramuestral tiene “ $m-n$ ” datos, al evaluar el funcionamiento del modelo de predicción se consideró sólo la proyección del valor inmediatamente cercano al último dato intramuestral. Posteriormente, el dato analizado sale del conjunto extramuestral (quedando éste con “ $m-n-1$ ” datos) y pasa a formar parte del conjunto intramuestral, por lo que la muestra de “ n ” datos que contiene los valores de entrada se incrementa a “ $n+1$ ”. Luego se realizó una nueva iteración, lo que implicó reestimar los pesos del modelo para cada una de las “ $m-n$ ” proyecciones. Este proceso se repitió hasta que en el conjunto extramuestral quedó sólo un dato, es decir, hasta el momento en que la última observación (correspondiente al período “ $m-1$ ”) es utilizada para proyectar el valor que la variable de salida podría alcanzar en el momento “ m ”, el cual representa el futuro inmediato. En la página siguiente se presenta un esquema del proceso recursivo.

Esquema del Proceso Recursivo



Fuente: En el proceso recursivo el conjunto intramensual aumenta su tamaño a través de las iteraciones, de manera de reflejar que los agentes van incorporando la nueva información a medida que esta está disponible en el mercado.

Anexos 5

Salidas Correspondientes a los Modelos Ingenuos, para cada una de las Acciones Estudiadas.

Salida GE

Dependent Variable: GE				
Method: Least Squares				
Date: 09/26/06 Time: 16:21				
Sample(adjusted): 2 312				
Included observations: 311 after adjusting endpoints				
GE=C(1)+C(2)*GE(-1)				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.064352	0.068933	-0.933547	0.3513
C(2)	-0.077655	0.056271	-1.380006	0.1686
R-squared	0.006125	Mean dependent var		-0.060257
Adjusted R-squared	0.002909	S.D. dependent var		1.21629
S.E. of regression	1.21452	Akaike info criterion		3.232985
Sum squared resid	455.7931	Schwarz criterion		3.257035
Log likelihood	-500.7291	F-statistic		1.904417
Durbin-Watson stat	1.998296	Prob(F-statistic)		0.168582

Salida GM

Dependent Variable: GM				
Method: Least Squares				
Date: 09/26/06 Time: 17:04				
Sample(adjusted): 2 312				
Included observations: 311 after adjusting endpoints				
GM=C(1)+C(2)*GM(-1)				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.0711	0.0926	-0.767822	0.4432
C(2)	0.071587	0.056696	1.262636	0.2077
R-squared	0.005133	Mean dependent var		-0.076334
Adjusted R-squared	0.001913	S.D. dependent var		1.632938
S.E. of regression	1.631376	Akaike info criterion		3.823134
Sum squared resid	822.3683	Schwarz criterion		3.847184
Log likelihood	-592.4974	F-statistic		1.594249
Durbin-Watson stat	2.0107	Prob(F-statistic)		0.207672

Fuente: Elaboración propia.

Salida IBM

Dependent Variable: IBM				
Method: Least Squares				
Date: 09/26/06 Time: 17:14				
Sample(adjusted): 2 312				
Included observations: 311 after adjusting endpoints				
IBM=C(1)+C(2)*IBM(-1)				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.145618	0.201067	-0.724227	0.4695
C(2)	0.024505	0.056386	0.434587	0.6642
R-squared	0.000611	Mean dependent var		-0.148714
Adjusted R-squared	-0.002623	S.D. dependent var		3.53899
S.E. of regression	3.543629	Akaike info criterion		5.37459
Sum squared resid	3880.207	Schwarz criterion		5.39864
Log likelihood	-833.7487	F-statistic		0.188866
Durbin-Watson stat	2.016134	Prob(F-statistic)		0.664166

Salida UTX

Dependent Variable: UTX				
Method: Least Squares				
Date: 09/26/06 Time: 17:24				
Sample(adjusted): 2 312				
Included observations: 311 after adjusting endpoints				
UTX=C(1)+C(2)*UTX(-1)				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.110383	0.079203	1.393659	0.1644
C(2)	-0.093698	0.056611	-1.655112	0.0989
R-squared	0.008787	Mean dependent var		0.100772
Adjusted R-squared	0.00558	S.D. dependent var		1.396912
S.E. of regression	1.393009	Akaike info criterion		3.507219
Sum squared resid	599.6065	Schwarz criterion		3.53127
Log likelihood	-543.3726	F-statistic		2.739395
Durbin-Watson stat	2.015573	Prob(F-statistic)		0.098917

Fuente: Elaboración propia.

Salida VZ

Dependent Variable: VZ				
Method: Least Squares				
Date: 09/26/06 Time: 17:34				
Sample(adjusted): 2 312				
Included observations: 311 after adjusting endpoints				
VZ=C(1)+C(2)*VZ(-1)				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.000173	0.070242	-0.002468	0.998
C(2)	0.036231	0.056835	0.637482	0.5243
R-squared	0.001313	Mean dependent var		-9.65E-05
Adjusted R-squared	-0.001919	S.D. dependent var		1.23755
S.E. of regression	1.238736	Akaike info criterion		3.272471
Sum squared resid	474.1505	Schwarz criterion		3.296521
Log likelihood	-506.8692	F-statistic		0.406383
Durbin-Watson stat	1.990404	Prob(F-statistic)		0.524283

Fuente: Elaboración propia.