

**UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA**  
**DEPARTAMENTO DE MÉTODOS CUANTITATIVOS EN ECONOMÍA Y**  
**GESTIÓN**



**TESIS DOCTORAL**

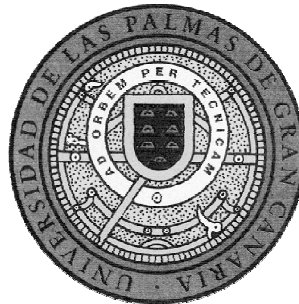
**NUEVAS PERSPECTIVAS DEL ANÁLISIS TÉCNICO DE LOS  
MERCADOS BURSÁTILES MEDIANTE EL APRENDIZAJE  
AUTOMÁTICO. APLICACIONES AL ÍNDICE GENERAL DE LA  
BOLSA DE MADRID.**

**CHRISTIAN GONZÁLEZ MARTEL**

Las Palmas de Gran Canaria, 2003

NUEVAS PERSPECTIVAS DEL ANÁLISIS TÉCNICO  
DE LOS MERCADOS BURSÁTILES MEDIANTE EL  
APRENDIZAJE AUTOMÁTICO. APLICACIONES AL  
ÍNDICE GENERAL DE LA BOLSA DE MADRID

Por  
Christian González Martel



EN CUMPLIMIENTO DE LOS REQUERIMIENTOS  
PARA LA SOLICITUD DEL GRADO DE DOCTOR EN  
ECONOMÍA CUANTITATIVA APLICADA  
POR LA  
UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA  
LAS PALMAS DE GRAN CANARIA  
MAYO 2003



UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA  
DEPARTAMENTO DE  
MÉTODOS CUANTITATIVOS EN ECONOMÍA Y GESTIÓN

Los abajo firmantes certifican que han leído y recomendado a la Facultad de Ciencias Económicas y Empresariales para su aceptación la tesis titulada “**Nuevas perspectivas del análisis técnico de los mercados bursátiles mediante el aprendizaje automático. Aplicaciones al Índice General de la Bolsa de Madrid**” realizada por **Christian González Martel** en cumplimiento de los requerimientos para el grado de doctor en:  
**Economía Cuantitativa Aplicada.**

Fecha: Mayo 2003

Director:

\_\_\_\_\_  
Fernando Fernández Rodríguez

Codirector:

\_\_\_\_\_  
Simón Sosvilla Rivero

UNIVERSIDAD DE LAS PALMAS DE GRAN  
CANARIA

Fecha: **Mayo 2003**

Autor: **Christian González Martel**

Título: **Nuevas perspectivas del análisis técnico de los  
mercados bursátiles mediante el aprendizaje  
automático. Aplicaciones al Índice General de  
la Bolsa de Madrid**

Dep.: **Métodos Cuantitativos en Economía y Gestión**

Grado: **Doctor**      Convocatoria: **Mayo**      Año: **2003**

---

Firma del Autor

*A mis padres.*

# Índice general

<b>Agradecimientos</b>	<b>ix</b>
<b>1. Introducción y Objetivos</b>	<b>1</b>
1.1. Introducción . . . . .	2
1.1.1. Explosión de las técnicas predictivas y la era de las TICS .	7
1.1.2. Desarrollo de la minería de datos . . . . .	9
1.1.3. El empleo de sistemas expertos e inteligencia artificial en la predicción financiera . . . . .	10
1.1.4. Empleo de sistemas inteligentes de optimización . . . . .	13
1.1.5. Irrupción de los modelos no lineales en Economía y Finanzas	15
1.1.6. Empleo de técnicas predictivas basadas en la regresión no paramétrica . . . . .	18
1.2. Objetivos . . . . .	20
<b>2. El objeto del análisis: el problema de la predicción financiera</b>	<b>29</b>
2.1. Introducción . . . . .	31
2.2. Análisis Fundamental . . . . .	33
2.3. Análisis Técnico . . . . .	36
2.3.1. Chartismo . . . . .	39
2.3.2. Reglas basadas en medias móviles . . . . .	46
2.3.3. Reglas basadas en órdenes estadísticos . . . . .	50
2.4. La hipótesis de los mercados eficientes . . . . .	54

2.4.1.	Anomalías e imperfecciones empíricas . . . . .	60
2.5.	Renacimiento académico del análisis técnico . . . . .	62
<b>3.</b>	<b>Los métodos de análisis: nuevas tecnologías predictivas basadas en el aprendizaje automático</b>	<b>69</b>
3.1.	Aprendizaje automático . . . . .	71
3.1.1.	Estado del arte del aprendizaje automático . . . . .	72
3.1.2.	Características del aprendizaje automático . . . . .	75
3.2.	Redes neuronales artificiales . . . . .	82
3.2.1.	Estado del arte de las redes neuronales artificiales . . . . .	82
3.2.2.	Estructura de una red neuronal biológica. . . . .	84
3.2.3.	Redes neuronales artificiales . . . . .	85
3.2.4.	Redes de retropropagación . . . . .	90
3.3.	Algoritmos genéticos . . . . .	101
3.3.1.	Evolución natural . . . . .	102
3.3.2.	Algoritmos Genéticos . . . . .	109
3.3.3.	Algoritmos genéticos con parámetros continuos . . . . .	129
3.4.	Predicciones por ocurrencias análogas . . . . .	131
3.4.1.	Algunos criterios de selección de los parámetros en la predicción NN . . . . .	139
3.4.2.	Selección automática de los parámetros en la predicción NN con AGs. . . . .	150
3.5.	Programación Genética . . . . .	155
3.5.1.	Proceso evolutivo de la programación genética . . . . .	155
3.5.2.	Estructura de los programas de computación en la programación genética . . . . .	156
3.5.3.	Inicialización de la población . . . . .	161
3.5.4.	Cruce . . . . .	163
3.5.5.	Mutación . . . . .	169



<b>4. Resultados empíricos: Aplicaciones a la predicción financiera del IGBM.</b>	<b>173</b>
4.1. Evaluación económica de las reglas técnicas . . . . .	174
4.2. Estrategias técnicas mediante redes neuronales . . . . .	185
4.2.1. Ejercicio experimental hasta 1997 . . . . .	185
4.2.2. Análisis empírico para los años 1997 a 2002 . . . . .	193
4.2.3. Comportamiento de una estrategia, con posiciones recomendadas para estar fuera o dentro del mercado, optimizada mediante redes neuronales. . . . .	196
4.3. Optimización de reglas técnicas mediante algoritmos genéticos . .	209
4.3.1. Ejercicio experimental hasta 1997 . . . . .	216
4.3.2. Análisis empírico para los años 1997 a 2002 . . . . .	221
4.4. Optimización de los parámetros utilizados en predicciones por ocurrencias análogas mediante algoritmos genéticos . . . . .	225
4.4.1. Ejercicio experimental hasta 1997 . . . . .	225
4.4.2. Análisis empírico para los años 1997 a 2002 . . . . .	283
4.5. Creación de nuevas reglas técnicas mediante la programación genética	316
4.5.1. Ejercicio experimental hasta 1997 . . . . .	319
4.5.2. Análisis empírico para los años 1997 a 2002 . . . . .	334
 <b>5. Conclusiones y desarrollos futuros</b>	 <b>341</b>
5.1. Conclusiones específicas sobre las metodologías predictivas aportadas a la serie IGBM . . . . .	342
5.1.1. Reglas técnicas obtenidas mediante redes neuronales artificiales . . . . .	342
5.1.2. Optimización de reglas técnicas clásicas mediante algoritmos genéticos . . . . .	343
5.1.3. Creación de nuevas reglas técnicas mediante el uso de la programación genética . . . . .	343

5.1.4.	Optimización de reglas técnicas mediante el uso de predictores por ocurrencias análogas . . . . .	344
5.2.	Sobre la hipótesis de mercados eficientes en el IGBM . . . . .	345
5.3.	Conclusiones generales . . . . .	346
5.3.1.	Sobre la necesidad de introducir modelos no lineales de predicción . . . . .	346
5.3.2.	Nueva visión de la Hipótesis de Mercados Eficientes . . . .	347
5.3.3.	La influencia de los costes de transacción en la búsqueda de reglas técnicas óptimas . . . . .	349
5.3.4.	Sobre el figoneo de datos y la sobreparametrización . . . .	351
5.3.5.	Evaluación del riesgo de las predicciones financieras . . . .	352
5.3.6.	Sobre el carácter de las reglas técnicas . . . . .	353
5.4.	Desarrollos futuros . . . . .	354

<b>Referencias Bibliográficas</b>	<b>357</b>
-----------------------------------	------------

# Agradecimientos

Quisiera agradecer a Fernando Fernández Rodríguez y Simón Sosvilla Rivero, por sus múltiples sugerencias y constante apoyo. Sin sus conocimientos sobre esta materia me hubiese sido imposible salir del caos en los que me vi sumido en los primeros años.

A Julian Andrada, por compartir conmigo su amplísima experiencia sobre mercados financieros y algo más importante para mi como es su amistad.

A los compañeros del departamento de Métodos Cuantitativos en Economía y Gestión porque siempre me han ayudado en los momentos en los que lo necesité.

Como no, a mis padres y hermanos, por todo lo que han soportado, por su paciencia y cariño.

José, Rene, Norberto y Juan Luis por las horas no dedicadas, por los buenos momentos ya que malos nunca ha habido.

Y a Vicky, porque siempre, siempre confió en mi. Siempre tuvo claro que esta tesis saldría adelante. Gracias Vicky.

Las Palmas de Gran Canaria, Las Palmas  
Mayo, 2003

Christian González Martel

# Capítulo 1

## Introducción y Objetivos

### Índice General

---

<b>1.1. Introducción</b>	<b>2</b>
1.1.1. Explosión de las técnicas predictivas y la era de las TICS	7
1.1.2. Desarrollo de la minería de datos	9
1.1.3. El empleo de sistemas expertos e inteligencia artificial en la predicción financiera	10
1.1.4. Empleo de sistemas inteligentes de optimización	13
1.1.5. Irrupción de los modelos no lineales en Economía y Finanzas	15
1.1.6. Empleo de técnicas predictivas basadas en la regresión no paramétrica	18
<b>1.2. Objetivos</b>	<b>20</b>

---

## 1.1. Introducción

La correcta predicción de los mercados financieros es una antigua quimera perseguida desde sus inicios por los inversores. Numerosos modelos han sido desarrollados con el fin de proporcionar predicciones de los precios o los rendimientos aunque, generalmente, el éxito ha sido bastante limitado. Tales modelos podrían clasificarse en dos grandes categorías:

- Los modelos basados en el Análisis Fundamental que pretenden inferir la evolución futura de las cotizaciones mediante el análisis de las variables fundamentales de la empresa, sector o la economía.
- Los modelos basados en el Análisis Técnico, que consiste en la utilización de la información histórica de las series financieras con fines predictivos.

La capacidad predictiva que se derivaría de estos modelos no puede justificarse desde el punto de vista de la Hipótesis de Mercados Eficientes en su forma semifuerte que afirma que los precios de los activos reflejan instantáneamente toda la información pública disponible (Fama, 1970). Bajo este punto de vista, cualquier oportunidad predecible de obtener beneficios en un mercado financiero tendría una existencia verdaderamente efímera, al ser explotada casi instantáneamente por los agentes. No obstante, se plantea la cuestión de por qué tantos individuos e instituciones se empeñan continuamente en obtener modelos capaces de batir al mercado. En este sentido, Grossman y Stiglitz (1980) sostienen que los inversores tienen poderosos incentivos para descubrir reglas de predicción financiera sólo si tal actividad es capaz de generar rendimientos más altos de las inversiones, de manera que los analistas financieros obtuviesen

compensaciones adecuadas por los gastos asociados con la construcción de los nuevos modelos predictivos. De esta forma, Grossman y Stiglitz (1980) plantearon la imposibilidad de eficiencia de los mercados financieros, desde el punto de vista de la información, atribuyendo la lentitud del proceso de ajuste al coste de adquirir y evaluar información a la necesidad de ajustarse a la nueva información. En un sentido similar Beja y Goldman (1980) señalan que los precios no siempre se ajustan rápidamente a la información, pudiendo encontrarse los mercados en situaciones de desequilibrio a corto plazo.

Tal como apuntan Min y Maddala (1999), otra explicación de los modelos predictivos descansa en la teoría financiera basada en el Modelo de Valoración de Activos de Capital (Capital Asset Pricing Model, CAPM). Según esta teoría, además de un rendimiento libre de riesgo, debe haber una compensación por el riesgo sistemático asociado con una posición en un mercado bursátil, de modo que cuanto más arriesgado sea la inversión en un activo deberá obtener unos mayores rendimientos esperados. De esta forma, los movimientos sistemáticos de los precios y el riesgo asociado pueden ser analizados por medio del comportamiento de los precios de los activos financieros y de los factores económicos subyacentes. Por tanto, la parte de los rendimientos asociados a la prima de riesgo puede ser predicha. Desde esta perspectiva, la predicción de los rendimientos bursátiles se basa en identificar los factores (económicos y financieros) subyacentes y en las relaciones funcionales entre tales rendimientos y los factores fundamentales. Así, desde el punto de vista del Análisis Fundamental una porción significativa de los rendimientos de los activos ha sido encontrada predecible usando la información públicamente disponible de datos financieros y variables económicas,

especialmente las relacionadas con el ciclo económico como son los dividendos obtenidos, los tipos de interés, las tasas de crecimiento de la producción industrial, las tasas de inflación, las tasas de crecimiento de los agregados monetarios y otras (Min y Maddala, 1999).

No obstante la mayoría de los operadores del mercado realiza sus predicciones recurriendo al Análisis Técnico, intentando buscar patrones de comportamiento predecibles en las series de precio y de volumen de contratación. El Análisis Técnico está basado en los principios establecidos por el analista financiero Charles Dow, y publicados en el Wall Street Journal algunos años antes de su muerte en 1902. A partir de tales principios han ido surgiendo multitud de reglas de carácter empírico destinadas a predecir la evolución de las cotizaciones. Aunque esta tecnología financiera se remonta a comienzos del siglo XX, durante los últimos años se ha producido un considerable desarrollo de la disciplina. Los cambios se iniciaron con la incorporación de ordenadores e información digital a las oficinas de contratación. Las viejas técnicas chartistas de análisis gráfico iban a ser sustituidas por ventanas e imágenes tridimensionales, que ayudaban al análisis del comportamiento de los mercados financieros desde las pantallas de los ordenadores. El siguiente paso ha sido el de crear máquinas de contratación inteligentes, susceptibles de ser usadas en el reconocimiento de patrones, y capaces de tomar decisiones autónomas sobre la contratación y la optimización de carteras.

El vertiginoso desarrollo de la tecnología de los mercados financieros se ha producido en un doble sentido. Por una parte, los economistas han flexibilizado su percepción del significado de la eficiencia del mercado, permitiendo algún

fundamento teórico que pudiese avalar la práctica del Análisis Técnico. Así, como señalan Campbell *et al.* (1997) en su célebre manual de Econometría Financiera, la literatura académica admite hoy día una teoría sobre los mercados eficientes mucho más flexible que hace veinticinco años, aceptando la posibilidad de cierta capacidad predictiva de los rendimientos de los mercados por medio de determinados algoritmos no lineales. Dicha capacidad predictiva se contempla hoy día asociada a un tipo de riesgo dinámico que es inherente al mercado. Así, durante los años 90 del siglo XX se produce un verdadero renacimiento académico del Análisis Técnico, demostrándose que las conclusiones acerca del nulo poder predictivo de las reglas chartistas eran, como mínimo, precipitadas. Brock *et al.* (1992) demostraron que los rendimientos obtenidos por algunas de las más populares estrategias técnicas no eran consistentes con los modelos más usados en finanzas como el modelo de paseo aleatorio o los modelos de la familia (G)ARCH. Corrado y Lee (1992) detectaron, igualmente, valor económico significativo a las reglas de filtro aplicadas a rendimientos bursátiles diarios. Para futuros de divisas, Taylor (1994) señaló que con la negociación de tales activos entre 1982 hasta 1990 se podrían haber obtenido beneficios usando una regla túnel chartista. Bessembinder y Chan (1995) sugirieron que algunos mercados asiáticos pueden no ser tan eficientes desde el punto de vista informativo como los europeos o norteamericanos, encontrando un gran éxito de las reglas técnicas en los mercados emergentes de Malasia, Tailandia y Taiwan. Igualmente ha aparecido toda una literatura teórica sobre los mercados con agentes heterogéneos que justifica la formación de patrones susceptibles de ser predichos por medio del Análisis Técnico (Bowden, 1990; Day y Huang, 1990; Chiarella, 1992; Lux, 1995; Brock y Hommes,



1997; Fernández Rodríguez *et al.*, 2002). Toda esta nueva literatura sugiere, modeliza y justifica que los agentes no reaccionan a la información de modo lineal ni homogéneo. Muy por el contrario, existen numerosos grupos de agentes con diferentes horizontes temporales de inversión, algunos de los cuales son propensos a las modas porque la experiencia y habilidad del conjunto no está uniformemente distribuida. Por lo tanto, de manera agregada, no hay por qué suponer que el conjunto de agentes actúen siempre de forma racional, incluso aunque tengan individualmente un comportamiento racional.

Por otra parte, el Análisis Técnico ha experimentado una importante transformación desde el punto de vista de las herramientas predictivas empleadas, ya que han ido recogiendo numerosos avances registrados en otras disciplinas. Así, en el actual panorama de la predicción financiera, el Análisis Técnico de los mercados ha ido incorporando tanto el desarrollo asociado a las nuevas tecnologías de la información y la comunicación (TICs), como el uso de métodos predictivos cada vez más sofisticados. Entre los nuevos avances en la investigación sobre predicción que comienzan a ejercer influencia en el Análisis Técnico cabría destacar los siguientes:

- Explosión de la actividad académica referente a los temas de predicción y al mejor uso de los conjuntos crecientes de información que proporcionan las nuevas TICs (y en particular el aumento de la velocidad de los microprocesadores).
- La utilización de sistemas inteligentes capaces de autoajustar determinados parámetros cruciales en los modelos estadísticos, dando paso a lo que ha dado en llamarse *data mining* o “minería de datos”.

- Amplio desarrollo y difusión de los sistemas expertos.
- El uso de herramientas predictivas que han sido tomadas prestadas de la inteligencia artificial y de los sistemas inteligentes de optimización.
- La irrupción de modelos de predicción no lineales que comienzan a competir con las teorías lineales clásicas de series temporales, gracias al desarrollo que se ha producido en el contexto de los procedimientos de selección de modelos.
- El empleo de técnicas predictivas basadas en desarrollos recientes de la regresión no paramétrica.

Por consiguiente, como introducción a esta Tesis Doctoral, sería conveniente desgarnar, de forma pormenorizada, cada uno de estos avances de la predicción y el entorno en que se han ido desarrollando.

### 1.1.1. Explosión de las técnicas predictivas y la era de las TICS

Hay que comenzar señalando que se ha producido un gran impulso de la actividad académica en el terreno de la predicción. En este sentido, ha ido apareciendo un creciente número de trabajos sobre predicción en numerosas revistas de estadística y econometría. Igualmente han surgido revistas destinadas específicamente al tema de la predicción como son *International Journal of Forecasting* y *Journal of Forecasting*. Todo ello ha animado el desarrollo de nuevos modelos y técnicas de predicción. Desde luego, en el terreno de la predicción económica, empresarial y financiera los mayores avances en el terreno predictivo

se han derivado del aumento del conjunto de información en que se basan las predicciones, aumento inequívocamente asociado a la era de las nuevas tecnologías de la información. Estas nuevas tecnologías han permitido, entre otras cosas, una creciente disponibilidad de la información generada por los mercados.

En este sentido, es indudable que los recientes avances en el Análisis Técnico han estado estrechamente relacionados con un importante desarrollo en las áreas relativas al almacenamiento, organización y búsqueda de datos, de modo que en la actualidad se cuenta con la existencia de cantidades masivas de datos tanto económicos como financieros. Gracias al aumento de la velocidad de los microprocesadores así como su menor precio ha producido una tendencia hacia el uso intensivo del ordenador y el aprendizaje empírico a partir de los datos ha tenido un paulatino impacto dentro de la literatura financiera donde el número de trabajos de finanzas empíricas, de alto contenido informático, ha crecido exponencialmente. Esta tendencia ha cristalizado igualmente en el nacimiento de nuevas revistas científicas altamente especializadas en el este tipo de investigaciones tales como pueden ser *Journal of Empirical Finance* o *Computational Economics*. Todo ello ha contribuido a la puesta en pie una nueva rama de la econometría que podría llamarse Econometría Financiera y Computacional. Debido al papel preponderante que ha tenido el ordenador en estos cambios no puede resultarnos sorprendente que muchas de las nuevas herramientas predictivas se hayan desarrollado en campos científicos exteriores a la estadística o la econometría clásicas, tales como son la ingeniería o las ciencias de la computación.

### 1.1.2. Desarrollo de la minería de datos

Con la llegada de los ordenadores y la era de las nuevas tecnologías de la información y la comunicación, han aparecido nuevas oportunidades para el análisis estadístico de datos financieros que el Análisis Técnico comienza a aprovechar desarrollando modelos econométricos cada vez más sofisticados. En este sentido hay que destacar la aparición de un nuevo campo de la modelización de datos conocido como “minerías de datos” o *data mining*. Aunque la calidad y habilidad para manejar los datos relacionados con el mercado que están hoy día en la mesa de los analistas de mercado es todavía limitada y raramente se examina con toda su complejidad, la minería de datos tiene un amplio futuro en su ayuda a las técnicas de dirección empresarial y puede convertirse en poco tiempo en la robótica de la gestión empresarial. Las vasta cantidad de información que está siendo generada en los numerosos campos de las finanzas permiten ahora la extracción de patrones y tendencias, así como la aparición de nuevas técnicas exploratorias que buscan el aprendizaje a partir de los datos, desbordando así la vieja idea de emplear como metodologías predictivas sólo y exclusivamente las que parten de la Teoría Económica previamente establecida. Durante los últimos años se ha puesto de manifiesto la eficacia de la aproximación a la modelización econométrica desde lo general a lo específico asociada a la London School of Economics, apareciendo toda una nueva literatura que podría encuadrarse bajo el título de “minería de datos reconsiderados”, *data mining reconsidered* donde se perfeccionan los algoritmos de minería de datos y se superaron las sombrías perspectivas que sobre este campo apuntaba Lovell (1983). En este sentido cabría citar diversos trabajos como son los de Hoover y Pérez (1999), Hendry y Krolzig

(1999) y Hansen (1999), entre otros.

### **1.1.3. El empleo de sistemas expertos e inteligencia artificial en la predicción financiera**

Los sistemas expertos han tenido, igualmente una amplia contribución en la predicción financiera. Un sistema experto suele definirse como un sistema de inferencia que cambia progresivamente por medio del uso de reglas de tipo IF - THEN para representar el conocimiento de un experto. La mayoría de los sistemas expertos tienen unos componentes básicos: base de conocimientos, motor de inferencia, base de datos e interfaz con el usuario. Muchos tienen, además, un módulo de explicación y un módulo de adquisición del conocimiento. La base de conocimientos contiene el conocimiento especializado extraído del experto en el dominio. Es decir, contiene conocimiento general sobre el dominio en el que se trabaja. El método más común para representar el conocimiento es mediante reglas de producción. El dominio de conocimiento representado se divide, pues, en pequeñas fracciones de conocimiento o reglas “SI . . . ENTONCES . . .”. Cada regla constará de una parte denominada condición y de una parte denominada acción, y tendrá la forma:

SI condición ENTONCES acción

Una característica muy importante consiste en que la base de conocimientos es independiente del mecanismo de inferencia que se utiliza para resolver los problemas. De esta forma, cuando los conocimientos almacenados se han quedado obsoletos, o cuando se dispone de nuevos conocimientos, es relativamente fácil añadir reglas nuevas, eliminar las antiguas o corregir errores en las existentes, no siendo necesario reprogramar todo el sistema experto. El motor de inferencias

es un programa que controla el proceso de razonamiento que seguirá el sistema experto. Utilizando los datos que se le suministran, recorre la base de conocimientos para alcanzar una solución.

Los primeros sistemas expertos fueron usados en medicina, ingeniería y ciencias básicas, pero a finales de los años ochenta comenzaron a ser aplicados a las finanzas, permitiendo la construcción de nuevas y más sofisticadas reglas técnicas. Según puede desprenderse de la experiencia acumulada durante largos años de investigación, la tecnología de sistemas expertos es capaz de alcanzar unos resultados más consistentes y reaccionar más rápidamente de lo que podría hacerlo un grupo humano. Los sistemas expertos se dividen generalmente en sistemas programados y sistemas de autoaprendizaje. Los sistemas programados se basan en la introducción explícita de reglas de decisión con el fin de adquirir conocimientos. Los sistemas expertos de uso más convencional hasta este momento en la inversión bursátil se sitúan dentro de esta categoría. Los sistemas de autoaprendizaje se programan ellos mismos mediante mecanismos de “entrenamiento” propios, siendo capaces de incorporar a su experiencia series históricas de datos. De esta forma, una nueva rama de la informática conocida como inteligencia artificial se abre paso, hoy día, dentro de la tecnología financiera. La inteligencia artificial es una disciplina que busca la creación de programas de ordenador inteligentes capaces de emular el razonamiento humano e imitar las funciones propias del cerebro. La tecnología que busca alcanzar esos resultados es conocida como neurocomputación o redes neuronales artificiales, o más genéricamente como inteligencia artificial. Hoy día las redes neuronales artificiales pueden ser consideradas desde el punto de vista estadístico como

modelos de regresión no paramétrica; no obstante la motivación histórica de las redes está asociada a los fenómenos biológicos, en particular en la fisiología de las células nerviosas. De este modo, las estructuras neuronales son imitadas por medio de un ordenador intentando con ello dotar a las máquinas de cualidades inéditas y consideradas hasta el momento patrimonio exclusivo del cerebro humano. Al ser capaces de usar el razonamiento inductivo, las redes neuronales pueden simular toda una serie de atributos típicos de la inteligencia humana a la hora de procesar información compleja como pueden ser la experiencia la intuición o la generalización.

Una forma común de revalidar la ejecución del proceso de autoaprendizaje de una red neuronal, antes de su implementación final, es considerar un subconjunto del total de datos disponibles que se empleará para realizar el entrenamiento, y medir la proporción de decisiones correctas realizadas sobre el resto de los datos. Este proceso de aprendizaje confiere a las redes neuronales una serie de ventajas e inconvenientes que las diferencian de los sistemas expertos programados más convencionales. Los sistemas expertos programables son ideales para simular el razonamiento lógico formal y reproducen, a la perfección, el razonamiento deductivo. Por el contrario, las redes neuronales han pretendido simular la inteligencia humana imitando la estructura y funciones del cerebro típicas del razonamiento inductivo.

En este sentido, las redes neuronales presentan diversas limitaciones como las siguientes:

- Las redes no pueden proporcionar respuestas precisas y sólo indicarán ciertos rangos.

- Las redes no pueden justificar sus respuestas y no tienen la facilidad para igualar el “cómo” y el “por qué” de los sistemas expertos programables, por lo que las redes pueden considerarse como cajas negras.
- Son muy sensibles a los datos con los que se les ha entrenado. (Véase Van Eyden, 1996, Páginas 73-77).

Pese a ello, las redes neuronales han demostrado tener gran poder y flexibilidad para detectar relaciones no lineales complejas. Su empleo en las finanzas es cada día más frecuente, llegando a incidir en temas trascendentales como pueden ser los productos derivados, cuyas fórmulas de valoración son altamente no lineales incluso cuando están disponibles en expresión analítica. De este modo, las redes neuronales permiten abordar el complejo problema del desarrollo de estrategias dinámicas de cobertura (Hutchinson *et al.*, 1994). Asimismo, las redes neuronales han sido sugeridas como herramientas valiosas para realizar predicciones sobre los rendimientos o la volatilidad de los mercados financieros, así como a la hora diseñar estrategias técnicas de contratación (Deboeck, 1994; Fernández Rodríguez *et al.*, 2001).

#### 1.1.4. Empleo de sistemas inteligentes de optimización

Además de las redes neuronales artificiales, la biología ha servido como fuente de inspiración de otros procedimientos algorítmicos de optimización inteligente que tienen un futuro prometedor en las finanzas. Se trata de los algoritmos genéticos.

Los algoritmos genéticos constituyen unos procedimientos de optimización que imitan los procesos empleados por los seres vivos en su lucha por la conservación



de la especie, haciendo uso de estrategias de optimización empleadas de modo ancestral en la naturaleza tales como son la selección natural o la transmisión de la herencia genética (Holland, 1975). El funcionamiento de un algoritmo genético es realmente simple: parte de un conjunto de candidatos a solución a los que aplica sucesivamente las operaciones de selección natural, recombinación genética y mutación, obteniendo entonces sucesivas generaciones de candidatos a solución mejor adaptados a la función objetivo que los de la generación precedente.

Los algoritmos genéticos poseen diversas características fundamentales que los convierten en instrumentos poderosos de optimización. Entre otras cabría señalar el que no se requieren propiedades especiales de la función objetivo a optimizar (tales como la continuidad o la derivabilidad de la función objetivo), o su capacidad para salir con facilidad de los óptimos locales cuando se trate de funciones objetivo muy complejas. Finalmente, hay que destacar que los algoritmos genéticos consisten en un procedimiento de optimización aleatorio, extremo éste que resulta fundamental si se piensa que una de las críticas que ha recibido, con frecuencia, el Análisis Técnico es la de estar bajo sospecha por prácticas cercanas al fisgoneo de datos. Por todo ello, los algoritmos genéticos constituyen una herramienta prometedora en el ámbito de las finanzas para la optimización de reglas técnicas.

Los algoritmos genéticos han experimentado también una poderosa evolución y adaptación que los han convertido en herramientas sumamente flexibles. En los algoritmos genéticos clásicos las estructuras genéticas son representadas por medio vectores numéricos de la misma longitud. Tal representación puede resultar inadecuada cuando el tamaño o la forma de solución del problema no puede ser

expresada de antemano. Koza (1992) desarrolló una extensión de tales algoritmos, denominada programación genética, con la intención de aliviar el problema de la longitud fija del material genético. En la programación genética los candidatos a solución son representados como una composición de funciones jerarquizadas, a modo de un determinado programa informático en un lenguaje adecuado. Como estos programas pueden representar tipos de reglas técnicas, la programación genética puede ser empleada con el fin de crear nuevas reglas técnicas con estructuras lógicas, diferentes a las empleadas de forma habitual por el Análisis Técnico.

### **1.1.5. Irrupción de los modelos no lineales en Economía y Finanzas**

La era actual de la exploración de datos también está caracterizada por el creciente uso de modelos no lineales de predicción. Como señala Saratis (2001), la Teoría Económica destaca diversas fuentes potenciales para la presencia de no linealidades y ciclos en los precios de los activos financieros. Las principales explicaciones incluyen la diversidad de creencias de los agentes (Brock y LeBaron, 1996; Brock y Hommes, 1998), heterogeneidad en los objetivos de los inversores que surgen de la existencia de horizontes de inversión y perfiles de riesgo variables (Peters, 1994; Guillaume *et al.*, 1995), comportamiento de rebaño (“herd behaviour”) (Lux, 1995), y cambios en las dotaciones presupuestarias entre situaciones de alto y bajo crecimiento económico (Cecchetti *et al.*, 1990). Por otra parte, existen numerosos estudios empíricos que han aportado evidencia empírica sobre el comportamiento no lineal de las cotizaciones bursátiles, entre los que cabría destacar los trabajos de Hsieh (1991), Brock *et al.* (1991), Peters (1994)

y Ryden *et al.* (1998), entre otros. Por todo ello, aunque la teoría financiera en sí misma no proporciona muchas motivaciones para el uso de modelos no lineales, algunos autores como Frances y Dijk (2000) consideran que los modelos lineales de series temporales no generan predicciones creíbles y, aunque esto no indica necesariamente que los modelos no lineales puedan hacerlo, los consideran más adecuados.

Durante los últimos años la actitud académica hacia el empleo de modelos no lineales en finanzas comienza a cambiar. Aunque la mayoría de los modelos empleados en las finanzas son lineales, muchos de los aspectos de su comportamiento económico pueden ser no lineales. Así, Campbell *et al.* (1997) apuntan que tanto la evidencia experimental como la introspección sugieren actitudes de los inversores hacia el riesgo y los rendimientos esperados de carácter no lineales. Asimismo, ponen de manifiesto cómo los términos de muchos contratos, tales como opciones y otros derivados, son no lineales. Finalmente, señalan que la interacción estratégica entre los participantes del mercado, el proceso por el cual se incorpora la información en el precio de las acciones y la dinámica de las grandes fluctuaciones económicas son inherentemente no lineales. Similar actitud muestran Frances y Dijk (2000) quienes consideran que la naturaleza de los datos financieros sugiere que los modelos no lineales pueden ser más apropiados para la predicción y la descripción precisa de los rendimientos y la volatilidad.

Hay que señalar, entonces, que existen numerosas razones que justifican el interés por el uso de modelos no lineales en Economía. En las finanzas empíricas, por ejemplo, es un hecho bien conocido que las series financieras presentan un

comportamiento asimétrico. Así, por ejemplo, los rendimientos negativos grandes aparecen mas frecuentemente que rendimientos positivos grandes. De este modo en el *crash* de los mercados bursátiles del 19 de Octubre de 1987 produjo un rendimiento negativo del orden del 23% del Índice Standard&Poor's 500. No obstante, para la mayoría de los mercados es extraño observar revalorizaciones en una jornada superiores al 10%. Otro curioso ejemplo muy sintomático de no linealidad es que rendimientos negativos grandes son, con frecuencia, preludio de periodos de alta volatilidad, mientras que tras grandes rendimientos positivos la volatilidad suele ser menor.

Los modelos lineales tienen la considerable ventaja de la fácil interpretación con la que cuentan sus elementos, interpretación que ha sido crucial en el desarrollo de muchas teorías. Por el contrario, los modelos no lineales suelen ser considerados con frecuencia como “cajas negras”, lo que dificulta su incorporación al mundo de la teoría económica pese a estar situados, con frecuencia, en conexión mucho más directa con el mundo empírico. A todo ello hay que unir la enorme cantidad de modelos no lineales susceptibles de ser empleados para modelizar y predecir series temporales, extremo éste que a veces ha dificultado la elección práctica e identificación del mejor modelo para una aplicación particular. En este sentido, hay que señalar que otro de los grandes avances en el terreno de la predicción se ha debido al importante desarrollo producido en el contexto de los procedimientos de selección de modelos. En efecto, los cambios acaecidos después de la era Box-Jenkins de los años setenta han sido importantes, apareciendo contrastes de especificación, abarcamiento, preocupación con la dinámica del modelo e integración entre el comportamiento a largo y corto plazo (Engle *et al.*,

1989). Sin embargo, tal como señala Dawes *et al.* (1994), por el momento existe poca evidencia de que la intensiva preocupación por tales temas, pese a constituir una mejor práctica de trabajo, haya conducido a más que una mejora marginal en la precisión de las predicciones. En el terreno de la selección de modelos no lineales los desarrollos también han sido cuantiosos y sustanciales. En este sentido, recientes avances en la estadística comienzan a propiciar la interpretación de los elementos no lineales, así como la aparición de técnicas tanto para combinar como para seleccionar modelos del máximo contenido informacional (Hastie *et al.*, 2001).

#### **1.1.6. Empleo de técnicas predictivas basadas en la regresión no paramétrica**

La gran proliferación de los modelos no lineales a la que asistimos en estos momentos está asociada a avances recientes en numerosas disciplinas cuantitativas tales como la teoría de los sistemas dinámicos, el análisis no lineal de series temporales, los modelos de volatilidad estocástica, la estadística no paramétrica y la inteligencia artificial. Todos y cada uno de estos campos ha impulsado el interés en el estudio de no linealidades en las series financieras.

Otra característica importante de muchas de las nuevas metodologías estadísticas y de regresión que han sido puestas en pie durante los últimos años, y que abren numerosas posibilidades en finanzas, es el empleo de modelos no paramétricos, concretamente modelos de regresión no paramétrica. Una curva de regresión describe una relación general entre una variable explicativa y una variable endógena. Así, el propósito del análisis de regresión es el de encontrar la dependencia media entre dos variables estadísticas ligadas por medio de una

función de respuesta desconocida, que describe la relación entre dichas variables. Esta aproximación a la función de dependencia media entre variables puede hacerse de dos formas alternativas. Por una parte, se tiene la aproximación paramétrica que supone que la curva de las dependencias medias tiene alguna forma funcional determinada, como por ejemplo una línea recta de coeficientes desconocidos (regresión lineal) o, más generalmente, una función polinómica de grado determinado con coeficientes desconocidos (regresión polinómica). Por otra parte, se encuentra la aproximación no paramétrica donde no se realiza ninguna hipótesis previa respecto a la forma funcional específica de la curva de dependencias medias, constituyendo así una herramienta flexible para analizar relaciones de regresión desconocidas. El término no paramétrico hace indudable alusión a la estadística no paramétrica, nombre con el que se designan los métodos estadísticos donde no se parte, a priori, de una determinada función de distribución. Tal dicotomía proviene de los años veinte donde se produjo una agria disputa entre Pearson y Fisher. Pearson, concentrado en los problemas de especificación del modelo, defendía que la aproximación paramétrica puede dar lugar a una mala especificación y a un gran sesgo. Por su parte, Fisher señalaba que la aproximación no paramétrica daba, en general, pobres resultados y centró su atención preferentemente en el problema de estimación de los parámetros de un modelo estadístico con una función de densidad especificada.

El objetivo de los métodos no paramétricos es el de relajar la suposición de la forma de la función de regresión, permitiendo con ello a los datos buscar una función adecuada que describa bien su dependencia media. Durante las pasadas tres décadas se han desarrollado, tras arduos esfuerzos, un gran cuerpo de doctrina

sobre el análisis de regresión no paramétrico y en particular sus aplicaciones a la predicción en series temporales. Ello ha traído consigo todo un creciente número de los trabajos centrados en la predicción de las cotizaciones de los mercados financieros por métodos no paramétricos y no lineales.

A modo de síntesis de cuanto se ha dicho, cabría afirmar que los numerosos desarrollos en el terreno de la predicción mencionados hasta el momento comienzan a hacer posible la realización de desarrollos y refinamientos del Análisis Técnico de los mercados bursátiles, que apuntan hacia la oportunidad de poner en pie sistemas casi automáticos de contratación, formación de carteras y cobertura de riesgos. Esta Tesis Doctoral tiene por objetivo profundizar en los diversos aspectos innovadores del Análisis Técnico planteándose una serie de objetivos que se enumerarán a continuación.

## 1.2. Objetivos

El Análisis Técnico clásico basado en el cruce de medias móviles, la ruptura de rangos o las figuras alcistas y bajistas tiene considerables defectos en la estructura del sistema de reglas que emplea para realizar sus predicciones. Tal sistema de análisis emplea suposiciones simplificadas con el fin de diseñar modelos de contratación menos complejos y más manejables (Deboeck, 1994). Así, muchos de los modelos de contratación tradicionales son reglas de estructura muy simple basadas en los escenarios “IF-THEN-ELSE” empleando varios tipos de medias móviles, indicadores técnicos y otros descriptores de patrones. Como señala Deboeck (1994), las reglas técnicas tradicionales son estáticas y tratan únicamente con información simbólica donde los inputs, umbrales y reglas de

decisión cambian en pasos discretos. Tales sistemas tienen dificultades para tratar con la no linealidad y por ello tales reglas deben ser frecuentemente moduladas de acuerdo con las circunstancias del mercado. En este sentido y con el propósito de profundizar y contribuir a este proceso evolutivo del Análisis Técnico, se plantean los siguientes objetivos que se pueden considerar aportaciones relevantes en el terreno de la Econometría Financiera No Lineal aplicada a los mercados bursátiles:

**Objetivo 1: Estudio de la capacidad de generar beneficios del Análisis Técnico en el Índice General de la Bolsa de Madrid (IGBM).**

Durante los últimos quince años se han realizado numerosos trabajos que se muestran optimistas sobre la predicción de los precios y los rendimientos de los rendimientos de los activos bursátiles (Granger, 1992; Saratis, 2001). No obstante este tipo de trabajos se ha centrado fundamentalmente en los grandes mercados bursátiles; por otra parte, para simplificar el problema que supone el reparto de dividendos suelen considerarse índices bursátiles de modo que ha sido ampliamente considerada la predicción de los grandes índices como el Dow-Jones, el Standard & Poor's 500, el New York Stock Exchange Equal Value o el Nikkei de la Bolsa de Tokyo. No obstante, existen factores relacionados con el "hábitat preferido de los inversores" que, pese a la creciente integración económica y financiera que vive el mundo actual, convierten a cada mercado en un mundo diferente. En este sentido son pocos los trabajos que se han realizado sobre la posibilidad de predicción y la capacidad de realizar beneficios de tales predicciones sobre índices de mercados de tamaño medio como ocurre con el Índice General de la Bolsa de Madrid. Entre los trabajos realizados hasta este



momento sobre predicción del IGBM cabría destacar Olmeda y Pérez (1995), Fernández Rodríguez *et al.* (1997) y Fernández Rodríguez *et al.* (2000). En todos ellos se estudia la capacidad predictiva de ciertas formas, más o menos sofisticadas de Análisis Técnico, pero sin considerar el problema crucial de los costes de transacción. Se plantea, por tanto, como uno de los objetivos de esta Tesis Doctoral el estudio de la capacidad de generar beneficios por parte del Análisis Técnico sobre el IGBM. Debido al carácter incipiente que poseen las metodologías sobre las que se han deducido las reglas técnicas que serán empleadas en esta Tesis, sólo se considerarán predicciones a corto plazo, concretamente a un día.

**Objetivo 2: Optimización de las reglas técnicas de contratación clásicas mediante algoritmos genéticos.**

Desde comienzos del siglo XX existe una ardua polémica sobre el poder predictivo del Análisis Técnico de los mercados financieros. Uno de los inconvenientes señalados más frecuentemente respecto al empleo de las reglas técnicas de contratación ha sido el tema de la selección objetiva de los parámetros que intervienen en tales reglas. La optimización de reglas técnicas de contratación es un problema de gran complejidad matemática por el carácter discreto de todas las variables que intervienen, pudiendo conducir a búsquedas enormemente tediosas en el espacio de posibles soluciones. Tales dificultades sugieren el empleo de algoritmos de optimización que no requieran la diferenciabilidad de la función objetivo tal como ocurre con los algoritmos genéticos, que se han mostrado de gran utilidad en numerosos problemas similares en otras áreas del conocimiento. Ocurre además que los algoritmos genéticos constituyen un procedimiento idóneo de selección de parámetros, libre de toda sospecha de fisgoneo de los datos, debido

a que se trata de un procedimiento de selección y optimización aleatorio. En esta Tesis Doctoral se plantea como uno de sus objetivos la optimización de reglas técnicas de contratación clásicas mediante algoritmos genéticos.

**Objetivo 3: Generación de nuevas reglas técnicas óptimas mediante la programación genética.**

Por otra parte una nueva generación de métodos de optimización genética conocidos genéricamente como Programación Genética, abren la posibilidad de la creación de nuevas reglas técnicas de contratación y no sólo la posibilidad de optimizar las ya existentes.

En los Algoritmos Genéticos (AGs) tradicionales las estructuras genéticas se representan por medio de tiras de caracteres de la misma longitud. Tal representación es adecuada para muchos problemas pero resulta demasiado restrictiva cuando el tamaño o la forma de la solución no puede ser expresada de antemano. Koza (Koza, 1992) desarrolló una extensión de los AGs denominada Programación Genética con la intención de aliviar el problema de la longitud fija de los AGs.

En la Programación Genética los candidatos a solución son representados como una composición de funciones jerárquicas. En estas estructuras en forma de árbol, el sucesor de cada nodo proporciona el argumento para la función identificada con el nodo. El nodo terminal corresponde al *input* observacional. El árbol en su conjunto se representa como una función, que será evaluada recursivamente por simple evaluación del nodo raíz del árbol. La estructura del candidato a solución no está especificado a *priori*. En su lugar, se define un conjunto de funciones como bloques de construcción que serán recombinados por

el algoritmo genético. El conjunto de funciones será elegido de modo adecuado para cada problema particular bajo estudio.

La Programación Genética también mantiene una población de estructuras genéticas. La población inicial está formada por un conjunto de árboles elegidos al azar y posteriormente se realizan entre los árboles las operaciones de selección, recombinación y mutación típicas de los AGs clásicos. Así pues, otro objetivo de esta Tesis Doctoral será el de investigar nuevas reglas técnicas obtenidas mediante evolución genética a partir de otras más simples, siguiendo los procedimientos iniciados por Allen y Karjalainen (1999) y basados en la Programación Genética de Koza (1992). Estas nuevas reglas parecen ser especialmente beneficiosas en mercados bursátiles periféricos o de dimensiones medianas como ocurre en el caso de Madrid.

#### **Objetivo 4: Reglas técnicas basadas en Redes Neuronales en el IGBM.**

Las reglas técnicas empleadas de forma más tradicional, tales como las de cruces de medias móviles o de ruptura de rango son ejemplos simples de lo que hoy día se conoce dentro de la informática como sistemas expertos.

Aunque los primeros sistemas expertos fueron diseñados pensando inicialmente en sus aplicaciones dentro de los campos de la medicina, la ingeniería y ciencias básicas, a finales de los años ochenta comenzaron a ser aplicados a las finanzas, permitiendo la construcción de nuevas y más sofisticadas reglas técnicas. Los sistemas programados se basan en la introducción explícita de reglas de decisión con el fin de adquirir conocimientos. Los sistemas expertos de uso más convencional hasta el momento en la inversión bursátil se sitúan dentro de esta

categoría y las reglas técnicas clásicas de cruce de medias móviles y de ruptura de rango forman un caso muy particular de esta categoría.

Los sistemas expertos basados en redes neuronales no emplean reglas explícitas. El mecanismo mediante el cual se recomiendan determinadas acciones, es más sutil. El sistema *input-output* que finalmente se establece, depende de la topología de la red, de las funciones de transferencia de nodos y de los pesos de interconexión que resultan del proceso de entrenamiento de la red.

La aplicación de las redes neuronales a las actividades como la gestión de inversiones y a las finanzas en general se encuentra aún en un estado de desarrollo relativamente temprano, de modo que mucho del trabajo que se efectúa en estos momentos se encuentra aún en fase de investigación. En cualquier caso, muchos especialistas en aplicaciones de los sistemas expertos aplicados en el área financiera consideran que los sistemas de autoaprendizaje tienen ciertas ventajas sobre los sistemas programados (Trippi y Lee, 1992).

La habilidad de las redes neuronales para clasificar datos con atributos que están altamente correlacionados las hace muy útiles para su aplicación en la valoración de riesgos, así como todo tipo de decisiones sobre clasificación de riesgos basados en datos históricos. Las redes neuronales, tras un entrenamiento adecuado, han sido empleadas en el análisis financiero para ayudar en la toma de decisiones en los siguientes terrenos:

- Investigaciones para la automatización en la concesión de créditos.
- Valoración de riesgo de hipotecas.
- Gestión de proyectos y licitación de estrategias.

- Predicciones económicas.
- Detección de irregularidades en los movimientos de los precios de las acciones.
- Identificación de factores económicos explicativos.

La posibilidad de diseñar cierto tipo de redes neuronales para el reconocimiento de irregularidades en series temporales de naturaleza altamente no lineal, posibilitan otro tipo de aplicaciones a las finanzas, todavía prácticamente casi inéditas en el IGBM. Se trata de emplear las redes para el diseño de reglas técnicas de contratación.

**Objetivo 5: Optimización de procedimientos de regresión no paramétrica mediante algoritmos genéticos en el contexto de la predicción financiera. Reglas técnicas basadas en predicciones por ocurrencias análogas.**

La “predicción por ocurrencias análogas” (*Nearest Neighbour*, NN según sus siglas en inglés), constituye una nueva técnica no paramétrica de predicción a corto plazo donde no se considera una forma funcional a *priori* que especifique la dependencia entre variables. Por el contrario, la única suposición que se realiza es que los patrones de comportamiento de una serie temporal, en su pasado, pueden tener parecido con determinados patrones de dicha serie en el futuro. Este principio puede ser aplicado con el fin de generar predicciones, de forma que, los patrones similares detectados en el pasado se emplearán para predecir el comportamiento futuro. Así, la metodología de predicción NN no intenta ajustar un modelo paramétrico global de la serie temporal a predecir, sino que únicamente

utiliza información local relativa a los patrones de comportamiento anteriores a las observaciones que van a ser predichas.

Dada una serie temporal, existen diversos parámetros que deben ser seleccionados previamente antes de realizar una predicción NN. Tales parámetros son el tiempo de retardo entre observaciones consecutivas, el número de observaciones entre los patrones y el número de patrones que se emplean en la predicción. Por otra parte, si queremos transformar las predicciones sobre una determinada serie bursátil en una regla técnica de contratación resulta esencial establecer algún tipo de filtro que deseche señales engañosas y evite que la estrategia técnica sea demasiado activa. De esta forma, para establecer una regla técnica es preciso optimizar una función de pérdida (por ejemplo el exceso de rendimientos sobre la estrategia de comprar y mantener) que depende de un conjunto de parámetros algunos de los cuales son de carácter discreto. Quizás el único procedimiento práctico factible para realizar este tipo de optimización sea por medio de un algoritmo genético.

Como otro objetivo de esta Tesis Doctoral se plantea, por tanto, el estudio de la capacidad de generar beneficios por medio de reglas técnicas óptimas basadas en predicciones NN por medio de algoritmos genéticos.



## Capítulo 2

# El objeto del análisis: el problema de la predicción financiera

### Índice General

---

<b>2.1. Introducción</b>	<b>31</b>
<b>2.2. Análisis Fundamental</b>	<b>33</b>
<b>2.3. Análisis Técnico</b>	<b>36</b>
2.3.1. Chartismo	39
2.3.2. Reglas basadas en medias móviles	46
2.3.3. Reglas basadas en órdenes estadísticos	50
<b>2.4. La hipótesis de los mercados eficientes</b>	<b>54</b>
2.4.1. Anomalías e imperfecciones empíricas	60
<b>2.5. Renacimiento académico del análisis técnico</b>	<b>62</b>

---



En los últimos años los mercados de valores mundiales han registrado un firme proceso de crecimiento sostenido. Por ejemplo, en el año 2000 el volumen anual de contratación en el NYSE superó los 11 mil millones de dólares y la capitalización en este mercado alcanzó los 17 billones de dólares. El crecimiento económico, logrado, en parte, gracias a una mayor liberalización del flujo de capitales y la aparición de nuevos productos financieros, como, por ejemplo, los fondos de inversión, ha contribuido a este incremento de la actividad en estos mercados. Los sistemas técnicos, operativos y de organización, en los que se apoya hoy el mercado han permitido que se hayan canalizado importantes volúmenes de inversión y que se haya dotado a los mercados de una mayor transparencia, liquidez y eficacia.

De esta forma, un mayor número de empresas han acudido a estos mercados, que se ha visto correspondido con una mayor presencia de inversores, que, por ejemplo, en el año 2000 en la Bolsa de Madrid, superaron los 390 millones de euros en volumen de contratación. El mismo proceso de globalización no es ajeno a las mismas Bolsas, muchas de las cuales están inmersas en un proceso de fusiones que hará que el nivel de actividad alcance proporciones significativas.

Como hemos visto anteriormente, avances recientes en las técnicas analíticas y computacionales en el tratamiento de series temporales han hecho posible nuevas formas de predicción financiera. Todas estas herramientas van a configurar un nuevo panorama en la predicción, modelización del riesgo y formación de carteras eficientes dentro de los mercados financieros.

## 2.1. Introducción

La posibilidad de predecir los cambios de precios de los activos financieros ha constituido un tema recurrente en la literatura financiera. Cualquier intento de predicción en el terreno de las finanzas constituye un reto importante, por cuanto que el cambio en el precio de cualquier activo es consecuencia de numerosas variables que hacen difícil y compleja la elaboración de un método de predicción basado en la información que se considere relevante. Debido al creciente peso que desempeñan las finanzas en el nuevo orden económico internacional, el interés en cualquier grado de predicción de los mercados financieros no se halla exclusivamente restringido al ámbito del mundo académico o a los profesionales del mercado, .

Por otra parte, las tremendas dificultades que entraña la predicción financiera, unidas en muchos casos a dudosas prácticas por parte de determinados “gurús”, han conducido a que la predicción de los mercados financieros sea una disciplina científica que sólo se ha reconocido, por parte del mundo académico, en épocas muy recientes. En efecto, aún cuando numerosos profesionales del mundo de las finanzas han venido empleando, desde comienzos del siglo pasado, las más variadas técnicas de predicción, toda su actividad siempre fue realizada de espaldas al mundo académico. Ha sido únicamente durante las últimas dos décadas cuando, en la literatura académica, comenzaron a aparecer algunos trabajos señalando que, en alguna medida, muchos de los mercados financieros son susceptibles de alguna forma de predicción.

Existen diversos factores que han favorecido, de forma decisiva, los recientes éxitos producidos en el ámbito de la predicción financiera, destacamos como

fundamentales los siguientes:

- Desarrollos, por parte del mundo académico, de nuevas herramientas matemáticas, estadísticas e informáticas capaces de analizar el mundo no lineal.
- Existencia de importantes bancos de datos de alta frecuencia que no se poseían en el pasado.
- Generalización del empleo de poderosos ordenadores cuya disponibilidad ha posibilitado la puesta a punto de las nuevas técnicas de análisis.

Hace treinta años, cualquier alusión relativa a la posibilidad de obtener alguna rentabilidad prediciendo el precio de activos financieros habría merecido la condena unánime del mundo académico por considerarla contraria al paradigma científico dominante de la eficiencia de los mercados. Durante la última década la mentalidad sobre el paradigma de la eficiencia ha cambiado considerablemente. En la actualidad comienza a darse un amplio consenso, entre autores de gran reputación en las finanzas, de que un cierto grado de habilidad predictiva en un mercado puede estar relacionado con un riesgo dinámico inherente al uso de la estrategia técnica que se emplee que no es incompatible, necesariamente, con la eficiencia de dicho mercado (Campbell *et al.*, 1997). Hoy comenzamos a adentrarnos en los sutiles matices que encierra la Hipótesis de los Mercados Eficientes tales como el análisis de los costes de transacción, o el riesgo que siempre entraña cualquier inversión. Tal como señala Granger (1992), la simple habilidad predictiva, en cierto grado, significa ineficiencia, a menos que esté acompañada de unos beneficios económicos después de haber descontado todo tipo de costes

y haber incluido alguna suerte de prima de riesgo por la inversión.

Históricamente han destacado tres escuelas que han abordado de forma completamente antagónica el problema de la predicción de las series financieras: análisis fundamental, análisis técnico y la hipótesis de los mercados eficientes.

## **2.2. Análisis Fundamental**

El Análisis Fundamental consiste en un procedimiento que valora cada activo a partir de las variables económicas fundamentales y la consiguiente capacidad de la empresa para generar beneficios en el futuro. El análisis fundamental considera que los cambios que se producen en el precio de un activo se deben esencialmente a factores económicos relativos, tanto a la empresa como al ambiente económico en general. Desde este punto de vista, el analista financiero decide qué posiciones de compra y venta tomar de acuerdo con información fundamental estrechamente ligada a los estudios financieros: informes de auditores cuentas de resultados, balances trimestrales, historial de dividendos, estrategia empresarial, etc. Con toda esta información interna analiza los datos de ventas, el talento directivo, la competencia, los índices de producción, las estadísticas de las cotizaciones y las previsiones de rendimiento, así como las noticias diarias sobre la evolución económica general para establecer una estimación de las futuras condiciones de la actividad de la empresa cuyas acciones son de interés.

El enfoque fundamental examina los factores más relevantes que afectan la corriente futura de beneficios de una empresa con el fin de determinar su valor intrínseco, que no es otra cosa que el valor que los fundamentos económicos le atribuyen según la ley de la oferta y la demanda. Dicho en lenguaje más técnico,

el valor intrínseco de una empresa es el valor actual descontado de la corriente de beneficios que se prevé que la empresa obtendrá en el futuro (Suárez, 1996).

El valor intrínseco se calcula considerando tanto los factores relativos a los fundamentos económicos (tales como la tasa de inflación, el nivel general de los tipos de interés, los tipos de cambio, el producto interior bruto, etc.), como considerando factores que atañen exclusivamente a la actividad de la propia empresa. Tales factores internos van desde la cuenta de resultados de la empresa a las sanciones a que pueda estar sometida por la Administración, el malestar laboral y, en general, cualquier otro factor que pueda influir sobre el mercado a la hora de determinar la oferta y la demanda de las acciones de la empresa.

La cuestión básica que pretende dilucidar el Análisis Fundamental es la capacidad de la empresa para generar beneficios en el futuro. Para ello focaliza su atención en el conjunto de fuerzas económicas que determinarán su corriente futura de beneficios, es decir, su valor intrínseco. Si el precio de dicha acción se encuentra por debajo del valor intrínseco entonces debe comprarse, y *viceversa*.

La aproximación fundamental realiza un exhaustivo análisis del ambiente económico, del sector y de la empresa con el fin de determinar el valor intrínseco de un activo. En general se consideran los siguientes factores:

- El ambiente económico general que se valora por medio de indicadores cíclicos y por los tipos de interés. El indicador cíclico (CIAM) suele ser empleado para predecir el estado general de la economía utilizando el principio de que una economía experimenta períodos observables de expansión y contracción, generalmente relacionados con el ciclo económico. El comportamiento de la economía se aproxima intentando determinar cuál

es su posición dentro del ciclo económico.

- Los cambios en los tipos de interés afectan directamente tanto a la economía como al mercado de capitales. Un aumento de los tipos de interés produce, por una parte, un efecto contractivo sobre la actividad económica haciendo menos rentables las diversas alternativas de inversión. Por otro parte, incrementa para un inversor el coste de oportunidad de mantener sus ahorros en activos de renta variable.
  
- El análisis sectorial, que se refiere a la evaluación de otras empresas que operan en el mismo sector industrial, con el fin de determinar la competitividad de una empresa y clasificar la etapa del ciclo de vida industrial en el que se encuentra.
  
- El análisis de la empresa, que está destinado a determinar el valor intrínseco de sus acciones en relación a los anteriores análisis que influyen en la generación de sus beneficios, identificando los factores individuales, únicos en cada empresa, tales como la tasa esperada de rendimientos, las expectativas de reparto de dividendos y el grado de riesgo.

Sobre la base de los análisis explicados, la estrategia de transacción del inversor puede sintetizarse por medio de las siguientes variables:

- La tasa de crecimiento esperada del rendimiento: el inversor debe estar dispuesto a pagar un precio más elevado por una acción cuanto más tiempo se espere que dure la tasa de crecimiento.

- Las expectativas de reparto de dividendos: el inversor racional debe estar dispuesto a pagar un precio más elevado por una acción cuanto más alta sea la tasa de crecimiento de los dividendos.
- El grado de riesgo: el inversor racional (y adverso al riesgo) debe estar dispuesto a pagar un precio más elevado por una acción, a igualdad de condiciones, cuanto menos arriesgadas sean las acciones de una empresa.
- El nivel de los tipos de interés del mercado: el inversor racional debe estar dispuesto a pagar un precio más elevado por una acción, a igualdad de condiciones, cuanto menor sea el tipo de interés.

Los principios citados son defendidos por los analistas de inversiones como vía para realizar las estimaciones que conduzcan a encontrar el anhelado valor intrínseco del precio de una acción, confiados en la existencia de una cierta racionalidad implícita en los mercados de valores. Una gestión de inversión segura consistiría sencillamente, en comprar cuando la cotización de un activo estuviera momentáneamente por debajo de su valor intrínseco, y vender cuando fuese más elevada que dicho valor intrínseco (Malkiel, 1992).

## 2.3. Análisis Técnico

El objetivo de todo inversor es obtener posiciones recomendadas de compra y venta. Muchos profesionales del mercado suelen utilizar varias técnicas de predicción basadas en datos históricos que le ayuden en su análisis y conocidas genéricamente como análisis técnico. Si atendiéramos a la HME la utilización de estas técnicas no conllevaría la obtención de beneficios. Sin embargo, el análisis

técnico está ampliamente extendido como método de predicción a corto plazo dentro de los profesionales del sector.

De modo genérico, el análisis técnico es el estudio de la información pasada del comportamiento del mercado con la finalidad de predecir tendencias futuras, independientemente del estudio de los activos que se negocian en ese mercado. El termino técnico, aplicado al mercado de valores, se refiere al estudio de la actividad del mercado en sí misma y, de forma independiente, al estudio de los activos de que se ocupa. El analista técnico se preocupa sólomente de la evolución de los precios y el volumen de negociación. Y esto se debe a que el análisis técnico se basa en tres premisas fundamentales (Neely *et al.*, 1997):

- El mercado lo descuenta todo, es decir, toda la información, sea fundamental, política, psicológica, etc., disponible hasta ese momento, está reflejada en el precio, que no hace mas que seguir el comportamiento de la oferta y la demanda. El analista técnico supone que si los precios suben es porque la cantidad demandada es mayor que la cantidad ofertada y *viceversa*, sin importarle las causas de esta conducta.
- El segundo principio establece que los precios se mueven, en el largo plazo, por tendencias de duraciones relativamente largas y delimitadas en el tiempo. El objetivo del analista es identificar tendencias que están en sus primeras etapas de desarrollo para tomar decisiones de inversión que sigan la misma línea.
- La tercera premisa establece que los patrones de comportamiento pasados se volverán a repetir en un futuro. Por lo que el mercado tiene memoria y los precios no se comportan como un camino aleatorio.



La idea básica es que los precios se mueven siguiendo tendencias que están determinadas por actitudes cambiantes de los inversores hacia una variedad de hechos económicos, monetarios políticos y psicológicos. (Pring, 1991)

Las herramientas más utilizadas dentro del análisis técnico son el chartismo y las reglas técnicas. Mientras las primeras se basan en los gráficos de las series de precios y volúmenes, las segundas intentan eliminar el componente subjetivo de la interpretación de los gráficos a través de funciones matemáticas que indicarán al inversor posiciones de compra o venta, en función de los precios pasados. Dentro de estas reglas técnicas podemos destacar las basadas en órdenes estadísticos y las basadas en medias móviles.

El Análisis Técnico está ampliamente extendido en los mercados financieros reales y ha constituido una frecuentísima maquinaria predictiva para los inversores financieros durante los últimos cien años (Plummer, 1989).

Taylor y Allen (1992) hicieron, a petición de la División Económica del Banco de Inglaterra, un cuestionario sobre el uso del Análisis técnico por parte de los agentes que operaban en el mercado de divisas de Londres. Como resultados más llamativos de la encuesta, los autores destacaron los siguientes:

- El 90% de los encuestados declararon conceder algún tipo de peso a los consejos técnicos cuando formaban sus expectativas sobre el tipo de cambio sobre horizontes de predicción de una semana.
- Existió un claro consenso entre los encuestados de que el análisis gráfico es usado principalmente como una guía del comportamiento de las divisas a corto plazo y de que las técnicas chartistas deberían ser usadas junto con recomendaciones fundamentales.

- Por último, la confianza de los agentes en el análisis fundamental aumenta cuando el horizonte de predicción aumenta.

Taylor y Allen (1992) concluyen que la pobreza predictiva de los modelos empíricos de tipos de cambio y su alta volatilidad pueden ser, en gran parte, explicadas atendiendo a dos hechos. Por una parte, la interacción que se produce en los mercados cambiarios entre fuerzas fundamentales y no fundamentales. Por otra parte, si la estructura de los mercados financieros es inherentemente no lineal, entonces los fundamentalistas pueden perder credibilidad en el mercado cuando se producen grandes movimientos de precios y sus modelos de predicción lineal dejan de proporcionar una buena aproximación local al comportamiento de los precios. Mientras tanto, el Análisis Técnico, al trabajar de forma muy cercana al día a día del mercado, es capaz de obtener, de forma intuitiva, una aproximación local más cercana a la estructura económica subyacente, ganando así popularidad entre los agentes carentes de un entendimiento profundo de las fuerzas del mercado. Por último, los autores proponen, como reto importante en la investigación de la economía financiera internacional, la obtención de modelos que puedan explicar los movimientos pasados de los tipos de cambio, predecir los movimientos futuros con algún grado de precisión y explicar la popularidad del Análisis Técnico.

### **2.3.1. Chartismo**

El chartismo es un sistema de análisis y predicción bursátil que constituye uno de los procedimientos más clásicos del análisis técnico. Intenta deducir la posible tendencia futura de los precios a través de un gráfico bursátil donde se ha dibujado la curva de cotizaciones históricas. Centra toda su atención en el

nivel de precios y, en menor medida, en el volumen de negociación, que vendría a reforzar el análisis realizado con la serie de precios. El gráfico más utilizado para el análisis chartista es el gráfico de barras diario. El eje horizontal muestra los días o sesiones de cotización. En el eje vertical se dibujan los precios máximos y mínimos diarios unidos por una línea vertical, y cruzando ésta por una raya horizontal que muestra el precio de cierre.

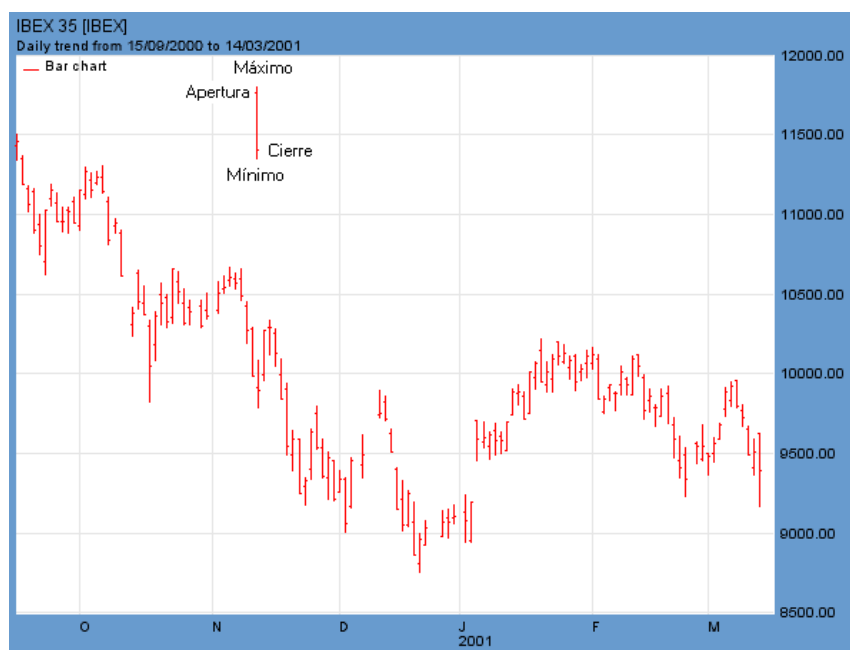


Figura 2.1: Gráficos de barras diario del índice IBEX-35 en el periodo comprendido entre 15/09/2000 y 14/03/2001

Los chartistas fundamentan sus decisiones en función del estudio de las figuras que se dibujan en estos gráficos. Estas figuras se hallan minuciosamente estudiadas y codificadas, indicando, con un cierto grado de cumplimiento, estrategias futuras de inversión.

El origen del chartismo se puede situar a principios del siglo pasado, pero

tuvo su verdadero asentamiento en los años treinta, después de la crisis bursátil de 1929, ya que generó intensas reflexiones sobre la utilización de nuevas técnicas de análisis en bolsa. Se difundió en Europa en los años cincuenta para asentarse en España en la década de los sesenta.

Observando un gráfico (*chart*), los chartistas aseguran que las cotizaciones se mueven siguiendo tendencias: que durante cierto tiempo, los precios tienden a seguir una trayectoria ascendente (tendencia alcista), descendente (trayectoria bajista) o lateral (tendencia estable).

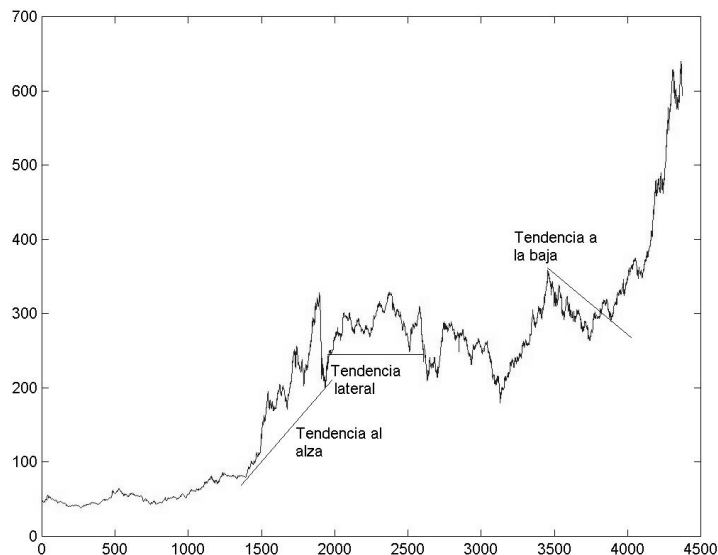


Figura 2.2: Gráfico de línea diario del IBEX35 donde se identifican los tres tipos de tendencia

La teoría de Dow identifica tres tipos de tendencias. La tendencia primaria, o principal, cubre oscilaciones entre nueve y dieciocho meses. La secundaria, o intermedia, abarca entre tres y doce semanas y la terciaria se considera a cualquier periodo inferior a las tres semanas. La Figura 2.2 muestra los tres tipos de tendencia en el IBEX35 de la Bolsa de Madrid. Cada grado de tendencia

es una corrección de su tendencia inmediatamente más larga. En una tendencia primaria ascendente de larga duración el mercado se autocorregirá durante un par de meses para continuar después con su línea de crecimiento.



Figura 2.3: Tres grados de tendencia

Una tendencia principal alcista se produce cuando, inicialmente, los inversores más informados han ido acumulando títulos. Posteriormente, cuando comienzan a incorporarse nuevos inversores, gracias a los informes de los buenos resultados empresariales, se produce un alza importante en los mercados y es aquí cuando entra en masa los pequeños inversores y se produce el *boom* bursátil, momento en el que los analistas más informados empiezan a vender sus títulos.

Dentro de una tendencia los precios suben y bajan, dejando entrever una serie de picos y valles. La Figura 2.3 muestra diversas correcciones de una tendencia alcista en un gráfico bursátil. Es la unión de esos picos y valles lo que determina la dirección de la tendencia. Cuando los picos y valles son superiores (inferiores)

en nivel a los inmediatamente anteriores se define una tendencia alcista (bajista). Una serie de picos y valles al mismo nivel identificaría una tendencia lateral, aunque los analistas llaman a este periodo “ausente de tendencia”. Se denomina directriz de soporte a la línea que enlaza todas las cotizaciones mínimas, actuando como barreras o límites de nuevas bajadas. Se denomina directriz de resistencia a la línea que enlaza todas las cotizaciones máximas, actuando como barreras o límites de resistencia frente a nuevas subidas.

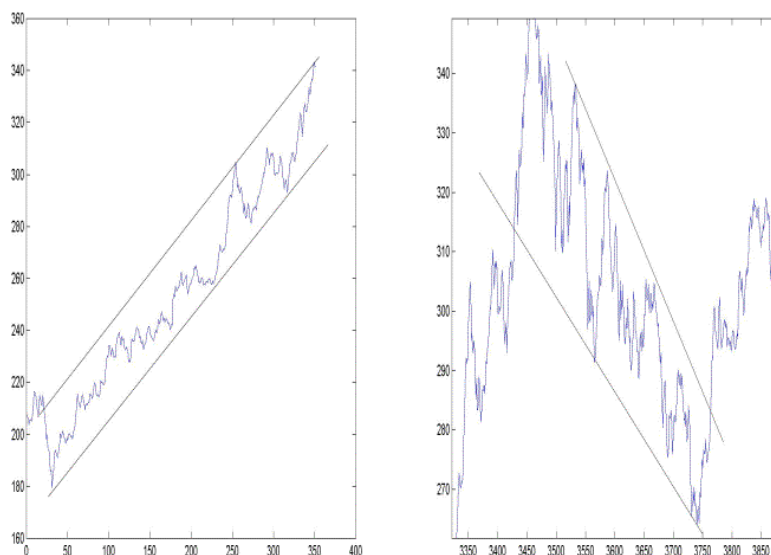


Figura 2.4: Tendencias al alza y a la baja situadas dentro de un túnel.

La fortaleza de una directriz alcista (bajista) es mayor cuanto mayor ha sido el número de veces que las cotizaciones tocan esta directriz sin lograr perforarla y también cuanto mayor es el tiempo que permanece vigente. En cambio, cuando logra perforarla mayor es el potencial de caída (subida). En general, a largo plazo las cotizaciones suelen situarse dentro del túnel formado por estas dos directrices, tal como muestra la Figura 2.4.

Los chartistas siempre operan en dirección de la tendencia. Sin embargo, tienen reconocidos patrones de precios de los valores que aparecen en los gráficos y tienen valor de predicción. Estos patrones pueden ser de cambio de tendencia o de confirmación de la tendencia. Entre los primeros podemos encontrarnos una de las figuras más típicas y fiables dentro del chartismo: “hombro cabeza hombro” [Figura 2.5].

Esta figura comienza con un pico (primer hombro), seguido de un pico más alto (cabeza), finalizando con un último pico de altura similar al primero (segundo hombro). La línea recta, generalmente horizontal, que une el punto más bajo del descenso del primer hombro con el punto más bajo de la caída tras la formación del pico de la cabeza se denomina línea de cuello (*neckline*). La figura de cabeza hombros queda completada cuando la caída del segundo hombro traspasa esta línea. Una línea de cuello descendente señala la debilidad técnica de un título. El volumen es muy importante en esta figura, ya que da una idea de la fuerza del mercado en el momento de la ruptura.

Generalmente, el volumen sigue la siguiente secuencia:

1. En el primer hombro el volumen de contratación crece con respecto a las sesiones anteriores.
2. Después del primer pico se produce un descenso del volumen hasta completar el primer hombro.
3. Al incremento de las cotizaciones en el momento de la formación de la cabeza le acompaña un incremento del volumen de negociación, aunque en menor proporción que en el primer hombro. Es decir, los nuevos máximos se consiguen con un volumen de negocio inferior respecto al impulso anterior.

4. En la fase de descenso, que terminaría de formar la cabeza, el volumen se contrae.
5. En el segundo hombro el volumen es débil. Sin embargo, si la ruptura se produce con un volumen ligeramente superior al de sesiones anteriores, la posibilidad de éxito de la figura es bastante elevado.
6. Posteriormente los precios suelen descender de forma acelerada.

Para determinar el objetivo de la caída, se traza una línea vertical desde el pico de la cabeza hasta la línea de cuello. Esa sería la magnitud de la caída desde el punto de la ruptura. Aunque, como es habitual en las figuras chartistas, se puede dar una reacción técnica (*pull back*) que hará que las cotizaciones se acerquen e incluso toquen el nivel de ruptura, pero finalmente la curva de cotizaciones descenderá hasta alcanzar su objetivo de caída. La reacción técnica, se produce en la confirmación de casi todas las figuras chartistas pero que no rompe de la directriz principal del movimiento.

Para un cambio de tendencia al alza el patrón sería exactamente el inverso al explicado anteriormente.

Existen otras figuras, como la V invertida, la W invertida, el doble y triple techo, que, al igual que la figura hombro-cabeza-hombro, indican cambios de tendencia [Figura 2.6].

Como contraposición a las figuras de cambio de tendencia, existen las figuras de continuidad. Estas formaciones técnicas nos indican una pausa en la tendencia de orden superior, de modo que al término de estas figuras se retomará la dirección de la tendencia inmediatamente superior. Podemos señalar como figuras técnicas





Figura 2.5: Figura hombro-cabeza-hombro.

de continuidad de tendencia los triángulos, las banderas y los rectángulos [Figura 2.7].

### 2.3.2. Reglas basadas en medias móviles

Uno de los métodos más extendidos, dentro del Análisis Técnico, es el de las reglas basadas en medias móviles, que resultan fácilmente programables para que generen señales de compra y de venta. Las medias móviles se utilizan, principalmente, para identificar tendencias. La media móvil suaviza la serie de precios, eliminando de ella las fluctuaciones que se producen a corto plazo, con el fin de reconocer la tendencia subyacente.

Sea la serie temporal de precios  $\{P_t\}_{t=1\dots N}$ , la media móvil de longitud  $\theta$  en el instante  $t$  se define como

$$M_t(\theta) = \frac{1}{\theta} \sum_{i=0}^{\theta-1} P_{t-i}, \quad t = \theta, \theta + 1, \dots, N \quad (2.3.1)$$

De esta forma podemos generar señales de compra y venta estableciendo una

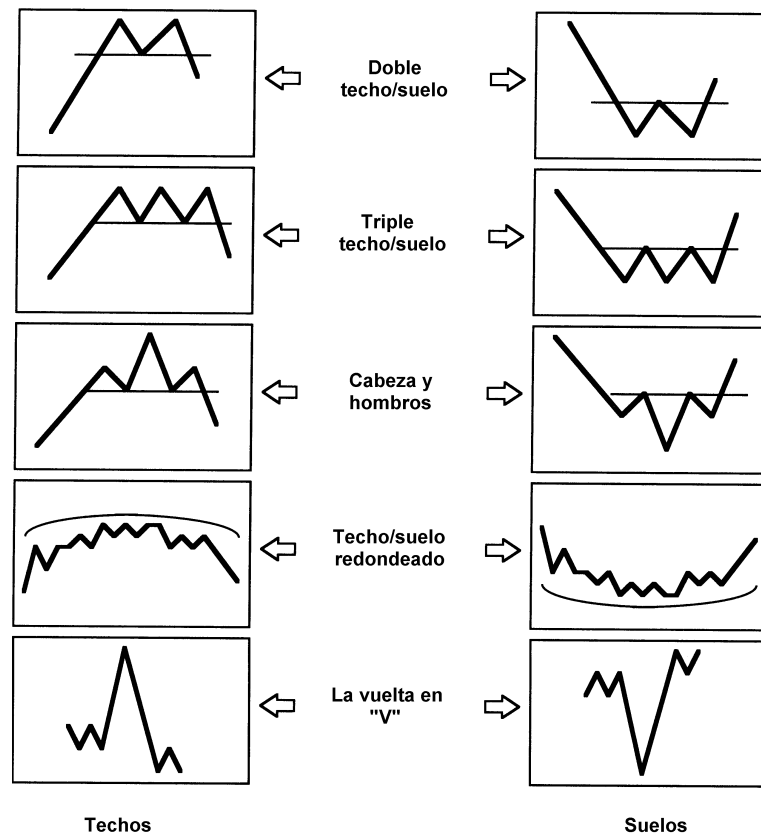


Figura 2.6: Figuras chartistas de vuelta.

regla técnica por la diferencia de dos medias móviles de distinta longitud  $\theta$ . Si la media móvil de longitud más corta (media móvil a corto) es superior a la media móvil de longitud más larga (media móvil a largo), se genera una señal de compra, mientras que, si ocurre lo contrario, se genera una señal de venta. Sin embargo, se pueden generar señales engañosas. Por este motivo se suele añadir un filtro para eliminar dichas señales, por lo que se genera una señal de compra si la media móvil a corto supera en un  $x\%$  la media móvil a largo, mientras que si cae por debajo de un  $x\%$  se genera una señal de venta. En la Figura 2.8 se muestra el gráfico de barras diario de las cotizaciones de ACS (Actividades de construcción

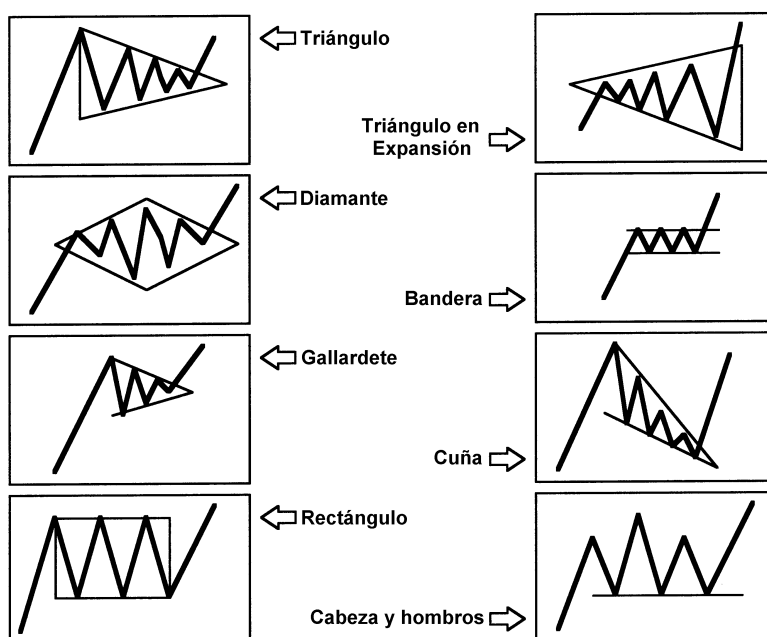


Figura 2.7: Figuras chartistas de continuidad.

y servicios) junto con las medias móviles de 20 y 50 días.

Esta regla técnica generalizada se puede representar mediante una función binaria, donde una salida igual a 1 indicaría una señal de compra, mientras que una salida igual a 0 reflejaría una señal de venta.

La expresión de la función generadora de señales es:

$$D(\Theta)_t = M(\theta_1) - (1 + (1 - 2S_{t-1})\theta_3)M(\theta_2)_t, \quad (2.3.2)$$

donde  $\Theta = [\theta_1, \theta_2, \theta_3]$  representa el vector de parámetros utilizados en la regla. El parámetro  $\theta_1$  es la longitud de la media móvil a corto, mientras que  $\theta_2$  es la longitud de la media móvil a largo. El parámetro  $\theta_3$  representa el porcentaje que se emplea para eliminar las señales erróneas.

Las diferencias  $D(\Theta)_t$  generan una función binaria que toma valores según el

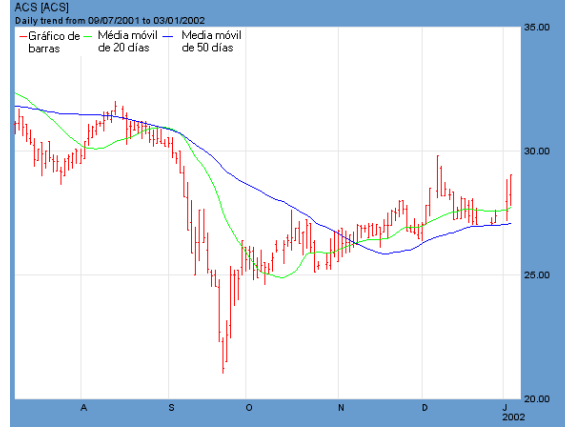


Figura 2.8: Gráfico de barras diario de ACS (Actividades de construcción y servicios) junto a las medias móviles de 20 y 50 días

signo de  $D(\Theta)_t$ .

$$S_t = \begin{cases} 1 & D(\Theta)_t \geq 0 \\ 0 & D(\Theta)_t < 0 \end{cases} . \quad (2.3.3)$$

Si  $S_t = 1$  se produce una señal de compra. Por el contrario,  $S_t = 0$  se genera una señal de venta.

A partir de esta regla generalizada de media móvil, podemos crear las reglas ampliamente utilizadas en la literatura, imponiendo restricciones a los distintos parámetros.

- Media móvil simple:  $\theta_1 = 1, \theta_2 > 1, \theta_3 = 0$

$$D(\Theta)_t = P_t - M(\theta_2) \quad (2.3.4)$$

- Media móvil filtrada:  $\theta_1 = 1, \theta_2 > 1, \theta_3 > 0$

$$D(\Theta)_t = P_t - (1 + (1 - 2S_{t-1})\theta_3)M(\theta_2)_t \quad (2.3.5)$$

- Doble media móvil:  $\theta_1 > 1, \theta_2 > 1, \theta_3 = 0$

$$D(\Theta) = M(\theta_1)_t - M(\theta_2)_t \quad (2.3.6)$$

Si se eligen dos medias móviles a muy corto plazo (de 5 o 10 días) se producen muchos cortes entre ellas. Esto provoca una estrategia muy activa que eleva tremendamente los costes de transacción. En tal caso, puede ocurrir también que se generen muchas señales falsas debido al movimiento aleatorio del precio a corto plazo. Sin embargo, tiene la ventaja de que genera señales tempranas. Si se eligen dos medias móviles a largo plazo (de 150 o 200 días) provocarían la aparición de señales tardías en el tiempo.

Así pues, existe un dilema entre el momento en que se generan las señales y los costes de transacción. Si se eligen medias móviles cortas se obtiene señales tempranas, pero por el contrario, se generan altos costes de transacción. Sin embargo, si se eligen medias móviles largas los costes de transacción serán bajos pero se generan señales tardías, lo que podría provocar pérdidas de rendimiento ya que no se ha seguido la tendencia desde su inicio.

La elección de los parámetros se viene haciendo de forma arbitraria. Sin embargo, trabajos recientes (Neely *et al.*, 1997; Allen y Karjalainen, 1999) han utilizado técnicas de optimización, como las proporcionadas por los algoritmos genéticos, para evitar el problema del fisgoneo de datos que provocaba la elección arbitraria de los parámetros de las reglas.

### 2.3.3. Reglas basadas en órdenes estadísticos

Las reglas técnicas basadas en órdenes estadísticos utilizan los precios máximos y mínimos durante un periodo temporal. Así podemos definir, dada una serie histórica de precios de  $\phi$  observaciones, el precio máximo y el precio mínimo en el instante  $t$  de la siguiente manera:

$$P_t^{\max}(\phi) = \text{Max}[P_{t-1}, \dots, P_{t-\phi}] \quad (2.3.7)$$

$$P_t^{\min}(\phi) = \text{Min}[P_{t-1}, \dots, P_{t-\phi}] \quad (2.3.8)$$

Las reglas técnicas que están basadas en índices estadísticos son las reglas de túnel (*channel rule*) y las reglas de filtro (*filter rule*).

Las ecuaciones [2.3.7] y [2.3.8] representan las expresiones de resistencia y soporte, respectivamente, que son conceptos empleados ampliamente en el análisis bursátil. La resistencia es el precio máximo durante un periodo de  $\phi$  observaciones, mientras que como soporte usaremos el precio mínimo.

La regla de túnel está configurada para que se produzca una señal de compra cuando la serie de precios rompe la resistencia de abajo hacia arriba, indicando que la demanda es tal que los precios suben por encima de la anterior barrera, y una señal de venta cuando rompe el soporte de arriba hacia abajo. Indica que la cantidad ofrecida de títulos es superior a la cantidad demandada, por lo que los precios terminan rompiendo esa barrera.

La regla de filtro fue introducida en la literatura financiera por Alexander (1961) y se basa en la idea de que los precios seguirán subiendo si sobrepasan un cierto nivel o, por el contrario, seguirán bajando si caen por debajo de un cierto nivel. De manera específica, la regla produce una señal de compra cuando los precios suben un  $x\%$  por encima del precio mínimo habido durante los  $\phi$  días anteriores. Asimismo genera una señal de venta cuando los precios bajan por debajo de un  $x\%$  de la cotización máxima marcada durante los  $\phi$  días anteriores. En ambos casos,  $\phi$  representa el número de observaciones previas dentro de las cuales se selecciona el máximo o mínimo de referencia. Cuando no se producen ni

señales de compra ni de venta el inversor estará dentro del mercado si la última transacción fue comprar o bien estará fuera del mercado si la última transacción fue vender.

En la regla filtro creada por Alexander (1961), el parámetro  $\phi$  era igual al número de días habidos desde la última transacción realizada. De esta forma, es el propio mercado quien de forma implícita determina  $\phi$ , por lo que no tiene por qué ser constante. Sin embargo, este parámetro puede ser elegido de forma explícita por el analista seleccionando un valor fijo para el número de observaciones. Estas dos reglas, de filtro y túnel, se pueden representar mediante una única regla de decisión. Utilizaremos la siguiente función como generador de señales:

$$D(\Phi)_t = P_t - (1 + (1 - 2S_{t-1})\phi_2)(P_t^{\max}(\phi_1))^a(P_t^{\min}(\phi_1))^b \quad (2.3.9)$$

donde

$$a = \phi_3 S_{t-1} + (1 - \phi_3)(1 - S_{t-1})$$

$$b = \phi_3(1 - S_{t-1}) + (1 - \phi_3)S_{t-1}$$

$$\forall \phi_1, \phi_2 \in \{1, 2, 3, \dots\}$$

Las diferencias  $D(\Phi)_t$  generan una función binaria  $S_t$  que toma valores según el signo de  $D(\Phi)_t$ . Así si  $D(\Phi)_t \geq 0$  implica que  $S_t = 1$ , produciéndose con esto una señal de compra. Si, por el contrario,  $D(\Phi)_t < 0$  implica que  $S_t = 0$  generándose una señal de venta. Así pues,  $\phi_1$  representa el número de días que se utilizan para decidir los precios máximos y mínimos a utilizar en la regla;  $\phi_2$  representa el porcentaje a partir del cual deberíamos tomar las decisiones de compra o venta;  $\phi_3$  es una variable binaria que toma valor 1 si estamos usando la regla de filtro y

0 si estamos usando la regla de túnel. Imponiendo restricciones a los parámetros podremos trabajar con las distintas reglas basadas en órdenes estadísticos:

- Regla filtro:  $\phi_3 = 1$ .

$$D(\Phi)_t = P_t - (1 + (1 - 2S_{t-1})\phi_2)(P_t^{\max}(\phi_1))^{S_{t-1}}(P_t^{\min}(\phi_1))^{(1-S_{t-1})} \quad (2.3.10)$$

- Si estamos en el mercado no saldremos de él (no se producirá una señal de venta) hasta que el precio caiga por debajo de  $\phi_2$  % de la resistencia.

$$S_{t-1} = 1 \quad D_t = P_t - (1 - \phi_2)P_t^{\max}$$

- Si estamos fuera del mercado no entraremos en él (no se producirá una señal de compra) hasta que el precio suba por encima de  $\phi_2$  % del soporte.

$$S_{t-1} = 0 \quad D_t = P_t - (1 + \phi_2)P_t^{\min}$$

- Regla túnel:  $\phi_2 = 0; \phi_3 = 0$ .

$$D(\Phi)_t = P_t - (P_t^{\max}(\phi_1))^{(1-S_{t-1})}(P_t^{\min}(\phi_1))^{S_{t-1}} \quad (2.3.11)$$

- Si estamos en el mercado no saldremos de él (no se producirá una señal de venta) hasta que el precio caiga por debajo del soporte.

$$S_{t-1} = 1 \quad D_t = P_t - P_t^{\min}$$

- Si estamos fuera del mercado no entraremos en él (no se producirá una señal de compra) hasta que el precio suba por encima de la resistencia.

$$S_{t-1} = 0 \quad D_t = P_t - P_t^{\max}$$



De esta forma, se crea un túnel entre soportes y resistencias entre el cual el inversor no toma decisiones y sigue estando en posiciones que habían sido tomadas previamente. Sólo cuando se traspasa ese túnel es cuando el inversor realiza una transacción.

## 2.4. La hipótesis de los mercados eficientes

Los académicos financieros optaron por la Hipótesis del Mercado Eficiente (HME) como una tercera vía en el enconado debate entre el chartismo y el fundamentalismo (Fama, 1970). De forma sintética, un mercado es eficiente cuando valora correctamente los activos que cotiza, constituyendo los precios señales eficientes en la asignación de recursos. La implicación básica de la HME es que ninguna técnica de predicción, ni fundamental ni técnica, puede ser empleada con éxito en la obtención de beneficios superiores a los normales.

En un mercado eficiente, la nueva información se absorbe instantáneamente y se incorpora a los precios del mercado, que reflejarán así toda la información relevante existente en cada momento. Este proceso implica que los sucesivos cambios en los precios de cada acción son independientes (es decir, obedecen a un modelo de “paseo aleatorio”), lo que implica que el mercado no tiene memoria: las series históricas de precios no tienen utilidad alguna para predecir los precios futuros. En un mercado eficiente no es posible la obtención de beneficios fáciles superiores a los normales. Por cuanto el arbitraje realizado por los inversores racionales, que intentarán aprovechar cualquier oportunidad para obtener beneficios, eliminará cualquier posible ganancia que pudiese derivarse de la existencia de patrones predecibles.

Aunque la formulación y fundamentación definitiva de la HME se produce en los años sesenta, tal hipótesis recoge, engloba y sistematiza numerosas aportaciones producidas desde principios del siglo pasado sobre la incapacidad de predecir los precios de los activos financieros.

Louis Bachelier en su Tesis Doctoral leída en 1900 desarrolló una elaborada teoría matemática de los precios especulativos, que contrastó en los bonos del gobierno francés, encontrando que tales precios eran consistentes con un modelo precursor del que hoy conocemos con el nombre de paseo aleatorio. Durante los años sesenta se registran avances cruciales en las investigaciones sobre la habilidad predictiva de los precios financieros (Mills, 1993) y el modelo de paseo aleatorio adquiere carta de naturaleza.

Esta hipótesis de mercado eficiente tiene su reflejo matemático en el modelo de camino aleatorio.

$$P_t = P_{t-1} + \epsilon_t, \tag{2.4.1}$$

donde  $P_t$ , es el precio observado al comienzo del instante  $t$  y  $\epsilon_t$  es un término de error que tiene media cero, varianza constante y cuyos valores son independientes unos de otros. La independencia implica no sólo que los incrementos son independientes ( $\Delta p_t = \epsilon_t$ ), sino que cualquier función no lineal de los incrementos es también incorrelada. Dicho término de error suele denominarse a veces ruido blanco estricto (Campbell *et al.*, 1997). Como  $\epsilon_t$  representa los cambios en los precios, dichos cambios serán independientes de su pasado. Haciendo sucesivas sustituciones en esta ecuación puede observarse que podemos escribir el precio corriente como una acumulación de todos los errores pasados es decir.

$$P_t = \sum_{i=1}^t \epsilon_t \quad (2.4.2)$$

de modo que el modelo de paseo aleatorio implica que los precios están generados por una acumulación de cambios puramente aleatorios.

La justificación de este modelo dio lugar a numerosas investigaciones de naturaleza teórica y empírica. Dos trabajos pioneros en este sentido fueron los de Roberts y Osborne. Roberts (1959) presentó numerosos y sofisticados argumentos heurísticos justificando la independencia de los cambios de los precios. Paralelamente Osborne (1959) propuso que la independencia debería ser una propiedad de la que gozasen, no los cambios absolutos de los precios, sino los cambios de los logaritmos de los precios, además tales cambios deberían estar normalmente distribuidos. Ello implica que los precios están generados por un movimiento browniano.

En este modelo, el logaritmo de los precios se comporta como:

$$\log P_t = \log P_{t-1} + \mu + \epsilon_t \quad (2.4.3)$$

donde  $\mu$  es la constante del modelo y  $\epsilon_t$  la perturbación independiente e idénticamente distribuida (IID). Definiendo el rendimiento como  $r_t = \log\left(\frac{P_t}{P_{t-1}}\right)$ , resultará que tales rendimientos son (IID) (Fama, 1970). Los rendimientos están en torno a  $\mu$ , que representa un rendimiento libre de riesgo más la prima de riesgo del inversor. Por lo que, según esta hipótesis, el mercado no tendría memoria y la serie de precios de los activos sería inútil para predecir posteriores variaciones en los precios. De esta forma y atendiendo a la estructura del paseo aleatorio, la mejor estrategia para el inversor es la de comprar y mantener posiciones.

Estos dos últimos trabajos estimularon numerosos estudios, en los años siguientes, acerca de la independencia de los precios o sus logaritmos y del modelo de paseo aleatorio. Fama (1965) realiza, en este sentido, una valiosa aportación empírica examinando las propiedades de autocorrelación de los cambios en los precios.

Durante los años sesenta muchas de las investigaciones fueron conducidas hacia los fundamentos teóricos de los mercados financieros, dando lugar al desarrollo de la teoría de los mercados eficientes. Aparecen entonces serias cuestiones sobre la hipótesis del paseo aleatorio como un modelo teórico de mercado financiero. La suposición de que los cambios en los precios  $\epsilon_t$  son independientes e idénticamente distribuidos se encontró demasiado restrictiva como para ser consistente con una clase razonablemente amplia de modelos de optimización, dentro del marco teórico del equilibrio (Mills, 1993). Surge así la necesidad de relajar tal hipótesis y aparece el concepto de *martingala*

Una martingala es un proceso estocástico  $P_t$  que puede ser escrito en la forma

$$P_t = P_{t-1} + a_t, \tag{2.4.4}$$

donde  $a_t$  se define como un proceso incorrelacionado, con una distribución común. La aparente identidad de este modelo con el paseo aleatorio es puramente superficial. La gran diferencia estriba en que en la martingala los incrementos  $a_t$  no son necesariamente independientes aunque están incorrelacionados linealmente. El modelo de martingala posibilita, por tanto, una forma más compleja de dependencia en las series financieras donde, junto con la independencia lineal, se admiten dependencias entre los momentos condicionales de orden superior.

Esto tiene importantes implicaciones predictivas pues, aunque cualquier técnica de predicción lineal sobre los cambios de precios futuros basada en los precios históricos será necesariamente inefectiva, no se niega la posibilidad de ganancias predictivas de momentos condicionales de órdenes superiores explotando al algún tipo de relaciones no lineales entre las series temporales.

Otra forma alternativa de representar matemáticamente una martingala es mediante la propiedad de la esperanza condicional

$$E(P_{t+1}/P_t, P_{t-1}, \dots) = P_t \quad (2.4.5)$$

que contiene la idea de “juego justo” (*fare game*): Si  $P_t$  representa las ganancias acumuladas en el período  $t$ , en un juego de azar que se realiza en cada período, las ganancias acumuladas que se esperan en el período siguiente serán simplemente iguales a las existentes en el período actual. Desde el punto de vista de la predicción, la hipótesis de martingala implica que la mejor predicción del precio de mañana es simplemente el precio de hoy. Tal propiedad se considera como una condición necesaria de mercado eficiente, en su forma débil, pues expresa que la información contenida en los precios pasados esta instantánea, plena y perpetuamente reflejada en los precios corrientes de los activos (Campbell *et al.*, 1997). Otro aspecto de la hipótesis de martingala es que los cambios de precios no solapados se encuentran incorrelacionados, lo que implica la inefectividad de cualquier regla lineal de predicción de los cambios futuros de precios basados solamente en su propia historia.

El ataque a las teorías precedentes, el Análisis Técnico y el Análisis Fundamental, dividió la HME en tres niveles distintos de eficiencia: la forma

“débil”, la forma “semifuerte” y la forma “fuerte”.

En su forma débil, la HME establece que los precios de los activos reflejan, en cada instante, toda la información contenida en el pasado de las series históricas de precios pasados y de volumen de transacción por lo que resulta imposible predecir el precio de un activo a través de los datos pasados. El futuro de las cotizaciones es independiente de su pasado y por tanto es imposible obtener de forma sistemática, por medio del Análisis Técnico, una rentabilidad superior a la que se obtendría con la sencilla estrategia de “comprar y mantener”, habiendo considerado costes de transacción y una prima de riesgo. Por tanto, según esta hipótesis resulta una pérdida de tiempo analizar gráficas de precios pasados y movimientos del volumen de transacción con el fin de predecir los precios futuros.

En su forma semifuerte, la HME determina que los precios reflejan toda la información pública que pudiese afectar al valor intrínseco de un activo. Es una pérdida de tiempo analizar informes anuales de las empresas o desarrollar reglas de transacción basadas en datos macroeconómicos ya disponibles por el público. De esta forma, sólo los inversores con información privilegiada pueden obtener unos beneficios superiores a los que se consideran normales.

En su forma fuerte, la HME estipula que los precios de los activos reflejan toda la información pública, privada y privilegiada (*inside information*). Incluso los agentes con información privilegiada son incapaces de obtener beneficios anormales. Los cambios en los precios son variables aleatorias independientes que incorporan cualquier información relevante. En definitiva, lo que viene a señalar es que, de existir una posibilidad de beneficio futuro, ésta sería inmediatamente conocida por los agentes, que harían modificar el precio a través

de los movimientos de oferta y demanda, hasta el punto de hacer desaparecer ese beneficio potencial. Como ilustración, basta decir que es la misma razón por la que no hay billetes de 500 euros tirados en la calle. Primero, porque es muy raro encontrárselos y, segundo, porque en ese caso la gente los cogería rápidamente.

### 2.4.1. Anomalías e imperfecciones empíricas

A mediados de los setenta, la HME, en su versión semifuerte, fue una hipótesis ampliamente aceptada en la literatura económica. Fama (1970) aporta tanto fundamentación teórica como evidencia empírica a favor de la eficiencia de los mercados, sentando las bases del modelo de paseo aleatorio antes mencionado.

No obstante, la teoría de los mercados eficientes no concuerda con numerosos acontecimientos como los sucedidos en los mercados bursátiles en octubre de 1929, en octubre de 1987 o en octubre de 1997. Tampoco es capaz de explicar las tormentas monetarias que sacudieron los mercados cambiarios europeos entre 1992 y 1994; en general la HME no es capaz de explicar por qué los precios de las acciones presentan más volatilidad que lo que cabría esperar de la evolución de sus variables fundamentales (West, 1988).

A partir de la década de los ochenta, ya eran muchos los investigadores que tenían serias dudas sobre la eficiencia de los mercados, existiendo numerosas evidencias empíricas contrarias con la HME (Chiarella, 1992). Algunos de estos estudios han mostrado las inconsistencias con el modelo puro de paseo aleatorio o forma débil de la HME. Otros trabajos advirtieron anomalías relativas al exceso de rendimientos debidos a una variedad de factores tales como tamaño de la firma, sentimientos del mercado, días de la semana, etc.

Keim (1986) y Lo y MacKinlay (1988) realizan un tratamiento empírico

de muchas de estas anomalías utilizando contrastes basados en estimadores de varianza, concluyendo que el modelo de paseo aleatorio no es en general consistente con el comportamiento de las rentabilidades bursátiles semanales.

Otros trabajos llegan a la conclusión de inconsistencia entre la hipótesis de eficiencia del mercado, que afirma que los beneficios de una inversión son serialmente independientes, y la reversión de beneficios. Fama y French (1988) encuentran evidencias de que, aunque existe una autocorrelación serial positiva para los rendimientos de las acciones en períodos cortos de tiempo, en períodos más largos muestran una autocorrelación serial negativa. A análogas conclusiones llegan Poterba y Summers (1987) que la explican por medio de una sobreacción especulativa de la tendencia de los precios, la cual es gradualmente eliminada más allá de cierto rango.

French y Roll (1986) revisan una de las ideas básicas en las que descansa la HME, que afirma que los movimientos de los precios en los mercados están ocasionados por la llegada de nueva información. Su trabajo pone de manifiesto que la difusión de información privada es la causa principal que explica la varianza del precio en el mercado, dejando abierta la posibilidad de influencias sobre la volatilidad del mercado procedente de las noticias relacionadas con la economía, la industria, empresas privadas, la información privilegiada y la actuación de los profesionales.

En West (1988) se encuentra una recopilación de diversas investigaciones relativas al exceso de volatilidad de los precios de las cotizaciones. En todos ellos se encuentran evidencias a favor de la hipótesis de exceso de volatilidad y se establece que dicho exceso no puede ser explicado adecuadamente por los



modelos estándar de rendimientos esperados o por medio de burbujas racionales. El exceso de volatilidad apunta hacia la presencia de fuerzas dinámicas intrínsecas relacionadas con la especulación en los mercados. West cuestiona la idea de que los precios de los activos estén determinados, exclusivamente, por los valores fundamentales esperados racionalmente que de acuerdo con Samuelson (1965), deberían obedecer a un modelo de paseo aleatorio.

Shiller (1989) propone un estadístico capaz de contrastar el exceso de volatilidad que presentan los precios de los títulos, con respecto a la volatilidad, que sería capaz de explicar el modelo de mercado eficiente. En su trabajo Shiller introduce diversas consideraciones acerca de la psicología de multitudes en el proceso de determinación de los precios en el mercado de valores.

Durante los últimos años se viene registrando un importante volumen de investigaciones que tratan de explicar los comportamientos observados en las series temporales de precios generados por los mercados: distribuciones con colas gruesas, agrupación de volatilidad, bajas autocorrelaciones pero sustancial dependencia no lineal, etc. En este sentido, han proliferado en la literatura las aplicaciones de los modelos ARCH, GARCH y sus derivados (Engle, 1982).

## **2.5. Renacimiento académico del análisis técnico**

Históricamente el análisis técnico ha sido frecuentemente menospreciado por los académicos de la comunidad financiera. Como trabajo pionero podría citarse a Alexander (1961) que aplicó una regla de filtro en la que se compra un activo cuando su precio se incrementa en un  $x\%$  desde su último soporte y se vende

cuando su precio cae un  $x\%$  desde su última resistencia. Alexander establecía los soportes y las resistencias como los mínimos y los máximos, respectivamente. Pretendía, mediante esta técnica, estudiar la posible existencia de tendencias en los precios que pudiesen estar enmascaradas entre las fluctuaciones del mercado. Concluyó que en los Índices Standard & Poor's 500 y Dow Jones de la Bolsa de Nueva York existen tendencias predecibles, mediante la comparación del rendimiento de esta estrategia con la estrategia de comprar y mantener. Fama (1965) y Fama y Blume (1966) presentaron un análisis empírico más detallado de dichas reglas, corregidas por dividendos y costes de transacción. Llegaron a la conclusión de que tales reglas no mejoran la simple estrategia de "comprar y mantener". Estos trabajos mantuvieron al análisis técnico fuera de los medios académicos durante casi treinta años.

Desde finales de los ochenta se ha puesto especial énfasis en contrastar la posibilidad de predecir los rendimientos bursátiles. Tal es el caso de diversos trabajos como los de Fama y French (1988), Lo y MacKinlay (1988) y Fama (1991).

No obstante, durante los años noventa, el análisis técnico experimentó un importante renacimiento en la literatura financiera, en la medida en que recientes estudios sobre los rendimientos de los activos han puesto de manifiesto la posibilidad de predecir hasta cierto grado en muchos mercados.

Tales aportaciones empíricas han demostrado que las conclusiones acerca del nulo poder predictivo de las reglas chartistas eran, como mínimo, precipitadas.

Un trabajo crucial en este sentido es el de Brock *et al.* (1992). En dicho trabajo se contrasta la capacidad de generar beneficios de las dos reglas técnicas

más populares (medias móviles y ruptura del rango) en el índice Dow Jones. Los autores concluyen que los rendimientos obtenidos mediante estas estrategias no son consistentes con los cuatro modelos más utilizados en finanzas: el Paseo Aleatorio, AR(1), el GARCH-M y el GARCH exponencial. Es decir, que las diversas reglas usadas en el Índice Dow Jones eran capaces, bajo ciertas hipótesis, de generar beneficios frente a los modelos econométricos más empleados en finanzas. Encuentran que las señales de compra generan rendimientos más altos que las señales de venta y, además, los rendimientos que proporcionan las señales de compra son menos volátiles que los rendimientos que proporcionan las señales de venta. Detectan claros síntomas de estructura no lineal que se manifiesta en la naturaleza asimétrica de los rendimientos. Además, los rendimientos obtenidos de las señales de venta son negativos lo que no es fácilmente explicable por cualquiera de los modelos de equilibrio existentes en la actualidad.

Uno de los aspectos metodológicos más innovadores del trabajo de Brock *et al.* (1992) consiste en la aplicación de la metodología de “simulaciones sucesivas” (*bootstrapping*), original de Efrom (1983). Siguiendo esta metodología se evalúa el poder predictivo de las reglas de mercado comparando su capacidad de generar beneficios en la serie original, con su aplicación en quinientas replicaciones artificiales de dicha serie que se generan a partir de los diversos modelos nulos considerados desordenando aleatoriamente y con reemplazamiento los residuos obtenidos en cada uno de dichos modelos.

En definitiva, Brock *et al.* (1992) concluyen que los altos rendimientos obtenidos mediante el análisis técnico no pueden ser explicados por los modelos econométricos más usados en la literatura financiera.

Igualmente, Bessembinder y Chan (1995), Mills (1996) y Fernández Rodríguez *et al.* (1999a) presentan evidencias similares para los mercados asiáticos, británico y español, respectivamente.

En sentido similar, Pesaran y Timmermann (1992) presentan evidencia sobre cierta posibilidad predictiva del exceso de rendimientos en carteras de los índices Standard & Poor´s 500 y el Dow Jones.

La metodología de Brock *et al.* (1992) ha sido, extendida a los mercados cambiarios por Levich y Thomas (1993), Osler y Chang (1995) y Mills (1996). Sullivan *et al.* (1997) corrigen dicha metodología del sesgo que puede introducir en los resultados el figoneo de datos.

No obstante, otros autores han obtenido resultados más pesimistas en relación a la capacidad predictiva del Análisis técnico. Así, Corrado y Lee (1992), Hudson *et al.* (1996), Bessembinder y Chan (1995), Brown *et al.* (1998) y Allen y Karjalainen (1999) concluyen que los modelos que utilizan reglas técnicas no son capaces de generar beneficios una vez descontado los costes de transacción y la prima de riesgo pero si que tienen capacidad predictiva.

Las discrepancias entre los resultados obtenidos, por los distintos autores, en relación al poder predictivo del Análisis técnico ha conducido a una profunda reconsideración sobre el modo de elaborar las reglas técnicas. En este sentido Lo y MacKinley (1990) concluyeron que las especificaciones *ad-hoc* de los parámetros que se emplean en las reglas técnicas, provocaba figoneo en los datos. Es decir, que la elección arbitraria, tanto de las reglas como de los parámetros seleccionados, producían desviaciones en los resultados. Asimismo, concluían que, de haber empleados técnicas de optimización para la elección de dichos

parámetros, estas desviaciones se podían haber reducido e incluso, en algunos casos eliminado. Neely *et al.* (1997) y Allen y Karjalainen (1999) se basaron en estas conclusiones para desarrollar modelos para los mercados cambiarios y de activos, respectivamente, basados en reglas técnicas que previamente habían sido optimizadas mediante programación genética.

En la actualidad han emergido otros enfoques alternativos dentro del Análisis Técnico y en la predicción de las series financieras, sobre la base de nuevas metodologías no lineales y un importante avance en las técnicas computacionales, tales como la reconstrucción de espacios de fases de series temporales (Farmer y Sidorowich, 1987; Fernández Rodríguez y Sosvilla Rivero, 1998; Fernández Rodríguez *et al.*, 1999b), animados por la literatura del caos determinista.

Por otra parte, recientes avances en las matemáticas y la inteligencia artificial, muchas de las cuales son aplicaciones directas del mundo de la biología, particularmente en las redes neuronales (Refenes, 1995; Gençay, 1998; Fernández Rodríguez *et al.*, 2000), los algoritmos genéticos (Koza, 1992; Allen y Karjalainen, 1999), la lógica difusa (Benachenhou, 1994) y los Wavelets (Daubechies, 1990), han abierto nuevos caminos en la predicción.

Cabe citar como trabajos a destacar en este sentido el de Gençay (1998) que mide la capacidad de generar beneficios con simples reglas técnicas basadas en modelos de redes neuronales. Encuentra que tales reglas proporcionan beneficios significativos frente a la estrategia de comprar y mantener. Además, concluye que las predicciones sobre los signos de los rendimientos en tales modelos de redes son estadísticamente significativos.

Más recientemente, Fernández Rodríguez *et al.* (1997); Fernández Rodríguez *et al.* (1999a) aplican técnicas locales no paramétricas de predicción conocidas como predicciones por ocurrencia análogas al mercado español y al Índice Nikkei 255. Encuentran, usando técnicas de filtro, que durante los episodios no alcistas del mercado la capacidad de generar beneficios de las predicciones no paramétricas es superior a la estrategia de “comprar y mantener”.

En este sentido, para el caso español existen algunos trabajos sobre el Índice General de la Bolsa de Madrid (IGBM). Olmeda y Pérez (1995) estudian la posibilidad de mejorar la capacidad predictiva, de un modelo de camino aleatorio por medio del uso de redes neuronales. Fernández Rodríguez *et al.* (2000) analizan, para varios periodos del mercado, la capacidad de generar beneficios de una regla técnica basada en redes neuronales artificiales frente a la estrategia de “comprar y mantener”.

El conjunto de estas técnicas abre amplias posibilidades en la predicción de series financieras mediante nuevas y sofisticadas formas del Análisis Técnico. Ante diversos resultados parciales muy prometedores que aparecen recientemente en la literatura, una importante tarea pendiente es la de perfeccionar, y contrastar empíricamente de forma sistemática, las nuevas formas predictivas tanto en mercados cambiarios como bursátiles, españoles y extranjeros, así como mercados de futuros.

Este renacimiento que ha vivido el análisis técnico durante los noventa ha cambiado también parte de la rígida filosofía que se tenía acerca de la HME. Como reconocen abiertamente Campbell *et al.* (1997) en su manual de Econometría Financiera, la literatura académica admite hoy una teoría sobre los mercados

eficientes mucho más flexible que hace veinticinco años, aceptando la posibilidad de predicciones no lineales asociadas a la existencia de un tipo de riesgo dinámico que es inherente al mercado.

# Capítulo 3

## Los métodos de análisis: nuevas tecnologías predictivas basadas en el aprendizaje automático

### Índice General

---

<b>3.1. Aprendizaje automático . . . . .</b>	<b>71</b>
3.1.1. Estado del arte del aprendizaje automático . . . . .	72
3.1.2. Características del aprendizaje automático . . . . .	75
<b>3.2. Redes neuronales artificiales . . . . .</b>	<b>82</b>
3.2.1. Estado del arte de las redes neuronales artificiales . . . . .	82
3.2.2. Estructura de una red neuronal biológica. . . . .	84
3.2.3. Redes neuronales artificiales . . . . .	85
3.2.4. Redes de retropropagación . . . . .	90
<b>3.3. Algoritmos genéticos . . . . .</b>	<b>101</b>
3.3.1. Evolución natural . . . . .	102
3.3.2. Algoritmos Genéticos . . . . .	109
3.3.3. Algoritmos genéticos con parámetros continuos . . . . .	129
<b>3.4. Predicciones por ocurrencias análogas . . . . .</b>	<b>131</b>
3.4.1. Algunos criterios de selección de los parámetros en la predicción NN . . . . .	139
3.4.2. Selección automática de los parámetros en la predicción NN con AGs. . . . .	150
<b>3.5. Programación Genética . . . . .</b>	<b>155</b>



3.5.1. Proceso evolutivo de la programación genética . . . .	155
3.5.2. Estructura de los programas de computación en la programación genética . . . . .	156
3.5.3. Inicialización de la población . . . . .	161
3.5.4. Cruce . . . . .	163
3.5.5. Mutación . . . . .	169

---

### **3.1. Aprendizaje automático**

Las técnicas computacionales empleadas en este trabajo pueden englobarse dentro de lo que la literatura ha dado en denominar “aprendizaje automático” o “algoritmos inteligentes”. Mediante esta acepción se pretende aglutinar aquellos procesos que persiguen que el propio algoritmo tenga la habilidad de aprender. Sin embargo, en una definición más exacta podríamos decir que se trata del estudio de los algoritmos computacionales que se mejoran automáticamente a través de la experiencia. (Mitchell, 1996)

Existe en estos últimos años un creciente interés por el uso de nuevas técnicas computacionales dentro del estudio de la economía financiera. Todas estas técnicas permiten el desarrollo de métodos de optimización y búsquedas alternativos a los tradicionalmente empleados. Dentro de estas técnicas computacionales cabría destacar las redes neuronales artificiales, las técnicas de regresión no paramétrica y la computación evolutiva.

Las redes neuronales artificiales son algoritmos que intentan simular la actividad que se desarrolla en redes neuronales biológicas. Las redes logran, a través de unidades interconectadas en paralelo y mediante un proceso de aprendizaje, que las respuestas proporcionadas se asemejen más al objetivo prefijado.

Las técnicas de regresión no paramétricas se basan en la idea de que patrones de comportamiento en el pasado de la serie pueden reproducirse en tramos futuros. Por lo que el comportamiento de la serie tras los días posteriores a esos tramos debe ser similar.

La computación evolutiva se basa en la aplicación de la teoría de la evolución

natural y la genética a la adaptación evolutiva de estructuras computacionales. Aunque los algoritmos genéticos desarrollados por Holland (1975) son los más populares de los métodos evolutivos, recientemente se está empleando la programación genética (GP). Esta técnica fue desarrollada por Koza (1992) y no es más que una extensión de los algoritmos genéticos. La principal diferencia se basa en que el espacio de soluciones son estructuras de programas computacionales, estructuras más complejas que las utilizadas por los algoritmos genéticos más tradicionales.

### 3.1.1. Estado del arte del aprendizaje automático

Como señalábamos anteriormente, el aprendizaje automático consiste en el diseño de algoritmos que tienen la habilidad de aprender. Un trabajo pionero en este campo fue el de Samuel (1963) que consiguió que un programa computacional realizara un aprendizaje serio. El objetivo era programar un ordenador digital que se comportase de tal manera que, si ese comportamiento fuese realizado por un ser humano, se dijese que su conducta implica un proceso de aprendizaje. El resultado fue un programa de ordenador que, aplicado al juego de las damas, fue capaz de mejorar sus resultados mediante la práctica. Samuel (1963) fue quien terminó por dar nombre a la disciplina.

Por otra parte, Friedberg *et al.* (1959) logró que el ordenador escribiese programas computacionales para resolver problemas relativamente sencillos. El objetivo era conseguir que la máquina realizase tareas por sí misma. Concluyó que había dos posibilidades para lograrlo. Por una parte, había que darle capacidad para resolver problemas sin necesidad de decirle exactamente cómo hacerlo. Por otra parte, debía especializar sus acciones diciéndole exactamente que hacer a

través del proceso.

Estas dos vertientes tuvieron su reflejo en dos disciplinas: El aprendizaje automático y la inteligencia artificial. El aprendizaje automático hacía especial énfasis en el aprendizaje, ya que si la máquina no sabía exactamente cómo resolver el problema debía aprender del entorno para hacerlo. La inteligencia artificial se basa principalmente en el conocimiento. Si la máquina obedecía a una serie de instrucciones que la guiaban en el proceso diciéndole exactamente qué hacer, entonces todas estas rutinas tenían que estar implementadas ya en la máquina.

Durante la década de los sesenta y setenta se trabajó, fundamentalmente, sobre inteligencia artificial. Básicamente se trataba de emular el conocimiento humano y transformarlo en conocimiento máquina. En los años ochenta tuvo un enorme auge los sistemas expertos. Así, por ejemplo, se preguntaba a expertos sobre cómo resolver una determinada tarea y esa información se codificaba en un sistema de toma de decisiones. El sistema sabe exactamente que hacer cuando se le presentaba una situación conocida. Sin embargo, tiene bastantes problemas cuando se encuentra ante entradas novedosas o con ruido. Así, los sistemas expertos son sistemas basados en el conocimiento, en este caso de los expertos, y no en el aprendizaje del propio sistema.

La forma de evitar los problemas de los sistemas expertos ante situaciones desconocidas consiste en obligar al sistema a aprender del entorno. El aprendizaje computacional recoge esta necesidad. Las redes neuronales artificiales, las técnicas de regresión no paramétricas, los algoritmos genéticos y la programación genética son sistemas que se programan para aprender y, por lo tanto, están incluidos

dentro del aprendizaje automático. Todas estas técnicas tienen puntos en común, como son la definición del entorno y el proceso de aprendizaje. Sin embargo, son las distintas representaciones y enfoques de cómo aprende el sistema del entorno lo que da lugar a las diferencias entre las distintas técnicas.

La característica común en estos sistemas es el proceso de aprendizaje. Comienza con la identificación del dominio de aprendizaje. El dominio de aprendizaje está constituido por las las entradas al sistema (*inputs*), que representan características del problema a tratar, y las salidas objetivos (*targets*), que caracteriza el resultado o resultados que debe predecir del sistema. Evidentemente, las entradas del problema deben estar relacionadas de alguna forma con las salidas objetivos a predecir para que el sistema tenga alguna habilidad predictiva. Así, por ejemplo, en la serie de precios al cierre del Índice General de la Bolsa de Madrid el precio de hoy y de los 9 días anteriores pueden constituir las entradas del modelo y el precio de mañana la salida que hay que predecir.

Los datos que recogen las relaciones entre las entradas y salidas del sistema representan el entorno donde el sistema debe aprender y constituyen el dominio de aprendizaje. Pero para ello se debe seleccionar parte de los datos de los que el sistema debe aprender las relaciones entre los *inputs* y los *targets*. Este periodo se denomina periodo de entrenamiento y constituye uno de los pilares dentro de los sistemas basados en el aprendizaje automático. El sistema debe entrenarse en este periodo con los casos seleccionados y debe aprender de estos ejemplos para producir un modelo que obtenga unas salidas apropiados. Se puede concluir que el aprendizaje se produce durante el entrenamiento.

Pero el proceso no puede concluir en el entrenamiento. Se debe verificar la calidad del aprendizaje producido en la etapa de entrenamiento, por lo que se debe contrastar la habilidad predictiva de la mejor solución encontrada en el periodo de entrenamiento con otros pares de entradas/salidas. Estos nuevos ejemplos constituyen el periodo de verificación o de contrastación y, aunque deben ser distintos, deben pertenecer al dominio de aprendizaje. Este periodo de verificación nos da una medida de la capacidad de generalización del sistema, ya que puede comportarse bien dentro del periodo de entrenamiento pero muy mal fuera de él, con lo que no se ha conseguido el objetivo buscado. De esta forma, el sistema habría conseguido localizar el problema, pero no sabe generalizarlo.

### **3.1.2. Características del aprendizaje automático**

Como se ha visto en la sección anterior, existen características comunes dentro de las diversas técnicas computacionales de ser empleadas en la econometría financiera Y que se engloban dentro del “aprendizaje automático”. Sin embargo, cada uno posee una serie de especificidades que los diferencia del resto. Podemos clasificar estas diferencia en base a varios criterios:

#### **Representación de las soluciones**

Cada sistema tiene una forma distintas de representar las soluciones. El espacio de soluciones estará configurado por aquellas soluciones que sean compatibles con la representación definida por cada sistema. Las distintas representaciones pueden afectar a:

- **La representación de las entradas y salidas** que puede ser números binarios, reales, valores booleanos, etc.

- **La representación de los conceptos que el sistema debe aprender** ya que se puede hacer a través de expresiones booleanas, redes neuronales, programas computacionales, cromosomas, etc.
- **La interpretación de los conceptos aprendidos como salidas**, ya que éstas pueden tomar las formas de representación booleana; umbral, como en las redes neuronales artificiales; las basadas en casos, como en los algoritmos basados en las predicciones por ocurrencias análogas; las representaciones en árboles y las representaciones basadas en la genética, como los algoritmos genéticos que utilizan una representación de carácter fijo y la programación genética que utiliza estructuras de tamaño variable.

### **Operadores de búsqueda**

Normalmente, en un problema medianamente complejo el espacio de soluciones se caracteriza por ser demasiado grande para evaluar todos los candidatos (búsqueda exhaustiva). Para ello, el sistema debe adoptar un algoritmo de búsqueda que sea capaz de moverse a través del espacio de soluciones de una forma eficiente. En qué candidato se debe empezar la búsqueda, cuál es el siguiente a evaluar y cuándo debe terminar el criterio de búsqueda son detalles que el sistema debe predefinir antes de ejecutarse. Todo ello está especificado en los operadores de búsqueda. Cada sistema utiliza distintos operadores.

- **Operadores genéricos/específicos.**

Se delimita el campo de soluciones a sectores más prometedores mediante especificaciones comunes de esos candidatos. Suelen ser los empleados en problemas cuya representación sea booleana o umbral. Por ejemplo,

en problemas de representación booleana podemos especificar que los candidatos más prometedores son los que cumplen una condición en particular. Si a través de la estrategia de búsqueda se encuentra una nueva condición se obtendrá un espacio más específico dentro del anterior y estará compuesto por aquellos candidatos que cumplan con las dos condiciones. Si hubiese una tercera se limitaría aun más el espacio de soluciones prometedoras.

- **Operadores basados en el gradiente descendente.**

Las redes neuronales artificiales (RNA) son los sistemas que con mayor frecuencia utilizan estos operadores. El algoritmo más empleado en las RNA es el de retropropagación, donde los pesos de las conexiones entre neuronas se ajustan del gradiente descendente. El error se retropropaga hacia atrás, empezando por la salida de la red, y ajusta las conexiones para que ese error sea mínimo. Es un algoritmo que intenta encontrar el mínimo a través de pequeños cambios en las ponderaciones para moverse sobre la superficie del error hasta encontrar ese mínimo o hasta que se cumpla un criterio de parada. Este algoritmo se verá con más detalle en el apartado 3.2.4

- **Operadores genéticos.**

Los algoritmos genéticos y la programación genética utilizan estos operadores para buscar nuevos individuos dentro del espacio de soluciones. Entre estos operadores hay que destacar la mutación y el cruce. En el cruce se elige a dos individuos de la población y parte de la información contenida en los dos padres se intercambia dando lugar a dos nuevos



individuos llamados “hijos”. De esta forma, la información contenida en los padres, que los caracterizan por ser soluciones prometedoras, se acumulan en su descendencia (Koza, 1992). La mutación, por contra, afecta a un sólo individuo y consiste en alterar aleatoriamente la información contenida en un individuo. La mutación permite que aparezcan nuevos individuos en la población y, por lo tanto, provoca que exista innovación dentro del sistema evitando que el sistema se pierda en un óptimo local.

### **Estrategias de búsqueda**

Mientras que los operadores de búsqueda definen los tipos de saltos que se han de dar dentro del espacio de soluciones, la estrategia de búsqueda nos muestra sobre que superficie del espacio de soluciones se realiza la búsqueda, por lo que, mide de alguna manera, el grado de la búsqueda guiada.

- **Búsqueda ciega.**

Este tipo de estrategias se ejecuta sin tener en cuenta la información recopilada en iteraciones anteriores. Se contraponen a las búsquedas heurísticas que utilizan la información del espacio de soluciones y guían la búsqueda en función de los resultados logrados en los pasos anteriores. Escencialmente, la búsqueda a ciegas se basa en la búsqueda a través de árbol, que es como se simbolizan el espacio de soluciones en este tipo de estrategias. La búsqueda ciega sólo tiene en cuenta qué ramas del árbol ya han sido previamente visitadas (Banzhaf *et al.*, 1998).

- **Gradiente descendente (*Hill climbing*).**

En este caso, la estrategia consiste en tomar un candidato y evaluarlo.

A continuación se toma otro, se evalúa y compara con el anterior. Si es mejor tomamos este nuevo candidato como referencia y si no es así se descarta (normalmente). Se continúa con este proceso iterativo hasta llegar a un criterio de finalización. El algoritmo de retropropagación utilizado en la mayoría de los sistemas que utilizan redes neuronales artificiales y el algoritmo “recocido simulado” (*simulated annealing*) (Kirkpatrick *et al.*, 1983) son algoritmos basados en esta estrategia de búsqueda. En concreto el algoritmo de retropropagación, que explicaremos con más detalle en la sección 3.2.4, parte de un punto en el espacio de soluciones seleccionado aleatoriamente. Posteriormente, a través de la modificación de las ponderaciones de las conexiones de la red, encuentra un nuevo punto pero en saltos de tamaño fijo. La dirección de estos saltos la marcan ciertas derivadas por un método de gradiente descendente y que guían la búsqueda hacia zonas de mínimos. A diferencia del algoritmo de recocido simulado, el algoritmo de retropropagación utiliza saltos de tamaño fijo aunque recientes variaciones de este algoritmo trabajan ya con saltos de tamaño variable.

- **Búsqueda prometedora.**

A diferencia de las estrategias vistas con anterioridad, que son búsquedas punto a punto, la búsqueda prometedora (*beam search*) trabaja con multitud de puntos. Estos puntos se seleccionan sobre la base de un valor, que en los algoritmos evolutivos lo da la función de adaptación. Sólo a éstos se le aplican los operadores de búsqueda. De esta forma, estos sistemas regulan los individuos que integran esa población prometedora a través de los operadores de búsqueda y el orden en función de la adaptación que

cada individuo consiga. El resto de los puntos del espacio de soluciones son descartados. La búsqueda prometedor es, por lo tanto, un compromiso entre el algoritmo de gradiente descendente y la búsqueda exhaustiva.

Para mostrar la diferencia entre los operadores de búsqueda y estrategias de búsqueda, lo mejor es verlo a través de un sencillo ejemplo (Díaz *et al.*, 1996). Nos sitúan en una ciudad que no conocemos y nos piden encontrar el edificio más alto. Podríamos emplear una búsqueda a ciegas para alcanzar ese objetivo. Consistiría en recorrer calles y tomar la altura de los edificios hasta llegar a cierto criterio de parada. Se podría emplear una estrategia basada en el gradiente descendente, que sería equivalente a pensar que los edificios altos se concentran en una cierta zona, por lo que al encontrar un edificio más bajo que el anterior, desecharíamos la ruta que hemos tomado y escogeríamos otra ruta. Esta sería la estrategia, los operadores de búsqueda nos daría si dada cierta intersección tomamos la calle de la derecha, de la izquierda o continuamos de frente.

## Aprendizaje

- Aprendizaje supervisado

El aprendizaje puede ser supervisado, si se presenta al sistema junto a los patrones de entrada los correspondientes patrones de salida deseada (*target*).

- Aprendizaje no supervisado

El aprendizaje puede ser no supervisado, si no se le presenta al sistema los patrones de salidas deseadas.

- Aprendizaje reforzado

Existe un tercer tipo de aprendizaje, a medias entre los dos anteriores, llamado aprendizaje reforzado. En éste, se le indica al sistema si la respuesta que ha proporcionado es correcta o no, pero no se le indica que respuesta debe dar.

En los apartados anteriores hemos visto cómo los distintos sistemas englobados dentro del aprendizaje automático, intentan aprender del entorno para mejorar su comportamiento, aunque posean distintas formas para lograr ese objetivo. Las redes neuronales artificiales, los algoritmos genéticos y la programación genética parten del mismo principio, conseguir que el algoritmo aprenda de la experiencia, pero enfocan el problema de formas distintas ya que cada técnica emplea distintas estrategias, operadores de búsqueda, aprendizajes, o incluso representación de soluciones. Pero dentro de cada técnica surgen nuevas incorporaciones que amplían el campo del aprendizaje automático. Por nombrar algunos ejemplos, en un principio los algoritmos sólo trabajaba con representación binaria de los individuos posteriormente se ha diseñado los algoritmos para que puedan trabajar con una representación continua del problema.

En esta tesis nos hemos centrado en las redes neuronales artificiales, la regresión no paramétrica, los algoritmos genéticos y la programación genética para abordar el problema de la predicción en las series financieras. En este apartado hemos presentado los rasgos principales de estas técnicas, profundizaremos en ellas en los apartados siguientes.

## 3.2. Redes neuronales artificiales

Las redes neuronales artificiales simulan una estructura computacional en paralelo donde ciertas unidades individuales están interconectadas. Se intenta, de esta forma, crear computacionalmente un modelo que simule la actividad de una red neuronal biológica como es el cerebro. Así, las redes neuronales artificiales imitan la capacidad que tiene el cerebro humano de dar respuestas a situaciones nuevas, sin necesidad de haber desarrollado instrucciones previas para dar respuesta a esas situaciones.

De modo genérico, la modelización e imitación de las actividades cognitivas cerebrales se conoce como Inteligencia Computacional, que se subdivide en dos ramas: la rama simbólico-deductiva y la que trata el aprendizaje automático. La primera se encarga del estudio y aplicación de sistemas expertos, basándose en sistemas formales de reglas y manipulación simbólica, es decir, el conocimiento. En este enfoque, el sistema resuelve las situaciones en función de una serie de reglas que ya posee el sistema.

Dentro del aprendizaje automático podemos diferenciar la rama conexionista, que está inspirada en las redes biológicas naturales. Estos modelos afrontan los problemas en función de unas situaciones de las que previamente ha aprendido cómo reaccionar, sin necesidad de indicarle unas reglas previas.

### 3.2.1. Estado del arte de las redes neuronales artificiales

El origen de los modelos conexionistas los podemos situar en los trabajos de McCulloch y Pitts (1943) quienes propusieron que una cadena de proposiciones lógicas (la suma, multiplicación, negación, implicación y equivalencia) podía

representarse mediante conexiones eléctricas (*hardware*) o mediante programas *software* que simulase el comportamiento neuronal.

Hebb (1949) introdujo la idea de que una percepción o concepto se representa en el cerebro por un conjunto de neuronas activadas simultáneamente y afirmó, además, que la memoria se localiza en las conexiones entre neuronas (sinapsis). Por lo que, las conexiones entre dos neuronas se refuerzan si ambas son activadas.

Rosenblatt (1958) ideó la primera red neuronal artificial que, en un primer momento, la desarrolló sobre un circuito eléctrico, pasándose más tarde a formas más flexibles como es la simulación en ordenador. Estas primeras redes sólo tenían una capa de neuronas y se las denominó “perceptrones”. En este tipo de redes tanto las entradas como las salidas son binarias.

Widrow y Hoff (1960) diseñaron redes similares que se denominaron “ADALINES” (Adaptative Linear Networks), donde los patrones ya podían tomar valores distintos a los binarios.

Sin embargo, Minsky y Papert (1969) demostraron, utilizando técnicas matemáticas, que los perceptrones simples no podían resolver problemas sencillos como la función lógica XOR (o exclusivo).

Esto provocó una crisis en las investigaciones sobre redes neuronales hasta que varios autores, de forma independiente, desarrollaron un nuevo método de aprendizaje para las redes de varias capas. Werbos (1974), Parker (1982), Rumelhart *et al.* (1986) introdujeron el algoritmo de retropropagación para entrenar la red en problemas arbitrarios, de forma que, pudiesen resolver los problemas planteados en el trabajo de Minsky y Papert.

### 3.2.2. Estructura de una red neuronal biológica.

La unidad básica de una red neuronal biológica la constituye la neurona. Éstas están interconectadas entre sí formando circuitos o redes que desarrollan funciones específicas. Las señales eléctricas producidas tras un estímulo pasan a través de las neuronas excitándolas o inhibiéndolas. El impulso puede ser más o menos intenso, dependiendo del aprendizaje de la red neuronal.

La neurona se divide en los siguientes elementos [Figura 3.1]:

- El soma, o célula nerviosa, que es el cuerpo central de la neurona.
- El axón, unido al soma, que es el que se activa eléctricamente y produce el impulso eléctrico emitido por la neurona y que se escinde en millares de ramificaciones.
- Las dendritas, que son las encargadas de recibir las señales que proceden de otras neuronas.

Las conexiones entre neuronas se producen a través de la sinapsis, que transforma el impulso eléctrico en un mensaje neuroquímico, ya que los axones jamás tocan las dendritas de otras neuronas. Las señales excitadoras o inhibitorias se combinan, y en función de la estimulación total recibida, toman un cierto nivel de activación, que se traduce en la generación de impulsos nerviosos que se producen con una determinada tasa de frecuencia o “tasa de disparo”. El aprendizaje se produce mediante la variación de la efectividad de la sinapsis, cambiando la influencia que tienen unas neuronas con otras.

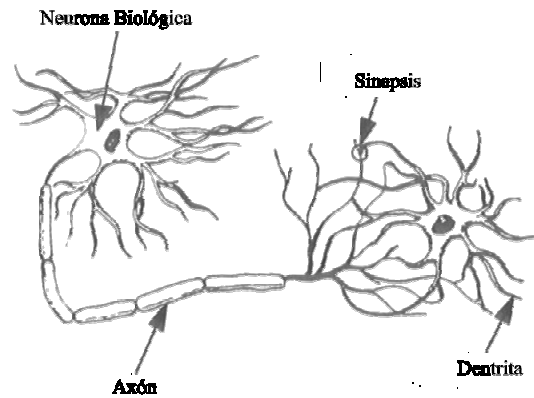


Figura 3.1: Red neuronal biológica

### 3.2.3. Redes neuronales artificiales

Todos los elementos descritos para las redes neuronales biológicas se pueden implementar en elementos *hardware*, como transistores y resistencias, o bien, en *software*. Se simula el soma de la neurona artificial como el sumatorio de las entradas que provienen bien del entorno, bien de las salidas de otras neuronas artificiales. El axón se simula mediante una función, que llamaremos de activación. Y las dendritas se simulan mediante ponderaciones a las entradas de la neurona [Figura 3.2]. Así, el impulso excitatorio se representa mediante ponderaciones positivas y los inhibitorios mediante ponderaciones negativas.

Aunque originalmente las redes se diseñaron con una sola capa, como se ha explicado en el apartado 3.2.1, pronto se demostró que no podían resolver problemas sencillos como la función lógica XOR (o exclusivo). En la Figura 3.3 podemos observar como un perceptrón simple puede aprender a discriminar entre las dos clases (0 ó 1) a través de un línea sencilla para la función OR. Sin embargo, en la función lógica XOR el algoritmo no puede discriminar linealmente entre los



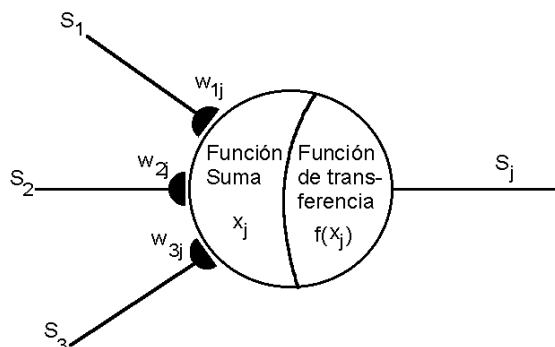


Figura 3.2: Perceptrón simple

distintos casos presentados a la red. Como mínimo habría que introducir dos líneas sencillas (si quisiésemos discriminar linealmente) para poder separar las distintas clases.

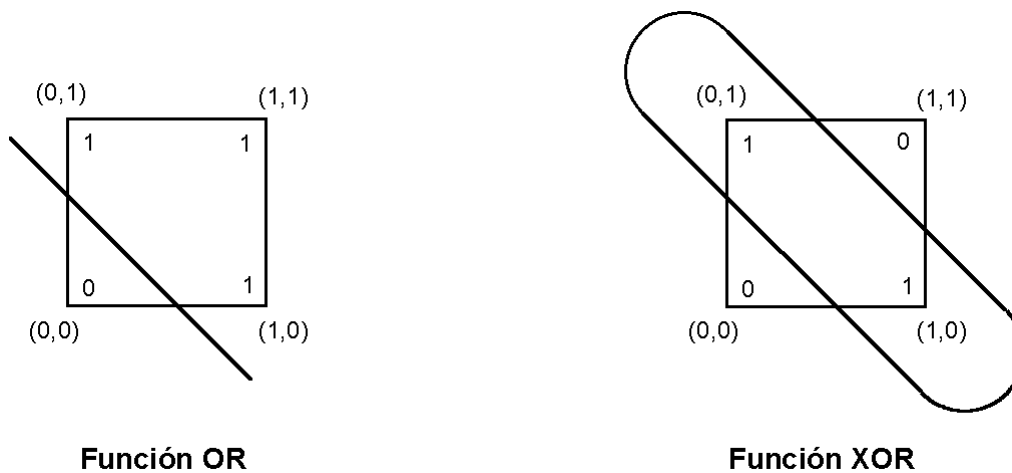


Figura 3.3: Funciones OR y XOR

Como ya señalábamos, esto provocó un estancamiento en las investigaciones hasta que se diseñaron las redes con varias capas. Como muestra la Figura 3.4, actualmente, las redes se suelen diseñar con una capa de entrada, que es la que recibe los inputs; una o varias capas ocultas o intermedias, que recibe las salidas

de la capa de entrada, las somete a las operaciones programadas y las transmite a la siguiente capa; y una capa de salida que se encarga de procesar las señales procedentes de la capa intermedia.

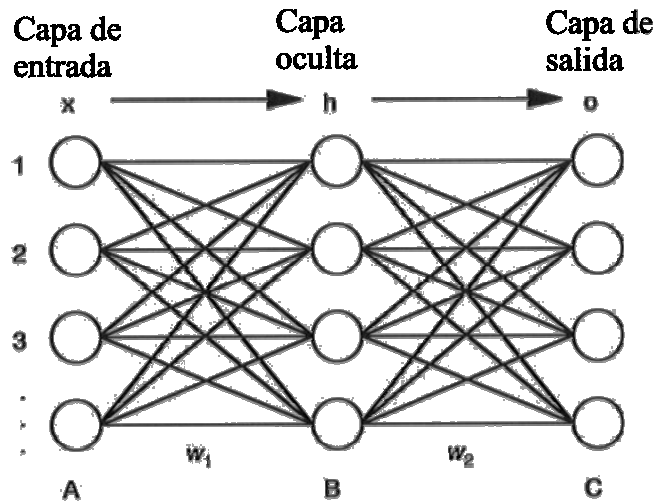


Figura 3.4: Red neuronal artificial

Las redes se pueden estructurar no sólo en función de las capas sino también teniendo en cuenta cómo se interconectan unas capas con otras. Las redes feedforward (redes en cascada) transforman las entradas (*inputs*) en salidas (*outputs*) con señales que viajan en un solo sentido. Las redes recurrentes pueden llevar información a unidades donde se ha procesado previamente formando, de esta manera, bucles en la red. En estas redes se permiten también las conexiones intracapa, incluso la de una unidad consigo misma.

Cada unidad de la capa oculta y de la capa de salida contiene una función de transferencia (llamada también función de activación), que transforma la señal recibida. La capa de entrada no tiene función de transferencia, pero se utiliza para distribuir las señales de entrada a la red. Por esa razón, muchos analistas

no la consideran como una capa en sí misma.

### Funciones de transferencia

Se suele distinguir entre tres tipos de funciones de transferencia:

- **Funciones lineales**, en las que las salidas son proporcionales a las entradas.
- **Funciones umbral**, en las que las salidas son valores binarios que dependen de si el sumatorio ponderado de las entradas supera o no cierto valor de umbral. Podemos destacar dentro de este tipo de funciones binarias de “límite estricto” (*hard-limit*).
- **Funciones no lineales**, no proporcionales a la entrada, entre las cuales la más usada es la función sigmoideal. Esta es, sin duda, una de las funciones más utilizadas dentro de las redes neuronales, ya que presenta bastantes similitud con los impulsos eléctricos que se producen en las redes biológicas. Su forma funcional es:

$$f(x) = \frac{1}{1 + e^x} \quad (3.2.1)$$

Se trata de una función continua creciente, no lineal y derivable. La salida de esta función está comprendida entre 0 y 1. Cuando el argumento de la función es 0, la salida es igual a 0.5, indicando que, incluso en periodos de no estimulación, la unidad tiene cierta actividad. Las razones por las que se suelen utilizar este tipo de función de transferencia son las que se citan a continuación:

- Amplifica las señales poco intensas.

- Admite señales muy intensas, pero adaptándolas para que el ruido (componente aleatorio) no perturbe la respuesta
- Es fácilmente derivable:

$$f'(x) = \frac{-e^x}{1 + e^x} = f(x)(1 - f(x)) \quad (3.2.2)$$

El inconveniente básico de esta función de transferencia es que no sirve para expresar valores binarios, por lo que en estos casos se trabaja con la tangente hiperbólica, que tiene cualidades parecidas pero el rango de salida varía entre 1 y -1.

$$f(x) = \frac{1 - e^{2x}}{1 + e^{2x}}; f'(x) = (1 - f(x)^2) \quad (3.2.3)$$

### Características de las redes neuronales

Las principales características de las redes neuronales artificiales son las siguientes:

- **Capacidad de aprendizaje:**

Presentándoles patrones de entrada y, eventualmente, unos patrones de respuestas deseadas, la red es capaz de modificar su estructura para extraer las reglas que consigan que sus respuestas se asemejen a las deseadas. Sin embargo, estas reglas no se representan explícitamente, como en los sistemas simbólicos deductivos (o sistemas expertos), sino que quedan almacenadas en las conexiones de la red.

- **Capacidad de generalización:**

La red es capaz de procesar patrones de entrada distintos de los del proceso de aprendizaje. La respuesta dependerá del parecido de estos nuevos patrones con los de entrenamiento.

- **Capacidad de abstracción o de tolerancia al ruido:**

Las redes son capaces de abstraer las características esenciales del patrón de entrada que se le presenta, pudiendo procesar patrones incompletos o distorsionados.

- **Memoria asociativa:**

Las redes son capaces de almacenar patrones de forma que el conocimiento acumulado se halla distribuido por numerosas conexiones por lo que es tolerante a los fallos. Basta con presentarles una información parcial del patrón o, incluso, a pesar de sufrir lesiones en sus conexiones, la red puede seguir funcionando adecuadamente. Esto no ocurre con procesos en línea, donde un fallo puede interrumpir el desarrollo del programa, dar una respuesta incoherente o, bien, no dar respuesta.

### 3.2.4. Redes de retropropagación

Para estudiar el comportamiento de una red de retropropagación, vamos a considerar una red con tres capas: la capa de entrada, la capa oculta que contiene  $k$  unidades y la de salida con una neurona.

En la Figura 3.5 se representa, esquemáticamente, una red neuronal de tres capas. La capa de entrada se puede representar por el vector  $X$ , donde  $x_0$  es una entrada que siempre va a ser igual a 1 y lo que hace es proporcionar más

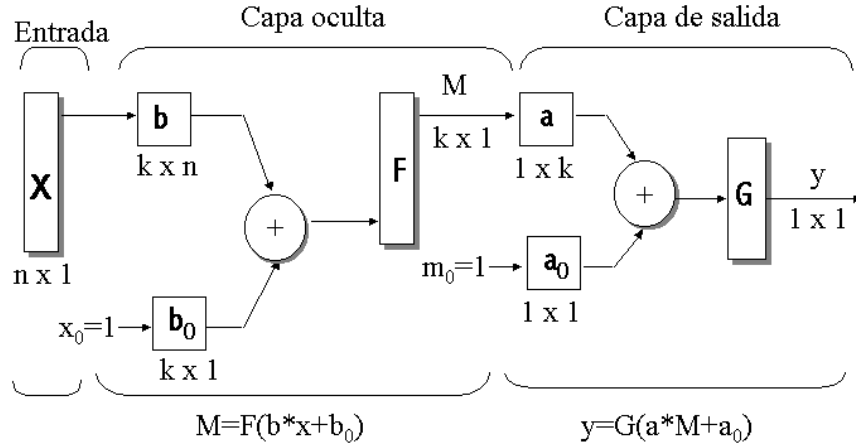


Figura 3.5: Red neuronal de tres capas

grados de libertad a la red para la posterior selección de las ponderaciones. A esta entrada se le denomina desviación para la primera capa.

La capa oculta se representa por el vector  $M = (m_1, \dots, m_k)'$ . La salida de la red está representada por  $y$ . Cada unidad de la capa oculta recibe la suma ponderada de todas las entradas más la desviación de la primera capa y produce la salida:

$$m_j = F\left(\sum b_{ji}x_i\right), \quad j = 1, 2, \dots, k \quad i = 0, 1, \dots, n \quad (3.2.4)$$

donde  $F$  es la función de transferencia de la capa oculta,  $x_i$  es la entrada  $i$ -ésima y  $b_{ji}$  es el peso de la conexión que va de la  $i$ -ésima unidad de entrada a la  $j$ -ésima unidad de la capa oculta.

De la misma forma, la unidad de salida recibe como señal la suma ponderada de las señales de salida de la capa intermedia y produce el resultado:

$$y = G\left(\sum a_j m_j\right), \quad j = 0, 1, 2, \dots, k \quad (3.2.5)$$

donde  $G$  es la función de transferencia de la capa de salida,  $a_j$  es el peso de la conexión que va desde la unidad  $j$ -ésima de la capa intermedia y donde la unidad

$m_0$  representa una desviación que siempre es igual a 1, por lo que obtenemos la siguiente representación algebraica de una red de retropropagación:

$$y = G(a_0 + \sum a_j F(\sum b_{ji} x_i)) \quad (3.2.6)$$

### Modo de operación

Como hemos indicado anteriormente, a la red se le deben presentar unos patrones para que ésta modifique las ponderaciones de las conexiones. A esta fase se le denomina proceso de entrenamiento o aprendizaje. Todo este proceso se realiza siguiendo una regla de aprendizaje. Posteriormente se puede aplicar una fase opcional que sería la de verificación, donde la red comprueba las ponderaciones alcanzadas en la fase anterior para los datos presentados en esta fase. Si la red no da una respuesta satisfactoria en esta fase, corregirá las ponderaciones en función de los inputs de esta fase. Finalmente, se implementa la fase de aplicación. Es en ésta última donde la red dará respuesta a los patrones de entrada en función de lo aprendido en el proceso de entrenamiento.

El aprendizaje puede ser supervisado, no supervisado o reforzado. Entre las redes con aprendizaje supervisado nos podemos encontrar las redes de retropropagación, mientras que de las de aprendizaje no supervisado podemos destacar las redes de *Kohonen*.

Independientemente del tipo de aprendizaje empleado, la red utiliza una regla de aprendizaje para modificar las ponderaciones usadas entre conexiones. Es aquí donde se ha investigado en mayor medida en los últimos años, ya que los analistas han intentado superar las limitaciones que tenían las anteriores reglas.

Las redes que emplean el aprendizaje supervisado [Figura 3.6] utilizan patrones

de entrada y salidas deseadas para modificar los pesos, por medio de la regla de aprendizaje, y de esta forma, lograr que la respuesta de la red sea lo más parecida a la salida deseada. La regla de aprendizaje, en el aprendizaje supervisado, debe seguir el siguiente proceso iterativo:

1. Tomar aleatoriamente los pesos de todas las conexiones.
2. Seleccionar el patrón de entrada y el patrón de salida deseada correspondiente.
3. Presentar a la red los patrones de entrada. La red obtendrá una salida en función de los pesos.
4. Calcular el error que se produce entre la salida deseada y la salida proporcionada por la red. Este error se mide como la suma de los cuadrados de las diferencias entre las salidas deseadas y obtenidas por la red para cada una de las unidades de salida de la red . Si el error es inferior a un cierto umbral el proceso finaliza.
5. Se aplica la regla de aprendizaje, para modificar los pesos de las conexiones y de esta manera reducir el error. Estos pesos varían en proporción a la derivada del error frente al peso.
6. Volver al paso 2.

La regla más utilizada actualmente en redes basadas en aprendizaje supervisado es la regla de retropropagación y la regla delta generalizada, que no es más que una extensión de la regla delta.



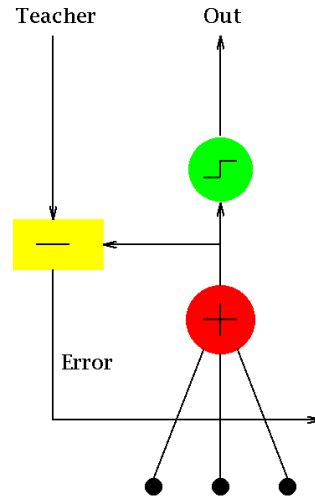


Figura 3.6: Aprendizaje supervisado

En la regla delta [Figura 3.7] la variación de cada peso se calcula sumándole la diferencia entre la salida deseada y la obtenida, multiplicada por la entrada y por un coeficiente de aprendizaje.

$$\Delta w = \eta \cdot x_i \cdot (T - y), \quad (3.2.7)$$

donde  $\eta$  es el coeficiente de aprendizaje,  $x_i$  es la entrada de la red,  $T$  es la salida deseada e  $y$  es la salida de la red. Hay que destacar que esta regla delta solamente se utiliza para redes de una sola capa.

En la Figura 3.8 podemos ver una representación gráfica de cómo funciona la regla delta para una red con dos inputs y una salida. Esta regla busca el mínimo local de la superficie del error cuadrático medio a través del gradiente descendente.

Sin embargo, el algoritmo más usado es el de retropropagación. Se trata de una regla para entrenar redes multicapa, lo que dio un impulso definitivo a la utilización de redes neuronales, ya que las redes con una sola capa estaban muy

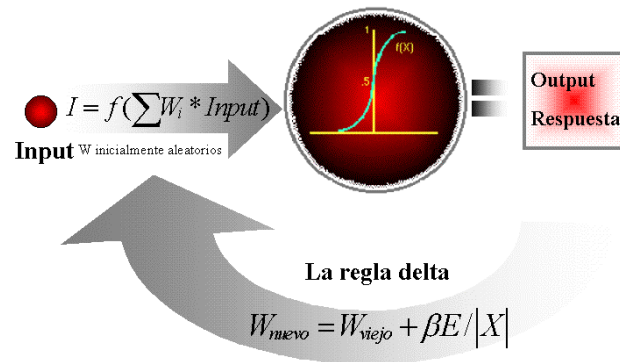


Figura 3.7: La regla delta

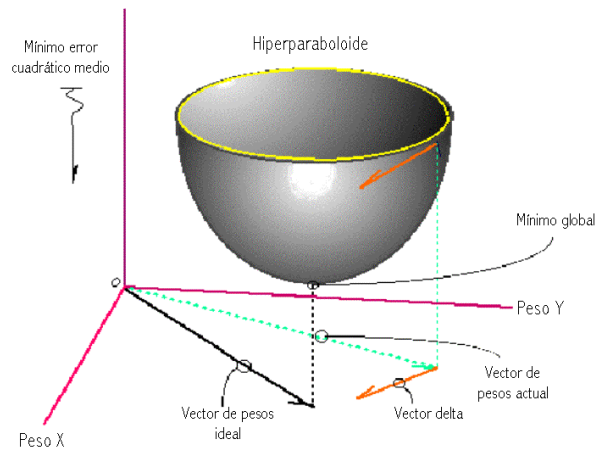


Figura 3.8: La regla delta

limitadas en cuanto a su aplicación. Esta regla fue descubierta de forma separada por investigadores de distintas disciplinas como Werbos (1974), Parker (1982) y Rumelhart *et al.* (1986). Esto da a entender el carácter multidisciplinar que tiene el campo de las redes neuronales y que provoca que los avances hechos en un campo no sean aplicados después en otros.

La regla de retropropagación mide los errores que existen entre las salidas de la red y los valores reales y, a través de una técnica de vector gradiente descendente, intenta ajustar los parámetros con el fin de encontrar el mínimo error cuadrático

medio.

Pero para poder aplicar este algoritmo es necesario que la red cumpla una serie de requisitos:

- La red no debe ser recurrente.
- No puede contener funciones de activación no derivables.
- Es aconsejable utilizar la función de activación sigmoideal, ya que es capaz de recoger no linealidades subyacentes teniendo una gran capacidad de representación y tiene derivadas fáciles de calcular.
- Hay que tener en cuenta que todas las unidades deben estar interconectadas con las unidades de la siguiente capa.

Para una explicación más detallada de este algoritmo utilizaremos la red diseñada en la Figura 3.5.

Al presentar un patrón de entrada  $\mu$ , la señal que le llega a la unidad  $j$  de la capa oculta es igual a

$$n1_j^\mu = \sum_k b_{jk} \cdot x_k^\mu \quad (3.2.8)$$

Tras aplicarle la función de activación a la señal, la salida producida es igual a  $m_j^\mu = F(n1_j^\mu)$ . La señal que le llega a las neuronas de la capa de salida es igual a  $n2_j^\mu = \sum_j a_{ij} \cdot m_j^\mu$  que produce la salida  $y_i^\mu = G(n2_i^\mu)$ .

La función de pérdida, error cuadrático medio de los errores, viene dada por

$$E[a, b] = \frac{1}{2} \sum_{\mu, i} (t_i^\mu - y_i^\mu)^2, \quad (3.2.9)$$

donde  $t_i^\mu$  e  $y_i^\mu$  son, respectivamente, la salida deseada y la salida de la red para la neurona  $i$  de la capa de salida al presentarle el patrón  $\mu$  como entrada de la

red. Por tanto, si aplicamos los resultados anteriores a [3.2.9] el error se puede presentar como:

$$E[a, b] = \frac{1}{2} \sum_{\mu, i} \left( t_i^\mu - G \left( \sum_j a_{ij} \cdot F \left( \sum_k b_{jk} \cdot x_k^\mu \right) \right) \right)^2. \quad (3.2.10)$$

Así, la función de pérdidas depende de los pesos de las conexiones de la red. Es continua y derivable en función de los pesos, por lo que se puede aplicar el método del gradiente descendente para hacer el reajuste de las ponderaciones.

Para la matriz  $a$ , el reajuste se haría siguiendo la siguiente regla:

$$\Delta a_{ij} = -\eta \frac{\partial E}{\partial a_{ij}}, \quad (3.2.11)$$

donde  $\eta$  es el coeficiente de aprendizaje.

Aplicando la regla de la cadena a la derivada parcial obtendríamos

$$\frac{\partial E}{\partial a_{ij}} = \sum_{\mu} \frac{\partial E}{\partial y_i^\mu} \cdot \frac{\partial E}{\partial y_i^\mu} \cdot \frac{\partial y_i^\mu}{\partial n2_j^\mu} \cdot \frac{\partial n2_j^\mu}{\partial a_{ij}} \quad (3.2.12)$$

por lo que

$$\frac{\partial E}{\partial a_{ij}} = - \sum_{\mu} (t_i^\mu - y_i^{\mu}) \cdot G'(n2_i^\mu). \quad (3.2.13)$$

Por consiguiente, la variación en los pesos de las conexiones entre la capa oculta y la capa de salida [Ecuación 3.2.11], se puede describir de la siguiente forma:

$$\Delta a_{ij} = \eta \sum \delta 2_i^\mu \cdot m_j^\mu, \quad (3.2.14)$$

donde  $\delta 2_i^\mu$  se ha definido como  $\delta 2_i^\mu = (t_i^\mu - y_i^{\mu}) \cdot G'(n2_i^\mu)$ .

Para la matriz de ponderaciones  $b$ , la variación en los pesos es viene dada por la siguiente expresión

$$\Delta b_{jk} = -\eta \frac{\partial E}{\partial b_{jk}}. \quad (3.2.15)$$

Aplicando nuevamente la regla de la cadena, pero esta vez para la matriz  $b$ , la derivada queda como sigue:

$$\frac{\partial E}{\partial b_{jk}} = \sum_{\mu} \frac{\partial E}{\partial y_i^{\mu}} \cdot \frac{\partial E}{\partial y_i^{\mu}} \cdot \frac{\partial y_i^{\mu}}{\partial n_2^{\mu}} \cdot \frac{\partial n_2^{\mu}}{\partial m_j^{\mu}} \cdot \frac{\partial m_j^{\mu}}{\partial n_1^{\mu}} \cdot \frac{\partial n_1^{\mu}}{\partial b_{jk}} \quad (3.2.16)$$

por lo que

$$\frac{\partial E}{\partial b_{jk}} = \sum_{\mu, i} (t_i^{\mu} - y_i^{\mu}) \cdot G'(n_2^{\mu}) \cdot a_{ij} \cdot F'(n_1^{\mu}) \cdot x_k^{\mu}. \quad (3.2.17)$$

Así pues, la variación de los pesos de las conexiones entre la capa de entrada y la capa oculta [Ecuación 3.2.15] es igual a:

$$\Delta b_{jk} = \eta \sum \delta 1_j^{\mu} \cdot x_k^{\mu}, \quad (3.2.18)$$

donde se ha definido:  $\delta 1_j^{\mu} = F'(n_1^{\mu}) \sum_i b_{ij} \cdot \delta 2_i^{\mu}$ .

Como se puede observar, la función de los términos  $\delta$  es la retroalimentación del error para producir una adecuada variación de las ponderaciones. La presentación de los patrones de entrada y salida deseada no tiene por qué seguir un orden fijo, sino que puede presentarse de forma aleatoria. Cuando se ha presentado el conjunto total de parejas de entradas y salidas deseadas, la red habrá completado un ciclo o iteración. Por consiguiente, la red debe completar un número de iteraciones para que el valor del error sea aceptablemente pequeño.

### Problemas con el algoritmo

Sin embargo, este algoritmo presenta varios problemas, entre los que destacan los siguientes:

- Su dudosa plausibilidad neurofisiológica.

- Las salidas de la red varían en función de las ponderaciones iniciales, por lo que, es necesario entrenar la red varias veces para elegir los pesos que den una mejor respuesta.
- El tiempo de computación en ocasiones es increíblemente alto, incluso para problemas sencillos.
- La incorporación de nuevos ejemplos de aprendizaje puede provocar un empeoramiento del rendimiento del sistema.
- Cuando los pesos quedan ajustados a valores muy altos, entonces las unidades dan respuestas muy próximas a 1, con lo que el gradiente tiende a 0.

En consecuencia no se producen variaciones en las ponderaciones y el aprendizaje queda detenido. Por eso es conveniente inicializar los pesos a valores pequeños y utilizar una tasa de aprendizaje también pequeña, aunque el aprendizaje sea más lento.

Si usamos una tasa de aprendizaje alta, los incrementos en los pesos serán, también, elevados. Esto puede provocar que en la siguiente iteración la red intente compensar ese exceso de incremento, provocando, de esta forma, oscilaciones continuas en los pesos. Para solucionar este problema se utilizan tasas pequeñas o bien, tasas adaptativas, que aumenten si el error disminuye o *viceversa*. Entre las técnicas más utilizadas dentro de las reglas adaptativas cabe destacar la del momento, creada por Plaut *et al.* (1986), que añade un momento en el cálculo de la variación de las ponderaciones, de forma que se logra darle cierta inercia a los pesos para que tiendan a cambiar en la dirección en la que se ha producido

mayor descenso en la superficie del error, de esta forma la nueva regla viene dada por la siguiente expresión:

$$\Delta w_{ij}(t+1) = -\eta \frac{\partial E}{\partial w_{ij}} + \xi \cdot \Delta w_{ij}(t), \quad \xi \in [0, 1]. \quad (3.2.19)$$

Sin embargo, el problema más serio de este algoritmo es su tendencia a caer en los mínimos locales. Al seguir la superficie del error en busca de un valle, suele estancarse en valles que no son mínimos globales. Esto no implica que no de resultados satisfactorios, aunque se ha intentado superar estas limitaciones empleando otros algoritmos como son el modo de operación asíncrono o probabilístico, el uso de métodos estadísticos, el equilibrio termodinámico simulado o añadir cierto nivel de ruido a las variaciones en los pesos o las conexiones.

### 3.3. Algoritmos genéticos

Los algoritmos basados en la evolución genética, entre los que se encuentran los algoritmos genéticos y la programación genética, intentan mimetizar los procesos optimizadores que se producen en la evolución de las especies. La naturaleza crea las múltiples estructuras vivas que pueblan la Tierra por medio de la selección natural, selección que se realiza por medio de la adaptación de estas estructuras al medio que les rodea. Dicha adaptación se produce por medio de complejos procesos evolutivos que tienen carácter incierto y no determinista. En ellos se involucran numerosas actividades asincrónicas, descoordinadas, e independientes que no están ni controladas centralmente ni orquestadas.

Como señala Koza (1992), la adaptación y no la parsimonia es el factor dominante en la evolución natural. La parsimonia desempeña algún papel sólo cuando interfiere en la adaptación. No existe evidencia en la naturaleza a favor de la parsimonia.

Si quisiéramos realizar una descripción en términos matemáticos de la evolución habría que concebirla como un proceso de optimización donde la función objetivo que se desea maximizar no es otra que la supervivencia, no del individuo sino de la especie. Pero la evolución es un proceso continuo, que no tiene un punto terminal bien definido, por lo que habría que asociarla entonces con un proceso de optimización dinámica, nunca estática.

Se requieren cuatro condiciones para que se produzcan los procesos evolutivos dentro de la naturaleza:

- La existencia de un ente que tenga la capacidad de reproducirse.



- La existencia de una población de dichos entes autorreproductores.
- La existencia de algún tipo de diversidad entre los entes autorreproductores.
- La existencia de alguna diferencia en la habilidad para sobrevivir en el medio ambiente asociada con la diversidad.

Por sorprendente que parezca, los procesos evolutivos de la naturaleza se producen igualmente en diversos ámbitos de la evolución de las sociedades. En particular no puede decirse que tales procesos puedan considerarse ajenos a la evolución de las instituciones económicas. Obsérvese, por ejemplo, el paralelismo que puede establecerse entre diversos procesos biológicos y económicos tales como la diversidad genética y la diversificación económica; la selección natural darwiniana y la competencia; la reproducción sexual y la fusión entre empresas. Incluso, el denominado proceso de destrucción creativa de Schumpeter puede asimilarse a un proceso evolutivo en el sentido biológico.

En esta sección introduciremos los conceptos relacionados con la evolución natural y cómo se desarrollan los procesos inherentes a la selección natural para poder extrapolarlos a los algoritmos basados en la evolución genética. A continuación, describiremos los algoritmos genéticos y cómo implementarlos en la obtención de óptimos en problemas de optimización.

### **3.3.1. Evolución natural**

La teoría de la evolución de las especies (Darwin, 1859) introdujo en el mundo científico de la era victoriana el concepto de evolución a través de la selección natural. Dió base científica a la idea de que la especie debe prevalecer sobre el individuo.

La naturaleza garantiza, mediante el mecanismo de la selección natural, el proceso evolutivo por el que se perpetúa la especie. Pero de forma indirecta, la selección natural actúa también sobre la información genética de cada uno de los individuos de una especie. Esta información genética está contenida en el *genoma* y codificada en el Ácido Desoxirribonucleico (ADN). En el genoma o genotipo están almacenadas las claves para la diferenciación de las células que forman los diferentes tejidos y órganos de un individuo. Parte de esa información pasa de padres a hijos a través de la herencia y constituye el principal mecanismo de variación dentro de la población. De esta forma, la diversidad se manifiesta en los cromosomas de las células de los seres vivos condicionando el comportamiento de los entes en su medio. Las variaciones en estructura y comportamiento se traducen, durante el proceso evolutivo, en diferencias en la tasa de supervivencia y reproducción.

Las propiedades visibles de un individuo (es decir, su apariencia) recibe el nombre de *fenotipo*. La selección natural actúa sobre el fenotipo ya que elimina a los individuos menos adaptados, es decir, aquellos cuyas propiedades o fenotipo no se adecúen apropiadamente al entorno en el que se desarrolla.

Finalmente, la *ontogenia* es el enlace entre éstos y el entorno, ya que representa el cambio del individuo desde la fertilización hasta la madurez, por lo que la ontogenia no debe considerarse evolución, ya que ésta no actúa sobre los individuos.

Los cambios en poblaciones, que consideramos evolutivos, son aquéllos que se heredan por medio del material genético de una generación a la siguiente (Futuyma, 1986). Es decir, cambios en el genotipo pueden provocar cambios en

las características de los individuos de una población, siempre y cuando éstos se hereden a la siguiente generación y se difundan a lo largo de varias generaciones.

Pero el proceso inverso no se da: cambios en la apariencia de un individuo no se transmite a sus descendiente, aunque a menudo el cambio genético es inferido a partir de cambios fenotípicos que son heredables.

Veamos con un sencillo ejemplo las implicaciones anteriores.

Consideremos una población de individuos localizados en un entorno en el que deben correr para librarse de un depredador natural. Supongamos que un individuo  $X$  posee un código genético que le ha permitido poseer garras en las patas con las que poder defenderse. Este individuo tiene más posibilidades de sobrevivir que el resto de la población. Al aparearse con otros individuos transmite parte de esa información a sus hijos, parte de los cuales tendrán garras con los que serán éstos los que tienen más posibilidades de transmitir su información genética. Es así como funciona la evolución.

Pero cambios en el individuo no se transmiten a sus descendientes. En el mismo ejemplo, si un individuo desarrollase patas más fuertes por el ejercicio provocado de huir constantemente del depredador, ésta cualidad no se transmitirá a sus hijos. Solamente lo haría si se debiese a un cambio genético.

Una definición más concreta de evolución la podemos encontrar en (Curtis y Barnes, 1989, pág. 974): “La evolución puede definirse cabalmente como cualquier cambio en la frecuencia de alelos dentro de un acervo genético entre una generación y la siguiente”, por lo que si se ha detectado un cambio en la frecuencia de genes en una población se dice que se ha observado una evolución.

Todos estos cambios se producen a nivel microbiológico (es decir, a nivel de

ADN). Es necesario describir el ADN y sus componentes, ya que parte de la nomenclatura usada en los algoritmos evolutivos proviene de su paralelismo con sus contrapartidas biológicas.

### El ADN

El ADN es una molécula polimérica, relativamente simple, que está formada por una cadena doble compuesta por nucleótidos bases que están químicamente enlazados. Cada uno de estos nucleótidos posee una base nitrogenada que únicamente puede estar constituida por cuatro bases diferentes: la Adenina, la Timina, la Citocina y la Guanina. Además estas bases sólo se pueden enlazar con unos pares determinados [Figura 3.9].

- Citocina con Guanina
- Adenina con Timina

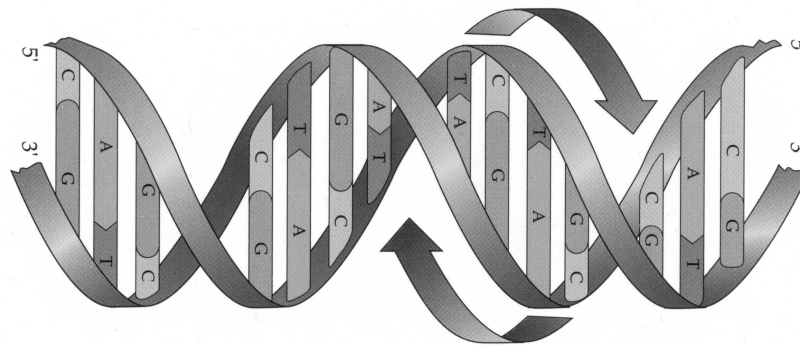


Figura 3.9: Estructura del ADN

De esta forma, al unirse los dos ejes, coinciden en el mismo plano las bases nitrogenadas formando los escalones de una escalera helicoidal o de doble hélice [Figura 3.10].

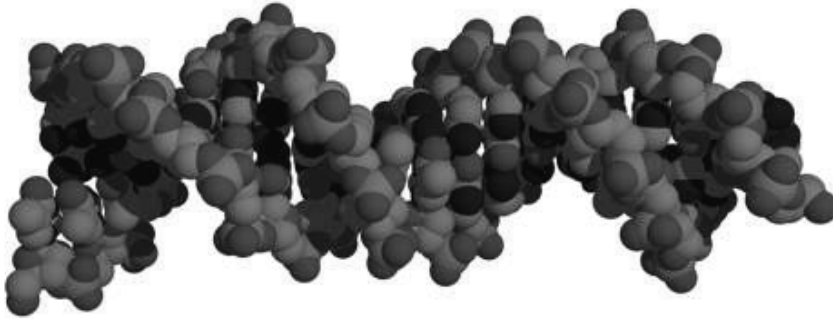


Figura 3.10: Doble hélice del ADN

La secuencia de bases nitrogenadas en uno de los ejes de la cadena, en grupo de tres, se denomina *codon*. Cada uno de ellos determina la información para la creación de un aminoácido, que constituyen los elementos principales de las proteínas. El otro eje, que será complementario al primero por las características de los enlaces, será el medidor que permite la duplicación fiel de la información.

Esta repetición de la información contenida en el ADN permite que, a pesar de la pérdida de una parte de este, se pueda reconstruir su contenido.

¿Pero cuáles son los elementos que dan las características al individuo? ¿Qué es lo que marca el color de los ojos de una persona? La respuesta está en el *gen*. Aunque existen definiciones mucho más precisas, puede decirse que el gen da los atributos específicos de los individuos, que de alguna forma están codificados en el cromosoma del ADN. Los valores particulares que puede adquirir el gen se denominan *alelos*. El alelo recoge las distintas posibilidades que puede tomar el gen. Aunque hay acepciones que definen el gen tanto como la posición que ocupa éste en el cromosoma y como el atributo de una característica específica, otros enfoques diferencian entre el gen y la posición.

Las características, que están codificadas en los genes, se transmiten de padres a hijos. Los padres deben transmitir a su descendencia nueva información que sea útil a efectos de engendrar individuos mejores adaptados al entorno. La naturaleza debe garantizar que parte de la información contenida en los padres pase a su descendencia para no perder las buenas cualidades del progenitor, que han logrado que no sea eliminado en el proceso evolutivo. De esta forma el hijo también tendrá altas posibilidades de sobrevivir en su entorno. Pero, de igual forma, también se debe garantizar cierta variabilidad en los individuos, que al fin y al cabo es el resultado de la evolución. Los mecanismos que intervienen para garantizar la estabilidad, pero también la variabilidad de la población son, principalmente, la recombinación sexual homóloga y la mutación. Existen otros mecanismos a nivel biológico, pero nos centraremos en los mencionados anteriormente, ya que son los que se implementan en el desarrollo de los algoritmos evolutivos.

- **Recombinación sexual homóloga**

La naturaleza utiliza una estrategia optimizadora encaminada hacia la supervivencia de las especies. Tal estrategia se llama reproducción sexual. Se trata de un proceso por el que los caracteres genéticos de un individuo constituyen una mezcla aleatoria de los de sus progenitores. La reproducción sexual tiene la ventaja de contribuir, de forma extraordinaria, a la diversidad de individuos de las poblaciones. La reproducción sexual implica, por tanto, que la selección natural se realice sobre una amplia diversidad de caracteres, de modo que una buena adaptación al medio por parte de alguno de los individuos siempre quede garantizada. El carácter

aleatorio de la transmisión de caracteres en la reproducción sexual y las primeras leyes de la genética fueron descubiertas originalmente por Mendel (1865).

La reproducción sexual constituye un mecanismo imprescindible para que se de la evolución ya que introduce variabilidad en la población pero a su vez garantiza cierta estabilidad en los cambios. Evita que mutaciones perjudiciales para la adaptación del individuo se transmita a la información genética de su descendencia (Manynard-Smith, 1994), pero también logra reparar segmentos de ADN dañado (Watson *et al.*, 1987). Pero como se ha dicho, la recombinación sexual también introduce variaciones en los nuevos individuos, ya que sobre la base de los ADNs paternos se crea un hijo con información genética completamente distinta. Pero para ello la recombinación sexual debe ser homóloga. Es decir, tiene que ocurrir entre segmentos iguales o parecidos de la cadena de ADN, ya que debe preservarse la funcionalidad del gen, y debe producirse en una posición que de lugar al mismo resultado funcional. Como veremos posteriormente Banzhaf *et al.* (1998) desarrolló un procedimiento para simular la recombinación homóloga en la programación genética.

#### ■ La mutación

Constituye la principal fuente de variabilidad en el código genético. Puede estar provocada por cambios en las bases, la desaparición o agregación de nuevas bases, o realineamientos de segmentos de ADN, aunque estas últimas suelen considerarse perjudiciales, ya que pueden provocar cambios drásticos en la funcionalidad de los genes. Por último, hay que considerar que la

recombinación sexual no homóloga es en realidad una macromutación de enormes consecuencias.

Es importante recalcar que los algoritmos evolutivos no intentan copiar el proceso evolutivo, sólo se inspiran en él. De ahí que la mayoría de las acepciones utilizadas en biología se mantengan al definir conceptos encuadrados dentro de los algoritmos evolutivos.

### **3.3.2. Algoritmos Genéticos**

Holland (1962, 1975) mostró cómo los procesos evolutivos pueden ser empleados en la optimización de los sistemas artificiales. Todo problema de adaptación puede ser formulado en términos genéticos y, a menudo, resuelto por una técnica de optimización que se conoce como Algoritmo Genético. Tales algoritmos simulan los procesos evolutivos darwinianos y las operaciones genéticas que se producen en la naturaleza dentro de los cromosomas y constituyen una forma de búsqueda adaptativa, mediante técnicas de optimización basadas en la evolución natural. A partir de los algoritmos genéticos se han desarrollado otros algoritmos evolutivos que incluyen las estrategias evolutivas (Rechenberg, 1973; Schwefel, 1981), programación evolutiva (Fogel *et al.*, 1966), sistemas de clasificación (Holland, 1976, 1980) y programación genética (Koza, 1992).

Los algoritmos genéticos utilizan un sencillo proceso iterativo que logra que el algoritmo alcance una solución satisfactoria y que intenta simular, de alguna manera, el proceso de selección natural que tiene lugar en la naturaleza. Basándose en una búsqueda prometedora utilizan una población de individuos que representan posibles soluciones del problema a optimizar. Se evalúan los



individuos de acuerdo a una función de pérdidas (también llamada función de adaptación), que define el entorno para la evolución. De entre los candidatos se seleccionan aquellos más adaptados, es decir, aquellos con menor valor en la función de pérdidas. Los individuos seleccionados se cruzan entre ellos y se mutan para crear una nueva población, que es la que sigue con el proceso iterativo de selección, cruce y mutación. Estos operadores son los que establecen las diferencias entre unos algoritmos y otros. Este proceso iterativo continúa hasta lograr una solución satisfactoria.

La principal característica de los algoritmos genéticos es que en la codificación de los individuos se utilizan estructuras de tamaño fijo. Los elementos de la población son codificados como cadenas de caracteres en un alfabeto binario (0 y 1). Dichas cadenas de caracteres binarios suelen denominarse *cromosomas*.

El proceso empleado por un AG está fundamentado en tres operaciones biológicas básicas: la *selección natural*, la *recombinación* y la *mutación*. Estas operaciones pueden ser consideradas como un procedimiento efectivo de búsqueda mediante la actualización de los candidatos a óptimo de cada generación a la siguiente. De cara a la aplicación de estas operaciones biológicas en los procesos matemáticos de optimización, es necesario definir dos nuevas operaciones relativas a la creación de la *población inicial* y al *emparejamiento* entre los cromosomas. La secuencia de operaciones que realiza un AG es como se describe en la Figura 3.11.

Existen numerosos tipos de AGs dependiendo del mecanismo de selección y las operaciones de recombinación. Los AGs evalúan la adaptación de los cromosomas creados en la nueva generación y reemplazan los miembros menos adaptados

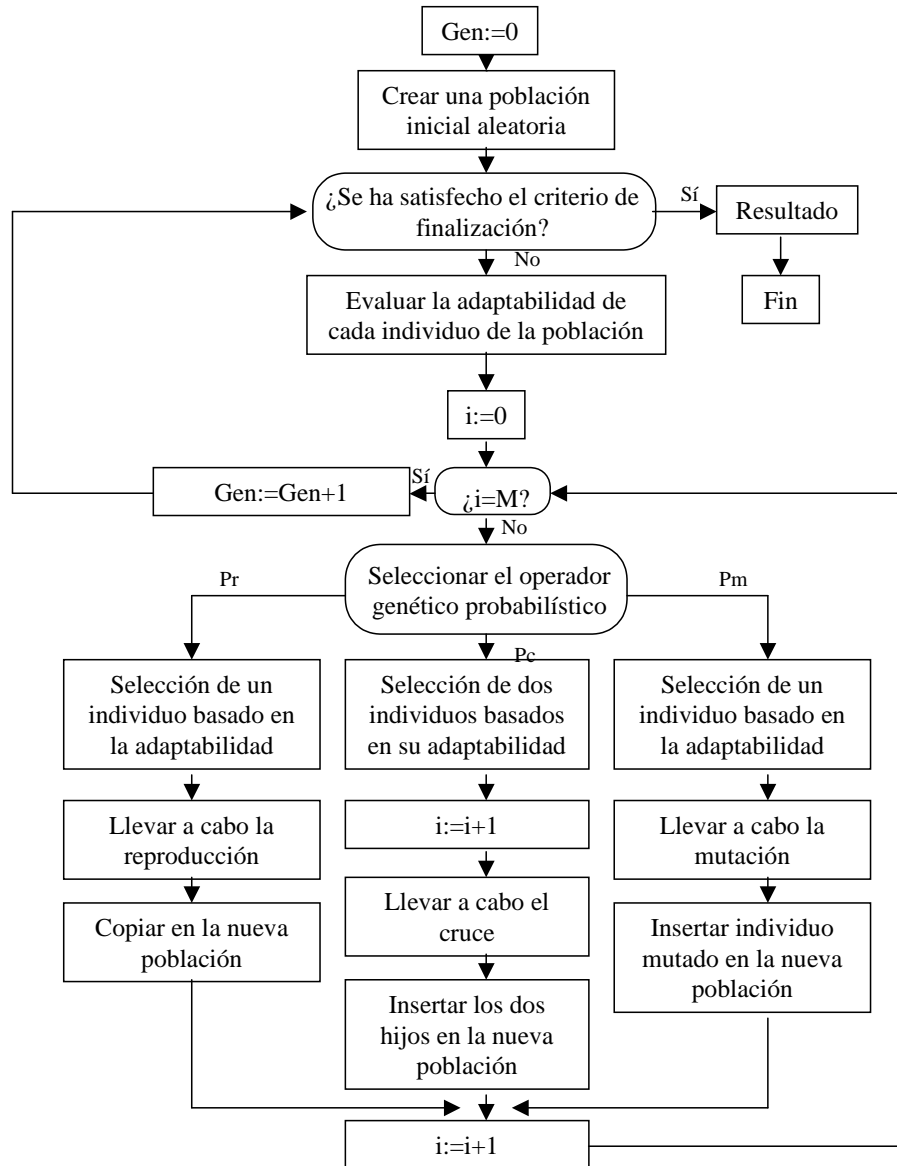


Figura 3.11: Diagrama de flujo de un AG donde “M” es igual al tamaño de la población.

por nuevos cromosomas obtenidos mediante recombinación y mutación de los mejor adaptados. Se crean, por este procedimiento, generaciones sucesivas de nuevos cromosomas hasta que se satisface un criterio de parada establecido de antemano. En este proceso, el cromosoma más adaptado de la última generación es considerado como la solución al problema de optimización.

De forma esquemática, un AG puede describirse por medio de la siguiente estructura secuencial de operaciones:

1. Crear una población inicial de cromosomas elegidos al azar.
2. Formar un *ranking* de dichos cromosomas de acuerdo a una determinada función objetivo.
3. Seleccionar los mejores candidatos para la recombinación.
4. Efectuar las operaciones de recombinación y mutación de los candidatos seleccionados.
5. Formar una nueva generación con los cromosomas mejor adaptados respecto a la función objetivo y su descendencia.
6. Realizar un nuevo *ranking* en la nueva generación.
7. Volver al paso 3 si no se satisface un determinado criterio de convergencia.

Existen diversas diferencias importantes entre los AG y otros métodos de optimización. Las principales ventajas de los AG pueden resumirse como sigue (Véase Haupt y Haupt, 1998):

- No se requiere ni continuidad ni derivabilidad de la función objetivo a optimizar.
- La optimización se realiza tanto para funciones de variables continuas o discretas.
- Se emplean exclusivamente valores de la función objetivo y no información relativa a sus derivadas.
- Es posible optimizar, sin dificultad, funciones de un gran número de variables.
- Al tratarse de un proceso de optimización aleatorio que emplea reglas probabilísticas de recombinación y mutación, el algoritmo puede salir, con facilidad, de los óptimos locales cuando se trata de una función objetivo muy compleja.
- Los AG pueden trabajar en contextos de datos generados numéricamente, datos experimentales o funciones analíticas.

No obstante, se pueden señalar ciertas dificultades potenciales y limitaciones de los AG como técnicas de optimización matemática, que señalamos a continuación.

- Es necesario elegir, *a priori*, diferentes características técnicas de los AG tales como el número de cromosomas, unos procedimientos de cruce y de mutación específicos, etc.
- Aunque el “Teorema de esquemas” (*schema theorem*) que propuso Goldberg (1989) asegura que los AGs sitúan automáticamente un número

especialmente creciente de cromosomas en las regiones más prometedoras del espacio de búsqueda, en general no hay un resultado teórico que garantice la convergencia de un AG al óptimo global.

- El proceso de convergencia hacia el óptimo de un AG puede suponer numerosas evaluaciones de la función objetivo obligando, por consiguiente, al empleo de mucho tiempo de ejecución.

Las dificultades y el éxito de los AG dependen mucho de la naturaleza del problema. Aunque no hay garantía absoluta de encontrar un óptimo global, un AG es capaz de encontrar una solución aceptable, con relativa rapidez, en una amplia variedad de problemas. Finalmente es preciso señalar que los AG han sido aplicados a una gran variedad de campos científicos. Dentro de la Economía destacan las aplicaciones a problemas en Finanzas e Inversión (Bauer, 1994), aprendizaje (Montana y Davis, 1989), control óptimo (Holly y Hughes-Hallet, 1989), estimación econométrica Dorsey y Mayer (1995), Análisis Técnico Neely *et al.* (1997); Allen y Karjalainen (1999), marketing Michalewicz (1992) y en otras áreas relacionadas con la investigación operativa, la organización y la planificación Downsland y Downsland (1992); Venogopal y Narendan (1992).

### **Estructura de los algoritmos genéticos**

Los algoritmos genéticos utilizan vectores de 0 y 1 para representar los distintos puntos del espacio de soluciones. Con la característica añadida de que el vector tiene la misma dimensión para todos los puntos. A estos vectores se les denomina cromosomas y cada dígito binario se le denomina *bit*. Por ello el paso previo a la utilización del algoritmo es codificar las posibles soluciones a código

binario para crear la estructura necesaria a la que poder aplicarle los operadores genéticos.

Supongamos un problema sencillo que consiste en optimizar una función con un sólo parámetro. Habrá que definir el rango en el que se encuentra el óptimo del problema, definiendo  $vmax$  el valor máximo que puede alcanzar el parámetro y  $vmin$  su valor mínimo. El óptimo debe encontrarse dentro de este rango de valores. El tamaño del cromosoma vendrá dado por la ecuación [3.3.1]

$$tc = \text{int} \frac{\ln(vmax - vmin + 1)}{\ln 2}, \quad (3.3.1)$$

siendo  $tc$  el tamaño del cromosoma e “int” la función que recoge la parte entera de un número.

Así, si el valor máximo del parámetro es 10 y 0 su valor mínimo, el tamaño del cromosoma debe ser de cuatro dígitos o *bits*, que tomarán valores 0 o 1. Siguiendo con el ejemplo, si el óptimo de esa función fuese ocho, [1000] sería el cromosoma que lo representa.

La Ecuación 3.3.2 permite decodificar un cromosoma  $[x_1, x_2, \dots, x_{tc}]$  de tamaño  $tc$ , en valores decimales para su posterior evaluación en la función.

$$x_1 \cdot 2^{tc} + x_2 \cdot 2^{tc-1} + \dots + x_{tc-1} \cdot 2^1 + x_{tc} \cdot 2^0 \quad (3.3.2)$$

Si una función depende de varios parámetros, el cromosoma tendrá un tamaño igual a la suma de dígitos que deben representar a cada parámetro. Sea un problema con  $N_{par}$  parámetros a optimizar, las soluciones se podrán representar como un vector  $[p_1, p_2, \dots, p_{N_{par}}]$ . Si  $tc_i$ ,  $i = 1, \dots, N_{par}$  es el número de bits necesario para representar el parámetro  $i$ , el cromosoma codificado tendrá un

tamaño de  $\sum_i^{N_{par}} tc_i$  y se representará como:

$$[x_1, \dots, x_{tc_1}, x_{tc_1+1}, \dots, x_{tc_1+tc_2}, \dots, x_{\sum_i tc_i}]$$

Una vez definido el largo de los cromosomas, se genera aleatoriamente una población inicial de cromosomas que constituyen cadenas de dígitos binarios elegidos al azar, de modo que la probabilidad de obtener un cero sea la misma que la de obtener un 1. Una vez decodificados a código decimal, estos cromosomas representan vectores de parámetros generados aleatoriamente y donde cada uno de ellos está comprendido entre los rangos de valores  $vmax$  y  $vmin$  previamente definidos.

El número de individuos que conforman la población es un parámetro vital en los algoritmos genéticos, ya que de elegir pocos individuos el espacio de búsqueda queda muy limitado y puede que existan puntos que queden sin explorar. Por otra parte, si la población es muy amplia puede ralentizar el tiempo de computación.

Una vez inicializada la población se evalúan los individuos en la función de pérdida que, debido a la similitud con el proceso de selección natural, se ha convenido en llamarla función de adaptación. Esta función no tiene porqué cumplir con requisitos que sí son exigidos en otros métodos de optimización, como puede ser la continuidad en las derivadas, la convexidad de la función, etc.

Posteriormente, los mejores individuos, los más adaptados, son seleccionados para ser mutados y cruzados con otros individuos.

## Selección

Una vez definida la función de adaptación hay que establecer cuáles de esos individuos son seleccionados para pasar a la siguiente iteración: la selección. El método de selección influye de forma determinante en la velocidad de convergencia del algoritmo. Una rápida convergencia puede provocar un algoritmo ineficiente y el estancamiento en soluciones no óptimas. Sin embargo, aunque hay diferentes métodos de selección (Goldberg, 1989), la idea básica en cada uno de ellos consiste en elegir a los candidatos de acuerdo a una distribución de probabilidad que debe estar sesgada hacia los individuos relativamente mejor adaptados. A continuación veremos cuáles son los métodos más usados como criterio de selección dentro de los algoritmos evolutivos.

- **Selección por adaptación ponderada.**

Se crea una probabilidad para cada individuo  $i$  y que viene dada por

$$p_i = \frac{f_i}{\sum_j f_j}$$

donde  $pob$  representa el total de individuos que hay en la población y  $f_j$  es el valor de la función de adaptación del  $j$ -ésimo individuo. La población se ordena en función de la adaptación de cada individuo. Se generan números aleatorios y se seleccionan aquellos individuos cuya probabilidad acumulada sea inmediatamente superior a los números aleatorios generados.

En el Cuadro 3.1 se muestra un ejemplo con una población de 6 cromosomas, junto al valor alcanzado por cada uno de ellos en la función de adaptación,



Cromosoma	$f_i$	$P_i$	$\sum_{i=1}^6 p_i$
1	20.778	0.3518	0.3518
2	12.36	0.2092	0.5611
3	10.338	0.1750	0.7361
4	9.255	0.1567	0.8928
5	5.166	0.0875	0.9803
6	1.164	0.0197	1

Cuadro 3.1: Ranking de los cromosomas ordenados en función de su adaptación.

que en este ejemplo coincide con el rendimiento obtenido por cada uno de los cromosomas en un tramo de una serie financiera.

Se generan seis números aleatorios: 0.1535, 0.6781, 0.0872, 0.1936, 0.7021, 0.3933. Siguiendo estos números, los cromosomas seleccionados para el cruce y la mutación son: cromosoma 1, cromosoma 3, cromosoma 1, cromosoma 1, cromosoma 3 y cromosoma 2.

#### ■ Selección por ruleta.

Es un caso especial del método de selección estudiado previamente. Se calcula la media de los valores de la función de adaptación de todos los individuos de la población. Este dato es necesario ya que un individuo tendrá tantas copias en la población seleccionada como número de veces supere el valor de la función objetivo de ese individuo a la media.

$$NC_i = \text{int} \left( \frac{f_i}{\text{media}} \right) \quad (3.3.3)$$

siendo *int* la parte entera de la fracción. Los individuos que no superen la media son eliminados.

Como se puede observar, ambos casos se puede asemejar a una ruleta [Figura 3.12] donde cada individuo tiene una sección circular directamente proporcional a su adaptabilidad en el primer caso y al número de veces que supera su adaptabilidad a la media de la población en el segundo caso.

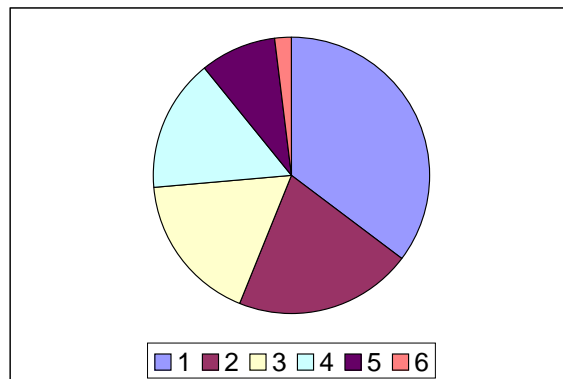


Figura 3.12: Selección por adaptación ponderada

- **Truncamiento o selección basada ( $\mu, \lambda$ )**

En este método se seleccionan los mejores  $\mu$  individuos de la población. Estos participarán en los operadores de reproducción, cruce y mutación para generar  $\lambda$  individuos que conforman la descendencia. Dentro de estos se seleccionan, otra vez, los  $\mu$  mejores individuos que constituirán los padres de la siguiente generación.

Una variante es la selección ( $\mu + \lambda$ ) donde los padres participan junto a su descendencia en el proceso de selección

- **Selección por Ranking**

En este caso se asigna a cada individuo una probabilidad pero en este caso, dicha probabilidad está asociada al rango que ocupa el individuo en la

ordenación que se hace de cada uno de ellos en función de su valor en la función de adaptación.

Se puede usar un ranking lineal o exponencial. En el caso lineal a cada individuo se le asigna la siguiente probabilidad:

$$p_i = \frac{1}{pob} \left[ p^- + (p^+ - p^-) \frac{i-1}{N-1} \right] \quad (3.3.4)$$

donde  $\frac{p^-}{pob}$  es la probabilidad del individuo peor adaptado y  $\frac{p^+}{pob}$  es la probabilidad del individuo mejor adaptado. Debe cumplirse que

$$p^- + p^+ = 2 \quad (3.3.5)$$

para garantizar que el tamaño de la población sea constante.

En la siguiente asignación de probabilidades

$$p_i = \frac{pob - i + 1}{\sum_j^{pob} j}, \quad (3.3.6)$$

comprobamos que se cumple con los requisitos mencionados anteriormente,

ya que siendo  $p_1 = \frac{pob}{\sum_j^{pob} j}$  la probabilidad del individuo mejor adaptado

y  $p_2 = \frac{1}{\sum_j^{pob} j}$ , la del peor adaptado podemos concluir que  $p^+ = \frac{pob^2}{\sum_j^{pob} j}$  y  $p^- = \frac{pob}{\sum_j^{pob} j}$ . Así,

$$p^+ + p^- = \frac{pob(pob+1)}{\sum_j^{pob} j} = 2, \quad (3.3.7)$$

ya que el denominador es una progresión aritmética de razón 1 y que da como resultado  $\frac{pob(pob+1)}{2}$ .

En el ejemplo usado en la selección por adaptación ponderada y calculando la probabilidad en función de la ecuación [3.3.6], la selección se hubiese realizado como:

Cromosoma	$P_i$	$\sum_{i=1}^6 p_i$
1	0.2857	0.2857
2	0.2381	0.5238
3	0.1905	0.7143
4	0.1429	0.8571
5	0.09524	0.9525
6	0.0476	1

Cuadro 3.2: Ranking de los cromosomas.

Si se generan los mismos seis números aleatorios que los utilizados atendiendo al anterior criterio (0.1535, 0.6781, 0.0872, 0.1936, 0.7021, 0.3933), los cromosomas seleccionados para el cruce y la mutación son: cromosoma 1, cromosoma 3, cromosoma 1, cromosoma 1, cromosoma 3 y cromosoma 2. En este caso se seleccionan los mismos programas que si atendiésemos al criterio de selección por adaptación ponderada.

Para ranking exponenciales, la probabilidad se asigna como:

$$p_i = \frac{c - 1}{c^{pob} - 1} c^{pob - i} \tag{3.3.8}$$

donde  $0 < c < 1$  es una constante de desviación.

Una selección basada en el valor de la función de adaptación tiende a ponderar el individuo más adaptado en mayor medida mientras mayor sea la diferencia entre él y el individuo menos adaptado [3.3.9]. En una selección

basada en el ranking la diferencia entre las probabilidades de un individuo y el siguiente más adaptado es constante [3.3.10].

$$\Delta p_i = \frac{\Delta f_i}{\sum_j^{pop} f_j} \quad (3.3.9)$$

$$\Delta p_i = \frac{1}{\sum_j^{pop} j} \quad (3.3.10)$$

- Selección por torneo

En primer lugar hay que determinar el tamaño del torneo. Una vez fijado este parámetro se seleccionarán aleatoriamente tantos individuos como tamaño tenga el torneo. El ganador de ese torneo es seleccionado para proseguir con los operadores genéticos. Los descendientes de los ganadores de cada torneo sustituirán al individuo peor adaptado de cada torneo.

En la figura 3.13 se detalla un torneo entre sólo dos individuos. Las figuras más oscuras representan los individuos más adaptados. Los hijos engendrados por los ganadores de los torneos serán los que reemplazarán a los perdedores de cada torneo, que coincide con el peor adaptado. Evidentemente, a mayor tamaño del torneo mayor presión selectiva habrá en la población.

### Cruce

Como indicamos en la sección 3.3.1 los individuos intercambian su información genética a través de la recombinación sexual. De esta forma los progenitores permutan parte de su código para engendrar individuos que tuviesen código de ambos padres.

## Torneo

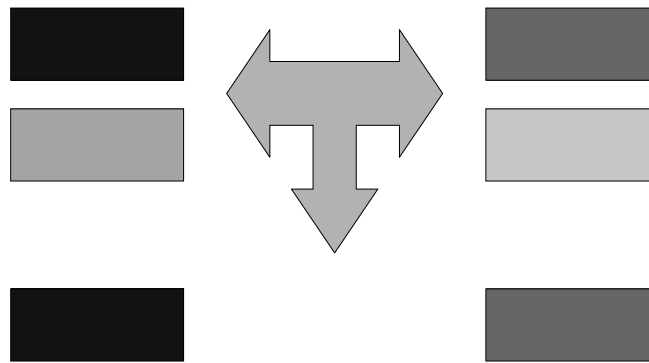


Figura 3.13: Selección por torneo

El operador “cruce” o *crossover* simula para los algoritmos evolutivos el proceso biológico de la recombinación sexual.

Sigue el siguiente proceso:

1. Se eligen dos cromosomas en función de un criterio de selección. A estos individuos les denominaremos padres. Si superan cierta probabilidad, llamada probabilidad de cruce, siguen al paso siguiente.
2. Se selecciona, aleatoriamente, una posición particular en la cadena de caracteres de la representación binaria de cada progenitor y se separan en dos subestructuras genéticas.
3. Se intercambian las subestructuras entre los dos padres. Se generan dos nuevos individuos, llamados hijos, que contiene información de los progenitores.

Con el fin de ilustrar el proceso de recombinación, consideremos dos vectores

$A = [10110]$  y  $B = [01001]$ , elegidos al azar. Supongamos que se tratan de vectores que contienen dos parámetros. El primero está representado en los tres primeros dígitos del vector, mientras el segundo parámetro se representa con el tercer y cuarto dígito. Así la codificación de los vectores son  $A = [5, 2]$  y  $B = [2, 1]$  (si el valor mínimo del rango es 0). Se elige al azar una posición de partición entre el segundo y tercer elemento de los vectores. Tales vectores pueden ser representados de la forma  $A = [10|110] = [A_1 A_2]$  y  $B = [01|001] = [B_1 B_2]$ . La recombinación se produce intercambiando los subvectores  $A_2$  y  $B_2$  de los vectores originales  $A$  y  $B$ , produciendo con ello una nueva generación de vectores  $C = [A_1, B_2] = [10001] = [4, 1]$  y  $D = [B_1 A_2] = [01110] = [3, 2]$ .

Gracias a ello el proceso de búsqueda, dentro del espacio de cadenas de caracteres no se realiza al azar sino en las regiones más prometedoras de dicho espacio de búsqueda.

### Mutación

La mutación es un operador que permite a un individuo cambiar aleatoriamente parte de su información genética. Esto permite que la población no se estanque con demasiada rapidez en determinadas zonas del espacio de soluciones y, de esta forma, evitar caer en óptimos locales.

Las mutaciones incrementan la diversidad en las poblaciones y redireccionan, al azar, la búsqueda dentro de zonas inexploradas del espacio de búsqueda evitando así que el algoritmo quede atrapado dentro de los mínimos locales. Una mutación consiste en seleccionar al azar cierto elemento binario en un determinado cromosoma. Si este elemento es igual a uno es mutado a cero y viceversa. Las mutaciones solo se producen con una probabilidad muy baja, con el fin de evitar

demasiadas distorsiones cuando el proceso de búsqueda se está realizando a través de zonas prometedoras dentro del espacio de búsqueda. Si la probabilidad de mutación es alta, entonces, el procedimiento se ajustaría a una búsqueda aleatoria. La mutación se ilustra fácilmente con el siguiente ejemplo. Consideremos el vector  $A = [10110]$  y supongamos que sufre una mutación en su cuarta cifra. En este caso el vector  $A$  se transformará en el nuevo vector  $E = [10100]$ .

### **El Teorema de Esquemas en los Algoritmos genéticos**

Desde el comienzo de la teoría de los algoritmos genéticos, el Teorema de Esquemas y la Hipótesis de los “Buenos Bloques” (*Building Blocks*) (Goldberg, 1989) han constituido los marcos teóricos mediante los cuáles se intentaba dar una explicación al funcionamiento de los algoritmos genéticos.

Definimos esquema (*schema*) como un cromosoma que contiene todos los bits similares (bits definidos) a través de los cromosomas altamente adaptados de la población, donde los bits no similares se representa con el símbolo “\*” “Símbolo no importa”. Así, el cromosoma  $[0, *, 1, *]$  representa un esquema de 4 bits donde dos de ellos están definidos y dos bits no están definidos.

Los cromosomas de la actual generación, que contienen todos los bits definidos del esquema se conocen como casos. Atendiendo al esquema del anterior ejemplo, podemos obtener cuatro casos.

$[0, 0, 1, 0]$

$[0, 0, 1, 1]$

$[0, 1, 1, 0]$

$[0, 1, 1, 1]$

El esquema recoge la información necesaria para analizar el efecto de la



reproducción y otros operadores genéticos en los buenos bloques contenidos en la población (Goldberg, 1989), ya que las similitudes de los cromosomas altamente adaptados pueden ayudar en la búsqueda guiada y que conducen al concepto de esquema.

Dada una generación de cromosomas, el Teorema de Esquemas no indica el número de casos de un cierto esquema que se darían en la siguiente generación.

Como sabemos que la probabilidad de selección de un cromosoma es directamente proporcional a su adaptabilidad, podemos calcular, para la siguiente generación el número esperado de casos de un determinado cromosoma, que puede expresarse como:

$$m_x(i+1) = \frac{F_x(i)}{\hat{F}_x(i)}, \quad (3.3.11)$$

donde  $m_x(i+1)$  es el número de casos del cromosoma  $x$  en la generación  $i+1$ ,  $F_x(i)$  es la adaptabilidad del cromosoma  $x$  en la generación  $i$  y  $\hat{F}_x(i)$  es la media de las adaptabilidades de todos los cromosomas de la generación  $i$ .

De la misma forma, si conocemos los cromosomas que pertenecen a un esquema en concreto podemos calcular el número de casos de ese esquema en la siguiente generación.

$$m_H(i+1) = \frac{\hat{F}_H(i)}{\hat{F}_i} m_H(i), \quad (3.3.12)$$

donde  $m_H(i+1)$  es el número de casos del esquema  $H$  en la generación  $i+1$ ,  $\hat{F}_H(i)$  es la media de la adaptabilidad de los cromosomas que contienen al esquema  $H$  en la generación  $i$ ,  $\hat{F}_i$  es la media de la adaptabilidad de los cromosomas de la generación  $i$  y  $m_H(i)$  es el número de casos del esquema  $H$  en la generación  $i$ .

La Ecuación 3.3.12 nos indica que un particular esquema crece al ratio que existe entre el valor de adaptación de un esquema y la media de adaptabilidad de la población. El proceso de selección asigna un número creciente de individuos a aquellos esquemas que estén por encima de la media.

Sin embargo, el proceso de selección no hace nada por sí solo para dirigir la exploración hacia nuevos espacios de búsqueda, sólo selecciona los cromosomas que ya están presentes en la población actual. Para evitarlo, se necesitan de los operadores de cruce y mutación. Ambos operadores pueden producir buenos *schemata*, pero también destruirlos. El Teorema de Esquemas considera sólo los efectos destructivos del cruce y la mutación, es decir, el efecto de reducir el número de casos de que un esquema se de en la siguiente generación.

La probabilidad para que un esquema  $H$  sobreviva al operador cruce se puede expresar como:

$$P_H(c) = 1 - p_c \left( \frac{l_d}{l-1} \right), \quad (3.3.13)$$

donde  $p_c$  es la probabilidad de cruce,  $l$  es el tamaño del *schema*  $H$  y  $l_d$  es el *tamaño definido*, que es la distancia que separa a los bits definidos (que no sean “\*”) más extremos. El número de puntos intermedios en los que puede haber un cruce es igual a  $l - 1$ , por lo que la probabilidad de que se rompa el *schema*  $H$  es igual a  $\left( \frac{l_d}{l-1} \right)$ . Por ejemplo, el esquema  $[1 * 1]$  ( $l_d = 2$ ), tiene una probabilidad 1 de que se rompa si existe el cruce, mientras el esquema  $[11*]$  ( $l_d = 1$ ) dicha probabilidad es de 0.5 si finalmente se produce el cruce.

La probabilidad de que un esquema  $H$  sobreviva a la mutación se puede expresar como:

$$P_H(m) = (1 - p_m)^n \approx 1 - p_m \cdot n, \quad (3.3.14)$$

donde  $p_m$  es la probabilidad de mutación y  $n$  es la especificidad de *schema*  $H$ , es decir, el número de bits definidos en el esquema  $H$ . Aquí,  $1 - P_m$  representa la probabilidad de que un bit no sea mutado. Por ejemplo, una mutación del cromosoma [011] tiene una mayor probabilidad de seguir perteneciendo al esquema [ $* * 1$ ] ( $n = 1$ ), que al *schema* [ $*11$ ] ( $n = 2$ ).

Tomando en cuenta los efectos destructivos de la mutación y del cruce podemos reescribir la Ecuación 3.3.12 como:

$$m_H(i+1) = \frac{\hat{F}_H(i)}{\hat{F}_i} m_H(i) \left[ 1 - p_c \left( \frac{l_d}{l-1} \right) \right] [(1 - p_m)^n]. \quad (3.3.15)$$

La Ecuación 3.3.15 es conocida como el Teorema de Esquemas, la cuál nos da una referencia del número esperado de los esquemas en la siguiente generación.

Analizando los componentes de la Ecuación 3.3.15 con más detalle, podemos ver que, en relación al efecto destructivo del cruce, si  $l_d < l$ , la probabilidad de que ocurra un cruce que destruya el esquema es muy baja. Podemos concluir que la probabilidad de supervivencia de un esquema es más alta para esquemas cortos. Analizando el efecto destructivo de la mutación, podemos ver que si el valor de  $n$  es bajo, la probabilidad de que ocurra una mutación entre los bits definidos del esquema  $H$  es baja. La probabilidad de supervivencia del esquema  $H$  después de la mutación, es mayor, para esquemas de especificidades bajas. Si  $\hat{F}_H(i) > \hat{F}_i$ , la probabilidad de que el esquema  $H$  sea seleccionada para la siguiente generación es alta, lo que implica que la probabilidad de selección está por encima de la media de la adaptabilidad de los esquemas. Se puede concluir que los esquemas

de tamaño pequeño, baja especificidad y adaptabilidad por encima de la media tienen más probabilidades de sobrevivir a la siguiente generación. Estos esquemas de tamaño pequeño, baja especificidad y adaptabilidad por encima de la media son conocidos como buenos bloques. La hipótesis de los buenos bloques establece que los algoritmos genéticos buscan comportamientos cercanos al óptimo por yuxtaposición de estos buenos bloques (Goldberg, 1989).

### **3.3.3. Algoritmos genéticos con parámetros continuos**

Cuando las variables que contiene la función objetivo de un problema de optimización son números reales, la resolución de este problema por medio de algoritmos genéticos requiere ligeras modificaciones para convertirlo en un algoritmo de parámetros continuos. Veamos seguidamente una descripción breve de estas modificaciones.

#### **Los parámetros y los cromosomas**

La primera modificación en el algoritmo se refiere a que los parámetros ya no estarán representados en código binario, sino por números reales dentro de un determinado rango que se considere apropiado. Con un código binario, cada parámetro continuo requeriría demasiados *bits* para ser representado. Si cada cromosoma contiene un número  $N_{par}$  de parámetros (que frecuentemente suelen también llamarse genes) el cromosoma tendrá una representación

$$cromosoma = [p_1, p_2, \dots, p_{N_{par}}]$$

donde  $p_i$ ,  $i = 1, \dots, n_{par}$  es un número en representación decimal.

### **Población inicial**

Si consideramos que los parámetros deben estar comprendidos dentro de un determinado rango, partiremos de una población inicial de cromosomas cada uno de los cuales está formado por un determinado conjunto de parámetros dentro de dicho rango.

### **Selección natural**

Al igual que en el caso binario es necesario decidir qué cromosomas de la población inicial están lo suficientemente adaptados para sobrevivir y posiblemente reproducirse, produciendo descendencia para la siguiente generación.

### **Emparejamiento**

Al igual que en el caso binario, existen diversas formas de seleccionar las parejas de cromosomas que, posteriormente, van a reproducirse. Una sencilla es por medio de la selección aleatoria sesgada a favor de los mejor adaptados.

### **Reproducción sexual**

Como en el caso binario es preciso asignar a los descendientes una combinación del material genético de sus padres. Existen diversas formas de realizar esta operación (Véase Michalewicz, 1992, como referencia general) como referencia general), pero un método sencillo sugerido por Haupt y Haupt (1998) es el siguiente. Partimos de una pareja de cromosomas a los que denominaremos padre y madre, cuyos subíndices contienen  $p$  y  $m$ . A continuación seleccionamos aleatoriamente un punto de cruce  $\alpha$  que será un número entero comprendido entre

l y  $N_{par}$ .

$$padre = [p_{p_1}, p_{p_2}, \dots, p_{p_\alpha}, \dots, p_{p_{N_{par}}}]$$

$$madre = [p_{m_1}, p_{m_2}, \dots, p_{m_\alpha}, \dots, p_{m_{N_{par}}}]$$

Los parámetros seleccionados se combinan para formar nuevos parámetros de la forma

$$p_{new1} = \beta p_{p_\alpha} + (1 - \beta) p_{m_\alpha}$$

$$p_{new2} = (1 - \beta) p_{p_\alpha} + \beta p_{m_\alpha}$$

donde  $\beta$  es un número aleatorio comprendido entre 0 y 1. La operación de recombinación produce entonces dos nuevas cromosomas de la siguiente forma

$$descendiente1 = [p_{p_1}, p_{p_2}, \dots, p_{new1}, \dots, p_{p_{N_{par}}}]$$

$$descendiente2 = [p_{m_1}, p_{m_2}, \dots, p_{new2}, \dots, p_{m_{N_{par}}}]$$

Esta operación suele aplicarse, no sólo al parámetro que ocupa el lugar  $\alpha$ , sino a todos los demás parámetros que están situados a la derecha de dicha posición.

### Mutaciones

En el caso continuo una mutación suele efectuarse seleccionando al azar un cromosoma y un determinado parámetro dentro del cromosoma. A continuación, reemplazamos dicho parámetro por un nuevo número elegido al azar en una distribución uniforme dentro del rango de parámetros.

## 3.4. Predicciones por ocurrencias análogas

La aparición de la teoría de los sistemas dinámicos no lineales caóticos ha puesto de manifiesto cómo sistemas deterministas muy simples son capaces de

generar comportamientos muy complejos que pueden ser confundidos con el azar desde el punto de vista de las técnicas estadísticas lineales. Una característica importante de los procesos caóticos, que los diferencia esencialmente de los procesos aleatorios, es la posibilidad de realizar predicciones a corto plazo. En este sentido han sido desarrolladas diversas técnicas de predicción, denominadas genéricamente predicciones por ocurrencias análogas (*Nearest Neighbour*, *NN en sus siglas inglesas*), que han dado lugar a diversos estudios empíricos sobre la habilidad predictiva de los mercados financieros.

La predicción NN, es una nueva técnica no paramétrica de predicción a corto plazo, que ha estado inspirada en la literatura sobre la predicción de sistemas dinámicos caóticos. La filosofía básica que subyace detrás de esta forma de predicción es que los tramos de una serie temporal, en su pasado, pueden tener parecido con determinados tramos de dicha serie en el futuro. Este principio puede ser aplicado con el fin de generar predicciones, seleccionando patrones de comportamiento en el pasado de la serie, por medio de su proximidad con el patrón de comportamiento reciente. Finalmente, los patrones similares detectados en el pasado se emplearán para predecir el comportamiento futuro. La metodología de predicción NN no intenta ajustar un modelo global de la serie temporal a predecir; por el contrario, sólo usa información local relativa a los patrones de comportamiento anteriores a las observaciones que van a ser predichas.

La historia de la predicción no lineal de series temporales tiene una doble raíz. Por un lado, la comunidad estadística ha construido una amplia variedad de estimadores mediante regresión no paramétrica (Ver Stone, 1997, como un trabajo pionero de recopilación), que han culminado con técnicas variadas como

la aproximación Kernel, la r-NN, la de splines y la recesión robusta ponderada localmente, entre otras (Veáanse Härdle, 1992; Eubank, 1999; Cleveland, 1979, como referencias generales).

Por otra parte, aproximadamente diez años después de muchas de las investigaciones cruciales sobre regresión no paramétrica llevadas por la comunidad estadística, y de forma completamente independiente, la comunidad de los sistemas dinámicos ha construido predictores para series temporales no lineales con el fin de ser empleados en las investigaciones sobre el caos determinista y sus posibilidades de predicción a corto plazo. Trabajos seminales en este sentido son los de Farmer y Sidorowich (1987), Casdagli (1992b,a) y Sugihara y May (1990), que dieron un importante impulso a la predicción NN. Las diversas metodologías no paramétricas de predicción han sido aplicadas en la predicción de series financieras, habiéndose obtenido diferentes grados de éxito en los resultados dependiendo de la serie temporal examinada, de que dicha serie sea una serie en nivel o una serie de rendimientos, del número de observaciones, de la frecuencia de los datos, del método de predicción, del horizonte de predicción, etc.

Las experiencias basadas en regresiones ponderadas localmente, realizadas empleando diversas variantes del método de Cleveland (1979), han sido, por lo general, pesimistas. Así los trabajos de Diebold y Nason (1990), Jaditz y Sayers (1995), Brock *et al.* (1992), Meese y Rose (1990), Hsieh (1991) y Mizrach (1992), no muestran mejoras significativas en la predicción respecto al modelo de paseo aleatorio en varias series económicas y financieras.

Otras experiencias empleando otras técnicas de predicción NN, más cercanas a la concepción de la línea de los sistemas dinámicos caóticos, han resultado



más prometedores. Tales son los trabajos de Canova (1993), MacDonald y Taylor (1994), Bajo Rubio *et al.* (1992a), Fernández Rodríguez *et al.* (1999a), Soofi y Cao (1999) o Lisi y Medio (1997); en todos ellos se concluye que las predicciones NN para datos de alta frecuencia son significativamente superiores a las del modelo de paseo aleatorio.

Como señalamos anteriormente, la metodología de predicción NN es una técnica de predicción a corto plazo inspirada en la literatura de la predicción de sistemas dinámicos caóticos. Con el nombre de sistemas caóticos se conocen numerosos sistemas dinámicos deterministas, de carácter no lineal, que son capaces de generar fluctuaciones bruscas e impredecibles por las metodologías estadísticas lineales como la de Box y Jenkins (1976), de modo que las series temporales que generan pueden ser confundidas con “ruido” pese a su indudable carácter determinista.

Una herramienta básica en este enfoque es la inmersión de la serie temporal dentro de un espacio de fases. Consideremos una serie temporal  $\{x_1, x_2, \dots, x_T\}$  de observaciones escalares. Con el fin de detectar patrones de comportamiento en esta serie, los segmentos de igual longitud son considerados como vectores  $x_{t,\tau}^d$  de  $d$  observaciones consecutivas muestreados a partir de la serie original. Esta operación se conoce en la literatura de los sistemas dinámicos como reconstrucción del espacio de fase de la serie temporal. Tal reconstrucción se realiza por medio de vectores de la forma:

$$x_{t,\tau}^d = (x_t, x_{t-\tau}, \dots, x_{t-(d-1)\tau}) \quad t = d, d+1, \dots, T \quad (3.4.1)$$

donde  $\tau$  es el tiempo de retardo y  $d$  la dimensión de inmersión. El tiempo de

retardo  $\tau$  representa el tiempo entre dos observaciones muestreadas sucesivamente que empleamos en la reconstrucción de un atractor  $d$ -dimensional por medio del proceso de inmersión de vectores. El parámetro  $d$  representa la dimensión del espacio de estados en el que se recreará la dinámica del subyacente sistema. A estos vectores  $d$ -dimensionales se les conoce habitualmente como  $d$ -historias, y al parámetro  $d$  se le conoce como dimensión de inmersión, mientras que al espacio  $d$ -dimensional  $R^d$  se le denomina el espacio de fases reconstruido de la serie temporal. El Teorema de inmersión de Takens (1981), pieza clave en todo este enfoque, establece que para una dimensión de inmersión  $d$  suficientemente grande, si la serie original está muestreada a partir de un sistema dinámico determinista, quizás caótico, la trayectoria de la  $d$ -historia  $x_t^d$  reproduce el proceso generador de datos. Dicho de otro modo, si la serie temporal está generada por medio de un sistema dinámico diferenciable, el Teorema de Taken (Takens, 1981) afirma que para una dimensión de inmersión  $d$  suficientemente elevada, existe una función

$$F : R^d \rightarrow R^d,$$

tal que  $x_{t+1,\tau}^d = F(x_{t,\tau}^d)$  y esta aplicación tiene el mismo comportamiento dinámico que el sistema original desconocido, en el sentido de la equivalencia topológica.

Por consiguiente, el sistema de predicción NN constituye una metodología basada en la detección de las  $d$ -historias  $x_t^d, x_{t_2}^d, \dots, x_{t_k}^d$  con un comportamiento dinámico similar a  $x_T^d$ , y se emplea en la predicción de la siguiente observación  $x_{T+1}$  al final de la serie  $\{x_1, x_2, \dots, x_T\}$ . La proximidad de dos  $d$ -historias en el espacio de fases  $R^d$  puede ser interpretada como signo de comportamiento dinámico similar y nos permitirá hablar de los *Nearest Neighbours* (vecinos más próximos) de un particular segmento  $x_t^d$  de la serie. La predicción  $\hat{x}_{T+1}$  se estima

como alguna extrapolación adecuada de las observaciones  $x_{t_1+1}, x_{t_2+1}, \dots, x_{t_k+1}$  a los puntos vecinos más próximos en el espacio de fases reconstruido de la serie temporal. Por tanto, la predicción de la observación  $x_{T+1}$  será estimada analizando las trayectorias históricas en torno a la última  $d$ -historia disponible

$$x_T^d = (x_T, x_{T-1}, \dots, x_{T-(d-1)}), \quad (3.4.2)$$

donde hemos tomado  $\tau = 1$  por simplicidad.

Con ese propósito, detectaremos en el espacio de fases  $R^d$  las  $k$   $d$ -historias

$$x_{t_1}^d, x_{t_2}^d, \dots, x_{t_k}^d, \quad (3.4.3)$$

con comportamiento dinámico similar a la  $d$ -historia final de la serie  $x_T^d$  que maximizan la función:  $\rho(x_t^d, x_T^d)$ . Por tanto, las  $k$   $d$ -historias [Ecuación 3.4.3] elegidas presentan la máxima correlación serial con respecto a la última  $d$ -historia  $x_T^d$ . Una vez que los NNs a  $x_T^d$  han sido establecidos la predicción  $\hat{x}_{T+1}^d$  de la observación  $x_{T+1}$ , puede ser estimada usando alguna extrapolación de las observaciones

$$x_{t_1+1}^d, x_{t_2+1}^d, \dots, x_{t_k+1}^d, \quad (3.4.4)$$

siguientes a las  $k$   $d$ -historias [Ecuación 3.4.3], es decir:

$$\hat{x}_{T+1} = \tilde{F}(x_{i_1+1}, x_{i_2+1}, \dots, x_{i_k+1}), \quad (3.4.5)$$

donde  $\tilde{F}(\cdot)$  es una función que aproxima la dinámica local del sistema. Hay varios métodos para seleccionar  $\tilde{F}(\cdot)$ ; entre ellos destacan las aproximaciones polinómicas de varios órdenes, las funciones radiales de base, redes neuronales y otras. Sugihara y May (1990) o Casdagli y Weigend (1994) ofrecen una exposición detallada de este tipo de predicciones. En numerosos trabajos de predicción

financiera han sido empleadas regresiones lineales locales con el fin de aproximar la función 3.4.5 que permite estimar la predicción. Véase, por ejemplo, Bajo Rubio *et al.* (1992a,b), Fernández Rodríguez *et al.* (1997), Fernández Rodríguez *et al.* (1997), Fernández Rodríguez y Sosvilla Rivero (1998) o Fernández Rodríguez *et al.* (1997).

La aproximación más simple a la hora de construir un predictor local es la aproximación de orden cero, donde el número de entornos es  $k = 1$ . En este caso  $\tilde{F}(\cdot)$  es la función proyección en cada uno de sus argumentos, es decir:

$$\hat{x}_{T+1}^{\text{orden cero}} = x_{t_r+1} \quad \text{para } r = 1, 2, \dots, k. \quad (3.4.6)$$

Una determinación más eficiente de la función  $\tilde{F}(\cdot)$  se obtiene usando la media de las aproximaciones de orden cero, es decir,

$$\hat{x}_{T+1}^{\text{bar}} = \frac{1}{k} \sum_{i=1}^k x_{t_r+1}. \quad (3.4.7)$$

Por razones geométricas, tal predictor  $\hat{x}_{T+1}^{\text{bar}}$  ha sido llamado predictor baricéntrico (Bajo Rubio *et al.*, 1992a; Fernández Rodríguez y Sosvilla Rivero, 1998)

Linsay (1991) y Mizrach (1992), en una aproximación similar procedente de la estadística no paramétrica, denominan a este tipo de predictor un estimador no paramético lineal básico (baseline).

Con el fin de mejorar las predicciones elementales o baricéntricas suele emplearse la Regresión Lineal Local (RLL en adelante) como forma más sofisticada de aproximar la función  $\tilde{F}(\cdot)$  de la Ecuación [3.4.5]. Este procedimiento implica realizar una regresión, por mínimos cuadrados de los  $k$  NN sobre sus  $d$ -historias precedentes, es decir, regresando  $x_{t_r}$ , de la Ecuación 3.4.4 sobre

$x_{t_r}^d = (x_{t_r}, x_{t_r-1}, \dots, x_{t_r-(d-1)})$  de la Ecuación [3.4.3], para  $r = 1, \dots, k$ . En tal caso, los coeficientes estimados pueden usarse para generar predicciones para cualquier  $x_{T+1}$  de forma

$$\hat{x}_{T+1}^{\text{RLL}} = \hat{a}_0 x_T + \hat{a}_1 x_{T-1} + \dots + \hat{a}_{d-1} x_{T-(d-1)} + \hat{a}_d, \quad (3.4.8)$$

donde los  $\hat{a}_i$  son los valores  $a_i$  que minimizan la expresión

$$\sum_{i=1}^k (x_{t_r+1} - a_0 x_{t_r} - a_1 x_{t_r-1} - \dots - a_{d-1} x_{t_r-(d-1)} - a_d)^2, \quad (3.4.9)$$

Otras aproximaciones para seleccionar el número de NN sugieren sistemas de ponderación que asignan, en la estimación de los coeficientes de la regresión lineal local, mayores pesos en las observaciones más cercanas. Así Cleveland (1979) y Cleveland y Devlin (1988), proponen ponderar un número  $k$  de los errores que figuran en la expresión [3.4.9] por medio de unos pesos que están inversamente relacionados con la distancia de cada entorno a la última  $d$ -historia  $x_T^d$  de la serie. En este caso, los  $\hat{a}_i$  son los valores de  $a_i$  que minimizan la expresión

$$\sum_{i=1}^k w(x_{t_r}) (x_{t_r+1} - a_0 x_{t_r} - a_1 x_{t_r-1} - \dots - a_{d-1} x_{t_r-(d-1)} - a_d)^2, \quad (3.4.10)$$

donde  $w(x_{t_r}) = W \frac{\|x_{t_r}^d - x_T^d\|}{\sum_{i=1}^k \|x_{t_r}^d - x_T^d\|}$ ,  $\|\cdot\|$  es la norma Euclídea y  $W(u) = (1 - u^3)^3$  para  $0 \leq u \leq 1$ .

Aunque tales sistemas de ponderación tienen atracción teórica porque permiten usar una cantidad arbitraria de entornos, los algoritmos que se emplean para estimar los coeficientes en la regresión ponderada [Ecuación 3.4.10] pueden ser numéricamente menos estables que los que se emplean para la regresión no

ponderada [ Ecuación 3.4.9]. Wayland *et al.* (1994) mostraron que, en presencia de ruido, los algoritmos con regresión no ponderada tienden a proporcionar mejores resultados. Ello es debido a que, cuando calculamos los parámetros de la regresión local, es necesario invertir una matriz  $X^T X$ , cuyas filas son altamente colineales porque están formadas por los entornos seleccionados que son muy similares a  $x_T^d$ . En tal caso los parámetros estimados son numéricamente inestables. Los experimentos numéricos revelan que este problema es especialmente agudo cuando la matriz ponderada  $X^T X$  presenta una alta multicolinealidad.

Cuando tomamos un número elevado de entornos, al añadir unos coeficientes de ponderación que se van haciendo muy pequeños para las  $d$ -historias que se alejan de la referencia  $x_T^d$ , el determinante de la matriz es prácticamente nulo y la matriz  $X^T X$  está más próxima a ser singular que en el caso no ponderado. Por esta razón, en nuestro procedimiento predictivo focalizaremos nuestra atención en los algoritmos no ponderados (Jaditz y Sayers, 1998).

### **3.4.1. Algunos criterios de selección de los parámetros en la predicción NN**

El Teorema de Takens, pese a su extraordinaria agudeza teórica y simplicidad conceptual, no ofrece información sobre la forma de escoger los diversos parámetros cruciales para la predicción NN como son el tiempo de retardo  $\tau$ , la dimensión de inmersión  $d$  y el número  $k$  de entornos para la reconstrucción de la función  $\tilde{F}(\cdot)$  en la Ecuación [3.4.5]. La única limitación que impone el teorema de Takens a los parámetros es la de tomar  $d \geq 2m + 1$ , donde  $m$  es la dimensión de la variedad compacta que ha de ser inmersa en el espacio Euclídeo  $R^d$  (Takens, 1981).

De cara a los propósitos prácticos, la determinación de la aplicación  $\tilde{F}(\cdot)$  en la Ecuación [3.4.5], es crucial para realizar predicciones con determinado grado de precisión. Existe una literatura fragmentaria que tiene como objetivo determinar, de forma independiente y por separado, cada uno de estos parámetros usando muy diversas reglas heurísticas. A continuación vamos a mostrar diversos métodos estándar que aparecen en la literatura con el fin de determinar estos parámetros claves en la predicción NN.

### Selección del parámetro $\tau$

En las ciencias de la naturaleza es posible obtener series temporales de muy alta frecuencia y elevado número de observaciones. Dicha situación se produce, por ejemplo, cuando un fenómeno físico está gobernado por un sistema dinámico continuo, tal como una ecuación diferencial, donde es posible obtener una observación para cualquier instante del tiempo. En tal caso se plantea el problema de seleccionar sólo un subconjunto de la serie temporal de cara a una adecuada reconstrucción del espacio de fases.

El parámetro  $\tau$  da cuenta de la frecuencia con que muestreamos en la serie temporal original con el fin de reconstruir el espacio de fases. Una buena selección de  $\tau$  es importante de cara a una correcta reconstrucción.

Si  $\tau$  es demasiado pequeño entonces  $x_t$  es muy similar a  $x_{t+1}$  y los vectores  $x_{t,\tau}^d$  no serán lo bastante independientes como para formar parte de un sistema de coordenadas. Esto significa que el sistema no ha evolucionado el tiempo suficiente como para producir nueva información sobre el espacio de fases en su conjunto.

Si, por el contrario,  $\tau$  es demasiado grande entonces las coordenadas son independientes. En tal caso, incluso una determinación muy precisa del valor

de  $x_t$  no puede evitar el crecimiento exponencial de los pequeños errores. Así, el propósito de la selección de  $\tau$  es obtener una buena frecuencia de muestreo que sea lo suficientemente grande para que  $x_t$  y  $x_{t-\tau}$  sean parcialmente independientes, pero no tan grande como para que sean completamente independientes desde el punto de vista estadístico Abarbanel (1996).

El valor de  $\tau$  suele elegirse, en la práctica, haciendo uso del concepto de información mutua entre dos series temporales, que intuitivamente se define como la información media que se gana, sobre una observación de una serie temporal, debido al conocimiento de una observación de la otra serie.

La noción de información mutua entre una medida  $a_i$ , tomada de un conjunto de observaciones  $A = \{a_i\}$ , y otra medida  $b_j$ , tomada de un conjunto  $B = \{b_j\}$ , es la cantidad de información ganada por el conocimiento de una de las dos medidas a la hora de predecir la otra. Tal ganancia de información se puede cuantificar, en bits, de la forma

$$\log_2 \left[ \frac{P_{AB}(a_i, b_j)}{P_A(a_i)P_B(b_j)} \right] \quad (3.4.11)$$

siendo  $P_{AB}(a, b)$  la densidad de probabilidad conjunta, mientras que  $P_A(a)$  y  $P_B(b)$  son las densidades de probabilidad individuales. Si la medida del valor  $a_i$  de  $A$  es completamente independiente de la medida del valor  $b_j$  de  $B$  entonces  $P_{AB}(a, b) = P_A(a)P_B(b)$ . En tal caso la información mutua entre ambas medidas es cero. Si realizamos un promedio del estadístico anterior, obtenemos la información mutua promedio entre las observaciones de  $A$  y de  $B$ , es decir,

$$I_{AB} = \sum P_{AB}(a_i, b_j) \log_2 \left[ \frac{P_{AB}(a_i, b_j)}{P_A(a_i)P_B(b_j)} \right]. \quad (3.4.12)$$

Esta idea puede ser empleada para cuantificar la información mutua existente



entre las observaciones  $s(t)$  y  $s(t + \tau)$  que distan  $\tau$  unidades de tiempo en una serie temporal. La información mutua promedio es en este caso,

$$I(\tau) = \sum P(s(t), s(t + \tau)) \log_2 \left[ \frac{P(s(t), s(t + \tau))}{P(s(t))P(s(t + \tau))} \right]. \quad (3.4.13)$$

Fraser y Swinney (1986) han sugerido que la función  $I(\tau)$  podría ser empleada como una forma de autocorrelación no lineal para determinar el retardo  $\tau$  que debería tomarse para que las observaciones  $s(t)$  y  $(s(t + \tau))$  fuesen lo suficientemente independientes una de otra, de modo que pudiesen ser empleadas como coordenadas independientes en la reconstrucción de un atractor, según establece el método de Takens. Como señala Abarbanel (1996) la selección del parámetro  $\tau$  suele elegirse como el primer mínimo de la información mutua promedio  $I(\tau)$  entre ambas series temporales.

Si bien estas consideraciones son de gran utilidad para series temporales muestreadas sobre un sistema dinámico continuo, en las series económicas muy rara vez puede preguntarse el problema de tener que proceder a una determinación de  $\tau$  porque la longitud de las series suele ser muy limitada. Quizás la única excepción digna de mención pueden ser las series financieras del mercado continuo.

Nosotros consideraremos, en este trabajo, que la frecuencia de muestreo de los datos es igual a uno ( $\tau = 1$ ), dada la naturaleza diaria de los datos utilizados en la aplicación empírica. No obstante nuestra metodología permite considerar otro tipo de situaciones donde el parámetro  $\tau$  se determine continuamente con todos los demás.

### **Selección de la dimensión de inmersión**

Un asunto capital en la predicción NN es la elección de una dimensión de inmersión de la serie temporal. Existen en la literatura del caos diversos métodos para realizar esta elección:

- Métodos basados en las propiedades dinámicas de los atractores extraños, que se basan en la estimación de algunos invariantes de la dinámica caótica tales como la dimensión de correlación o los exponentes de Lyapunov (Grassberger y Procaccia, 1983). Estos métodos no son prácticos en las series financieras donde una eventual naturaleza caótica de una serie temporal estaría contaminada por el ruido.
- El método de la descomposición singular de la matriz  $X^T X$ , donde  $X$  es una matriz cuyas columnas son las  $d$ -historias construidas a partir de la serie temporal (Broomhead y King, 1986), (Ver Lisi y Medio, 1997, para una aplicación del método en la predicción de tipos de cambio).
- El método de los falsos entornos (Kennel *et al.*, 1992) que consiste en determinar aquel valor de la dimensión de inmersión a partir del cual la proximidad de dos  $d$ -historias trae consigo la proximidad de las observaciones que siguen a dichas  $d$ -historias.
- El método del promedio de falsos entornos que constituye una modificación, de Cao (1997) del método de falsos entornos, y que no contiene ningún parámetro subjetivo una vez que la frecuencia de muestreo  $\tau$  ha sido elegida.
- El método de aproximación de orden cero de Cao y Soofi (1999), que consiste

en encontrar la dimensión de inmersión minimizando el promedio absoluto de los errores de predicción, un paso hacia delante, usando un modelo predictivo aproximado de orden cero (predicciones con un solo entorno); (Ver Cao y Soofi, 1999, para una aplicación en la predicción del tipo de cambio del dólar).

- En muchos casos se suele comprobar la robustez de las predicciones, frente a la dimensión de inmersión, haciendo que dicho parámetro varíe en un rango que se considere adecuado. Así en un modelo NN de repuesta del mercado, propuesto por Mulhern y Caprara (1994), estos autores emplean una dimensión de inmersión entre uno y diez, limitando el análisis a estas diez dimensiones como un razonable equilibrio entre tiempo de computación y poder del modelo. De forma similar Fernández Rodríguez *et al.* (1999a), en sus predicciones sobre diversos episodios del índice Nikkey de la Bolsa de Tokyo, consideran una parrilla de parámetros  $d$  y  $k$  consistente en tomar la dimensión comprendida entre tres y siete, mientras que el número de puntos próximos varía entre 100 y 150, con un paso de diez unidades.

### Selección del número de puntos próximos

Para series caóticas, el número ideal de entornos depende de la complejidad del atractor y del número de observaciones que forman la serie. La elección del número adecuado de NNs pretende constituir un equilibrio entre la ganancia de estabilidad de los parámetros estimados que se alcanza cuando empleamos muchas observaciones y la pérdida de unicidad de la región del atractor donde realizamos el muestreo.

En general no se dispone de un criterio uniforme para seleccionar el número de NNs en la predicción de series temporales. Asimismo, muchos autores sostienen que los errores estándar de predicción disminuyen con el número de entornos elegidos (Casdagli, 1992b,a; Jaditz y Sayers, 1998). Igualmente, algunos autores (Mulhern y Caprara, 1994) han observado que la sensibilidad de la predicción respecto al número de NNs se reduce con el empleo de un algoritmo de ponderación que regule la contribución de cada  $d$ -historia del espacio de fases de acuerdo con su distancia a la  $d$ -historia de referencia. Existen numerosas aproximaciones a este problema, especialmente cuando se trata de la elección de puntos próximos en la predicción financiera, dependiendo de la longitud de la serie y la naturaleza de las observaciones a predecir. En el ajuste de la Ecuación [3.4.8], el número de NNs debe ser, al menos de orden  $d + 1$ . Cuando  $k = d + 1$ , este método es equivalente a la interpolación lineal y el problema de mínimos cuadrados tiene una solución única. En la práctica, con el fin de asegurar la estabilidad de la solución y mejorar el error de predicción, resulta conveniente tomar  $k > d + 1$ . Se han registrado diferentes contribuciones en la literatura sobre la selección del parámetro  $d$ , de entre las que destacan las siguientes:

- Casdagli (1989) aconseja tomar una dimensión de inmersión en torno a  $2(d + 1)$ .
- Empleo de la descomposición singular de Broomhead y King (1986); Lisi y Medio (1997).
- Otros autores han considerado como número de puntos próximos adecuado, el dos por ciento de los puntos de la serie. Tal es el caso de

Fernández Rodríguez *et al.* (1999a) en el contexto de la predicción de las series de tipos de cambio del Sistema Monetario Europeo.

- Casdagli (1992b,a) ha propuesto un algoritmo de predicción basado en el estudio del comportamiento del error cuadrático medio (ECM) de predicción, normalizado por la desviación estándar de la serie a predecir, cuando la dimensión de inmersión y el número de NNs varía. Casdagli usa este algoritmo con el fin de distinguir el comportamiento caótico de baja dimensión del comportamiento estocástico lineal, comparando la precisión de la predicción a corto plazo. No obstante, tal procedimiento puede ser empleado con el fin de ajustar la dimensión de inmersión y el número de NNs. Consideremos la raíz del ECM (RCECM) de predicción en un determinado período de predicción un paso hacia delante

$$E_d(k) = \frac{\sqrt{(\sum_i (\hat{x}_i(k) - x_i))^2}}{\sigma} \quad (3.4.14)$$

donde  $\sigma$  es la desviación estándar de la serie temporal,  $d$  representa la dimensión de inmersión y  $k$  el número de NNs. La selección adecuada del número de NNs puede realizarse de la siguiente forma. Fijada una dimensión de inmersión  $d$  seleccionaremos un número de NNs tales que hagan mínima la función cuadrática de errores  $E_d(k)$ .

Aquella pareja de valores  $(d, k)$  donde la función  $E_d(k)$  tome un mínimo absoluto proporcionan, en teoría, los parámetros adecuados para el ajuste de la función 3.4.5 que proporciona las predicciones de la serie. Tal razonamiento presenta, en cambio, diversos problemas como el de generalización (nadie garantiza que fuera del período de prueba el RCECM

sea mínimo para estos parámetros) y el problema de la sobreparametrización de los datos, fenómeno que se podría producir si tomásemos dimensiones demasiado elevadas.

En el contexto de la regresión no paramétrica, la elección del número  $k$  de entornos, en la predicción NN o del parámetro de anchura de banda en la aproximación Kernel, está asociado con el problema de armonizar dos objetivos conflictivos como son el de una adecuada reducción del ruido observacional o el de una buena aproximación a la función de regresión generadora de los datos. Varias aportaciones significativas en el contexto de la regresión ponderada localmente, son las siguientes:

- Cleveland (1979) para estimar la curva de regresión no paramétrica a partir de un conjunto de  $T$  puntos  $\{(x_i, y_i)\}_{i=1}^T$  propone tomar un número de NNs  $k \approx \frac{T}{2}$ .
- Cleveland y Devlin (1988) proponen, en contexto similar, tomar  $k \approx \frac{2}{3}(T + d + 1)$ .
- Diebold y Nason (1990) analizan el comportamiento, tanto del error cuadrático medio de predicción como del error absoluto medio de predicción dentro de la muestra, para dimensiones de inmersión 1, 3 y 5, y un amplio rango del número de entornos. Para las dimensiones de inmersión 3 y 5 obtienen un porcentaje óptimo de puntos próximos alrededor del 90% en diversas series semanales de tipos de cambio.
- Meese y Rose (1990) consideran como número de entornos una ventana

formada por el 80 % del tamaño muestral, alegando robustez de los resultados derivados de dicha elección.

- Mizrach (1992) en el contexto de las series cambiarias (con 3063 observaciones) correspondientes a diversas divisas del Sistema Monetario Europeo, considera en la predicción dimensiones de inmersión comprendidas entre 1 y 3, y un rango de entornos de 1, 5, 30, 60 y 90. Tanto en el caso univariante como en el multivariante los resultados óptimos, de cara a la predicción, se obtienen para dimensiones 1 y 3, y para un número de entornos entre 1 y 5.
- Hsieh (1991), en series pertenecientes al índice Standard & Poor's 500, realiza todas las predicciones en una ventana que comprende entre el 10 % y el 90 % de entornos (avanzando de 10 % en 10 %) y para las dimensiones de inmersión comprendidas entre 1 y 5.
- Brock *et al.* (1992), en la predicción de rendimientos semanales del índice Standard & Poor's 500 y para la serie de tipo de cambio marco alemán/dólar estadounidense estudia un sistema de predicción unidimensional que se apoya en un índice de volatilidad para obtener información adicional. Emplea diversos números de entornos tales como 10, 100, 200, 300, 350 y 500.

En el contexto de las predicciones de series caóticas, donde se han elaborado predictores muy similares a los empleados en la regresión no paramétrica, se ha propuesto, igualmente, un amplio abanico de criterios a la hora de seleccionar del número de entornos considerados en la predicción.

- Farmer y Sidorowich (1987) proponen emplear como número de entornos dos veces el número de puntos necesitado para ajustar el polinomio de regresión local por mínimos cuadrados:  $2(d+1)$  para un polinomio de regresión lineal en dimensión de inmersión  $d$ , y  $2d(d+1)$  para un polinomio en regresión cuadrática.
- Fernández Rodríguez *et al.* (1999a), en la predicción de las diferentes monedas que forman el Sistema Monetario Europeo, proponen emplear un número de entornos del orden del 2% de las observaciones de la serie y una dimensión de inmersión alrededor de 6.
- Sugihara y May (1990) eligen, para series biosanitarias, una dimensión de inmersión que minimiza la correlación serial entre los datos reales y predichos. Entonces, el número de puntos próximos elegidos es  $d+1$ .
- Linsay (1991), en diversas predicciones sobre el atractor de Lorenz (1963) recomienda tomar un número de entornos entre 2 y 5 veces el número mínimo  $d+1$  de puntos necesario para hacer una predicción en dimensión  $d$  mediante un predictor lineal, estimando sus coeficiente por mínimo cuadrados.
- Jaditz y Sayers (1998), usando el algoritmo de Casdagli (1992b,a), proponen para series diarias de 4442 observaciones de diversos índices monetarios Divisia, una dimensión de inmersión entre 8 y 12 y un número de entornos comprendido entre 176 y 204.
- Soofi y Cao Soofi y Cao (1999) minimizan el error absoluto de predicción, en una aproximación de orden cero (predicciones con un solo entorno).



Para diversas series del dólar frente a otras monedas, con un total de 1009 observaciones disponibles, obtienen dimensiones de inmersión óptimas muy elevadas, comprendidas entre 27 y 56. Para elegir el número de entornos consideran el número óptimo que minimiza una función de pérdida, en un conjunto de validación anterior al conjunto de predicción. El número de entornos que utilizan para la predicción de cada moneda es sustancialmente diferente; van desde 1 entorno, para el caso marco alemán/dólar estadounidense, hasta 896, para el dólar canadiense/dólar estadounidense.

### 3.4.2. Selección automática de los parámetros en la predicción NN con AGs.

La nueva metodología que proponemos para estimar los parámetros  $\tau$ ,  $d$  y  $k$  parte de una filosofía que descansa en varios principios básicos.

En primer lugar pretendemos que la determinación de estos parámetros se realice, conjuntamente, utilizando un único principio común y no haciendo uso de reglas heurísticas independientes, como las que aparecen en la literatura. En este sentido, proponemos el uso de AG para la determinación de dichos parámetros. Los AG constituyen unos sistemas de optimización que tienen la ventaja de ser capaces de tratar funciones de pérdidas, dependientes de los parámetros del predictor [3.4.5], sin necesidad de invocar propiedades de continuidad o diferenciabilidad para la función de pérdida.

En segundo lugar, nuestra metodología pretende evitar el fisgoneo de datos, tan común en los problemas de predicción y en su evaluación. En este sentido emplearemos el principio de validación cruzada (*cross-validation*) a la hora

de determinar los parámetros que intervienen dentro de nuestro predictor. El criterio de validación cruzada nace en la literatura sobre regresión no paramétrica (Efrom, 1983) y consiste en permitir que sean los datos quienes elijan los *inputs* paramétricos fundamentales del modelo. El propósito de la validación cruzada es el de evitar un modelo que se ajusta bien a un determinado conjunto de datos, pero que es incapaz de generalizar sus buenos resultados predictivos para datos extramuestrales. La validación cruzada pretende perturbar, en alguna medida, el conjunto de observaciones muestrales, con el fin de estudiar la robustez de la inferencia del modelo estimado. La forma más directa de validación cruzada consiste en descomponer la muestra que se va a predecir en dos submuestras I y II. La submuestra I suele denominarse “período de entrenamiento” y en ella se ajustan los valores de los parámetros del modelo con el fin de que se minimice cierta función de pérdida que depende de los errores de predicción. Finalmente la evaluación del modelo, mediante algún tipo de función de pérdida adecuado, se realiza sobre la submuestra II que suele denominarse “período de validación” o “período de predicción”. El comportamiento del modelo sólo puede juzgarse y compararse con otro modelo por su comportamiento en el conjunto de validación, nunca en el de entrenamiento.

Igualmente nuestra metodología pretende abordar el problema de la sobreparametrización de los modelos, problema también relacionado con el del fisgoneo. Cuando un modelo contiene demasiados parámetros puede reproducir muy fielmente los datos de un conjunto de entrenamiento aunque será incapaz de “generalizar” los buenos resultados a predicciones extramuestrales. El problema de la sobreparametrización puede mitigarse tomando funciones de pérdida más

complejas que el error cuadrático medio que penalicen, en el modelo predictivo, el empleo de un elevado número de parámetros. Tal es el caso del criterio de información de Akaike (AIC) (Akaike, 1974), o el criterio de información de Schwarz (SIC) (Sawa, 1978; Schwarz, 1978).

Finalmente observemos, tal como señalan Campbell *et al.* (1997), que tanto el fisgoneo de datos como la sobreparametrización son problemas muy relacionados que se caracterizan porque el modelo predice “demasiado bien”, en el sentido en que ha capturado tanto el ruido aleatorio como las genuinas no linealidades de la función de regresión. Con el fin de salvaguardar los tres principios anteriormente mencionados procederemos en la forma siguiente. Una vez que hayamos seleccionado una especificación concreta del modelo [3.4.5] eligiendo, por ejemplo, un modelo lineal local como [3.4.8] ajustaremos los parámetros de dicho modelo sobre la base del único objetivo de la eficiencia predictiva que se obtiene al minimizar una cierta función de pérdida. Con este propósito es necesario seleccionar un “período de predicción”  $P = \{x_t : N < t \leq N\}$  para algún  $N < T$ . Para cada  $x_t \in T$  nosotros obtendremos la predicción, un paso hacia delante  $\hat{x}_{t+1}$  de la observación  $x_{t+1}$  usando solamente información previa a dicha información, es decir  $\hat{x}_{t+1} = E(x_{t+1}/F_t)$ , donde  $F_t = \{x_s : 1 \leq s \leq t\}$ . De este modo los parámetros del modelo pueden ser ajustados usando algún tipo de medida de eficiencia predictiva.

La medida más clásica de eficiencia predictiva es la raíz cuadrada del error cuadrático medio de predicción, que viene dado por

$$R_k(d) = \sqrt{\sum_{x_t \in P} \frac{(\hat{x}_t - x_t)^2}{P}} \quad (3.4.15)$$

Con el fin de evitar la sobreparametrización de una especificación determinada del modelo [3.4.5] es posible asignar una penalización a la introducción de cada unidad adicional de dimensión de inmersión. En este sentido pueden usarse diversas medidas de verosimilitud penalizando la complejidad tales como las siguientes:

El criterio de información de Akaike (Akaike, 1974) (al que nos referiremos en adelante como AIC, según sus siglas en lengua inglesa Akaike Information Criterion), considera la siguiente función de pérdida

$$AIC = \ln \left( \frac{1}{T - N - 1} \sum_{x_t \in P} (\hat{x}_t - x_t)^2 \right) + \frac{2(d + 1)}{T - N - 1}, \quad (3.4.16)$$

donde  $T - N - 1$  es la longitud del período de predicción y  $d$  es la dimensión de inmersión y, por tanto,  $d + 1$  es el número de parámetro del modelo.

El *criterio de información de Schwarz* (SIC) (Schwarz, 1978) considera la siguiente función de pérdida

$$SIC = \ln \left( \frac{1}{T - N - 1} \sum_{x_t \in P} (\hat{x}_t - x_t)^2 \right) + \frac{(d + 1) \log(T - N - 1)}{T - N - 1}. \quad (3.4.17)$$

El problema anterior puede considerarse, igualmente en el contexto de la regresión no paramétrica. En tal caso el objetivo es el de encontrar una aproximación razonable a la función de respuesta desconocida  $m(\cdot)$ , que describe la dependencia media entre las variables  $Y$  y  $X$ ,  $Y_i = m(X_i) + \epsilon_i \quad i = 1, \dots, n$ , donde  $\epsilon_i$  constituyen una familia de variables independientes e idénticamente distribuidas  $N(0, 1)$ , de la que sólo conocemos  $n$  observaciones  $\{x_i, y_i\}_{i=1}^n$ .

En este caso la curva de regresión produce lo que suele llamarse un suavizado o alisado (*smoothing*). Un problema común de las diversas técnicas empleadas

en la regresión no paramétrica Kernel (Nadaraya, 1964), k-NN (Loftsgaarden y Quesenberry, 1965), regresión robusta ponderada localmente (Cleveland, 1979) y otras, es el conflicto de objetivos que se produce entre una “buena reducción” del ruido observacional y una “buena aproximación” a la función de regresión (Härdle, 1992). La variable capaz de equilibrar este conflicto de objetivos es la anchura de banda en la versión *kernel* o el número de entornos en las versiones NN.

Los AGs pueden emplearse, igualmente, con el propósito de ajustar parámetros tales como la anchura de banda, el número de NNs o el grado de aproximación polinomial que se emplea en la metodología de Cleveland (1979). Esta nueva metodología permite añadir en la función de pérdida del AG un término destinado a penalizar la variabilidad local de la función de regresión estimada, tal como hace la metodología *spline*.

En tal caso pueden considerarse las siguientes funciones de pérdida de Akaike y Schwarz modificadas,

$$AIC(k) = \ln \left( \frac{1}{T} \sum_{i=1}^T (\hat{y}_i - y_i)^2 \right) + \frac{2(d+1)}{T} + \ln (Var(\Delta^{(k)}(\hat{y}_i))), \quad (3.4.18)$$

$$SIC(k) = \ln \left( \frac{1}{T} \sum_{i=1}^T (\hat{y}_i - y_i)^2 \right) + \frac{(d+1) \ln(T)}{T} + \ln (Var(\Delta^{(k)}(\hat{y}_i))), \quad (3.4.19)$$

donde  $\Delta(k)$  representa el operador de  $k$ -ésimas diferencias.

## **3.5. Programación Genética**

La programación genética, desarrollada por Koza (1992), constituye una extensión de los algoritmos genéticos que permite considerar estructuras más complejas que los patrones formados por sucesiones de caracteres como son los cromosomas permitiendo además que las estructuras genéticas tengan una longitud variable. Con este objetivo se consideran operaciones genéticas entre programas de computación que no son mas que una composición jerárquica de funciones. Dichas estructuras genéticas se pueden representar por programas de computación utilizando el lenguaje de programación “LISP”, como un árbol de funciones y terminales, mediante programas lineales o bien grafos.

El proceso que sigue es idéntico al de los algoritmos genéticos. Se elige una población inicial de individuos de forma aleatoria. Posteriormente se seleccionan los individuos mejores adaptados. El proceso de recombinación consiste en escoger a dos individuos (llamados padres) y generar dos nuevos individuos (llamados hijos) mediante el intercambio de un bloque de información del primer padre con un bloque de información del segundo padre.

La mutación se produce mediante la generación aleatoria de un bloque de información en uno de los individuos.

### **3.5.1. Proceso evolutivo de la programación genética**

La programación genética, se basa en los principios de la evolución genética, como se explicó en la sección 3.3. Describiremos a continuación más detalladamente los procesos de generación de una población, la elección de individuos y los operadores de selección, mutación y reproducción. Aunque

se basa en el mismo algoritmo que los algoritmos genéticos [Apartado 3.3.2], las peculiaridades de la programación genética merecen que detallemos con detenimiento cada uno de estos componentes.

### 3.5.2. Estructura de los programas de computación en la programación genética

La programación genética utiliza estructuras genéticas de tamaño variable donde el genotipo está compuesto por programas computacionales. Las representaciones clásicas de estos programas son a través de programas creados mediante el lenguaje LISP o bien de estructuras en árbol. Aunque existen otras representaciones, como son las lineales y los grafos, por simplicidad, usaremos en este capítulo solamente las dos primeras.

#### Estructuras en LISP

LISP (*LIS*t *Pro*cessing) es un lenguaje de programación de alto nivel que por diversas ventajas está muy extendido en la programación genética.

Una lista en LISP es un conjunto ordenado de items dentro de un paréntesis. A esta lista se le denomina *S-expression* o “Expresión simbólica”. Las expresiones simbólicas pueden estar constituidas solamente por unidades llamadas átomos, que suelen estar compuestas por las constantes del sistema, variables que obtienen los datos del dominio de aprendizaje o de funciones sin argumentos.

En LISP los individuos se pueden escribir como expresiones simbólicas. Así, un programa sencillo de programación genética se podría escribir como:

$$(* \ 2 \ 3) \tag{3.5.1}$$

que tiene su equivalente algebraico en:

$$(2 * 3) \tag{3.5.2}$$

Los paréntesis pueden anidarse dentro de otros para dar expresiones más complejas. De esta forma, la expresión debe evaluarse recursivamente partiendo desde la izquierda de la expresión, como se muestra en el ejemplo siguiente:

$$(- (* 2 3) 4) \tag{3.5.3}$$

cuya expresión algebraica viene dada por:

$$(4 - (2 * 3)) = 4 - 6 = -2 \tag{3.5.4}$$

LISP tiene la ventaja de que tanto los programas como los datos tienen la misma forma, expresiones simbólicas. Facilita la utilización de los operadores genéticos como el cruce basados en estructuras en árbol. Es más, una expresión simbólica no es más que una estructura en árbol, como veremos en la siguiente sección. El programador no necesita ningún código que interprete la información simbólica del programa en código máquina, ya que el intérprete LISP ya lo hace por él.

Sin embargo, posee una serie de inconvenientes como son los problemas de memoria, su velocidad y que no es eficiente en la representación lineal o de grafos.

Veremos en la siguiente sección, cómo las expresiones simbólicas en LISP están íntimamente relacionadas con la representación en árbol de los programas genéticos.



### Estructuras en árbol

Los terminales y las funciones constituyen las bases de los programas computacionales. El terminal proporciona un valor al sistema, mientras que la función procesa estos valores para dar otro valor que puede ser la salida del sistema o bien, el argumento de otra función.

El grupo de terminales está compuesto de los *inputs* del sistema, constantes y aquellas funciones que no tienen argumentos. Si siguiéramos con la representación en árbol, constituirían el final de cada una de las ramas del árbol. Las entradas o *inputs* deberán escogerse dentro del dominio de aprendizaje (es decir, del conjunto de datos a partir de lo cual el sistema debe aprender). Las constantes se eligen al azar entre una serie de números reales al comienzo de la ejecución. Posteriormente estas constantes irán cambiando a medida que se produzcan los operadores de combinación y mutación. Otras constantes se irán construyendo en el sistema a medida que sean ejecutadas con otras constantes y funciones.

El grupo de funciones está compuesto por operadores y funciones que se pueden procesar dentro del sistema. Los argumentos de las funciones vienen dados por los terminales. Entre las distintas funciones podemos destacar:

- **Funciones aritméticas**

Suma (+), resta (-), multiplicación (\*), división (/).

- **Funciones booleanas**

Y (AND), o (O), o exclusivo (OR), no (NOT).

- **Funciones trigonométricas**

Seno, Coseno, tangente.

- **Funciones condicionales**

Si..entonces. (IF...THEN)

- **Funciones de transferencia de datos**

Ir a, saltar a. (GOTO)

- **Funciones de bucle**

FOR, WHILE.

- **Funciones de asignación**

Al seleccionar el número de funciones a introducir en el sistema existe una relación de sustitución entre potencia en la representación de la solución del problema y tiempo de computación y complejidad de la solución. Lo mismo ocurre con la elección de las constantes, aunque no se requiere incluir todas las constantes como terminales del sistema, ya que éste puede crear nuevas constantes como se explicó más arriba.

Una vez establecidos los elementos con los que va a trabajar el algoritmo, es necesario crear una estructura que combine los elementos anteriormente descritos. La estructura más habitual es la estructura en árbol.

En este tipo de estructuras cada nodo es un terminal o una función. En el primer caso, no saldrán más ramas de ese nodo, mientras que si es una función saldrán tantas ramas como argumentos necesite esa función. Pero estas estructuras deben también recoger reglas o convenciones para saber cuándo y cómo se ejecuta cada nodo. En las estructuras en árbol se han establecido dos mecanismos de conversión. El *postfijado* donde se evalúa el nodo que está más a

la izquierda y para el cuál todos los *inputs* están disponibles; y el *prefijado* donde los nodos que se ejecutan primero son los más próximos a la raíz.

Así, la Figura 3.14, siguiendo el método de postfijado, representa la expresión:

$$((5 + 6)/7) * (8 + 9) \quad (3.5.5)$$

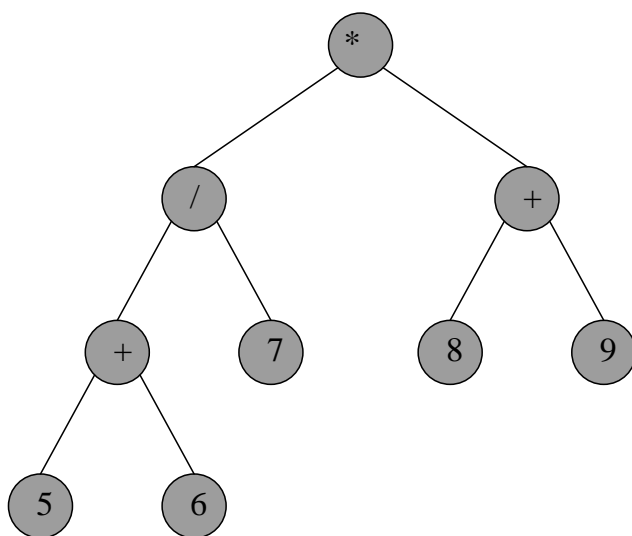


Figura 3.14: Programa computacional estructurado en forma de árbol

Como dijimos en la sección anterior, toda expresión simbólica podía expresarse como un programa computacional con estructura en árbol. Por lo que, el programa computacional representado en árbol en la Figura 3.14 y en la expresión algebraica [3.5.5] es equivalente a la siguiente expresión simbólica:

$$(* (+ 8 9) (/ 7 (+ 5 6))) \quad (3.5.6)$$

Se puede observar cómo, prácticamente, las estructuras en árbol y las expresiones simbólicas en LISP son intercambiables.

Hay que señalar que el uso de la memoria en este tipo de estructuras es local. Los valores de los argumentos están disponibles sólo para la función de las que son argumentos y no de forma global a todas las funciones del árbol.

### **3.5.3. Inicialización de la población**

Una vez definido la estructura de los individuos de la población, el primer paso que debe realizar el algoritmo es generar una población aleatoria de individuos. Este proceso debe garantizar la variedad de individuos, entendiéndose por ello que se alcance un rango de valores amplios de la función de adaptación. De esta forma, se garantiza que el proceso no se estanque en un óptimo local y se asegura la diversidad del proceso a través de los operadores genéticos.

Como se señaló en la introducción de este epígrafe la programación genética posee estructuras de tamaño variable, a diferencia de los algoritmos genéticos donde las estructuras son de tamaño fijo, por lo que hay que considerar un parámetro nuevo en la inicialización de la población. Este parámetro es el “tamaño máximo del programa”. En estructuras en árbol, mide la profundidad máxima del árbol o el número de total de nodos del árbol. La profundidad de un nodo es el número mínimo de nodos que se deben pasar entre el nodo raíz del árbol y el nodo seleccionado. Por lo que el tamaño máximo del programa es el máximo número de nodos que debe haber entre el nodo raíz y el nodo más alejado de él.

Para generar una población aleatoria de individuos se utilizan tres métodos (Koza, 1992).

- **Método creciente (“Grow Method”)**

Se selecciona la máxima profundidad del programa. El algoritmo escoge

aleatoriamente un elemento entre los terminales y funciones que se han definido para el programa. Si es una función elige nuevos elementos y si es un terminal la estructura de árbol termina en este punto. Para puntos en que se ha alcanzado la máxima profundidad entonces la selección se restringe al conjunto de terminales. De esta forma, las ramas del árbol tendrán profundidades distintas ya que en ciertas ramas el algoritmo pudo seleccionar terminales antes de llegar a la profundidad máxima del programa.

- **Método completo (“Full method”)**

En este método no se seleccionan terminales hasta que no se haya llegado hasta la profundidad máxima del programa. Así todos los árboles tienen la misma profundidad. Si nos encontramos en un nodo intermedio, es decir, cuya profundidad sea menor a la especificada, entonces la selección se realizará dentro del conjunto de funciones. Si el nodo es de profundidad máxima entonces la selección se restringirá al conjunto de terminales.

- **“Ramped half and Half method”**

Se crearán programas de profundidades entre dos y la profundidad máxima seleccionada. De esta forma, si la profundidad máxima es 6, se generarán árboles de 2,3,4,5 y 6 de profundidad. El número de individuos se repartirá equitativamente entre estas profundidades máximas. Así, para cada profundidad se generarán la mitad con el método creciente y la otra con el método completo. De esta forma, se garantiza la diversidad de individuos dentro de la población

### **3.5.4. Cruce**

El caso de la programación genética no se diferencia mucho de los demás algoritmos evolutivos.

1. Se seleccionan dos individuos en función de un criterio de selección. A estos individuos les denominaremos padres. Si superan cierta probabilidad, llamada probabilidad de cruce, siguen al paso siguiente.
2. Se selecciona aleatoriamente de cada padre dos subárboles. La selección puede ponderarse para que los subárboles constituidos por terminales tengan menos probabilidad de ser elegidos.
3. Intercambiar los subárboles seleccionados entre los dos padres. Se generan dos nuevos individuos, llamados hijos, que contiene información de los progenitores.

Este es el algoritmo básico del operador cruce. Pero como los programas genéticos son de estructura variable, los puntos de corte y los árboles a intercambiar pueden provocar efectos que han dado lugar a cruces muy determinados. Se pueden intercambiar subárboles entre dos individuos de la población [Figura 3.15], pueden intercambiarse dos subárboles de un mismo individuo [Figura 3.16]. A diferencia de los algoritmos genéticos, el cruce de un individuo consigo mismo puede derivar en dos hijos completamente distintos. Como hemos dicho, esto es debido, a las estructuras de tamaño variable con las que trabaja la programación genética. Finalmente, pueden intercambiarse dos módulos de dos individuos [Figura 3.17]

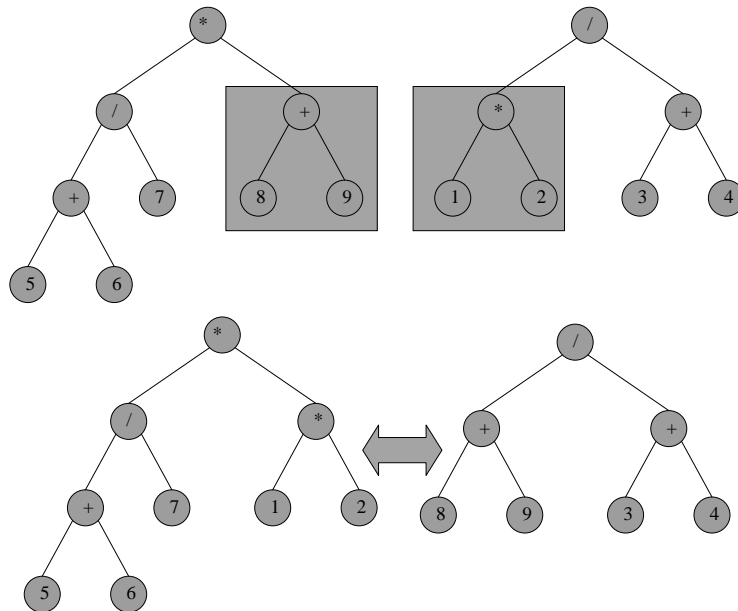


Figura 3.15: Operador cruce. Intercambio de dos árboles.

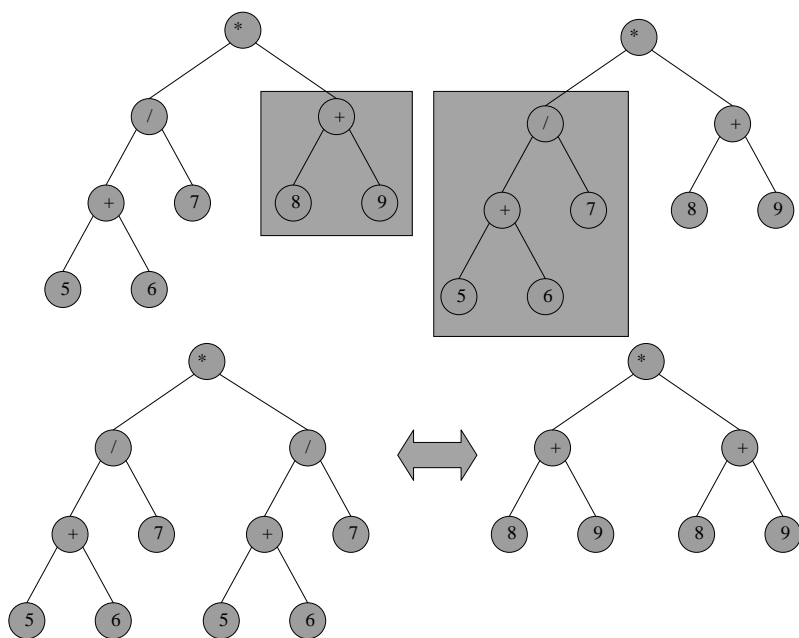


Figura 3.16: Operador cruce. Intercambio de dos árboles en un mismo individuo.

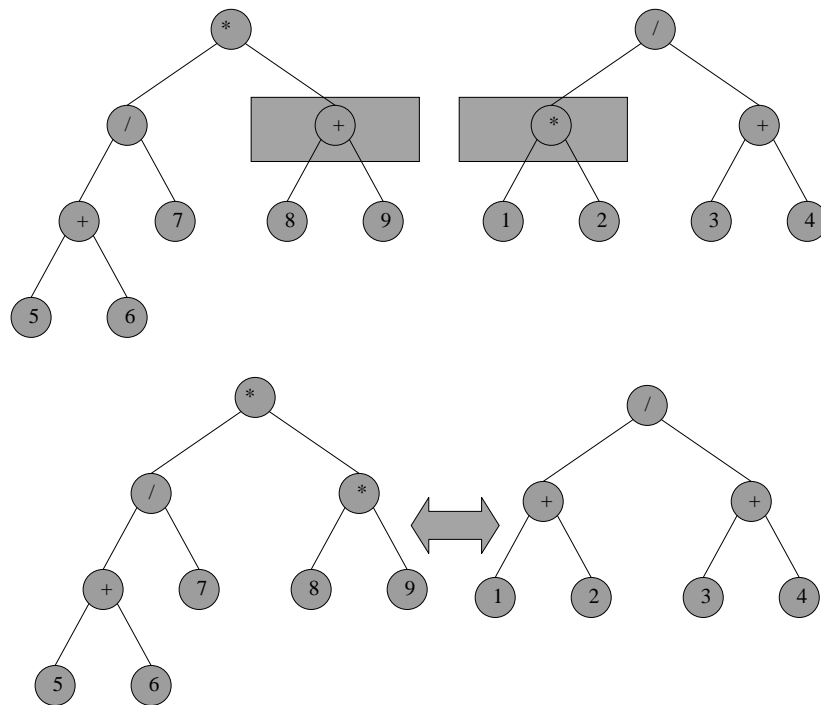


Figura 3.17: Operador cruce. Intercambio de dos módulos.



La base teórica por la que el proceso evolutivo en la programación llega a resultados en los que se mejora la adaptación de la población, es la misma que para los algoritmos genéticos [3.3.2]. En la programación genética, fue Koza (1992) quien enfocó por primera vez el Teorema de Esquemas para la programación genética. Al igual que los algoritmos genéticos, los programas, que son los individuos de la población con la que trabaja la programación genética, están compuestos por subárboles, llamados buenos bloques, y alguno de ellos consiguen un mejor comportamiento del programa. Según Koza, estos buenos subárboles eran transmitidos a las siguientes generaciones a través del cruce, proporcionando a través de la evolución del sistema un comportamiento global bueno de la población.

Sin embargo, en la programación genética las estructuras son de tamaño variable y, por lo tanto, el tratamiento que hace el operador cruce con estos bloques es más complejo. Varios autores (Lang, 1995; Banzhaf *et al.*, 1998), han demostrado que el cruce tradicional no garantiza la transmisión de los buenos bloques e incluso puede provocar la ruptura de estos bloques. Los distintos trabajos realizados sobre el comportamiento del operador cruce en la programación genética (Banzhaf *et al.*, 1998) concluyen que no existen pruebas concluyentes para afirmar que este método sea más eficiente que otros operadores de búsqueda, lo que permite deducir que el cruce en la programación genética actúa como un operador macromutador.

Aunque la macromutación puede mejorar la adaptación total de la población, el operador cruce, sin embargo, es susceptible de varias mejoras. La primera de ellas hace referencia a la extrema facilidad que tiene el cruce en inflar

los programas con subárboles que no ofrecen mejoras al comportamiento del programa. Es lo que se ha dado en llamar *introns*. A continuación se muestran varios ejemplos de *introns*.

(NOT (NOT X))  
(AND (OR X X))  
(+ (- X X))  
(+ X 0)  
(\* X 1)  
(\* ( X X))  
(IF (2=1) X)

Como se puede observar son estructuras que no aportan información alguna al programa y que por lo tanto pueden ser eliminadas sin menoscabo del valor de la función de adaptación. Esto podría hacer más eficiente y extensiva la búsqueda por el espacio de soluciones (Banzhaf *et al.*, 1996).

Otras mejoras que están inspiradas en una observación más detallada de la recombinación sexual biológica son las siguientes:

- Los subárboles de un programa padre pueden cruzarse con cualquier otro subárbol del otro progenitor. No se comprueba en ningún momento si los subárboles representan funciones similares. La reproducción biológica es homóloga, lo que quiere decir que diferentes alelos de un gen producen pequeños cambios en las funciones básicas.
- Un subárbol después del cruce puede situarse en un contexto que no tenga nada que ver con el subárbol con el que se acaba de intercambiar. El cruce

biológico sólo se produce entre genes coincidentes por parte de los dos padres, como se vio en la sección 3.3.1.

Computacionalmente, es muy probable que se destruyan buenos “bloques” durante el cruce en la programación genética. Biológicamente el intercambio se produce entre material genético similar con el propósito de preservar un funcionamiento similar de la función genética del codón que se ha cruzado.

- Finalmente, no hay razón para suponer que los programas creados aleatoriamente pertenezcan a la misma especie. En la naturaleza la reproducción generalmente se produce entre miembros de la misma especie.

Varios autores han perfeccionado el método de cruzamiento con el propósito de asemejarlo más con el cruce biológico. Así Tackett (1994), crea una población de hijos de  $N > 2$  individuos. Se evalúan estos nuevos individuos y se escogen los dos mejores. Simula de esta forma el comportamiento de varias especies que producen mayor descendencia de las que sobreviven.

Otros trabajos de gran trascendencia dentro de la programación genética sobre el algoritmo de cruce han sido los siguientes: Teller y Veloso (1995, 1996), Zannoni y Reynolds (1996) e Iba y de Garis (1996) diseñaron varios algoritmos de cruce que aprendían a seleccionar los buenos puntos de corte. Encontraban regularidades y que estas reglas se encargaban de aprovecharlos.

D’haeseleer (1994) trabajo en un algoritmo de cruce que solo intercambiase subárboles que estén situados en la misma posición en los dos programas progenitores.

Altenberg (1995) creó el marco teórico, y Hinchliffe *et al.* (1996) lo implementó en procesos industriales, de un algoritmo de cruce donde la

adaptación de un individuo fuese la suma de la adaptación de cada uno de sus bloques. Se le iban añadiendo genes al individuo sólo si mejoraban la adaptación total del individuo. De otra forma era descartado.

Nordin *et al.* (1996) ideó un sistema para eliminar los “introns” residentes en el programa genético. Se creaba y almacenaba un valor entero entre todos los nodos del programa. Cambiaron la probabilidad de cruce para que fuese proporcional a ese valor. De esta forma, el algoritmo aprende a reconocer los buenos bloques y mantenerlos.

Banzhaf *et al.* (1998) diseñó un cruce que intentaba reproducir el cruce homólogo que ocurre en la naturaleza. Para ello, se selecciona un subárbol del primer progenitor. Se tiene en cuenta el tamaño y posición del subárbol porque si supera la probabilidad de cruce entonces se intercambia con un subárbol del otro progenitor que tenga el mismo tamaño y posición. De esta forma los hijos tendrán el mismo tamaño que los padres, simulando de esta forma el cruce de bases similares en el ADN.

### **3.5.5. Mutación**

Se selecciona un subárbol al azar y se sustituye por otro generado siguiendo las mismas rutinas que con la generación de los individuos de la población inicial. Una vez alterado, el individuo mutado es el que formará parte de la población que pasará a la siguiente generación. Hay que señalar, que la probabilidad de mutación, suele ser baja ya que de otra forma podría producirse una búsqueda totalmente aleatoria.

Al igual que con el cruce, podemos clasificar los distintos tipos de mutación en función del efecto que produce tras la aplicación de este operador.

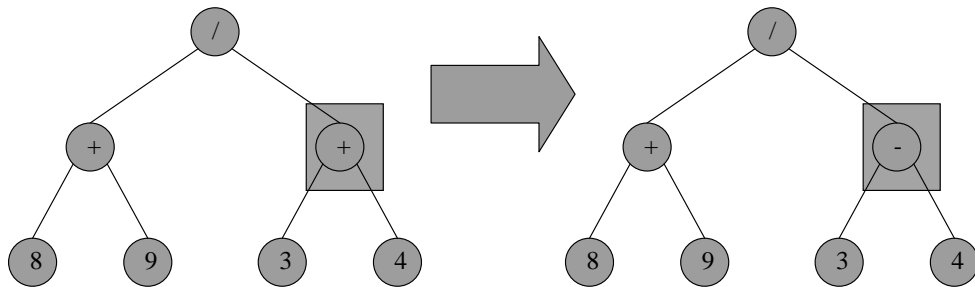


Figura 3.18: Mutación puntual

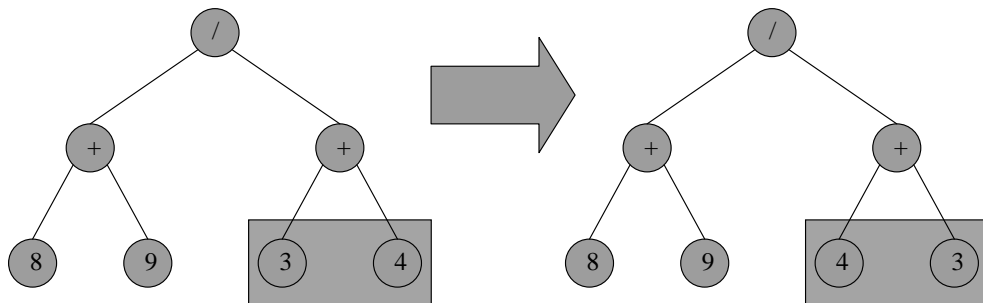


Figura 3.19: Permutación

- **Mutación puntual**

Se cambia un nodo aleatoriamente por otro de la misma categoría [Figura 3.18].

- **Permutación**

Los argumentos de una función se permutan [Figura 3.19].

- ***Hoist***

Se selecciona un subárbol del individuo que constituye ahora el nuevo individuo mutado [Figura 3.20].

- **Mutación expansiva**

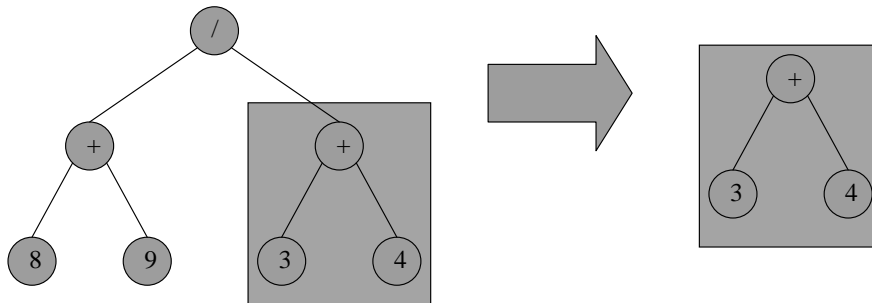


Figura 3.20: "Hoist"

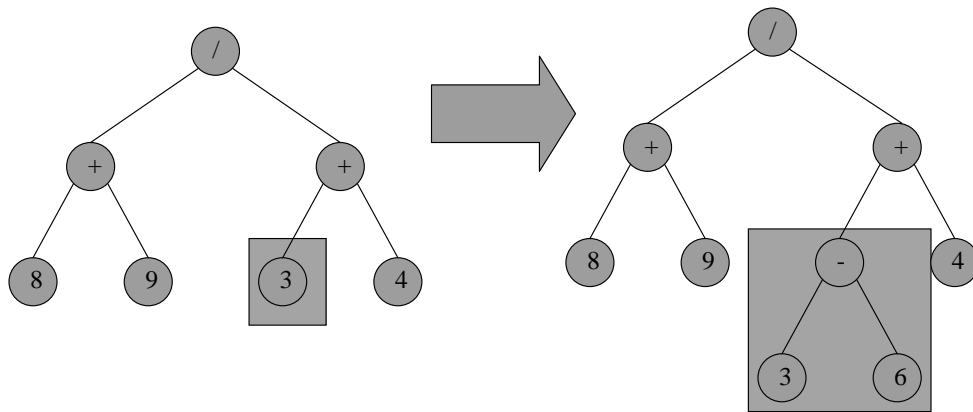


Figura 3.21: Mutación expansiva

Un terminal se sustituye por un subárbol creado aleatoriamente [Figura 3.21].

- **Mutación contractiva**

Es justo el proceso inverso al anterior. Se selecciona un subárbol que es sustituido por un terminal [Figura 3.22].

- **Mutación del subárbol**

Se selecciona un subárbol y se cambia por otro generado aleatoriamente [Figura 3.23].

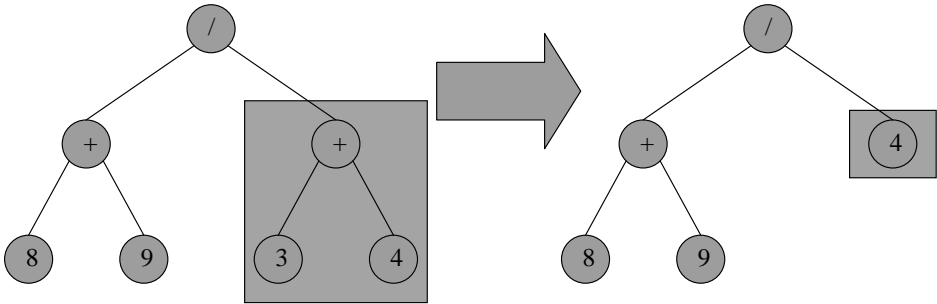


Figura 3.22: Mutación contractiva

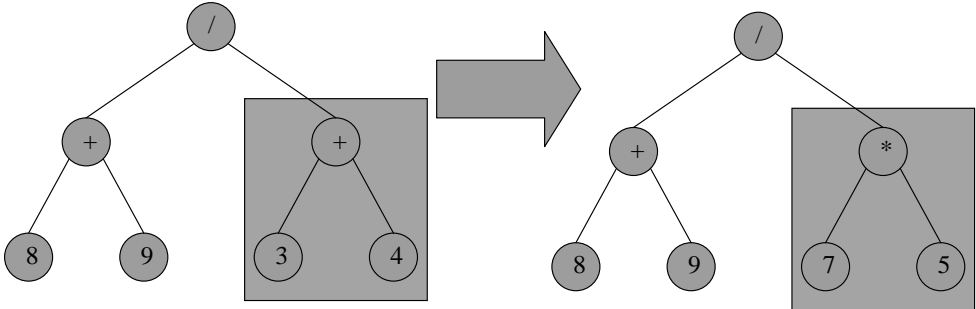


Figura 3.23: Mutación del subárbol

# Capítulo 4

## Resultados empíricos: Aplicaciones a la predicción financiera del IGBM.

### Índice General

---

<b>4.1. Evaluación económica de las reglas técnicas . . . . .</b>	<b>174</b>
<b>4.2. Estrategias técnicas mediante redes neuronales . . .</b>	<b>185</b>
4.2.1. Ejercicio experimental hasta 1997 . . . . .	185
4.2.2. Análisis empírico para los años 1997 a 2002 . . . . .	193
4.2.3. Comportamiento de una estrategia, con posiciones recomendadas para estar fuera o dentro del mercado, optimizada mediante redes neuronales. . . . .	196
<b>4.3. Optimización de reglas técnicas mediante algoritmos     genéticos . . . . .</b>	<b>209</b>
4.3.1. Ejercicio experimental hasta 1997 . . . . .	216
4.3.2. Análisis empírico para los años 1997 a 2002 . . . . .	221
<b>4.4. Optimización de los parámetros utilizados en     predicciones por ocurrencias análogas mediante     algoritmos genéticos . . . . .</b>	<b>225</b>
4.4.1. Ejercicio experimental hasta 1997 . . . . .	225
4.4.2. Análisis empírico para los años 1997 a 2002 . . . . .	283
<b>4.5. Creación de nuevas reglas técnicas mediante la     programación genética . . . . .</b>	<b>316</b>
4.5.1. Ejercicio experimental hasta 1997 . . . . .	319



4.5.2. Análisis empírico para los años 1997 a 2002 . . . . .	334
--	-----

---

## 4.1. Evaluación económica de las reglas técnicas

En el terreno de la predicción financiera la importancia de un sistema de predicciones no radica necesariamente en el hecho de que sean capaces de minimizar un criterio estadístico como cualquiera de las funciones de pérdida consideradas hasta el momento. El factor fundamental que avala un sistema de predicciones financieras es su capacidad para generar beneficios cuando se utiliza como una regla técnica de contratación (Satchell y Timmermann, 1995). Esta Tesis la hemos enfocado desde una óptica de rentabilidad financiera, aunque en algunas ocasiones analicemos la habilidad predictiva de cada sistema.

En este sentido, los indicadores económicos que emplearemos para analizar la rentabilidad de las reglas técnicas obtenidas a partir de nuestras predicciones serán el rendimiento total, rendimiento medio anualizado y el exceso de rendimientos sobre la estrategia “comprar y mantener” (*Buy and Hold, B&H en sus siglas inglesas*), consistente con un mercado eficiente caracterizado por un paseo aleatorio. Cada sistema emplea distintos tipos de reglas de contratación. La elección del rendimiento empleado estaba en función de estas reglas. Si consideramos una serie de precios al cierre  $\{P_t\}_{t=1}^T$  de un determinado activo financiero,  $r_{m,t} = \frac{P_t - P_{t-1}}{P_{t-1}} \approx \log P_t - \log P_{t-1}$  representa el rendimiento del activo el día  $t$ .

Sea  $R$  la suma de los rendimientos obtenidos por una regla técnica. Definimos el beneficio  $\pi$  de esa regla como

$$\pi = \exp \{R\} - 1 \quad (4.1.1)$$

Analíticamente compararemos las distintas reglas a través de sus rendimientos  $R$ , sin embargo, en los resultados siempre calcularemos  $\pi$ , aunque hagamos referencia a ellos como rendimientos.

Así, por ejemplo, si definimos una estrategia donde compramos un activo a precio de mercado  $P$  y donde no vendemos el activo hasta pasados  $N$  días, que es nuestro periodo de predicción, el rendimiento asociado a una estrategia B&H sería:

$$r_{bh} = \log\left(\frac{P_N}{P}\right) - 2c \quad (4.1.2)$$

donde  $c \in [0,1\%, 0,5\%]$  representa el coste de transacción unidireccional.

Definido el rendimiento, el beneficio de esta regla es

$$\pi_{bh} = \exp\{r_{bh}\} - 1 \quad (4.1.3)$$

Esta estrategia B&H representará la estrategia de referencia adecuada con la que comparar la capacidad de generar señales de compra/venta en la mayoría de los modelos. Otra forma de expresar este beneficio, al que llamaremos indistintamente como rentabilidad o rendimiento, se puede expresar a través de

$$\begin{aligned} 1 + \pi_{bh} &= \frac{P_N - cP_N}{P + cP} = \frac{P_N}{P} \frac{1 - c}{1 + c} = \exp(\log P_N - \log P) \frac{1 - c}{1 + c} = \\ &= \exp\left(\sum_{t=1}^N r_{m,t}\right) \frac{1 - c}{1 + c} = \exp\left(\sum_{t=1}^N r_{m,t} + \log \frac{1 - c}{1 + c}\right), \end{aligned} \quad (4.1.4)$$

Un primer objetivo de esta Tesis fue el de diseñar una red que predijese el signo del rendimiento un día hacia delante. Usamos esa información para implementar dos estrategias.

La primera de ellas consiste en comprar y vender cada día de mercado. La rentabilidad de esta regla se ha medido a través del rendimiento total estimado, que podemos definirlo como

$$\hat{R} = \sum_{t=n+1}^{n+N+1} \hat{y}_t r_{m,t} \quad (4.1.5)$$

donde  $N$  es el horizonte de predicción,  $r_{m,t} = \log\left(\frac{P_t}{P_{t-1}}\right)$  es el rendimiento en el periodo  $t$ ,  $\hat{y}_t$  es la posición recomendada en el periodo  $t$  que se modelizará como función de los rendimientos pasados y  $n$  representa el número de observaciones.

En la segunda estrategia, usamos los rendimientos predichos proporcionados por la red para decidir si permanecemos en el mercado o fuera de él. Para decidir cuando tomar una posición u otra, empleamos tanto una regla simple como una regla filtrada. Ambas se pueden definir como sigue:

#### ▪ Regla Técnica Simple

- Si  $\hat{P}_t > P_{t-1}$  y estamos fuera del mercado se produce una señal de compra. Si estamos dentro del mercado, la regla aconseja mantener la posición.
- Si  $\hat{P}_t \leq P_{t-1}$  y estamos dentro del mercado se produce una señal de venta. Si estamos fuera del mercado, la regla aconseja seguir fuera del mercado.

Esta regla técnica simple puede hacerse más sofisticada introduciendo un filtro con el fin de eliminar señales débiles o engañosas. En este sentido, para que se produzca una señal de compra (señal de venta) el día  $t$  será necesario que la predicción del día  $t$  supere (sea inferior) al precio del día anterior en un determinado porcentaje  $\alpha$  (que denominaremos banda), de la desviación típica  $\sigma = std(\{\Delta_1, \dots, \Delta_{t-1}\})$  de las primeras diferencias de la serie original, considerada hasta el día inmediatamente anterior. Dicha regla técnica ajustada con la banda  $\beta$ , puede ser formulada tal como indicamos a continuación

■ **Regla Técnica Filtrada**

- Si  $\hat{P}_t > P_{t-1} + \beta\sigma$  y estamos fuera del mercado se produce una señal de compra. Si estamos dentro del mercado, la regla aconseja seguir dentro del mercado.
- Si  $\hat{P}_t \leq P_{t-1} - \beta\sigma$  y estamos dentro del mercado se produce una señal de venta. Si estamos fuera del mercado, la regla aconseja seguir fuera del mercado.

A esta regla técnica la denominaremos “filtrada” en contraposición con la regla simple anterior.

En este caso emplearemos el rendimiento medio anualizado para medir la rentabilidad de la regla.

El rendimiento compuesto obtenido por esta regla técnica en el periodo  $[1, N]$

puede ser evaluado como

$$r = \sum_{t=1}^N S_{t-1} r_{m,t} + \sum_{t=1}^N (1 - S_{t-1}) r_{f,t} + d \cdot \log \frac{1 - c}{1 + c}, \quad (4.1.6)$$

donde  $r_{f,t}$  es el rendimiento de un activo libre de riesgo el día  $t$ ,  $N$  es el número de días de mercado,  $d$  representa el número de transacciones (es decir, una señal de compra seguida de una señal de venta o *viceversa*), y  $S_t$  es una variable indicador cuyo valor es igual a uno si la regla produce señal de compra y cero en caso de señal de venta. En tal caso, el rendimiento medio anualizado viene dado por  $\pi_{rma} = e^r - 1$ .

Pero para referenciarlo con una medida de rentabilidad apropiada que recoja el mismo número de días que está en el mercado habrá que compararlo con el rendimiento neto de la estrategia B&H ajustada por riesgo. Recoge la media ponderada del rendimiento neto de la estrategia B&H con el rendimiento obtenido por el activo libre de riesgo. Dicha ponderación se establece mediante unos coeficientes que representan los días dentro y fuera del mercado que establece la regla técnica. Si  $N$  representa el número de días del período de predicción que estamos evaluando,  $r_{bh}$  el rendimiento de la estrategia B&H,  $r_{f,t}$  el rendimiento del activo libre de riesgo,  $c$  el coste de transacción unidireccional y  $\alpha$  la proporción de días en los que la regla está dentro del mercado el rendimiento neto de la estrategia B&H ajustada por riesgo se define como

$$r_{bha} = (1 - \alpha) \sum_{t=1}^N r_{f,t} + \alpha \sum_{t=1}^N r_{m,t} + \log \frac{1 - c}{1 + c}. \quad (4.1.7)$$

Nótese que la marca a batir por cada regla de transacción no es necesariamente el rendimiento de la estrategia B&H sino su rendimiento ajustado por riesgo. La razón estriba en que, empleando la estrategia B&H, estaremos siempre dentro

del mercado, soportando por ello un riesgo superior. En cambio, haciendo uso de nuestra estrategia, solamente permaneceremos en el mercado una fracción  $\alpha$  del total de días, mientras que durante la fracción  $1 - \alpha$  restante estaremos fuera del mercado, obteniendo rendimientos derivados del activo libre de riesgo. El rendimiento neto de B&H, ajustado por riesgo, puede asimilarse con el rendimiento de una cartera en que el riesgo se diversifica colocando una fracción de su capital  $1 - \alpha$  en un activo libre de riesgo mientras que la otra fracción  $\alpha$  se coloca en el activo con riesgo “de mercado”.

Para medir la rentabilidad de las reglas técnicas basadas en medias móviles, las reglas de contratación basadas en órdenes estadísticos (descritas en la sección 2) y las reglas creadas mediante programación genética también se ha usado el rendimiento medio anualizado.

Sin embargo, para medir la rentabilidad de las reglas técnicas que se han elaborado a partir de las predicciones NN se ha usado el exceso de rendimiento que se define como

$$\text{Exceso (B\&H)} = \pi_{rma} - \pi_{rbha} \quad (4.1.8)$$

Observemos que el exceso de rendimiento sobre la estrategia B&H, ajustada por riesgo, no depende del tipo de interés, si éste se considera fijo sobre el período muestral. Ello es debido a que comparamos nuestra regla con un promedio ponderado de la estrategia B&H, que permanece fuera del mercado el mismo número de días que nuestra regla técnica. En tal caso, cualquier tipo de interés fijo, durante todo el período muestral, influye por igual en ambas reglas técnicas.

Emplearemos diversos criterios para valorar la bondad de las predicciones

realizadas por un sistema de predicción. Algunos de ellos son estadísticos y reflejan el hecho de que las predicciones obtenidas por la red sean significativamente mejores que las realizadas por algún otro procedimiento estándar de predicción (en finanzas suele ser el modelo de paseo aleatorio).

Pero además emplearemos criterios de carácter económico, ya que el objetivo de cualquier inversión es maximizar el beneficio y minimizar el riesgo. Por ello, el criterio más adecuado para la valoración de la bondad de unas predicciones bursátiles no puede ser otro que la evaluación de los beneficios obtenidos por tales predicciones al ser convertidas en una regla técnica. Con este objetivo definimos varios estadísticos que describen el comportamiento de las reglas de contratación.

Hemos empleado la ratio de beneficio ideal que mide el cociente entre el rendimiento total estimado por la red y un predictor perfecto (es decir, un predictor en el que se ha acertado correctamente todas las posiciones):

$$R_i = \frac{\sum_{t=n+1}^{n+N+1} \hat{y}_t r_{m,t} + d \cdot \left\{ \ln \frac{1-c}{1+c} \right\}}{\sum_{t=n+1}^{n+N+1} |r_{m,t}| d \cdot \left\{ \ln \frac{1-c}{1+c} \right\}} \quad (4.1.9)$$

donde  $\hat{y}_t$  representa la “posición recomendada” que nos ofrece la regla en cada instante. Esta posición recomendada será +1 si la regla técnica predice, para ese instante, rendimientos positivos. Será -1 cuando prediga rendimientos negativos.

De acuerdo con esto, si  $R_i = 1$  las predicciones han sido perfectas, si  $R_i = -1$  todas las predicciones han sido calculadas de forma incorrecta. Por lo que  $R = 0$  se utilizará como referencia para evaluar la bondad de la estrategia de inversión.

La *ratio de Sharpe* representa el rendimiento medio neto, anualizado, de una estrategia de inversión por unidad de riesgo. Se obtiene a partir de la siguiente

expresión:

$$RS = \frac{\bar{r}}{\sigma} \quad (4.1.10)$$

donde  $\bar{r}$  es el rendimiento medio anualizado obtenido a través de las reglas técnicas,  $\sigma$  es la desviación típica de los rendimientos diarios. La ratio de Sharpe nos sirve para establecer comparaciones entre nuestra estrategia técnica y la estrategia B&H.

Así, la ratio de Sharpe del B&H divide los rendimientos de la estrategia de comprar y mantener entre la desviación típica de los rendimientos.

$$RS_{bh} = \frac{\overline{r_{bh}}}{\sigma} \quad (4.1.11)$$

La ratio de Sharpe de las predicciones medirá el cociente entre el rendimiento total obtenido por nuestra regla técnica, dentro y fuera del mercado, dividido entre la desviación típica de dichos rendimientos.

Sin embargo, este ratio se suele comparar con la ratio de Sharpe del B&H ajustado por riesgo que divide el rendimiento entre la desviación típica de la estrategia B&H.

$$RS_{bha} = \frac{\alpha \cdot \overline{r_{m,t}} + (1 - \alpha) \cdot \overline{r_{f,t}}}{\sigma_{bh}} \quad (4.1.12)$$

Por lo que, una estrategia con mayor ratio de Sharpe es preferida a otra, ya que de esta forma, para el mismo riesgo se está obteniendo mayores rendimientos.

En la aplicación empírica de las distintas reglas técnicas examinadas en esta Tesis Doctoral hemos realizado dos ejercicios complementarios. Por una parte, y dado el relativo éxito obtenido por Fernández Rodríguez *et al.* (1997) aplicando predictores por ocurrencia análogas al IGBM, hemos realizado un primer ejercicio



experimental utilizando el período muestral examinado por estos autores (4 de enero de 1966 a 15 de octubre de 1997), explorando el comportamiento de cada una de las reglas técnicas en distintas submuestras. Por otra parte, y dado la espectacular evolución de los mercados bursátiles en los últimos años, con intensos episodios alcistas al calor de una burbuja especulativa que al estallar dio lugar a una larga etapa bajista, hemos llevado a cabo un segundo ejercicio evaluando el comportamiento anual de las reglas técnicas durante el período 1997-2002 [Figura 4.1], estudiando de esta forma su capacidad relativa de generar rendimientos en episodios extremadamente volátiles cuyo análisis ha suscitado un elevado interés entre los especialistas de la Economía Financiera.

En efecto, el periodo 1997-2002 constituye un verdadero desafío intelectual donde poner a prueba cualquier regla técnica, ya que dentro de esta sucesión general de una tendencia alcista y otra bajista, se registraron episodios recurrentes de subidas y bajadas dentro de cada año, manifestando su carácter heterogéneo, por lo que a continuación se realiza un breve resumen de la evolución del IGBM durante cada uno de estos años.

Durante 1997, el IGBM registró un comportamiento muy favorable, cerrando el ejercicio con una rentabilidad de un 35.4%, gracias a la senda a la baja en los tipos de interés, el crecimiento de la producción y los beneficios empresariales, el importante empuje de Wall Street y la convicción de los mercados de que España entraría en la Unión Monetaria Europea. Aún así, si bien el primer trimestre registró un importante sesgo alcista, se registraron sendas bajadas en julio y octubre, relacionadas, respectivamente, con rumores infundados sobre una posible deflación en Estados Unidos y con la tormenta monetaria del sureste asiático.

Por su parte, el año 1998 supuso un buen año para la Bolsa de Madrid, gracias al ritmo de crecimiento de la economía española y el contexto de bajos tipos de interés: el IGBM cerraba el año con una rentabilidad del 31.8 %, si bien la volatilidad fue muy elevada (33.3 %), debido al impacto de la crisis rusa.

El año 1999 fue testigo de la euforia de los valores tecnológicos y del nacimiento exitoso del euro como moneda común, que llevaron al IGBM a una rentabilidad del 16.22 %. A lo largo del ejercicio, se registró una secuencia de pequeñas subidas y bajadas, hasta que en el mes de octubre se produce una recuperación paulatina que llevó a máximos anuales e históricos en diciembre.

Respecto al año 2000, el IGBM registró pérdidas del 10.4 % en un contexto de inquietud por parte de los inversores sobre ciertos aspectos de la coyuntura económica y bursátil. Por una parte, la alta valoración que los activos tecnológicos habían logrado en 1999 comenzó a resentirse a partir del primer trimestre de 2000. Además, se comenzó a temer una ralentización de la economía estadounidense, en un contexto de subida del petróleo y de bajada del euro que hizo temer por el futuro de la fase expansiva del ciclo que experimentaban Europa y España. Fruto de estos desarrollos, sólo en tres meses se producen rentabilidades positivas (febrero, agosto y septiembre), registrándose en el resto retrocesos, destacando mayo y particularmente noviembre, cuando se registran mínimos anuales.

La evolución del mercado a lo largo del año 2001 tuvo dos períodos marcadamente diferenciados. Durante los primeros nueve meses, el IGBM mantuvo un comportamiento sostenidamente bajista (con una caída del 30 %), como reflejo de la intensa desaceleración de la economía mundial, que culminó con los trágicos acontecimientos del 11 de septiembre. A partir de ese momento, se

inicia una progresiva recuperación que lleva, al inicio de diciembre, a niveles cercanos al inicio del año, como consecuencia tanto de unas cotizaciones que habían alcanzado unos valores muy atractivos como de los efectos positivos de la agresiva reducción de tipos de interés llevada a cabo por las distintas autoridades monetarias. Al final del año, el IGBM perdía un 6.4 %.

Por último, el 2002 fue un año de gran dificultad e incertidumbre, debido a la ausencia de signos de confirmación de las expectativas de recuperación económica mundial y a la crisis generalizada de confianza de los accionistas debido a los escándalos financiero-contables. El IGBM registró una pérdida del 23.1 %, pudiéndose señalar tres períodos diferenciados en su evolución. Hasta el mes de mayo, el mercado mantuvo una tendencia plana con pequeñas oscilaciones. A partir de junio, la desconfianza generada por los escándalos contables, las dudas sobre la economía norteamericana, la desaceleración europea, la crisis de las grandes economías latinoamericanas y la incertidumbre sobre los beneficios de las grandes empresas forzaron una fase bajista. En el último trimestre se produjo un intento de recuperación sustentado por la mejora de los beneficios y las expectativas de recorte de los tipos de interés en Estados Unidos y Europa, que finalmente se produjeron en diciembre, si bien se frustró parcialmente en los últimos compases del año por el temor de los inversores a un conflicto bélico en Irak.

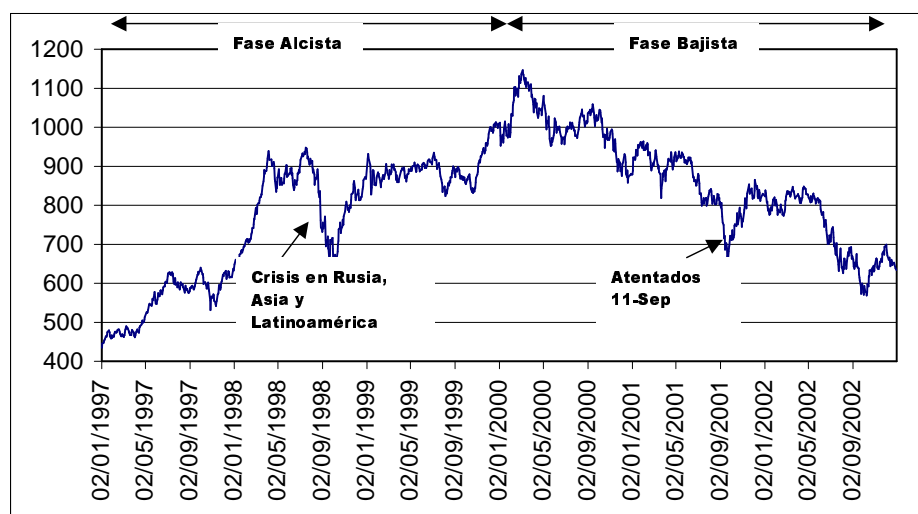


Figura 4.1: Periodo para el análisis empírico

## 4.2. Estrategias técnicas mediante redes neuronales

### 4.2.1. Ejercicio experimental hasta 1997

Uno de los objetivos de esta Tesis es el de formular reglas técnicas de contratación basadas en redes neuronales. Como se ha señalado anteriormente, el uso de las redes neuronales es cada vez mayor dentro del campo de las finanzas, ya que se ha demostrado que pueden recoger complejas relaciones no lineales como las que se producen en los comportamiento de los mercados financieros.

Con el fin de recoger esas no linealidades subyacentes se ha diseñado una red neuronal que permita realizar predicciones sobre la rentabilidad del Índice General de la Bolsa de Madrid (IGBM). Para ello, hemos considerado una serie del IGBM que contiene 6931 observaciones diarias, de precios al cierre, para el periodo comprendido entre el 4 de enero de 1966 y el 15 de octubre de 1997, lo que constituye una muestra representativa de este mercado de valores.

Con el objetivo de comparar los resultados en diferentes episodios del mercado, hemos realizado las predicciones en tres submuestras de 250 observaciones diarias cada una, que corresponden a un tramo alcista, otro bajista y uno estable de la serie. En cada caso se tiene un periodo suficientemente largo para evitar el fisgoneo de datos. Para el entrenamiento de la red se utilizará, en cada caso, todo el conjunto de observaciones anteriores disponibles a cada una de dichas submuestras. Tras realizar el proceso de aprendizaje, empleando sólo la información anterior, se realizarán las predicciones en cada submuestra.

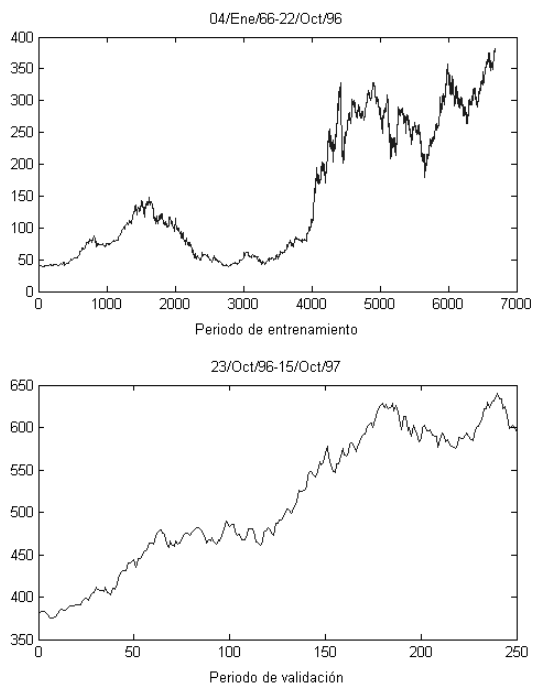


Figura 4.2: Tramos de entrenamiento y validación para el periodo alcista.

En las figuras 4.2, 4.3 y 4.4 se presentan los diferentes tramos de la serie, en cuyas 250 últimas observaciones hemos realizado la predicción y que se corresponde con un periodo de alzas, bajas y estable, respectivamente.

Como hemos indicado en la sección anterior la regla que emplearemos

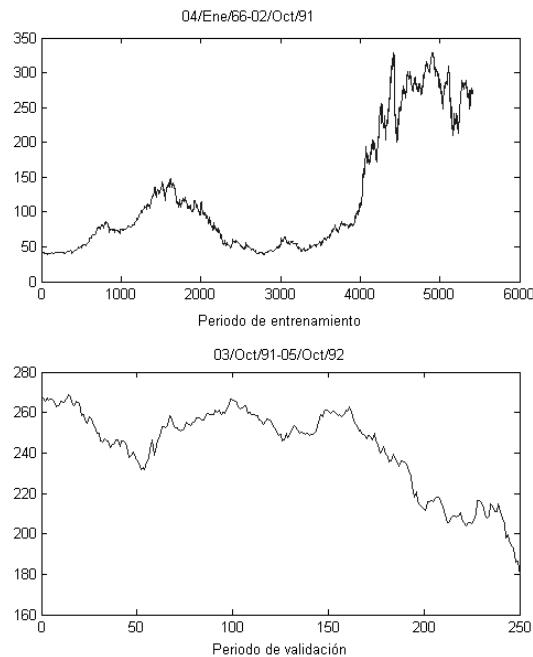


Figura 4.3: Tramos de entrenamiento y validación para el periodo bajista.

consistirá en comprar títulos de los activos que componen el IGBM al comienzo de la sesión y venderlos al finalizar el día de sesión. Es una regla que permite la venta a crédito, es decir, podemos obtener un beneficio de la caída de un valor sin más que vender el título al principio de la sesión para después comprarlo al terminar ésta. Por lo que si acertamos el signo del rendimiento obtendremos el valor absoluto del rendimiento de ese día. El objetivo de la red es predecir ese signo, que representa la posición recomendada para esta regla. Esta posición recomendada será +1 si la red predice, para ese instante, rendimientos positivos. Será -1 cuando prediga rendimientos negativos.

En este apartado no consideraremos la existencia de costes de transacción. Nótese que se producen dos transacciones cada día, por lo que la inclusión de costes de transacción haría bajar significativamente los beneficios de la regla.

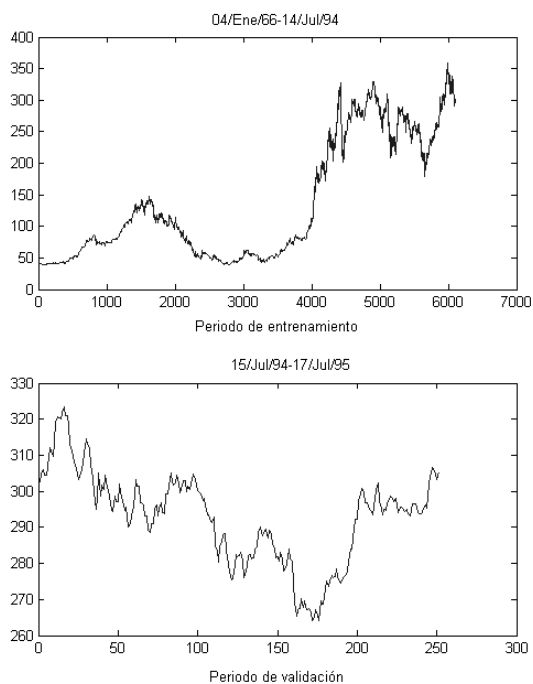


Figura 4.4: Tramos de entrenamiento y validación para el periodo estable.

Para estas reglas emplearemos el rendimiento total estimado en la Ecuación 4.1.5, ya que al no haber días fuera del mercado no es aplicable el rendimiento medio anualizado. Si analizamos con mayor detalle la expresión del rendimiento total estimado podemos ver que cuando acertamos la posición recomendada obtenemos un rendimiento positivo, lo que hace aumentar nuestro beneficio. Por el contrario, si fallamos en la predicción obtenemos un rendimiento negativo, lo que provoca un descenso de la rentabilidad. Por último, cabe señalar que como estamos tratando de predecir un año de sesiones este rendimiento total ya está anualizado.

Hemos comparado los beneficios obtenidos por la regla técnica que proporciona nuestra red con los beneficios obtenidos por la sencilla estrategia de “comprar y

mantener” que usaremos como referencia

$$r_{bh} = \log \left( \frac{P_{t+N}}{P_t} \right) \quad (4.2.1)$$

siendo  $N$  el periodo en que se mantiene la posición en bolsa . De nuevo, no tiene sentido en esta regla el B&H ajustado en el riesgo ya que esta regla siempre está dentro del mercado.

Usaremos, además, diversos criterios para valorar la bondad de las predicciones realizadas por medio de una red neuronal. Para medir la bondad predictiva del ajuste emplearemos el contraste de Pesaran y Timmermann (1992) y el porcentaje en la predicción de los signos. El contraste direccional de Pesaran y Timmerman evalúa la significatividad en la predicción de los signos. Bajo la hipótesis nula de independencia entre los valores reales y los predichos, el estadístico de dicho contraste se distribuye asintóticamente como una  $\mathcal{N}(0, 1)$ , por lo que los valores críticos a niveles de significación del 1 %, 5 % y 10 % serán 2.33, 1.645 y 1.82, respectivamente. El porcentaje en los signos mide el porcentaje de veces que la predicción acierta con la posición correcta de compra o venta.

Desde un punto de económico analizaremos tanto la ratio de beneficio ideal [Ecuación 4.1.9] como la ratio de Sharpe [Ecuación 4.1.10], donde  $c = 0$ .

Aunque la selección de un diseño adecuado de la red es un aspecto crucial para su correcto funcionamiento, la literatura al respecto es más bien escasa. Partimos de una red que emplea funciones logísticas que puedan recoger los procesos no lineales del mercado. Con este tipo de funciones, el mejor algoritmo de aprendizaje que podemos emplear es el de retropropagación.

Sin embargo, debido a las características de este algoritmo, el proceso de aprendizaje es sensible al número de neuronas de la capa oculta. En general,



mientras mayor número de neuronas haya en la capa intermedia mejor será el ajuste de la red. Esto puede provocar un efecto dañino para la predicción que se conoce como sobredimensionamiento o sobreparametrización. El resultado sería que la red se comporta aparentemente bien para los datos incluidos en el periodo de entrenamiento pero no produciría resultados razonables para los datos del periodo de validación. Por el contrario, la existencia de pocas neuronas en la capa oculta provoca salidas con altos errores que se incrementan con el entrenamiento. El resultado sería equivalente a hacer desaparecer la capa oculta.

Para el diseño de la red hemos empleado un método de validación cruzada Gençay (1996). Trabajaremos con una red de tres capas, con una capa de entrada de nueve neuronas. La predicción será a un día vista. Para la predicción del signo del rendimiento del día  $t$ , usaremos como inputs de esta capa de entrada los nueve retardos anteriores,  $(r_{t-1}, r_{t-2}, \dots, r_{t-9})$ , del rendimiento de las acciones que componen el IGBM . Además, una capa oculta con cuatro unidades, en donde se emplea como función de transferencia a la función logística y una capa de salida con una única neurona que usa la función tangente hiperbólica. Debido a la característica de la función tangente hiperbólica, la salida está comprendida entre -1 y 1. Por lo que si la salida de la red es inferior a 0, se genera una posición recomendada igual a -1. Si la salida es superior o igual a 0, se genera una posición recomendada igual a 1. Este dato es el que se comparará con el signo del rendimiento y en el proceso de aprendizaje se minimizará el error cuadrático medio entre la estimación proporcionada por la red y el valor realmente observado. Expresado matemáticamente, el modelo de red neuronal que utilizamos en este

trabajo es:

$$y_t = G(a_0 + \sum_{j=1}^4 a_j F(b_{0j} + \sum_{i=1}^9 b_{ji} r_{t-i})) \quad (4.2.2)$$

donde  $F$  sería la función logística y  $G$  la función tangente hiperbólica y cuya representación se corresponde con la Figura 4.5

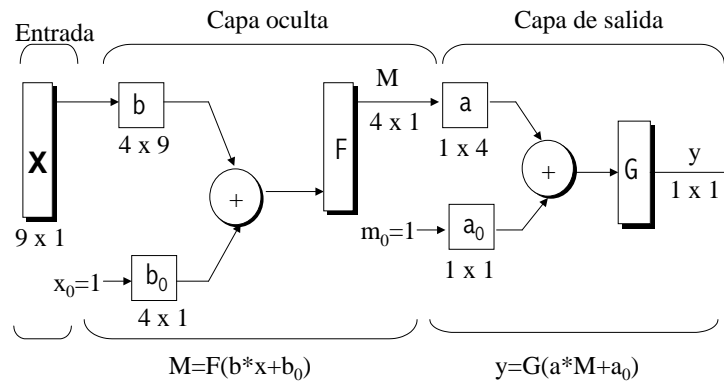


Figura 4.5: Red utilizada en el análisis

Los resultados obtenidos se han resumido en el Cuadro 4.1.

	Tramo bajista	Tramo estable	Tramo alcista
Rendimiento total estimado	0.48	0.27	0.29
Ratio de beneficio ideal	0.25	0.17	0.14
Ratio de Sharpe	0.19	0.13	0.11
Porcentaje en la predicción en signos	0.54	0.57	0.58
Test de Pesaran y Timmerman	0.97	2,24*	2,26*
Rendimiento estrategia "Comprar y mantener"	-0.40	0.0019	0.44

Cuadro 4.1: Principales estadísticos relativos a la capacidad predictiva de la red neuronal en los tres episodios del mercado. \* significativo al 5%

Se puede observar cómo el tramo en donde se ha conseguido un mayor rendimiento es el tramo bajista que coincide con el peor resultado que se podría

obtener con la estrategia comprar y mantener. En este tramo vemos que hemos obtenido casi la mitad de lo invertido mientras que siguiendo la estrategia comprar y mantener se hubiese perdido el 40 % de la inversión. Aún así, se estaría muy lejos de la rentabilidad ideal ya que hemos conseguido un 25 % de lo que se podría haber obtenido si se hubiese acertado todas las posiciones correctamente.

En los tramos estable y alcista el rendimiento estimado está en torno al 30 %, aunque en el tramo alcista se hubiese conseguido un 44 % de la inversión si se hubiese optado por la estrategia comprar y mantener. En ambos tramos, se ha logrado un porcentaje de aciertos entre el 57 % y el 58 % de los signos (es decir, se ha conseguido predecir correctamente la posición del inversor), y además significativamente como muestra el test de Pesaran y Timmerman. No obstante la ratio de beneficio ideal es inferior al 20 % en dichos tramos estable y alcista, por lo que los rendimientos medios anualizados se encuentran lejos de ese rendimiento ideal.

El tramo donde mejor se ha comportado la red es en este tramo bajista ya que, aparte de ser donde más rentabilidad se ha obtenido y más cerca se ha estado de la rentabilidad ideal, es el que presenta un Ratio Sharpe más alto. El tramo bajista es, por tanto, de los tres, donde se ha obtenido una mayor rentabilidad por unidad de volatilidad (riesgo). El tramo de peor comportamiento de las predicciones de la red ha sido el alcista; no sólo por los altos beneficios aportados por la estrategia de “comprar y mantener”, sino también porque se ha conseguido los peores ratios de beneficio ideal y de Sharpe.

Recapitulando, hemos estudiado la posibilidad de implementar reglas técnicas de contratación bursátiles en el IGBM basadas en modelos no

paramétricos como las redes neuronales. Nuestras conclusiones apuntan a que resulta imposible batir, con nuestro modelo, a la estrategia de “comprar y mantener” en un periodo de alzas generalizadas tan fuertes como el examinado. No obstante, nuestras reglas técnicas son capaces de producir rendimientos superiores a la estrategia de “comprar y mantener” en tramos bajistas o estables del IGBM. Los periodos de predicción empleados en todos los casos son de 250 observaciones, un intervalo de tiempo considerable que evita la posibilidad de fisgoneo de datos, como señalábamos anteriormente.

Sin embargo, estas conclusiones deben tomarse con cautela. Para examinar tanto la capacidad de generar beneficios por estas reglas, como las consiguientes implicaciones en el terreno de la eficiencia del mercado, deben tenerse en cuenta los costes de transacción. Esto es algo que realizaremos como futuras líneas de investigación.

#### 4.2.2. Análisis empírico para los años 1997 a 2002

Igualmente hemos examinado la capacidad que tienen las reglas técnicas basadas en redes neuronales para generar beneficios en el periodo comprendido entre 1997 y 2002. La base de datos empleada ha sido actualizada añadiendo a los datos que teníamos hasta 1997 los comprendidos entre 16 de octubre de 1997 a 30 de diciembre de 2002. Así, para cada año se han utilizado los datos disponibles hasta ese momento para entrenar la red. Una vez calculado los pesos se ha procedido a validarlos con los datos del año a examinar.

El cuadro 4.2 recoge los principales estadísticos empleados para medir la capacidad de las reglas técnicas basadas en redes neuronales para los años 1997 a 2002. Podemos observar cómo se han obtenido rendimientos positivos en todos los

años excepto en el año 2002. Asimismo, se ha superado la marca del 0% para la ratio de beneficio ideal excepto para el último año. Sin embargo, en la fase alcista (la que comprende desde 1997 hasta principios del 2000), nuestra regla no supera a la estrategia de “comprar y mantener”, salvo durante el año 1998, donde se obtiene un diferencia entre el rendimiento obtenido por la regla y la estrategia comprar y mantener de más del 10%. En la fase bajista se han obtenido, por parte de las reglas técnicas, unos rendimientos totales de -0.2361, mientras que el obtenido por la estrategia comprar y mantener es de -0.4635. De esta forma, se obtienen en estos periodos ganancias relativas (menores pérdidas) en torno al 20%, confirmándonos las conclusiones obtenidas en el ejercicio experimental. Incluso, podemos observar cómo durante los dos primeros años de esta etapa bajista (2000-2001) se han obtenido rendimientos positivos a pesar del estallido de la burbuja tecnológica. También hemos de señalar, que no estamos considerando costes de transacción para estas reglas y que de hacerlo, podría bajar considerablemente el rendimiento obtenido.

	Año 1997	Año 1998	Año 1999	Año 2000	Año 2001	Año 2002
Rendimiento total estimado	0.0772	0.4281	0.0457	0.0444	0.0653	-0.3458
Ratio beneficio ideal	0.0314	0.1332	0.0214	0.016	0.0213	-0.1045
Ratio de Sharpe	0.0241	0.0989	0.0156	0.0129	0.0168	-0.0816
Porcentaje en la predicción de los signos	0.8845	0.5121	0.5181	0.498	0.472	0.464
Test de Pesaran y Timmerman	0.5354	0.2826	0.5333	-0.0652	0	-1.2704
Rendimiento estrategia “comprar y mantener”	0.3522	0.3162	0.1503	-0.1356	-0.0653	-0.2626

Cuadro 4.2: Principales estadísticos relativos a la capacidad predictiva de la neuronal durante los años 1997 a 2002

### 4.2.3. Comportamiento de una estrategia, con posiciones recomendadas para estar fuera o dentro del mercado, optimizada mediante redes neuronales.

En el apartado 4.2.1 se llevó a cabo un estudio de la rentabilidad de un regla técnica donde se compraba y se vendía el mismo día, permitiendo obtener ganancias los días de caída del mercado sin más que vender a crédito. En este apartado emplearemos una estrategia en la que esta venta a crédito no está permitida. Así, si prevemos que el índice va a descender entonces permanecemos fuera del mercado para evitar esa pérdida, y obtenemos, además, un rendimiento libre de riesgo.

De esta forma, modificamos las señales creadas por la red. Si la salida para un día  $t$  es mayor que 0 crearemos una señal  $S_t = 1$ . Si se prevé que los rendimientos son positivos debemos permanecer en el mercado si ya estamos en él, o entrar en el mercado si estamos fuera. Si la salida para un día  $t$  es menor que 0 crearemos una señal  $S_t = 0$ , indicando que debemos permanecer fuera del mercado si estamos fuera de él, o salir del mercado si estamos dentro.

Con el objetivo de eliminar señales débiles o engañosas se ha introducido un filtro. En este sentido, una salida de la red mayor que ese filtro implica que debemos entrar en el mercado si no estamos ya en él. Si la salida es menor que ese filtro entonces debemos salir del mercado, si no estamos fuera de él. En otro caso, permanecemos manteniendo la posición del día anterior. El filtro irá desde el 0 hasta el 0.9, en saltos de 0.01, por lo que, estudiaremos la rentabilidad frente a 91 valores del filtro.

La medida de rentabilidad que emplearemos en este caso es el rendimiento medio anualizado deducido en la Ecuación 4.1.6 del apartado 4.1.

$$r = \sum_{t=1}^N S_{t-1} r_{m,t} + \sum_{t=1}^N (1 - S_{t-1}) r_{f,t} + d \cdot \log \frac{1 - c}{1 + c},$$

donde  $r_{f,t}$  es el rendimiento de un activo libre de riesgo el día  $t$ ,  $N$  es el número de días de mercado,  $d$  representa el número de transacciones (es decir, una señal de compra seguida de una señal de venta o *viceversa*), y  $S_t$  es una variable indicador cuyo valor es igual a uno si la regla produce señal de compra y cero en caso de señal de venta. En tal caso, el rendimiento medio anualizado viene dado por  $\pi_{rma} = e^r - 1$ .

Analizando más a fondo la Ecuación 4.1.6, observamos que si  $S_{t-1} = 1$ , donde la posición recomendada es la de estar en el mercado, el segundo sumando de la Ecuación [4.1.6] desaparece y se obtiene los rendimientos del mercado. Si  $S_{t-1} = 0$  la posición recomendada es la de estar fuera del mercado, el primer sumando desaparece por lo que se obtendrán rendimientos sólo por medio del activo libre de riesgo. Se ha estudiado la capacidad de generar beneficios de la regla ante distintos costes de transacción. Para ello, se seleccionó un abanico de costes que van desde el 0 %, hasta el 0.4 %, en saltos de 0.01 %. Por tanto la rentabilidad se ha medido enfrentando la regla a 41 costes de transacción.

Los rendimientos de esta estrategia se comparan con los rendimientos medios anualizados que se obtendría si utilizamos la estrategia B&H ajustada en el riesgo deducidos en la Ecuación 4.1.7.

$$r_{bha} = (1 - \alpha) \sum_{t=1}^N r_{f,t} + \alpha \sum_{t=1}^N r_{m,t} + \log \frac{1 - c}{1 + c}.$$

En esta fórmula  $\alpha$  expresa la proporción de días que la regla ha estado dentro del mercado. Este rendimiento es una media ponderada entre el rendimiento de



tomar una posición a largo en el mercado y el rendimiento libre de riesgo obtenido por tomar una posición no inversora en una proporción de  $(1 - \alpha)$  días. Así este rendimiento de la estrategia comprar y mantener ajustada en el riesgo expresa el rendimiento obtenido de invertir, tanto en el mercado como en el activo libre de riesgo, ponderado por los días que se ha estado dentro del mercado y fuera de él, respectivamente.

De esta forma, podremos obtener el exceso de rendimiento, ajustado en el riesgo según la Ecuación 4.1.8, que constituirá la variable de referencia.

$$\text{Exceso (B\&H)} = \pi_{rma} - \pi_{rbha}$$

De forma adicional, compararemos los excesos de rendimientos ajustados en el riesgo, con los proporcionados por un modelo alternativo ampliamente utilizado en la literatura financiera como es el modelo AR(1), que equivale a un modelo ARIMA (1,1,0) del IGBM.

Para evaluar estas reglas hemos considerado una serie del IGBM que contiene 4376 observaciones diarias, de precios al cierre, para el periodo comprendido entre el 2 de enero de 1979 y el 15 de octubre de 1997. Hemos realizado las predicciones en tres submuestras de 250 observaciones cada una, que corresponden a un tramo alcista, otro bajista y uno estable de la serie. Para el entrenamiento de la red se utilizará, en cada caso, todo el conjunto de observaciones anteriores disponibles a cada una de dichas submuestras. Tras realizar el proceso de aprendizaje, empleando sólo la información anterior, se realizarán las predicciones en cada submuestra. En cada caso se tiene un periodo de predicción de 250 observaciones diarias (aproximadamente un año).

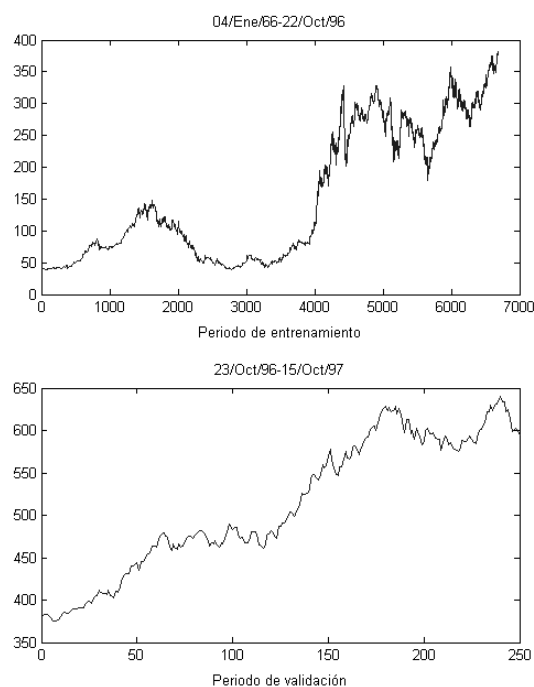


Figura 4.6: Tramos de entrenamiento y validación para el periodo alcista.

En las figuras 4.6, 4.7 y 4.8 se presentan los diferentes tramos de la serie, en cuyas 250 últimas observaciones hemos realizado la predicción y que se corresponde con un periodo de alzas, bajas y estable, respectivamente.

Los resultados se muestran en las siguientes figuras:

Las figuras 4.9, 4.13 y 4.17 muestran el exceso de rendimientos, ajustado en el riesgo, obtenido en función de los distintos costes y filtros aplicados a la reglas. Cada una de las 41 curvas de las figuras 4.10a, 4.14a y 4.18a reflejan el exceso de rendimientos, ajustados en el riesgo, para cada coste en función del filtro seleccionado. Las figuras 4.10b, 4.14b y 4.18b indican el coste máximo a partir del cual el exceso de rendimiento, ajustado por riesgo, es negativo (es decir, si se consideran costes por debajo del señalado se obtendrá excesos de rendimiento, ajustado por riesgo, positivos).

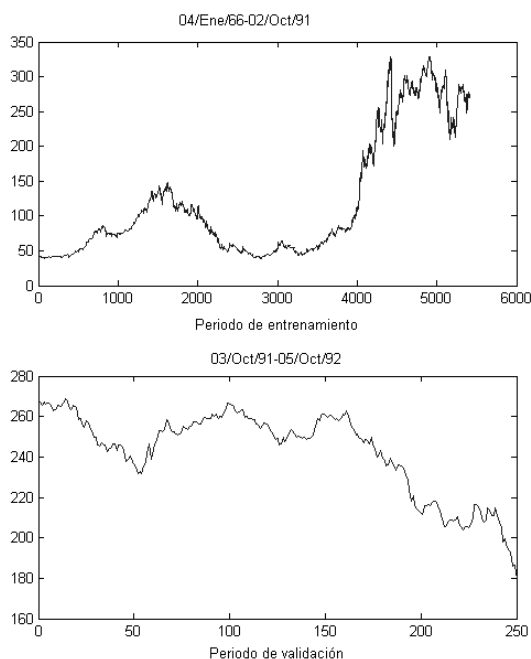


Figura 4.7: Tramos de entrenamiento y validación para el periodo bajista.

Como se puede observar en las Figuras 4.9, 4.13 y 4.17, al considerar filtros bajos el exceso de rendimiento ajustado por riesgo, es mucho más sensible al coste. Al existir mucho más días de transacción, el beneficio de la regla descenderá de forma drástica al incorporarle costes de transacción altos, provocando que, para algunos costes, el exceso de rendimiento ajustado por riesgo sea incluso negativo.

Sin embargo, cuando aumentamos el valor del filtro el exceso de rendimiento ajustado por riesgo tiende a igualarse para los distintos costes. Al aumentar el filtro, disminuyen los días totales de transacción, por lo que el beneficio no se ve tan influenciado por los costes. Cuando trabajamos con valores altos de los filtros, existen tan pocos días de transacción como los producidos por la estrategia B&H, ajustada por riesgo. De esta manera, los rendimientos obtenidos por la regla son prácticamente iguales a los obtenidos por la estrategia B&H, ajustada por riesgo,

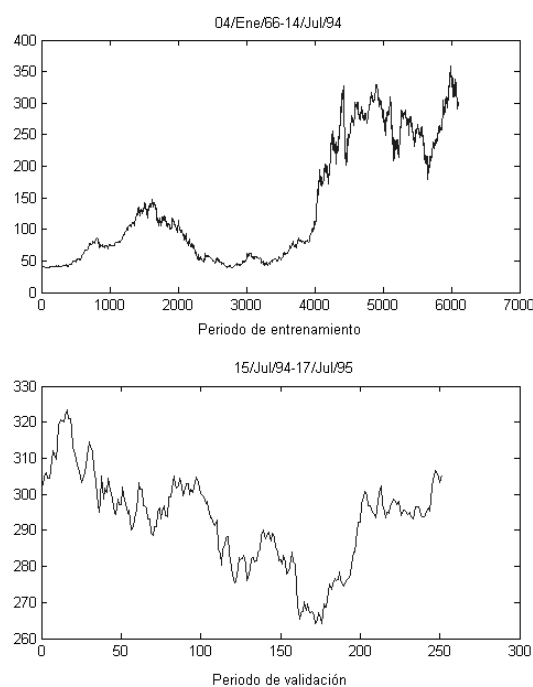


Figura 4.8: Tramos de entrenamiento y validación para el periodo estable.

produciéndose un exceso de rendimiento, ajustado por riesgo, muy cercano a cero.

La red posiciona al inversor en función de los nueve rendimientos anteriores al día a predecir. Actúa como si su filtro y coste de transacción fuesen iguales a 0. Es decir, intenta predecir el signo del rendimiento. Pero al no incluir una función objetivo que incluya los costes de transacción, puede que no seleccione la combinación adecuada de señales ante cambios en los costes. Esto se puede comprobar observando cómo los mayores excesos de rendimiento, ajustados en el riesgo, se producen en valores intermedios del filtro. Ello sugiere, probablemente, que para un coste dado, se obtiene una mejor combinación de señales y un mayor exceso de rendimiento.

Las figuras 4.11, 4.15 y 4.19 muestran la diferencia entre el exceso de rendimiento, ajustado por riesgo, de la regla y el exceso de rendimiento de un

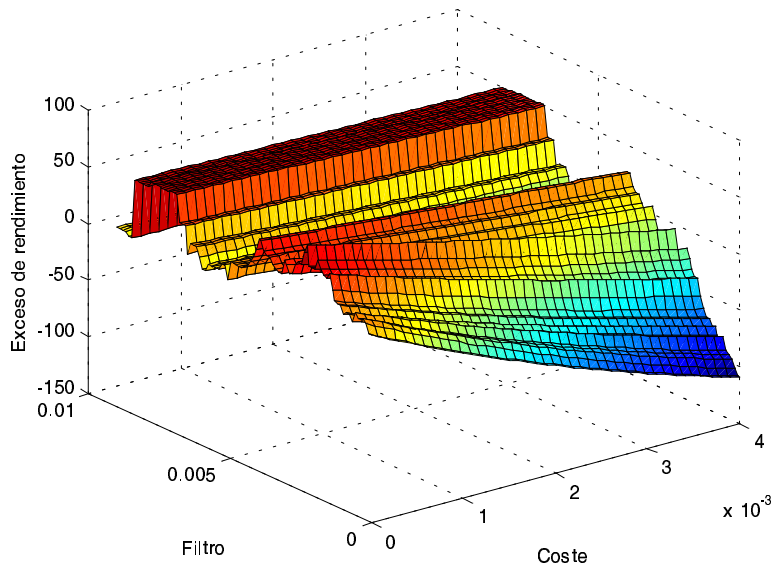


Figura 4.9: Exceso de rendimiento, ajustado en el riesgo, de la regla entrenada con redes neuronales. Tramo alcista.

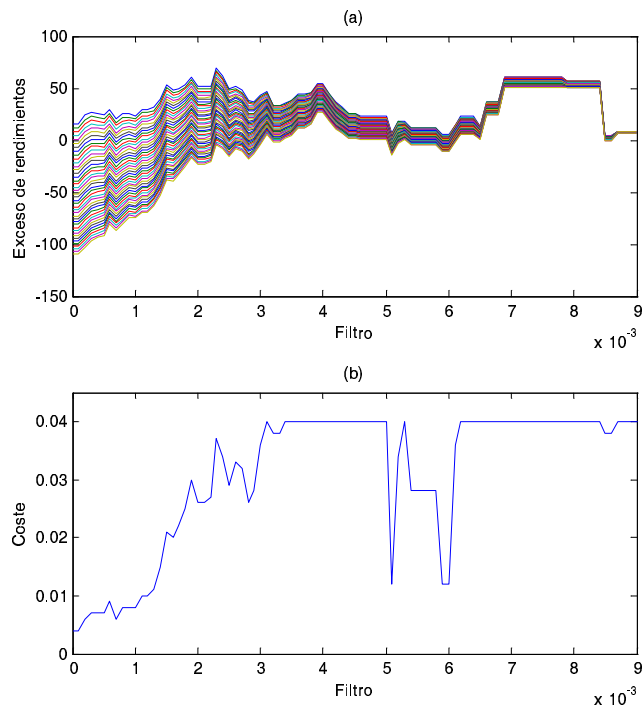


Figura 4.10: Exceso de rendimiento, ajustado en el riesgo, de la regla entrenada con redes neuronales (a). Coste por debajo del cual el exceso de rendimiento, ajustado en el riesgo es positivo (b). Tramo alcista.

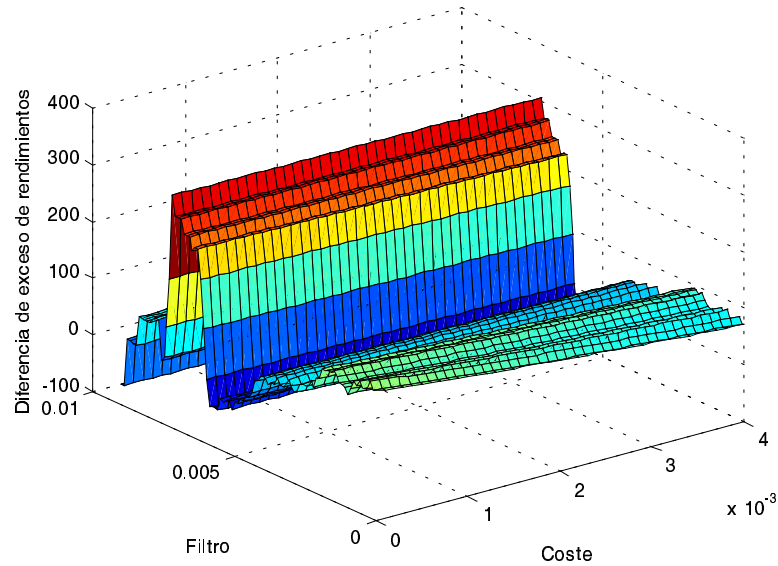


Figura 4.11: Diferencia entre el exceso de rendimiento obtenido a través de la regla y el modelo AR(1). Tramo alcista.

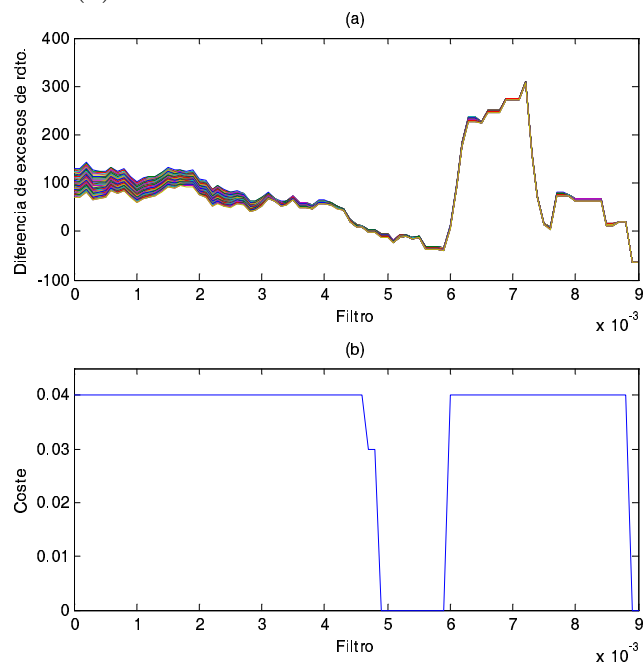


Figura 4.12: Diferencia entre el exceso de rendimiento obtenido a través de la regla y el modelo AR(1) (a). Coste por debajo del cual la diferencia del exceso de rendimiento, ajustado en el riesgo es positiva (b). Tramo alcista.

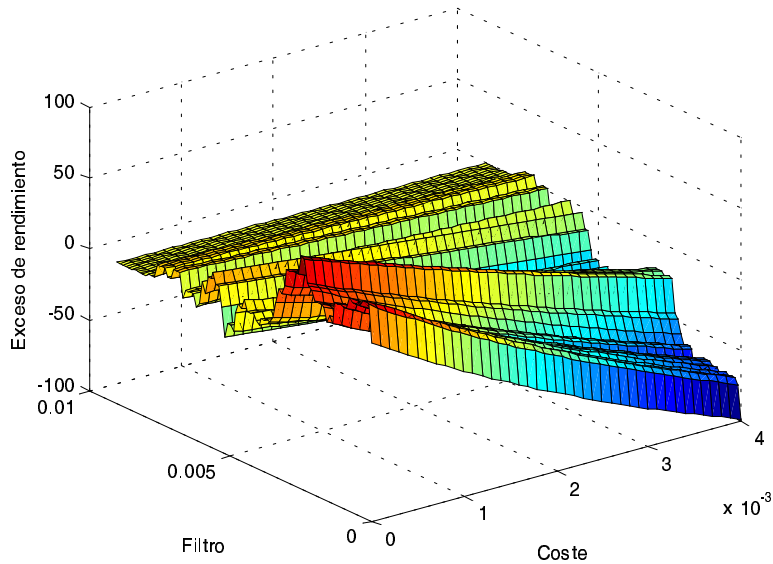


Figura 4.13: Exceso de rendimiento, ajustado en el riesgo, de la regla entrenada con redes neuronales. Tramo estable.

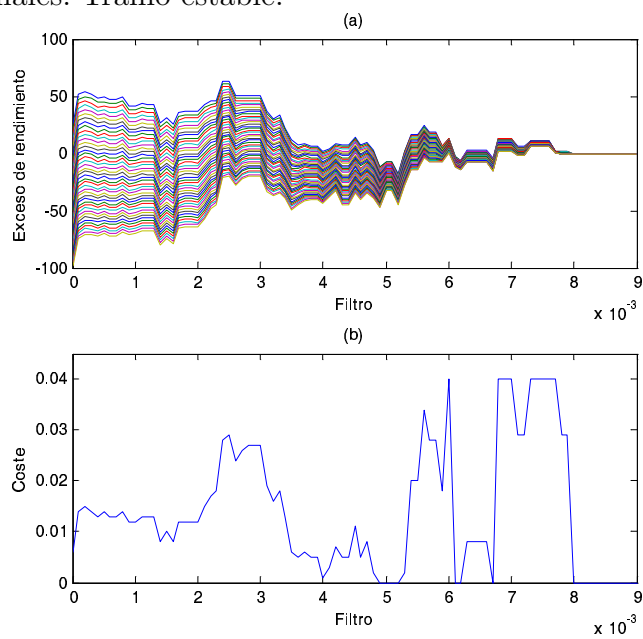


Figura 4.14: Exceso de rendimiento, ajustado en el riesgo, de la regla entrenada con redes neuronales (a). Coste por debajo del cual el exceso de rendimiento, ajustado en el riesgo es positivo (b). Tramo estable.

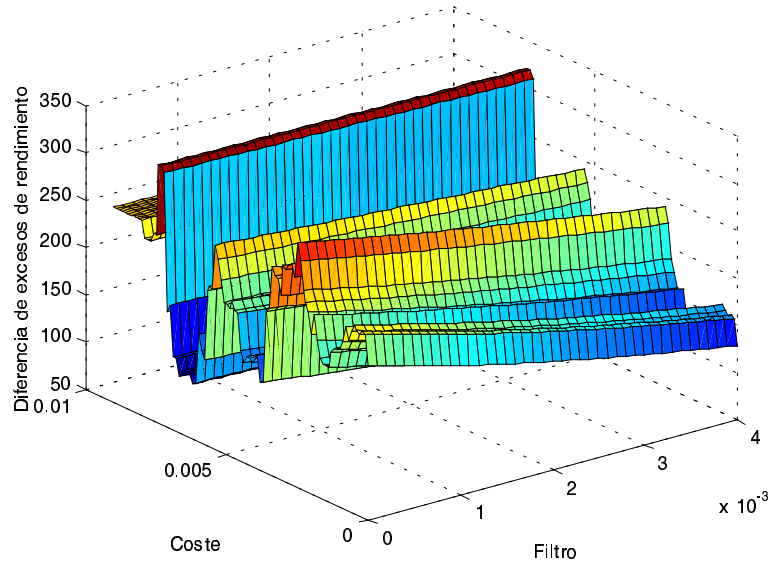


Figura 4.15: Diferencia entre el exceso de rendimiento obtenido a través de la regla y el modelo AR(1). Tramo estable.

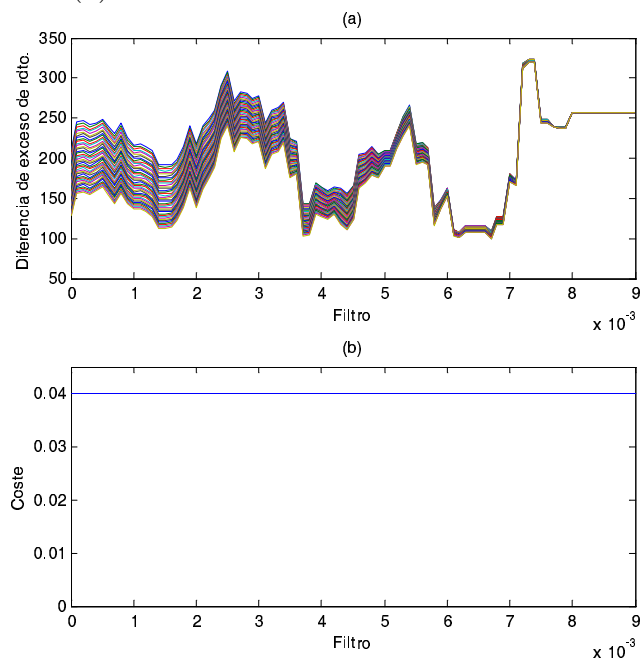


Figura 4.16: Diferencia entre el exceso de rendimiento obtenido a través de la regla y el modelo AR(1) (a). Coste por debajo del cual la diferencia del exceso de rendimiento, ajustado en el riesgo es positiva (b). Tramo estable.



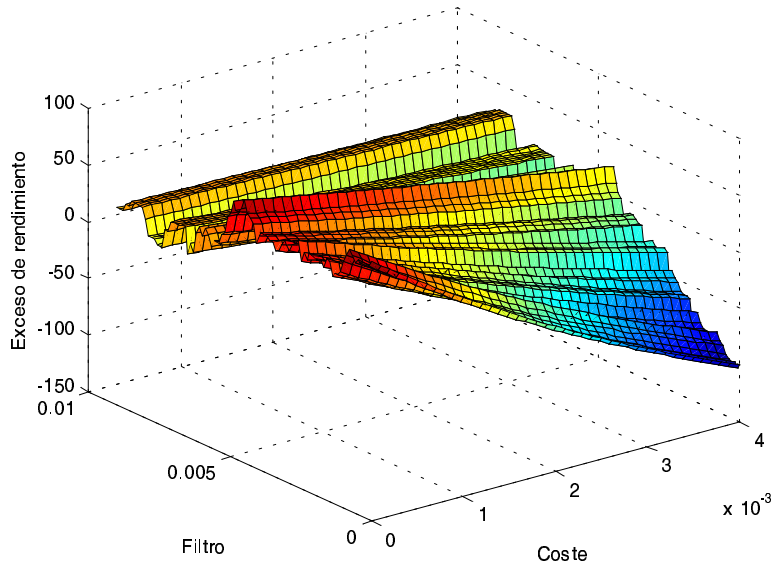


Figura 4.17: Exceso de rendimiento, ajustado en el riesgo, de la regla entrenada con redes neuronales. Tramo bajista.

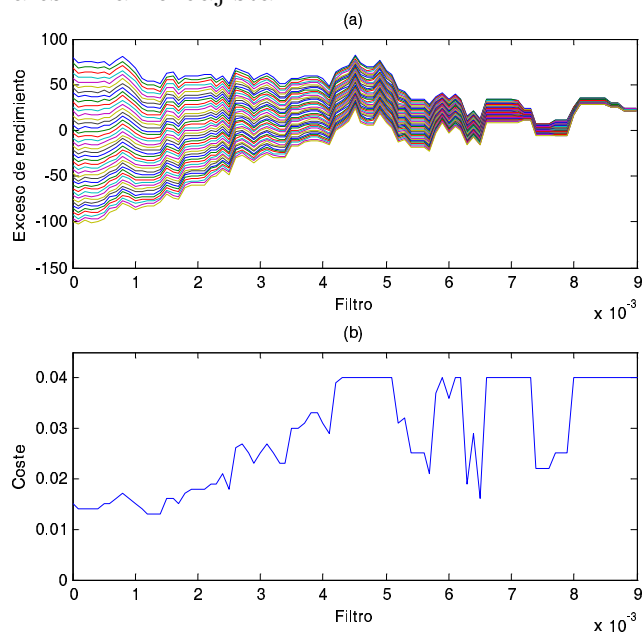


Figura 4.18: Exceso de rendimiento, ajustado en el riesgo, de la regla entrenada con redes neuronales (a). Coste por debajo del cual el exceso de rendimiento, ajustado en el riesgo es positivo (b). Tramo bajista.

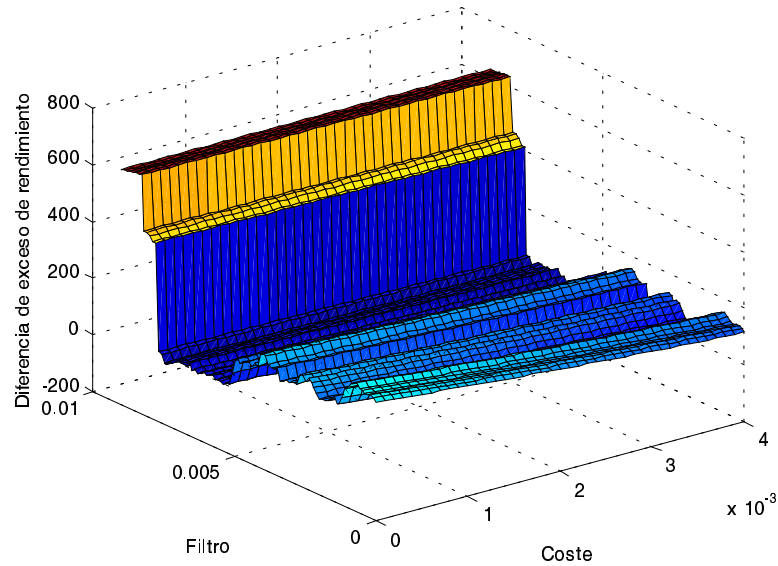


Figura 4.19: Diferencia entre el exceso de rendimiento obtenido a través de la regla y el modelo AR(1). Tramo bajista.

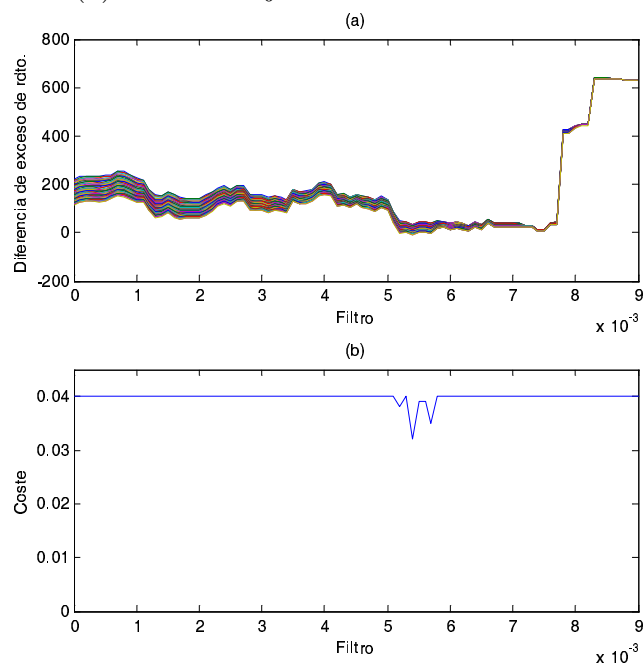


Figura 4.20: Diferencia entre el exceso de rendimiento obtenido a través de la regla y el modelo AR(1) (a). Coste por debajo del cual la diferencia del exceso de rendimiento, ajustado en el riesgo es positiva (b). Tramo bajista.

modelo AR(1). Cada una de las 41 curvas de las figuras 4.12a, 4.16a y 4.20a reflejan las diferencias entre el exceso de rendimientos, ajustados en el riesgo, para cada coste en función del filtro seleccionado. Las figuras 4.12b, 4.16b y 4.20b indican el coste por debajo del cual se obtendrán diferencias positivas.

El análisis en este caso es mucho más contundente. Se puede determinar que, excepto en casos puntuales del filtro, las diferencias siempre son positivas acentuándose dichas diferencias frente a valores altos del filtro.

En consecuencia, si las señales de compra y venta generadas por la red neuronal son filtradas, de modo que se eliminen las señales engañosas, el exceso de rendimientos sobre la estrategia de comprar y mantener generada por las predicciones de la red neuronal es superior al exceso de rendimiento sobre la estrategia de comprar y mantener generada por las predicciones del modelo AR(1).

Además, esos excesos de rendimientos generados por la red son consistentemente positivos para valores intermedios del filtro.

Finalmente, hay que señalar que el carácter positivo de los excesos de rendimientos sobre la estrategia de comprar y mantener para valores intermedios del filtro se produce indistintamente para los tramos alcistas, bajistas y estable.

### 4.3. Optimización de reglas técnicas mediante algoritmos genéticos

Como indicamos en la sección 3.3.2, los algoritmos genéticos (AGs) constituyen una forma de búsqueda adaptativa, mediante técnicas de optimización basadas en la evolución natural. Fueron desarrollados por Holland (1962, 1975) y han sido implementados en multitud de aplicaciones.

Nos apoyaremos en el trabajo de Lo y MacKinlay (1988), que establecía en sus conclusiones que las especificaciones *ad-hoc* de los modelos empleados en la predicción financiera podían provocar *figoneo* en los datos. Es decir, que la elección arbitraria, tanto de las reglas como de los parámetros seleccionados, producían desviaciones en los resultados. Asimismo, concluían que de haber empleados técnicas de optimización para la elección de dichos parámetros, estas desviaciones se podían haber reducido e incluso, en algunos casos, eliminado.

Utilizaremos reglas ya conocidas y empleadas ampliamente en los mercados bursátiles y cambiarios. Estas reglas son las basadas en medias móviles y en órdenes estadísticos y que han sido descritas en el apartado 2.3.2. Hasta ahora la elección de los parámetros de estas reglas había sido realizada de forma arbitraria por parte del inversor. En este apartado emplearemos un procedimiento de optimización para encontrar dichos parámetros. Concretamente, utilizaremos algoritmos genéticos para hallar los valores de los parámetros de las reglas técnicas, de forma que se optimice su capacidad de generar beneficios.

Así, seguiremos los pasos descritos en la sección 3.3.2 para implementar estos algoritmos.

Antes de generar la población inicial se debe definir la función a optimizar y la

### 2104.3. Optimización de reglas técnicas mediante algoritmos genéticos

estructura de los cromosomas que serán los sujetos de las operaciones genéticas. Como hemos indicado en el apartado 3.3.2, estos cromosomas contienen dígitos binarios, que no son más que la codificación de los parámetros del modelo. Esta representación binaria es necesaria para después proceder a los procesos de cruce y mutación. El rango de variación que puede tomar un parámetro determina el número necesario de dígitos binarios que se necesitan para codificarlo. Por lo tanto, si estamos tratando con un problema de  $n$  variables, la longitud del cromosoma vendrá dado como la suma de los dígitos binarios necesarios para los  $n$  parámetros.

#### **Reglas basadas en medias móviles**

En las reglas basadas en las medias móviles, y descritas en el apartado 2.3.2, el problema a maximizar sería el rendimiento de la regla. Esto conlleva que trabajemos con tres parámetros: el tamaño de la media móvil a corto, el tamaño de la media móvil a largo, y el ancho de banda. El rango de variación de  $\theta_1$  oscila entre 1 día y 250 días ( $1 \leq \theta_1 \leq 250$ ). Tal como se vio en el apartado 3.3.2 el tamaño de un cromosoma viene dado por la Ecuación 3.3.1

$$tc = \text{int} \frac{\ln(vmax - vmin + 1)}{\ln 2},$$

siendo  $tc$  el tamaño del cromosoma e “int” la función que recoge la parte entera de un número.

Por lo que se puede determinar que se necesitan 8 dígitos binarios para codificar los números comprendidos entre 1 y 250. El rango de variación de  $\theta_2$  oscila entre 2 y 250, por lo que también se necesitan 8 dígitos binarios para representar este parámetro. Sin embargo,  $\theta_3$  sólo oscila entre 1 y 100. De esta

forma, se necesitan 7 dígitos para codificarlo. Por tanto, se utilizarán cromosomas de  $8+8+7=23$  elementos para representar los parámetros de las reglas técnicas basadas en medias móviles. Sus ocho primeros elementos son los dígitos binarios necesarios para representar todos los posibles valores de  $\theta_1$ , los ocho siguientes los que representan  $\theta_2$  y los siete últimos representan los posibles valores de  $\theta_3$ . Así, el vector se simbolizaría mediante el cromosoma  $\Theta = [\theta_1|\theta_2|\theta_3]$ .

### **Reglas basadas en órdenes estadísticos**

Aunque establecíamos en el apartado 2.3.2, que las reglas técnicas basadas en órdenes estadísticos operaban con tres parámetros, en este trabajo emplearemos cuatro. El parámetro  $\phi_1$  representa el número de observaciones previas dentro del cual se selecciona el máximo o mínimo de referencia. Así, si  $\phi_1 = 7$ , el precio máximo y mínimo se escogerán de entre los precios al cierre de sesión ocurridos en la semana anterior al periodo estudiado.

Existen dos criterios para determinar este parámetro.

- Que sea el propio analista quien de forma arbitraria lo fije.
- Que sea el mercado quien lo determine tomando la serie de precios registrada desde la última transacción, para que se seleccione el precio máximo y el precio mínimo (Alexander, 1961).

La elección de este criterio también puede ser realizada por el algoritmo genético. Creamos un parámetro  $\phi_4$  que puede tomar dos valores: 0 y 1. Si  $\phi_4 = 0$ , entonces el parámetro  $\phi_1$  es igual al número de días habidos desde la última transacción (determinado de forma implícita). El algoritmo no maximizará  $\phi_1$  porque está determinado por el mercado. Si  $\phi_4 = 1$ , esto indicaría que es el

### 2124.3. Optimización de reglas técnicas mediante algoritmos genéticos

analista quien elige el número de días  $y$ , por lo tanto, el algoritmo tiene control sobre el parámetro  $\phi_1$  e intentará encontrar aquel que maximice la función de rendimientos.

El parámetro  $\phi_2$  representa el porcentaje a partir del cual deberíamos tomar las decisiones de compra o venta. Si el precio se mueve dentro de este porcentaje con respecto al precio máximo o mínimo, dependiendo si se trata de una regla filtro o una regla túnel, entonces el analista no tomará ninguna decisión y mantendrá la posición en la que se encontraba el día anterior. Si por el contrario el precio supera  $\phi_2$ , entonces cambiará de posición.

El parámetro  $\phi_3$  determinará si estamos trabajando con una regla filtro  $\phi_3 = 1$  o una regla túnel  $\phi_3 = 0$ .

Se necesitan cromosomas de 17 bits para representar los parámetros de las reglas técnicas basadas en órdenes estadísticos. Sus ocho primeros bits son los dígitos binarios necesarios para representar todos los posibles valores de  $\phi_1$ , ya que puede oscilar entre 1 y 250 días. Los siete siguientes representan  $\phi_2$ , ya que el rango de variación está entre 1 y 100. Sólo se necesita un dígito para representar todo el rango de valores de  $\phi_3$  y otro para el de  $\phi_4$ . Así, el vector se representaría mediante el cromosoma  $\Phi = [\phi_1|\phi_2|\phi_3|\phi_4]$ .

Una vez establecidas las estructuras de los cromosomas entonces se podrán iniciar los algoritmos. Generamos una población inicial de cromosomas de longitud 23 para el caso de las reglas basadas en medias móviles, elegidos al azar, y cromosomas de longitud 17 para las reglas basadas en órdenes estadísticos. Hay que señalar que se tratan de dos algoritmos independientes, uno para cada tipo de reglas, ya que los cromosomas son de tamaño fijo y no pueden cohabitar dentro

de un mismo algoritmo cromosomas de distinto tamaño. En nuestro trabajo comenzaremos con una población inicial de 150 individuos, número que Bauer (1994) también utiliza en otro trabajo similar. La elección del parámetro “número de cromosomas” siempre debe realizarse atendiendo al dilema existente entre la exploración adecuada del espacio de soluciones y el tiempo deseado de las iteraciones.

A continuación seleccionamos a los individuos que participarán en las operaciones de cruce y mutación. En este proceso de selección empleamos un método de selección de progenitores, basado en el *ranking* de los candidatos con respecto a la función objetivo, desarrollada por Whitley (1989). Se trata de una transformación de los algoritmos de selección  $(\mu, \lambda)$ . Se seleccionan todos los candidatos excepto el que peor valor ha obtenido en la función de adaptación, que se sustituye por el mejor.

Posteriormente estos 150 individuos se cruzan entre ellos para dar lugar a descendencia a través del proceso de cruce. Se escogen dos individuos al azar y si superan la probabilidad del 0.6 se cruzan entre si. La probabilidad de cruce también es un parámetro fundamental en los algoritmos genéticos, ya que si la probabilidad es alta el algoritmo convergerá rápidamente, pero puede provocar una búsqueda ineficiente. Por el contrario, si la probabilidad de cruce es baja tardará más tiempo en converger, pero la búsqueda se realizará en todo el espacio de soluciones, evitando caer en óptimos locales. Una vez superada la probabilidad de cruce, se selecciona un punto del cromosoma a partir del cual los dos progenitores intercambiarán sus genes. Es el punto de cruce simple.

Así, a través de los procesos de selección y cruce, los mejores individuos



### 2144.3. Optimización de reglas técnicas mediante algoritmos genéticos

van retransmitiendo sus genes a las siguientes generaciones. El último proceso que interviene en la generación de nuevos individuos es de la mutación. Ésta garantiza la diversidad de la población y, de esta forma, la búsqueda puede encaminarse hacia regiones más prometedoras del espacio de soluciones, evitando el estancamiento en zonas menos satisfactorias. La mutación no es más que la introducción de shocks aleatorios en los cromosomas a través de cambios en sus genes. Sin embargo, la probabilidad de mutación en la población debe ser muy baja, con el fin de no destruir las buenas estructuras que han sido previamente obtenidas y evitar que la búsqueda se convierta en una búsqueda aleatoria. En nuestro caso, escogimos una probabilidad de mutación del 0.05. De nuevo, hay que seleccionar con cuidado este parámetro. De ser muy elevado provocaríamos una búsqueda aleatoria, de ser pequeño no lograríamos el objetivo de la mutación.

La nueva población creada tras los procesos de selección, cruce y mutación se evalúa mediante la función objetivo y vuelve a pasar por los tres procesos citados anteriormente. Este proceso iterativo continúa hasta que se satisface una de las siguientes tres condiciones:

- La población converge hasta un único individuo.
- Se llega a un número máximo de iteraciones determinado.
- Durante un número determinado de iteraciones no hay mejoras en el mejor individuo de la población.

Finalmente, se han seleccionado los siguientes parámetros para la ejecución de los algoritmos. La población inicial será de 150 individuos y se estableció la probabilidad de cruce y mutación al 0.6 y 0.05, respectivamente. El número de

iteraciones y de iteraciones sin mejoras se fijó en 300 y 300, respectivamente. Si la población seleccionada, la probabilidad de mutación, el número de iteraciones máxima y el número de iteraciones máxima sin repetición son pequeñas y la probabilidad de cruce alta, el algoritmo convergerá más rápidamente. Sin embargo, esto puede provocar una búsqueda ineficiente en el espacio de soluciones y que el algoritmo finalice en un óptimo local (Goldberg, 1989; Bauer, 1994), por lo que, como hemos dicho, para la elección de estos parámetros hay que establecer un equilibrio entre tiempo de computación y exploración del espacio de soluciones. La función objetivo a maximizar es el rendimiento medio anualizado, que viene dada por la Ecuación 4.1.6.

$$r = \sum_{t=1}^N S_{t-1} r_{m,t} + \sum_{t=1}^N (1 - S_{t-1}) r_{f,t} + d \cdot \log \frac{1 - c}{1 + c},$$

Hemos empleado costes de transacción del 0%, del 0.25% y del 0.45%, para medir la capacidad que tienen las reglas optimizadas mediante algoritmos genéticos de generar beneficios ante distintos costes.

Esta estrategia se compara con los rendimientos medios anualizados que se obtendría si utilizamos la estrategia B&H ajustada en el riesgo [Ecuación 4.1.7].

$$r_{bha} = (1 - \alpha) \sum_{t=1}^N r_{f,t} + \alpha \sum_{t=1}^N r_{m,t} + \log \frac{1 - c}{1 + c}.$$

Emplearemos, asimismo, una medida que estime el riesgo de la inversión. Para estas reglas utilizaremos la ratio de Sharpe.

## 2164.3. Optimización de reglas técnicas mediante algoritmos genéticos

### **4.3.1. Ejercicio experimental hasta 1997**

En este apartado se ha utilizado una muestra de 4376 observaciones diarias del IGBM, al cierre, desde el desde el 2 de Enero de 1979 hasta el 15 de Noviembre de 1997. El algoritmo se entrenará para un periodo de 2188 datos (desde 02/01/79 hasta el 16/12/88) que posteriormente se validará con los 2188 datos posteriores. En la Figura 4.21 se muestra la serie de datos utilizados. Igualmente, con el fin de tener en cuenta el coste de oportunidad hemos hecho uso de la serie de tipo de interés a un día en el mercado interbancario, durante ese mismo periodo, proporcionada por el Banco de España.

Se trataría en definitiva de simular lo más realista posible el comportamiento de un operador de mercado que gestiona activos financieros para sus clientes. Por una parte, puede asegurar sus recursos en activos en el mercado de renta variable cuya evolución se seguirá mediante el IGBM. Por otra parte, puede recurrir al mercado interbancario como fuente alternativa para la obtención de un rendimiento para sus recursos. Em ambos casos se fija el horizonte de la inversión en un día, revisándose continuamente las decisiones relativas a acudir a uno u otro mercado.

Este algoritmo lo hemos ejecutado nueve veces, donde en las tres primeras ejecuciones se consideraron unos costes de transacción del 0.45%, que lo equiparamos al coste aplicado a operaciones con particulares. En las tres segundas los costes son del 0.25%, que es el que se aplicaría a clientes institucionales y por último las tres últimas donde se aplican costes de transacción nulos.

Los cuadros 4.3 y 4.4 muestran las mejores iteraciones calculadas para cada coste de transacción en el periodo de entrenamiento y validación respectivamente.

coste	mmc	mml	filtro	rs	rma	dt	bha
0%	2	21	0	0.0104	38.2122	117	19.2140
0.25%	207	242	0	0.0072	33.3044	5	25.3620
0.45%	207	242	0	0.0072	33.1304	5	25.3620

Cuadro 4.3: Parámetros hallados en el periodo de entrenamiento y principales estadísticos utilizados para la evaluación en el periodo de entrenamiento de las reglas técnicas basadas en medias móviles optimizadas mediante algoritmos genéticos. Media móvil a corto (mmc); Media móvil a largo (mml); Filtro; Ratio de Sharpe (rs); Rendimiento medio anualizado (rma); Días de transacción (dt); B&H ajustado en el riesgo (bha);

coste	mmc	mml	filtro	rsval	rmaval	dtval	bhaval
0%	2	21	0	0.0074	22.0491	139	10.055
0.25%	207	242	0	0.0047	14.6343	6	10.8609
0.45%	207	242	0	0.0046	14.4539	6	10.8609

Cuadro 4.4: Parámetros hallados en el periodo de entrenamiento y principales estadísticos utilizados para la evaluación en el periodo de validación de las reglas técnicas basadas en medias móviles optimizadas mediante algoritmos genéticos. Media móvil a corto (mmc); Media móvil a largo (mml); Filtro; Ratio de Sharpe en el periodo de validación (rsval); Rendimiento medio anualizado en el periodo de validación (rmaval); Días de transacción en el periodo de validación (dtval); B&H ajustado en el riesgo en el periodo de validación (bhaval)

### 2184.3. Optimización de reglas técnicas mediante algoritmos genéticos

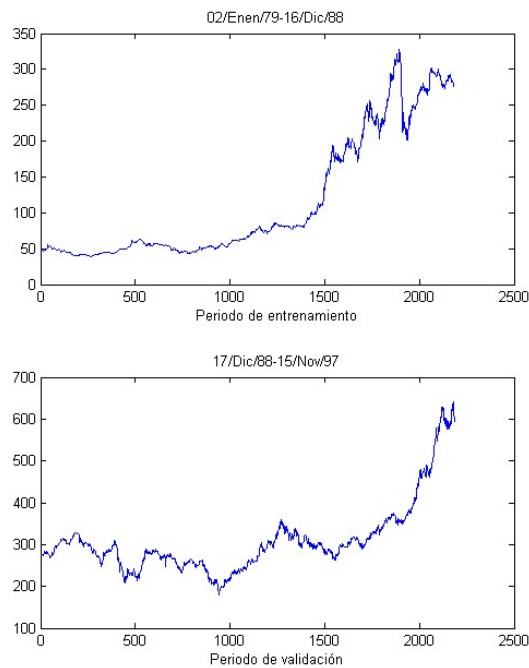


Figura 4.21: Tramo de entrenamiento y predicción del algoritmo genético

En las columnas de la 2 a 4 aparecen los parámetros optimizados para cada coste. En la quinta columna se muestra los ratios de sharpe, los rendimientos medios anualizados para cada coste en la sexta y en la séptima los días de transacción que hubo en el periodo analizado. La octava columna contiene los rendimientos medios anualizados que se obtendría si utilizamos la estrategia “buy and hold” ajustada en el riesgo en lugar de la regla técnica.

Como puede observarse en los cuadros 4.3 y 4.4, las medias móviles superan, para todos los costes, el rendimiento de la estrategia B&H, tanto en el periodo de entrenamiento como en el periodo de validación. Para costes altos el algoritmo ha seleccionado reglas muy poco activas donde se efectúa un número reducido de transacciones, mientras que para costes bajos ha seleccionado reglas mucho más activas.

#### 4. Resultados empíricos: Aplicaciones a la predicción financiera del IGBM.

219

Coste	N Obs.	Filtro	RF/RC	Imp/Exp	rs	rma	dt	bha
0 %	243	75	1	0	0.0139	18.972	208	16.1507
0.25 %	235*	53	1	1	0.047	15.035	6	16.0795
0.45 %	234*	54	1	1	0.0502	14.856	6	16.0706

Cuadro 4.5: Parámetros hallados en el periodo de entrenamiento y principales estadísticos utilizados para la evaluación en el periodo de entrenamiento de las reglas técnicas basadas en órdenes estadísticos optimizadas mediante algoritmos genéticos. Número de observaciones (N Obs.), Porcentaje de filtro (filtro), Regla Filtro o Regla túnel (RF/RC), Determinación del número de observaciones por el mercado o de forma explícita (Imp/Exp), Ratio de Sharpe en el periodo de entrenamiento (rs), Rendimiento medio anualizado en el periodo de entrenamiento (rma), Días de transacción en el periodo de entrenamiento (dt), B&H ajustado en el riesgo en el periodo de entrenamiento (bha)

Podemos observar que el algoritmo es sensible a los cambios en los costes. Cuando los costes son altos y, por lo tanto, una estrategia muy activa podría hacer mella en los beneficios, el algoritmo es capaz de seleccionar reglas técnicas con parámetros altos. De esta forma se suavizan las media móviles, lo que implica que existen pocos puntos de corte y, por lo tanto, pocos días de transacción. Para costes bajos, donde no importa tanto el número de transacciones, la regla selecciona reglas con parámetros bajos y, por lo tanto muy activas. El hecho de que existan muchas transacciones se ve compensado por los bajos costes, lo que no redunda en un descenso del beneficio.

Los cuadros 4.5 y 4.6 muestran los parámetros optimizados para las reglas técnicas basadas en órdenes estadísticos para los periodos de entrenamiento y validación respectivamente. La columna 3 expresa si la regla seleccionada está basada en la regla filtro (1) o bien en la regla túnel (0). Se puede observar cómo para los tres costes de transacción se escogieron reglas filtros. La columna 4 indica si el número de observaciones para la determinación de los precios

### 2204.3. Optimización de reglas técnicas mediante algoritmos genéticos

Coste	N Obs.	Filtro	RF/RC	Imp/Exp	rsva	rmava	dtva	bhava
0 %	243	75	1	0	0.2149	10.499	0	12.3861
0.25 %	235*	53	1	1	0.0434	10.958	4	12.332
0.45 %	234*	54	1	1	0.0467	10.845	4	12.3253

Cuadro 4.6: Parámetros hallados en el periodo de entrenamiento y principales estadísticos utilizados para la evaluación en el periodo de validación de las reglas técnicas basadas en órdenes estadísticos optimizadas mediante algoritmos genéticos. Número de observaciones (N Obs.), Porcentaje de filtro (filtro), Regla Filtro o Regla túnel (RF/RC), Determinación del número de observaciones por el mercado o de forma explícita (Imp/Exp), Ratio de Sharpe en el periodo de validación (rsva), Rendimiento medio anualizado en el periodo de validación (rmava), Días de transacción en el periodo de validación (dtva), B&H ajustado en el riesgo en el periodo de entrenamiento (bhava)

máximos y mínimos ha sido hecha explícitamente por el algoritmo, o bien, ha sido determinado implícitamente en el mercado. Para los costes de transacción del 0.25 % y 0.45 % el número de observaciones se determinó por el mercado (y se remarca con un asterisco en la primera columna) y se escogió como número de observaciones el último día de transacción. Para un coste del 0 % el periodo por el cual se determinan el soporte y la resistencia es de 235 días, solo que en este caso, fue el propio algoritmo quien lo determinó. Las columnas 6,7 y 8 expresan la ratio de Sharpe, el rendimiento medio anualizado y los días de transacción dentro del periodo de entrenamiento. La columna 9 indica el rendimiento B&H ajustado en el riesgo.

En este caso, los resultados son menos prometedores ya que el rendimiento de la regla no superó a la estrategia B&H en el periodo de validación. Es significativo como la regla obtenida con un coste del 0 % es muy activa en el periodo de entrenamiento y, sin embargo, muy poco activa en el periodo de validación.

Las reglas técnicas optimizadas, por medio de AGs, consiguen mejorar los

rendimientos de la estrategia B&H ajustada en el riesgo, en los periodos de entrenamiento y para todos los escenarios de costes de transacción, cuando se emplean reglas técnicas basadas en medias móviles.

Las reglas técnicas basadas en órdenes estadísticos superaron el rendimiento de la estrategia B&H ajustado en el riesgo cuando los costes de transacción eran del 0%. Sin embargo, eso no fue posible en el periodo de validación.

A modo de resumen podríamos decir que, nuestros resultados no son capaces de confirmar ni desmentir la afirmación, que hacen muchos teóricos del mercado, de que las medias móviles son indicadores que siguen la curva de cotizaciones confirmando la tendencia en vigor, en ese momento, pero que no son capaces de anticipar los cambios de tendencia. Los costes de transacción influyen decisivamente en la formación de las reglas técnicas. Con costes altos el AG suele encontrar reglas con pocos días de transacción, mientras que con costes bajos encuentra reglas mucho más activas, lo que se traduce en un incremento del rendimiento.

#### 4.3.2. Análisis empírico para los años 1997 a 2002

Con una muestra comprendida entre 2 de enero de 1979 hasta 31 de diciembre del 2002 hemos intentado medir la capacidad de generar beneficios en el IGBM de las reglas técnicas optimizadas mediante algoritmos genéticos para los años 1997 a 2002. Cada año constituye un periodo de predicción. El AG optimizará los parámetros de la regla tomando en cuenta los datos anteriores a dicho periodo de predicción. Se consideraron costes del 0.10% y 0.25% para medir la respuesta del algoritmo ante variaciones en los costes.

Los estadísticos empleados se corresponde a los anteriormente analizados y



### 2224.3. Optimización de reglas técnicas mediante algoritmos genéticos

los resultados se muestran en los Cuadros 4.7 y 4.8

De alguna forma, los resultados confirman las observaciones que hacíamos cuando analizábamos estas técnicas desde una perspectiva metodológica. Al aumentar los costes el algoritmo es capaz de adaptarse antes esos cambios para reducir los días de transacción y, de esta manera, intentar que el impacto de los costes sobre el beneficio sea menor. Además para los años 1997, 1998 y 2001 el rendimiento medio anualizado ha sido capaz de superar a la estrategia B&H ajustada en el riesgo. No obstante, en periodos de bajas generalizadas (2001), el rendimiento obtenido por la regla es negativo. Pero podemos observar cómo con un coste del 0.25 % el rendimiento total obtenido por nuestra regla técnica para los 6 años (51.402) es superior al obtenido por la estrategia B&H ajustada por riesgo (33.984), dando a entender que el algoritmo ha conseguido ajustarse para proporcionar beneficios por encima del la estrategia B&H ajustada en el riesgo para costes altos.

Año	Coste	Media móvil a corto	Media móvil a largo	Filtro	Rendimiento medio anualizado	Ratio de Sharpe	Rendimiento B&H ajustado en el riesgo	Ratio de Sharpe B&H ajustado en el riesgo	Dt	Por.
Año 1997	0.10 % 0.25 %	2 2	3 21	1 1	32.5324 28.1644	0.0101 0.0087	13.7167 13.8503	0.0062 0.0056	1139 235	51.20 % 57.08 %
Año 1998	0.10 % 0.25 %	2 2	3 20	1 52	31.2992 28.2024	0.0096 0.0085	14.2599 14.5729	0.0063 0.0056	1223 169	51.62 % 58.55 %
Año 1999	0.10 % 0.25 %	3 2	22 21	3 1	32.1367 30.0012	0.0093 0.0088	14.9096 14.9537	0.0055 0.0056	217 263	58.02 % 57.91 %
Año 2000	0.10 % 0.25 %	2 2	20 19	4 72	30.6914 28.0298	0.0088 0.0080	14.7390 14.8650	0.0055 0.0053	295 181	57.70 % 59.98 %
Año 2001	0.10 % 0.25 %	2 3	19 22	3 6	27.8551 24.7043	0.0080 0.0072	13.5984 13.5552	0.0051 0.0050	332 254	56.95 % 57.50 %
Año 2002	0.10 % 0.25 %	2 2	21 32	1 25	26.1677 23.4819	0.0075 0.0068	12.8275 12.8088	0.0047 0.0045	330 194	56.97 % 58.63 %

Cuadro 4.7: Parámetros hallados por el algoritmo genético en el periodo de entrenamiento para los años 1997 a 2002 en el IGBM y valor de los estadísticos estudiados evaluados en el periodo de entrenamiento. Dt: Días de transacción, Por: Porcentaje de días en el mercado

Año	Coste	Media móvil a corto	Media móvil a largo	Filtro	Rendimiento medio anualizado	Ratio de Shrape	Rendimiento B&H ajustado en el riesgo	Ratio de Sharpe B&H ajustado en el riesgo	Dt	Por.
Año 1997	0.10 %	2	3	1	15.1997	0.0040	25.7932	0.0077	83	59.11 %
	0.25 %	2	21	1	26.3018	0.0064	24.7053	0.0073	19	59.51 %
Año 1998	0.10 %	2	3	1	2.7046	0.0005	20.9426	0.0045	82	61.38 %
	0.25 %	2	20	52	54.5081	0.0104	16.3513	0.0039	9	55.69 %
Año 1999	0.10 %	3	22	3	7.2064	0.0024	9.8938	0.0044	17	58.63 %
	0.25 %	2	21	1	2.2383	0.0008	10.0093	0.0044	21	59.44 %
Año 2000	0.10 %	2	20	4	-15.1781	-0.0054	-2.5149	-0.0009	22	39.76 %
	0.25 %	2	19	72	-16.9011	-0.0055	-3.9841	-0.0014	16	44.98 %
Año 2001	0.10 %	2	19	3	-6.5724	-0.0016	-4.5035	-0.0016	24	40.73 %
	0.25 %	3	22	6	-2.2834	-0.0006	-4.6684	-0.0016	14	40.32 %
Año 2002	0.10 %	2	21	1	-12.0080	-0.0034	-7.4922	-0.0029	22	35.89 %
	0.25 %	2	32	25	-12.4617	-0.0038	-8.4294	-0.0032	14	37.10 %

Cuadro 4.8: Parámetros hallados por el algoritmo genético en el periodo de entrenamiento para los años 1997 a 2002 en el IGBM y valor de los estadísticos en el periodo de validación. Dt: Días de transacción, Por: Porcentaje de días en el mercado

## 4.4. Optimización de los parámetros utilizados en predicciones por ocurrencias análogas mediante algoritmos genéticos

### 4.4.1. Ejercicio experimental hasta 1997

A continuación vamos a estudiar la capacidad de generar beneficios de varias reglas técnicas de contratación, que hemos elaborado empleando predicciones NN, sobre la serie del Índice General de la Bolsa de Madrid (IGBM). En este trabajo se ha utilizado una muestra de 4376 observaciones diarias del IGBM, al cierre, desde el desde el 2 de Enero de 1979 hasta el 15 de Noviembre de 1997, cuya representación aparece en la Figura 4.21. Igualmente hemos hecho uso de la serie de tipo de interés a un día de la peseta, durante ese mismo período, que ha sido facilitada por el Banco de España.

Con de fin de evaluar el comportamiento de las reglas previamente definidas sobre el IGBM, hemos considerado diez períodos de predicción, consecutivos, de 250 observaciones que corresponden aproximadamente a un año de duración. Con el fin de estimar los parámetros óptimos de cada una de las reglas, se considera un período de entrenamiento consistente en los 250 días inmediatamente anteriores a cada período de predicción. El AGs optimizará los valores de los parámetros que la definen para el período de entrenamiento; finalmente los parámetros óptimos obtenidos en cada conjunto de entrenamiento son empleados en el correspondiente conjunto de predicción, para obtener diversos indicadores que revelarán la capacidad de obtener beneficios de cada regla técnica.

Para el caso de la regla simple, los parámetros esenciales de la predicción  $\hat{P}(d, k, \alpha)$  son la dimensión de inmersión  $d$ , el número de entornos  $k$  y el ancho de

banda  $\alpha$ , que empleamos para desechar señales engañosas. Tales parámetros serán determinados minimizando, mediante un algoritmo genético la función de exceso de rendimientos sobre la estrategia B&H ajustada por riesgo que ya encontramos en la Ecuación 4.1.8.

$$\text{Exceso (B\&H)} = \pi_{rma} - \pi_{rbha}$$

Para tener un abanico de respuestas ante distintos costes, se consideraron tres tipos de costes de transacción unidireccionales del 0 %, 0.10 % y del 0.25 %. Para nuestra regla técnica basada en las predicciones NN, el rendimiento del activo libre de riesgo se consideró igual al del tipo de interés a un día.

Con el fin de estudiar el comportamiento de ambas reglas de transacción hemos considerado diversos estadísticos que describen su comportamiento.

- El *rendimiento neto medio anualizado* [Ecuación 4.1.6]

$$r = \sum_{t=1}^N S_{t-1} r_{m,t} + \sum_{t=1}^N (1 - S_{t-1}) r_{f,t} + d \cdot \log \frac{1 - c}{1 + c},$$

de la regla técnica de contratación consiste en su rendimiento bruto menos los costes de transacción asociados a las sucesivas entradas y salidas del mercado.

- El *rendimiento neto de la estrategia B&H, ajustada por riesgo* [Ecuación 4.1.7]

$$r_{bha} = (1 - \alpha) \sum_{t=1}^N r_{f,t} + \alpha \sum_{t=1}^N r_{m,t} + \log \frac{1 - c}{1 + c},$$

es la media ponderada del rendimiento neto de la estrategia B&H con el rendimiento obtenido por el activo libre de riesgo. Dicha ponderación se

establece mediante unos coeficientes que representan los días dentro y fuera del mercado que establece la regla técnica.

- La *ratio de Sharpe* [Ecuación 4.1.10] es el rendimiento medio neto, anualizado, de una estrategia de inversión por unidad de riesgo. Se obtiene a partir de la siguiente expresión:  $RS = \frac{\bar{r}}{\sigma}$  donde  $\bar{r}$  es el rendimiento medio anualizado obtenido a través de las reglas técnicas,  $\sigma$  es la desviación típica de los rendimientos diarios.
- La *ratio de Sharpe de la estrategia B&H ajustada en el riesgo* [Ecuación 4.1.12] que representa el rendimiento medio de la dicha estrategia por unidad de riesgo.

$$RS_{bha} = \frac{\alpha \cdot \overline{r_{m,t}} + (1 - \alpha) \cdot \overline{r_{f,t}}}{\sigma_{bh}}$$

- Por último, la variable *exceso de rendimientos sobre la estrategia B&H, ajustada por riesgo* [Ecuación 4.1.8], que constituye un indicador clave para comprobar el comportamiento de nuestras reglas técnicas.

Sin embargo, para medir la rentabilidad de las reglas técnicas que se han elaborado a partir de las predicciones NN se ha usado el exceso de rendimiento que se define como

$$\text{Exceso (B\&H)} = \pi_{rma} - \pi_{rbha}$$

Si dicho exceso es positivo nuestra regla vence a la estrategia *Buy & Hold*, asumiendo igualdad de riesgos. Si dicho exceso es negativo nuestra estrategia es inferior a la de B&H.

Asimismo, se ha utilizado un modelo AR(1) para comparar el exceso de rendimiento producido por este modelo y el producido por nuestra regla. El modelo AR(1) representa uno de los modelos econométricos más utilizados en finanzas debido a la característica de los rendimientos de los mercados financieros. Además, se han empleado reglas optimizadas por algoritmos genéticos pero cuya función objetivo es la raíz cuadrada del error cuadrático medio de predicción.

$$R_k(d) = \sqrt{\sum_{x_t \in N} \frac{(\hat{x}_t - x_t)^2}{N}}. \quad (4.4.1)$$

De esta forma, intentamos comparar los rendimientos obtenidos por reglas optimizadas por medio de un algoritmos genético cuya función objetivo incluya los costes de transacción frente a otra, también optimizada mediante algoritmos genéticos, pero cuya función objetivo no contemple los costes de transacción.

En este apartado hemos considerado diez períodos de predicción, consecutivos y no solapados, de 250 días cada uno. Cada una de las reglas técnicas consideradas se obtiene optimizando sus parámetros en otro período de 250 días inmediatamente anterior al período de predicción. Finalmente, en cada uno de los diez períodos de predicción se evalúan los diferentes tipos de rendimientos que señalan el comportamiento de cada una de las reglas consideradas. Los Cuadros 4.9 a 4.55 muestran los resultados de las reglas en cada uno de los diez periodos de predicción.

Tramo1			
Coste	0 %	0.10 %	0.25 %
Dimensión de inmersión	6	1	1
Porcentaje de puntos próximos	3	1.7	3.5
Ancho de banda	5	43	87
Número de puntos próximos	123	70	144

Cuadro 4.9: Parámetros obtenidos mediante AGs para las reglas basadas en predicciones NN en el tramo 1

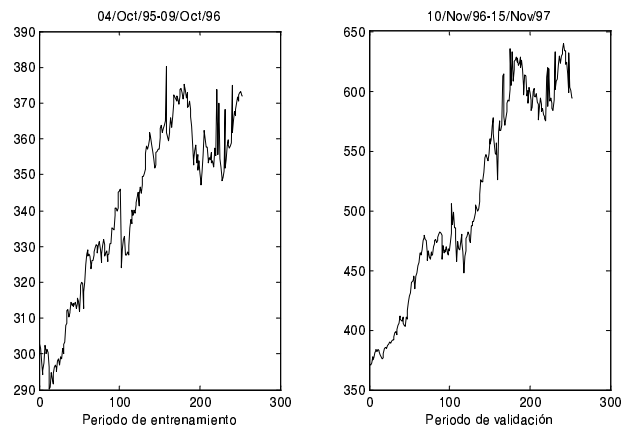


Figura 4.22: Periodo de entrenamiento y validación para el tramo 1.



230 **4.4. Optimización de los parámetros utilizados en predicciones por  
ocurrencias análogas mediante algoritmos genéticos**

Tramo1					
	Coste	Periodo de entrenamiento		Periodo de evaluación	
		Modelo NN	AR(1)	Modelo NN	AR(1)
Exceso de rendimiento ajustado en el riesgo	0 %	19.1504	3.56	3.7761	18.71
	0.10 %	6.5645	-9.81	0.8496	3.48
	0.25 %	5.7574	-27.05	0.859	-16.49
Rendimiento medio anualizado	0 %	35.9246	20.5025	41.5238	54.3732
	0.10 %	17.85	6.8881	48.0428	38.8702
	0.25 %	14.6074	-10.7046	54.104	18.4854
Ratio de Sharpe	0 %	0.0129	0.0079	0.0107	0.0140
	0.10 %	0.0132	0.0028	0.0102	0.0106
	0.25 %	0.0139	-0.0048	0.0102	0.0054
Rendimiento B&H ajustado en el riesgo*	0 %	16.7742	16.94	37.7476	35.66
	0.10 %	11.2855	16.70	47.1932	35.39
	0.25 %	8.85	16.35	53.245	34.98
Ratio de Sharpe estrategia B&H ajustada en el riesgo*	0 %	0.0088	0.0086	0.0118	0.0125
	0.10 %	0.0156	0.0086	0.0114	0.0125
	0.25 %	0.0348	0.0086	0.0112	0.125
Porcentaje de días en el mercado	0 %	56.75 %	58.80 %	62.95 %	56.75 %
	0.10 %	22.22 %	58.80 %	79.28 %	56.75 %
	0.25 %	7.94 %	58.80 %	89.64 %	56.75 %
Días de transacción	0 %	97	118	104	105
	0.10 %	28	118	11	105
	0.25 %	8	118	1	105

Cuadro 4.10: Estadísticos obtenidos para las reglas encontradas para el tramo 1.  
\*: A la hora de calcular estos estadísticos se ha ponderado por el porcentaje de días de cada uno de los modelos.

Dimensión de inmersión	6					
	25					
Porcentaje de puntos próximos	0.0000	0.2000	0.4000	0.6000	0.8000	1.0000
Anchura del filtro	0.0000	0.2000	0.4000	0.6000	0.8000	1.0000
Rend. Neto de la regla	38.8381	39.8754	37.3436	36.6564	41.0805	42.3068
Rend. B&H	30.5501	30.7125	31.3623	31.8497	32.9868	32.8244
Exceso de rend. sobre B&H	8.2880	9.1629	5.9813	4.8068	8.0937	9.4824
Ratio de Sharpe	0.1861	0.1935	0.1800	0.1761	0.1944	0.2008
Ratio de Sharpe B&H	0.1904	0.1901	0.1892	0.1885	0.1871	0.1873
Porcentaje de aciertos	57.1428 %	57.1428 %	57.1428 %	57.1428 %	57.1428 %	57.1428 %
Porcentaje de días en el mercado	60.3174 %	60.7142 %	62.3015 %	63.4920 %	66.2698 %	65.8730 %

Cuadro 4.11: Regla técnica con predicciones NN y función objetivo Raíz cuadrada de los error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 1 y coste del 0 %

Dimensión de inmersión	6					
Porcentaje de puntos próximos	25					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	29.2381	30.8754	28.9436	28.4564	34.0805	35.9068
Rend. B&H	30.4294	30.5911	31.2377	31.7227	32.8543	32.6926
Exceso de rend. sobre B&H	-1.1913	0.2843	-2.2941	-3.2663	1.2262	3.2142
Ratio de Sharpe	0.1400	0.1498	0.1393	0.1365	0.1611	0.1704
Ratio de Sharpe B&H	0.1896	0.1894	0.1885	0.1878	0.1863	0.1865
Porcentaje de aciertos	57.1428 %	57.1428 %	57.1428 %	57.1428 %	57.1428 %	57.1428 %
Porcentaje de días en el mercado	60.3174 %	60.7142 %	62.3015 %	63.4920 %	66.2698 %	65.8730 %

Cuadro 4.12: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadráticos medio. Parámetros obtenidos mediante AGs para el tramo 1 y coste del 0.10 %

Dimensión de inmersión Porcentaje de puntos próximos Anchura del filtro	6 25					
	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	14.8381	17.3754	16.3436	16.1564	23.5805	26.3068
Rend. B&H	30.2485	30.4090	31.0508	31.5322	32.6555	32.4950
Exceso de rend. sobre B&H	-15.4104	-13.0336	-14.7073	-15.3758	-9.0750	-6.1882
Ratio de Sharpe	0.0706	0.0838	0.0780	0.0768	0.1107	0.1241
Ratio de Sharpe B&H	0.1885	0.1883	0.1873	0.1867	0.1852	0.1854
Porcentaje de aciertos	57.1428 %	57.1428 %	57.1428 %	57.1428 %	57.1428 %	57.1428 %
Porcentaje de días en el mercado	60.3174 %	60.7142 %	62.3015 %	63.4920 %	66.2698 %	65.8730 %

Cuadro 4.13: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 1 y coste del 0.25 %

234 **4.4. Optimización de los parámetros utilizados en predicciones por  
ocurrencias análogas mediante algoritmos genéticos**

---

Tramo2			
Coste	0 %	0.10 %	0.25 %
Dimensión de inmersión	7	1	1
Porcentaje de puntos próximos	3.3	3.7	3.7
Ancho de banda	43	73	73
Número de puntos próximos	127	143	143

Cuadro 4.14: Parámetros obtenidos mediante AGs para las reglas basadas en predicciones NN en el tramo 2

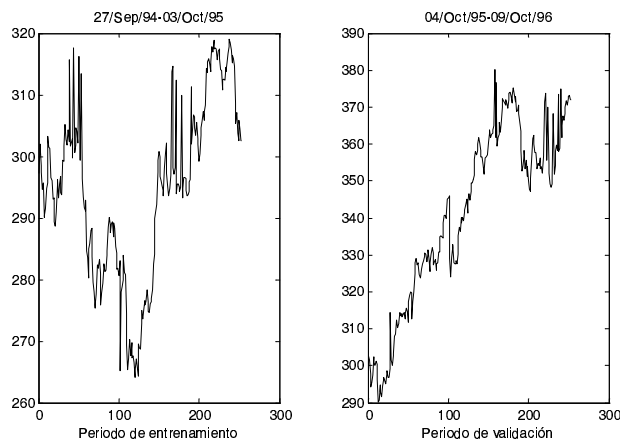


Figura 4.23: Periodo de entrenamiento y validación para el tramo 2.

Tramo2					
	Coste	Periodo de entrenamiento		Periodo de evaluación	
		Modelo NN	AR(1)	Modelo NN	AR(1)
Exceso de rendimiento ajustado en el riesgo	0 %	21.1342	10.9697	3.4849	18.71
	0.10 %	18.1832	-1.4685	-2.432	-9.8160
	0.25 %	16.1016	-17.5805	-3.6198	-26.9544
Rendimiento medio anualizado	0 %	25.585	15.6633	-1.1035	19.7553
	0.10 %	22.2569	3.0134	-2.432	6.2193
	0.25 %	19.8586	-13.4153	7.1662	-11.2708
Ratio de Sharpe	0 %	0.011	0.0064	0.0066	0.0077
	0.10 %	0.0083	0.0013	0.0056	0.0026
	0.25 %	0.0076	-0.0062	0.0046	-0.0051
Rendimiento B&H ajustado en el riesgo*	0 %	4.4509	4.6935	18.8812	16.2705
	0.10 %	4.0737	4.4819	11.1256	16.0354
	0.25 %	3.757	4.1652	10.786	15.6836
Ratio de Sharpe estrategia B&H ajustada en el riesgo*	0 %	0.0028	0.0029	0.0088	0.0082
	0.10 %	0.0026	0.0029	0.0179	0.0082
	0.25 %	0.0026	0.0029	0.0179	0.0082
Porcentaje de días en el mercado	0 %	47.62 %	48.80 %	63.35 %	58.73 %
	0.10 %	49.60 %	48.80 %	19.12 %	58.73 %
	0.25 %	49.60 %	48.80 %	19.12 %	58.73 %
Días de transacción	0 %	65	114	73	119
	0.10 %	12	114	9	119
	0.25 %	12	114	9	119

Cuadro 4.15: Estadísticos obtenidos para las reglas encontradas para el tramo 2.  
\*: A la hora de calcular estos estadísticos se ha ponderado por el porcentaje de días de cada uno de los modelos.

Dimensión de inmersión	5					
Porcentaje de puntos próximos	25					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	22.9556	23.4102	23.0297	25.3497	22.7730	22.4948
Rend. B&H	15.3055	15.4040	15.5518	15.4040	15.4040	15.5518
Exceso de rend. sobre B&H	7.6501	8.0062	7.4779	9.9457	7.3690	6.9430
Ratio de Sharpe	0.1595	0.1620	0.1566	0.1743	0.1583	0.1548
Ratio de Sharpe B&H	0.1401	0.1390	0.1375	0.1390	0.1390	0.1375
Porcentaje de aciertos	54.7619 %	54.7619 %	54.7619 %	54.7619 %	54.7619 %	54.7619 %
Porcentaje de días en el mercado	56.7460 %	57.5396 %	58.7301 %	57.5396 %	57.5396 %	58.7301 %

Cuadro 4.16: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 2 y coste del 0 %

Dimensión de inmersión	5					
	25					
Porcentaje de puntos próximos	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	12.1556	12.8102	12.8297	15.9497	13.5730	13.4948
Rend. B&H	15.1920	15.2889	15.4343	15.2889	15.2889	15.4343
Exceso de rend. sobre B&H	-3.0364	-2.4787	-2.6046	0.6608	-1.7159	-1.9395
Ratio de Sharpe	0.0846	0.0888	0.0874	0.1099	0.0945	0.0929
Ratio de Sharpe B&H	0.1390	0.1380	0.1365	0.1380	0.1380	0.1365
Porcentaje de aciertos	54.7619 %	54.7619 %	54.7619 %	54.7619 %	54.7619 %	54.7619 %
Porcentaje de días en el mercado	56.7460 %	57.5396 %	58.7301 %	57.5396 %	57.5396 %	58.7301 %

Cuadro 4.17: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 2 y coste del 0.10 %



Dimensión de inmersión	5					
	25					
Porcentaje de puntos próximos						
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	-4.0445	-3.0898	-2.4704	1.8496	-0.2271	-0.0053
Rend. B&H	15.0217	15.1163	15.2581	15.1163	15.1163	15.2581
Exceso de rend. sobre B&H	-19.0662	-18.2061	-17.7285	-13.2667	-15.3433	-15.2634
Ratio de Sharpe	-0.0279	-0.0212	-0.0167	0.0126	-0.0016	0.0000
Ratio de Sharpe B&H	0.1375	0.1364	0.1349	0.1364	0.1364	0.1349
Porcentaje de aciertos	54.7619 %	54.7619 %	54.7619 %	54.7619 %	54.7619 %	54.7619 %
Porcentaje de días en el mercado	56.7460 %	57.5396 %	58.7301 %	57.5396 %	57.5396 %	58.7301 %

Cuadro 4.18: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 2 y coste del 0.25 %

Tramo3			
Coste	0 %	0.10 %	0.25 %
Dimensión de inmersión	5	5	5
Porcentaje de puntos próximos	1.3	3.2	3.5
Ancho de banda	62	78	83
Número de puntos próximos	47	115	126

Cuadro 4.19: Parámetros obtenidos mediante AGs para las reglas basadas en predicciones NN en el tramo 3

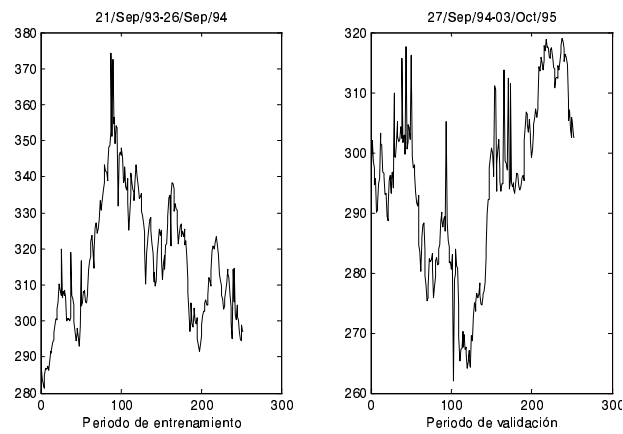


Figura 4.24: Periodo de entrenamiento y validación para el tramo 3.

4.4. Optimización de los parámetros utilizados en predicciones por  
240 ocurrencias análogas mediante algoritmos genéticos

Tramo3					
	Coste	Periodo de entrenamiento		Periodo de evaluación	
		Modelo NN	AR(1)	Modelo NN	AR(1)
Exceso de rendimiento ajustado en el riesgo	0 %	36.7492	22.8096	5.6148	11.1786
	0.10 %	26.1336	9.7897	-5.265	-1.4578
	0.25 %	13.7472	-7.2508	-14.7207	-17.8246
Rendimiento medio anualizado	0 %	43.5248	29.6371	10.9873	17.4188
	0.10 %	32.6338	16.4012	-0.3354	4.5677
	0.25 %	19.894	-0.9624	-10.0549	-12.1206
Ratio de Sharpe	0 %	0.0139	0.0086	0.0049	0.0070
	0.10 %	0.011	0.0050	-0.0002	0.0019
	0.25 %	0.0071	-0.0003	-0.0046	-0.0055
Rendimiento B&H ajustado en el riesgo*	0 %	6.7756	6.8275	5.3725	6.2403
	0.10 %	6.5002	6.6115	4.9296	6.0255
	0.25 %	6.1468	6.2884	4.6658	5.7041
Ratio de Sharpe estrategia B&H ajustado en el riesgo*	0 %	0.0033	0.0030	0.0037	0.0037
	0.10 %	0.0032	0.0030	0.0033	0.0037
	0.25 %	0.0031	0.0030	0.0034	0.0037
Porcentaje de días en el mercado	0 %	44.84 %	51.60 %	43.03 %	48.81 %
	0.10 %	46.43 %	51.60 %	46.22 %	48.81 %
	0.25 %	47.22 %	51.60 %	45.42 %	48.81 %
Días de transacción	0 %	85	106	90	115
	0.10 %	69	106	79	115
	0.25 %	65	106	77	115

Cuadro 4.20: Estadísticos obtenidos para las reglas encontradas para el tramo 3.  
\*: A la hora de calcular estos estadísticos se ha ponderado por el porcentaje de días de cada uno de los modelos.

Dimensión de inmersión	3					
Porcentaje de puntos próximos	65					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	16.3440	14.5278	15.0660	14.3991	13.2160	15.4882
Rend. B&H	5.0255	5.1055	4.9988	5.0521	4.9454	5.0255
Exceso de rend. sobre B&H	11.3186	9.4223	10.0672	9.3470	8.2706	10.4628
Ratio de Sharpe	0.1148	0.1007	0.1027	0.0998	0.0899	0.1038
Ratio de Sharpe B&H	0.0479	0.0498	0.0473	0.0485	0.0461	0.0479
Porcentaje de aciertos	53.5714 %	53.5714 %	53.5714 %	53.5714 %	53.5714 %	53.5714 %
Porcentaje de días en el mercado	51.5873 %	50.3968 %	51.9841 %	51.1904 %	52.7777 %	51.5873 %

Cuadro 4.21: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 3 y coste del 0 %

Dimensión de inmersión	3					
Porcentaje de puntos próximos	65					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	4.7440	3.3278	4.2660	3.7991	3.2160	6.2882
Rend. B&H	4.9223	5.0047	4.8948	4.9498	4.8399	4.9223
Exceso de rend. sobre B&H	-0.1783	-1.6769	-0.6288	-1.1507	-1.6239	1.3659
Ratio de Sharpe	0.0330	0.0228	0.0287	0.0260	0.0216	0.0417
Ratio de Sharpe B&H	0.0469	0.0488	0.0463	0.0475	0.0451	0.0469
Porcentaje de aciertos	53.5714 %	53.5714 %	53.5714 %	53.5714 %	53.5714 %	53.5714 %
Porcentaje de días en el mercado	51.5873 %	50.3968 %	51.9841 %	51.1904 %	52.7777 %	51.5873 %

Cuadro 4.22: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 3 y coste del 0.10 %

Dimensión de inmersión	3					
Porcentaje de puntos próximos	65					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	-12.6560	-13.4723	-11.9340	-12.1010	-11.7840	-7.5118
Rend. B&H	4.7675	4.8535	4.7389	4.7962	4.6815	4.7675
Exceso de rend. sobre B&H	-17.4236	-18.3258	-16.6729	-16.8971	-16.4656	-12.2794
Ratio de Sharpe	-0.0859	-0.0896	-0.0776	-0.0803	-0.0771	-0.0486
Ratio de Sharpe B&H	0.0454	0.0474	0.0448	0.0461	0.0436	0.0454
Porcentaje de aciertos	53.5714 %	53.5714 %	53.5714 %	53.5714 %	53.5714 %	53.5714 %
Porcentaje de días en el mercado	51.5873 %	50.3968 %	51.9841 %	51.1904 %	52.7777 %	51.5873 %

Cuadro 4.23: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 3 y coste del 0.25 %

4.4. Optimización de los parámetros utilizados en predicciones por  
 244 ocurrencias análogas mediante algoritmos genéticos

Tramo4			
Coste	0 %	0.10 %	0.25 %
Dimensión de inmersión	4	4	4
Porcentaje de puntos próximos	2.3	2.3	2.3
Ancho de banda	78	84	86
Número de puntos próximos	77	77	77

Cuadro 4.24: Parámetros obtenidos mediante AGs para las reglas basadas en predicciones NN en el tramo 4

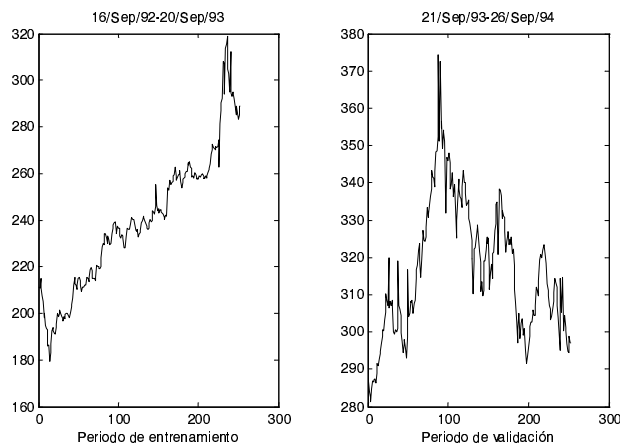


Figura 4.25: Periodo de entrenamiento y validación para el tramo 4.

Tramo4					
	Coste	Periodo de entrenamiento		Periodo de evaluación	
		Modelo NN	AR(1)	Modelo NN	AR(1)
Exceso de rendimiento ajustado en el riesgo	0 %	53.8128	19.8083	27.9669	22.7080
	0.10 %	40.8637	3.9612	17.2092	9.8681
	0.25 %	23.2333	-16.5174	1.7253	-6.9547
Rendimiento medio anualizado	0 %	79.9833	44.8142	34.9145	28.8290
	0.10 %	66.7781	28.7143	23.9265	15.7746
	0.25 %	48.7645	7.8576	8.1042	-1.3693
Ratio de Sharpe	0 %	0.0139	0.0120	0.0049	0.0084
	0.10 %	0.017	0.0081	0.0082	0.0049
	0.25 %	0.0131	0.0024	0.003	-0.0026
Rendimiento B&H ajustado en el riesgo*	0 %	26.1705	25.0059	6.9476	6.1211
	0.10 %	25.9144	24.7531	6.7172	5.9065
	0.25 %	25.5312	24.3750	6.3788	5.5855
Ratio de Sharpe estrategia B&H ajustado en el riesgo*	0 %	0.0101	0.0099	0.0032	0.0026
	0.10 %	0.0101	0.0099	0.0032	0.0026
	0.25 %	-0.0017	0.0099	0.0097	0.0026
Porcentaje de días en el mercado	0 %	55.16 %	54.40 %	47.81 %	51.59 %
	0.10 %	46.43 %	55.16 %	46.22 %	0.4821
	0.25 %	47.22 %	55.16 %	45.42 %	48.61 %
Días de transacción	0 %	74	116	81	106
	0.10 %	74	116	81	106
	0.25 %	74	116	81	106

Cuadro 4.25: Estadísticos obtenidos para las reglas encontradas para el tramo 4.  
\*: A la hora de calcular estos estadísticos se ha ponderado por el porcentaje de días de cada uno de los modelos.



Dimensión de inmersión	2					
Porcentaje de puntos próximos	51					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	20.0171	19.1484	21.0320	22.4126	21.3994	20.5876
Rend. B&H	5.4773	5.3432	5.3879	5.3208	5.2761	5.2091
Exceso de rend. sobre B&H	14.5398	13.8052	15.6441	17.0918	16.1233	15.3785
Ratio de Sharpe	0.1062	0.1015	0.1124	0.1176	0.1120	0.1073
Ratio de Sharpe B&H	0.0381	0.0356	0.0364	0.0352	0.0344	0.0333
Porcentaje de aciertos	54.7619 %	54.7619 %	54.7619 %	54.7619 %	54.7619 %	54.7619 %
Porcentaje de días en el mercado	53.1746 %	55.5555 %	54.7619 %	55.9523 %	56.7460 %	57.9365 %

Cuadro 4.26: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 4 y coste del 0 %

Dimensión de inmersión	2					
Porcentaje de puntos próximos	51					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	8.8171	8.9484	11.4320	13.2126	12.1994	11.7876
Rend. B&H	5.3709	5.2321	5.2784	5.2089	5.1626	5.0932
Exceso de rend. sobre B&H	3.4462	3.7163	6.1536	8.0037	7.0367	6.6944
Ratio de Sharpe	0.0468	0.0477	0.0614	0.0695	0.0641	0.0616
Ratio de Sharpe B&H	0.0374	0.0349	0.0357	0.0345	0.0337	0.0325
Porcentaje de aciertos	54.7619 %	54.7619 %	54.7619 %	54.7619 %	54.7619 %	54.7619 %
Porcentaje de días en el mercado	53.1746 %	55.5555 %	54.7619 %	55.9523 %	56.7460 %	57.9365 %

Cuadro 4.27: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 4 y coste del 0.10 %

Dimensión de inmersión	2					
	51					
Porcentaje de puntos próximos						
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	-7.9829	-6.3517	-2.9680	-0.5874	-1.6007	-1.4125
Rend. B&H	5.2114	5.0654	5.1141	5.0411	4.9924	4.9194
Exceso de rend. sobre B&H	-13.1943	-11.4171	-8.0821	-5.6285	-6.5931	-6.3319
Ratio de Sharpe	-0.0421	-0.0338	-0.0159	-0.0031	-0.0084	-0.0074
Ratio de Sharpe B&H	0.0363	0.0337	0.0346	0.0333	0.0326	0.0314
Porcentaje de aciertos	54.7619 %	54.7619 %	54.7619 %	54.7619 %	54.7619 %	54.7619 %
Porcentaje de días en el mercado	53.1746 %	55.5555 %	54.7619 %	55.9523 %	56.7460 %	57.9365 %

Cuadro 4.28: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 4 y coste del 0.25 %

Tramo5			
Coste	0 %	0.10 %	0.25 %
Dimensión de inmersión	5	5	5
Porcentaje de puntos próximos	2.1	2.1	2.2
Ancho de banda	37	36	57
Número de puntos próximos	65	65	68

Cuadro 4.29: Parámetros obtenidos mediante AGs para las reglas basadas en predicciones NN en el tramo 5

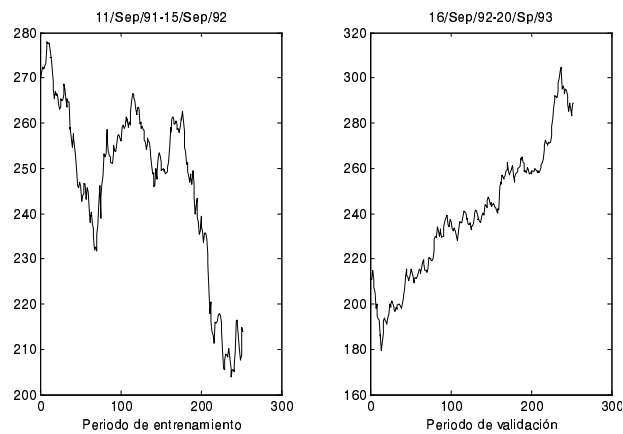


Figura 4.26: Periodo de entrenamiento y validación para el tramo 5.

250 **4.4. Optimización de los parámetros utilizados en predicciones por  
ocurrencias análogas mediante algoritmos genéticos**

Tramo5					
	Coste	Periodo de entrenamiento		Periodo de evaluación	
		Modelo NN	AR(1)	Modelo NN	AR(1)
Exceso de rendimiento ajustado en el riesgo	0 %	33.9271	24.7406	35.3526	19.2603
	0.10 %	24.2186	11.7825	22.7318	3.5653
	0.25 %	8.5264	-5.0295	9.0112	-16.7402
Rendimiento medio anualizado	0 %	31.6521	21.2699	59.7274	44.5329
	0.10 %	21.7453	8.1167	47.0268	28.5845
	0.25 %	5.6174	-8.9873	33.1866	7.9001
Ratio de Sharpe	0 %	0.0117	0.0078	0.0151	0.0120
	0.10 %	0.0084	0.0032	0.0124	0.0081
	0.25 %	0.0023	-0.0037	0.009	0.0024
Rendimiento B&H ajustado en el riesgo*	0 %	-2.2749	-3.4707	24.3747	25.2726
	0.10 %	-2.4733	-3.6658	24.295	25.0193
	0.25 %	-2.909	-3.9578	24.1754	24.6403
Ratio de Sharpe estrategia B&H ajustada en el riesgo*	0 %	-0.0016	-0.0023	0.0099	0.0098
	0.10 %	-0.0016	-0.0023	0.0099	0.0098
	0.25 %	-0.0017	-0.0023	0.0097	0.0098
Porcentaje de días en el mercado	0 %	39.29 %	43.60 %	52.59 %	53.97 %
	0.10 %	39.29 %	43.60 %	53.39 %	53.97 %
	0.25 %	39.68 %	43.60 %	54.58 %	53.97 %
Días de transacción	0 %	76	113	83	116
	0.10 %	76	113	83	116
	0.25 %	74	113	79	116

Cuadro 4.30: Estadísticos obtenidos para las reglas encontradas para el tramo 5.  
\*: A la hora de calcular estos estadísticos se ha ponderado por el porcentaje de días de cada uno de los modelos.

Dimensión de inmersión	4					
Porcentaje de puntos próximos	42					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	37.8743	38.2110	43.0046	45.9896	45.8437	46.0931
Rend. B&H	22.4598	22.5918	22.7238	22.9218	22.7898	23.0538
Exceso de rend. sobre B&H	15.4145	15.6192	20.2808	23.0678	23.0539	23.0393
Ratio de Sharpe	0.1993	0.2008	0.2224	0.2330	0.2333	0.2348
Ratio de Sharpe B&H	0.1593	0.1580	0.1566	0.1547	0.1560	0.1535
Porcentaje de aciertos	54.3650 %	54.3650 %	54.3650 %	54.3650 %	54.3650 %	54.3650 %
Porcentaje de días en el mercado	54.3650 %	55.1587 %	55.9523 %	57.1428 %	56.3492 %	57.9365 %

Cuadro 4.31: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 5 y coste del 0 %

Dimensión de inmersión	4					
Porcentaje de puntos próximos	42					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	26.2743	26.8110	32.6046	36.3896	36.6437	37.2931
Rend. B&H	22.3511	22.4815	22.6119	22.8075	22.6771	22.9379
Exceso de rend. sobre B&H	3.9233	4.3295	9.9927	13.5821	13.9666	14.3552
Ratio de Sharpe	0.1372	0.1398	0.1674	0.1828	0.1853	0.1884
Ratio de Sharpe B&H	0.1586	0.1572	0.1559	0.1539	0.1552	0.1527
Porcentaje de aciertos	54.3650 %	54.3650 %	54.3650 %	54.3650 %	54.3650 %	54.3650 %
Porcentaje de días en el mercado	54.3650 %	55.1587 %	55.9523 %	57.1428 %	56.3492 %	57.9365 %

Cuadro 4.32: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 5 y coste del 0.10 %

Dimensión de inmersión	4					
Porcentaje de puntos próximos	42					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	-7.9829	-6.3517	-2.9680	-0.5874	-1.6007	-1.4125
Rend. B&H	5.2114	5.0654	5.1141	5.0411	4.9924	4.9194
Exceso de rend. sobre B&H	-13.1943	-11.4171	-8.0821	-5.6285	-6.5931	-6.3319
Ratio de Sharpe	-0.0421	-0.0338	-0.0159	-0.0031	-0.0084	-0.0074
Ratio de Sharpe B&H	0.0363	0.0337	0.0346	0.0333	0.0326	0.0314
Porcentaje de aciertos	54.7619 %	54.7619 %	54.7619 %	54.7619 %	54.7619 %	54.7619 %
Porcentaje de días en el mercado	53.1746 %	55.5555 %	54.7619 %	55.9523 %	56.7460 %	57.9365 %

Cuadro 4.33: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 5 y coste del 0.25 %



254 **4.4. Optimización de los parámetros utilizados en predicciones por  
ocurrencias análogas mediante algoritmos genéticos**

---

Tramo6			
Coste	0 %	0.10 %	0.25 %
Dimensión de inmersión	3	3	3
Porcentaje de puntos próximos	3	2.9	3
Ancho de banda	11	13	12
Número de puntos próximos	85	83	85

Cuadro 4.34: Parámetros obtenidos mediante AGs para las reglas basadas en predicciones NN en el tramo 6

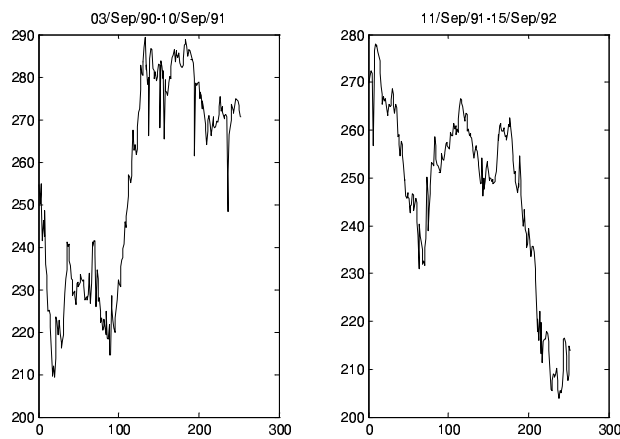


Figura 4.27: Periodo de entrenamiento y validación para el tramo 6.

	Tramo6				
	Coste	Periodo de entrenamiento		Periodo de evaluación	
		Modelo NN	AR(1)	Modelo NN	AR(1)
Exceso de rendimiento ajustado en el riesgo	0 %	60.8242	-16.0153	12.5526	-11.3392
	0.10 %	45.0309	-26.303	3.0642	-19.92
	0.25 %	25.2249	-39.6022	-14.0113	-30.9986
Rendimiento medio anualizado	0 %	71.8785	-4.6652	9.7548	-21.2244
	0.10 %	55.8818	-15.178	0.0683	-29.9874
	0.25 %	35.7164	-28.8141	-17.3036	-41.3386
Ratio de Sharpe	0 %	0.0149	-0.0011	0.0038	-0.0065
	0.10 %	0.0122	-0.0039	0	-0.0097
	0.25 %	0.0084	-0.0090	-0.0077	-0.0145
Rendimiento B&H ajustado en el riesgo*	0 %	11.0543	11.3501	-2.7978	-9.8852
	10.8509	11.1249	-2.9959	-10.0674	
	0.25 %	10.4916	10.7881	-3.2923	-10.3400
Ratio de Sharpe estrategia B&H ajustada en el riesgo*	0 %	0.004	-0.0035	-0.0019	-0.0038
	0.10 %	0.0041	-0.0035	-0.0019	-0.0038
	0.25 %	0.004	-0.0035	-0.0019	-0.0038
Porcentaje de días en el mercado	0 %	49.60 %	50.40 %	40.24 %	56.75 %
	0.10 %	49.21 %	50.40 %	40.24 %	56.75 %
	0.25 %	49.60 %	50.40 %	40.24 %	56.75 %
Días de transacción	0 %	92	115	110	117
	0.10 %	92	115	108	117
	0.25 %	92	115	110	117

Cuadro 4.35: Estadísticos obtenidos para las reglas encontradas para el tramo 6.  
 \*: A la hora de calcular estos estadísticos se ha ponderado por el porcentaje de días de cada uno de los modelos.

Dimensión de inmersión	7					
Porcentaje de puntos próximos	9					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	14.0198	12.5958	11.4275	15.1994	13.7018	7.0546
Rend. B&H	-4.6138	-4.7570	-5.0435	-5.4733	-5.4733	-6.9058
Exceso de rend. sobre B&H	18.6336	17.3528	16.4710	20.6727	19.1751	13.9604
Ratio de Sharpe	0.0848	0.0755	0.0685	0.0934	0.0845	0.0414
Ratio de Sharpe B&H	-0.0432	-0.0442	-0.0461	-0.0489	-0.0489	-0.0571
Porcentaje de aciertos	53.9682 %	53.9682 %	53.9682 %	53.9682 %	53.9682 %	53.9682 %
Porcentaje de días en el mercado	47.6190 %	48.0158 %	48.8095 %	50 %	50 %	53.9682 %

Cuadro 4.36: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 6 y coste del 0 %

Dimensión de inmersión	7					
Porcentaje de puntos próximos	9					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	2.6198	1.3958	0.8275	5.9994	4.9018	-0.7454
Rend. B&H	-4.7090	-4.8531	-5.1411	-5.5733	-5.5733	-7.0137
Exceso de rend. sobre B&H	7.3288	6.2489	5.9686	11.5727	10.4751	6.2683
Ratio de Sharpe	0.0157	0.0083	0.0049	0.0368	0.0301	-0.0044
Ratio de Sharpe B&H	-0.0441	-0.0451	-0.0470	-0.0498	-0.0498	-0.0580
Porcentaje de aciertos	53.9682 %	53.9682 %	53.9682 %	53.9682 %	53.9682 %	53.9682 %
Porcentaje de días en el mercado	47.6190 %	48.0158 %	48.8095 %	50 %	50 %	53.9682 %

Cuadro 4.37: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 6 y coste del 0.10 %

Dimensión de inmersión	7					
Porcentaje de puntos próximos	9					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	-14.4803	-15.4043	-15.0726	-7.8006	-8.2982	-12.4454
Rend. B&H	-4.8519	-4.9971	-5.2876	-5.7233	-5.7233	-7.1756
Exceso de rend. sobre B&H	-9.6284	-10.4072	-9.7850	-2.0774	-2.5750	-5.2698
Ratio de Sharpe	-0.0852	-0.0903	-0.0887	-0.0471	-0.0501	-0.0722
Ratio de Sharpe B&H	-0.0455	-0.0465	-0.0484	-0.0511	-0.0511	-0.0593
Porcentaje de aciertos	53.9682 %	53.9682 %	53.9682 %	53.9682 %	53.9682 %	53.9682 %
Porcentaje de días en el mercado	47.6190 %	48.0158 %	48.8095 %	50 %	50 %	53.9682 %

Cuadro 4.38: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 6 y coste del 0.25 %

Tramo7			
Coste	0 %	0.10 %	0.25 %
Dimensión de inmersión	4	4	2
Porcentaje de puntos próximos	1.8	1.8	2.6
Ancho de banda	48	46	73
Número de puntos próximos	47	47	67

Cuadro 4.39: Parámetros obtenidos mediante AGs para las reglas basadas en predicciones NN en el tramo 7

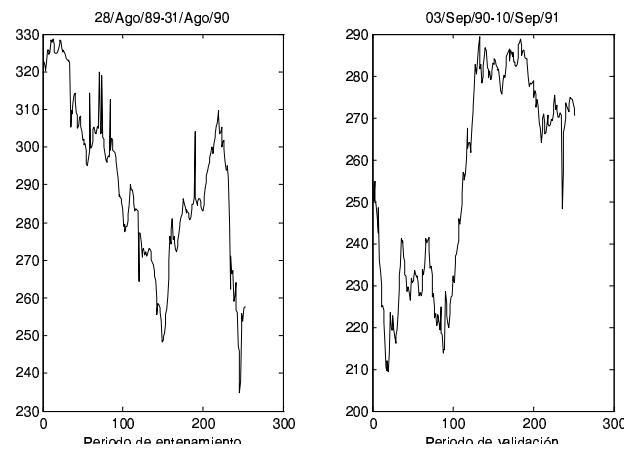


Figura 4.28: Periodo de entrenamiento y validación para el tramo 7.

260 **4.4. Optimización de los parámetros utilizados en predicciones por  
ocurrencias análogas mediante algoritmos genéticos**

Tramo7					
	Coste	Periodo de entrenamiento		Periodo de evaluación	
		Modelo NN	AR(1)	Modelo NN	AR(1)
Exceso de rendimiento ajustado en el riesgo	0 % 0.10 % 0.25 %	53.863 42.0524 27.5776	41.1273 28.6842 12.1652	36.0343 22.862 6.0163	30.8175 16.0422 -3.1979
Rendimiento medio anualizado	0 % 0.10 % 0.25 %	50.1967 38.1906 23.0055	37.3533 24.7156 7.9055	48.0618 34.6612 17.4104	41.7059 26.7065 7.1309
Ratio de Sharpe	0 % 0.10 % 0.25 %	0.0163 0.0129 0.0082	0.0124 0.0086 0.0029	0.0123 0.0093 0.0049	0.0107 0.0072 0.0021
Rendimiento B&H ajustado en el riesgo*	0 % -3.8618 0.25 %	-3.6663 -3.9686 -4.5722	-3.7741 -3.9686 -4.2597	12.0275 10.6643 11.3941	10.8885 10.7643 10.3288
Ratio de Sharpe estrategia B&H ajustada en el riesgo*	0 % 0.10 % 0.25 %	-0.0017 -0.0017 -0.0019	-0.0018 -0.0018 -0.0018	0.0048 0.0048 0.0046	0.0040 0.0040 0.0040
Porcentaje de días en el mercado	0 % 0.10 % 0.25 %	48.02 % 48.02 % 49.21 %	48.00 % 48.00 % 48.00 %	45.02 % 45.02 % 46.61 %	48.81 % 48.81 % 48.81 %
Días de transacción	0 % 0.10 % 0.25 %	81 81 66	95 95 95	92 92 88	111 111 111

Cuadro 4.40: Estadísticos obtenidos para las reglas encontradas para el tramo 7.  
\*: A la hora de calcular estos estadísticos se ha ponderado por el porcentaje de días de cada uno de los modelos.

Dimensión de inmersión	1					
Porcentaje de puntos próximos	14					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	36.6050	36.7863	40.4117	39.6426	40.5279	41.8759
Rend. B&H	9.4733	9.4733	9.5089	9.5802	9.6158	9.1882
Exceso de rend. sobre B&H	27.1317	27.3130	30.9028	30.0624	30.9121	32.6876
Ratio de Sharpe	0.1673	0.1682	0.1886	0.1870	0.1914	0.1762
Ratio de Sharpe B&H	0.0594	0.0594	0.0601	0.0615	0.0623	0.0541
Porcentaje de aciertos	54.3650 %	54.3650 %	54.3650 %	54.3650 %	54.3650 %	54.3650 %
Porcentaje de días en el mercado	48.8095 %	48.8095 %	48.4126 %	47.6190 %	47.2222 %	51.9841 %

Cuadro 4.41: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medios. Parámetros obtenidos mediante AGs para el tramo 7 y coste del 0 %



Dimensión de inmersión	1					
Porcentaje de puntos próximos	14					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	24.8050	25.1863	29.6117	29.4426	30.9279	33.0759
Rend. B&H	9.3756	9.3756	9.4121	9.4849	9.5213	9.0842
Exceso de rend. sobre B&H	15.4294	15.8107	20.1997	19.9577	21.4065	23.9916
Ratio de Sharpe	0.1130	0.1148	0.1378	0.1386	0.1456	0.1390
Ratio de Sharpe B&H	0.0587	0.0587	0.0595	0.0609	0.0617	0.0534
Porcentaje de aciertos	54.3650 %	54.3650 %	54.3650 %	54.3650 %	54.3650 %	54.3650 %
Porcentaje de días en el mercado	48.8095 %	48.8095 %	48.4126 %	47.6190 %	47.2222 %	51.9841 %

Cuadro 4.42: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 7 y coste del 0.10 %

Dimensión de inmersión	1					
Porcentaje de puntos próximos	14					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	7.1050	7.7862	13.4117	14.1425	16.5278	19.8758
Rend. B&H	9.2292	9.2292	9.2668	9.3421	9.3797	8.9283
Exceso de rend. sobre B&H	-2.1243	-1.4430	4.1448	4.8005	7.1482	10.9475
Ratio de Sharpe	0.0320	0.0351	0.0618	0.0659	0.0770	0.0830
Ratio de Sharpe B&H	0.0578	0.0578	0.0585	0.0600	0.0607	0.0525
Porcentaje de aciertos	54.3650 %	54.3650 %	54.3650 %	54.3650 %	54.3650 %	54.3650 %
Porcentaje de días en el mercado	48.8095 %	48.8095 %	48.4126 %	47.6190 %	47.2222 %	51.9841 %

Cuadro 4.43: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 7 y coste del 0.25 %

264 **4.4. Optimización de los parámetros utilizados en predicciones por  
ocurrencias análogas mediante algoritmos genéticos**

Tramo8			
Coste	0 %	0.10 %	0.25 %
Dimensión de inmersión	4	6	6
Porcentaje de puntos próximos	3.9	4	4.7
Ancho de banda	5	35	34
Número de puntos próximos	92	94	110

Cuadro 4.44: Parámetros obtenidos mediante AGs para las reglas basadas en predicciones NN en el tramo 8

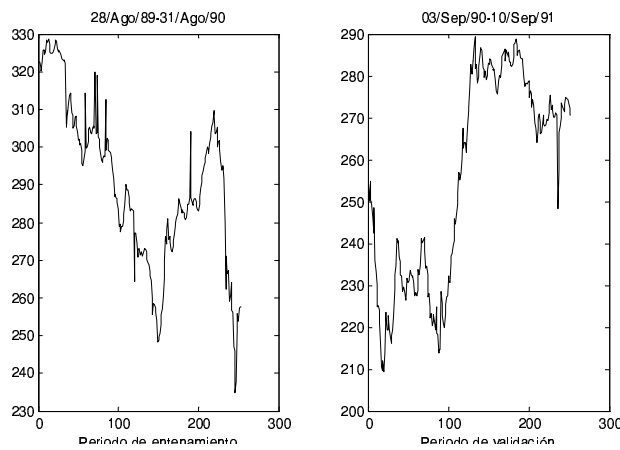


Figura 4.29: Periodo de entrenamiento y validación para el tramo 8.

Tramo8					
	Coste	Periodo de entrenamiento		Periodo de evaluación	
		Modelo NN	AR(1)	Modelo NN	AR(1)
Exceso de rendimiento ajustado en el riesgo	0 %	28.2907	35.9767	46.3396	41.7095
	0.10 %	17.6853	25.5314	8.228	29.4139
	0.25 %	5.9274	11.2939	4.9762	13.0544
Rendimiento medio anualizado	0 %	39.7779	47.3272	42.404	38.1777
	0.10 %	28.7306	36.6567	3.5447	25.687
	0.25 %	16.6461	22.0823	0.414	9.0357
Ratio de Sharpe	0 %	0.0227	0.0279	0.0138	0.0127
	0.10 %	0.0156	0.0220	0.0012	0.0090
	0.25 %	0.0093	0.0138	0.0002	0.0033
Rendimiento B&H ajustado en el riesgo*	0 %	11.4872	11.3504	-3.9355	-3.5319
	0.10 %	11.0453	11.1253	-4.6832	-3.7269
	0.25 %	10.7187	10.7885	-4.5621	-4.0187
Ratio de Sharpe estrategia B&H ajustada en el riesgo*	0 %	0.0098		-0.0018	
	0.10 %	0.0085		-0.002	
	0.25 %	0.0085		-0.0019	
Porcentaje de días en el mercado	0 %	53.57 %	50.80 %	49.40 %	48.41 %
	0.10 %	61.11 %	50.80 %	51.00 %	48.41 %
	0.25 %	60.71 %	50.80 %	49.80 %	48.41 %
Días de transacción	0 %	87	74	85	94
	0.10 %	65	74	77	94
	0.25 %	63	74	73	94

Cuadro 4.45: Estadísticos obtenidos para las reglas encontradas para el tramo 8. \*: A la hora de calcular estos estadísticos se ha ponderado por el porcentaje de días de cada uno de los modelos.

Dimensión de inmersión	3					
Porcentaje de puntos próximos	34					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	28.8266	31.7005	32.0612	33.9230	34.1712	33.0867
Rend. B&H	-3.6475	-3.3579	-3.2130	-3.0682	-3.0682	-2.6337
Exceso de rend. sobre B&H	32.4742	35.0584	35.2742	36.9912	37.2393	35.7204
Ratio de Sharpe	0.1549	0.1722	0.1740	0.1936	0.1950	0.1892
Ratio de Sharpe B&H	-0.0260	-0.0244	-0.0235	-0.0226	-0.0226	-0.0199
Porcentaje de aciertos	64.2857 %	64.2857 %	64.2857 %	64.2857 %	64.2857 %	64.2857 %
Porcentaje de días en el mercado	50.7936 %	50 %	49.6031 %	49.2063 %	49.2063 %	48.0158 %

Cuadro 4.46: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 8 y coste del 0 %

Dimensión de inmersión	3					
Porcentaje de puntos próximos	34					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	18.4266	22.1005	23.4612	25.5230	26.1711	25.4867
Rend. B&H	-3.7491	-3.4579	-3.3122	-3.1666	-3.1666	-2.7297
Exceso de rend. sobre B&H	22.1758	25.5584	26.7734	28.6896	29.3378	28.2164
Ratio de Sharpe	0.0982	0.1192	0.1265	0.1451	0.1489	0.1454
Ratio de Sharpe B&H	-0.0268	-0.0251	-0.0242	-0.0233	-0.0233	-0.0206
Porcentaje de aciertos	64.2857 %	64.2857 %	64.2857 %	64.2857 %	64.2857 %	64.2857 %
Porcentaje de días en el mercado	50.7936 %	50 %	49.6031 %	49.2063 %	49.2063 %	48.0158 %

Cuadro 4.47: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medios. Parámetros obtenidos mediante AGs para el tramo 8 y coste del 0.10 %

Dimensión de inmersión	3					
Porcentaje de puntos próximos	34					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	2.8266	7.7005	10.5611	12.9230	14.1711	14.0866
Rend. B&H	-3.9015	-3.6079	-3.4610	-3.3142	-3.3142	-2.8738
Exceso de rend. sobre B&H	6.7281	11.3083	14.0222	16.2372	17.4853	16.9604
Ratio de Sharpe	0.0148	0.0407	0.0560	0.0724	0.0795	0.0795
Ratio de Sharpe B&H	-0.0279	-0.0262	-0.0253	-0.0244	-0.0244	-0.0217
Porcentaje de aciertos	64.2857 %	64.2857 %	64.2857 %	64.2857 %	64.2857 %	64.2857 %
Porcentaje de días en el mercado	50.7936 %	50 %	49.6031 %	49.2063 %	49.2063 %	48.0158 %

Cuadro 4.48: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medios. Parámetros obtenidos mediante AGs para el tramo 8 y coste del 0.25 %

Tramo9			
Coste	0 %	0.10 %	0.25 %
Dimensión de inmersión	3	3	2
Porcentaje de puntos próximos	0.5	0.5	3.6
Ancho de banda	40	40	81
Número de puntos próximos	10	10	75

Cuadro 4.49: Parámetros obtenidos mediante AGs para las reglas basadas en predicciones NN en el tramo 9

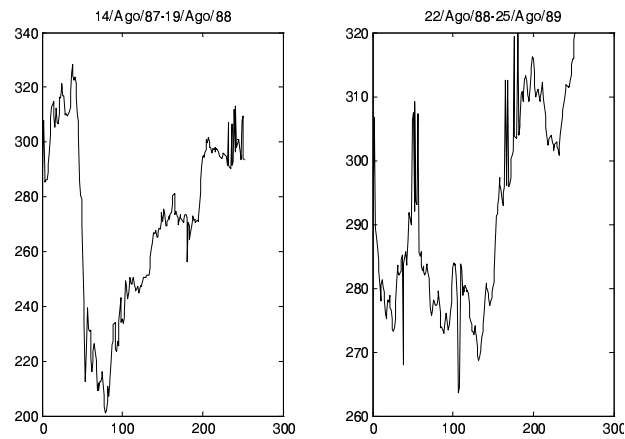


Figura 4.30: Periodo de entrenamiento y validación para el tramo 9.



270 **4.4. Optimización de los parámetros utilizados en predicciones por  
ocurrencias análogas mediante algoritmos genéticos**

Tramo9					
	Coste	Periodo de entrenamiento		Periodo de evaluación	
		Modelo NN	AR(1)	Modelo NN	AR(1)
Exceso de rendimiento ajustado en el riesgo	0 %	77.1716	49.7619	14.69	36.6644
	0.10 %	58.9892	35.3413	2.405	26.3226
	0.25 %	36.9317	16.2443	-15.4573	12.2157
Rendimiento medio anualizado	0 %	83.4815	55.9406	26.6284	46.9916
	0.10 %	65.0833	41.3054	14.1152	36.4268
	0.25 %	42.1683	21.8871	-4.1969	21.9861
Ratio de Sharpe	0 %	0.0163	0.0103	0.0182	0.0279
	0.10 %	0.0135	0.0080	0.0103	0.0221
	0.25 %	0.0091	0.0046	-0.0027	0.0133
Rendimiento B&H ajustado en el riesgo*	0 %	6.3099	6.1787	11.9384	10.3272
	0.10 %	6.0941	5.9641	11.7102	10.1042
	0.25 %	5.2366	5.6429	11.2604	9.7704
Ratio de Sharpe estrategia B&H ajustada en el riesgo*	0 %	0.0018	0.0017	0.0113	0.0095
	0.10 %	0.0018	0.0017	0.0113	0.0095
	0.25 %	0.0015	0.0017	0.0102	0.0095
Porcentaje de días en el mercado	0 %	52.38 %	54.80 %	48.21 %	50.40 %
	0.10 %	52.38 %	54.80 %	48.21 %	50.40 %
	0.25 %	56.75 %	54.80 %	52.99 %	50.40 %
Días de transacción	0 %	103	97	101	74
	0.10 %	103	97	101	74
	0.25 %	77	97	103	74

Cuadro 4.50: Estadísticos obtenidos para las reglas encontradas para el tramo 9.  
\*: A la hora de calcular estos estadísticos se ha ponderado por el porcentaje de días de cada uno de los modelos.

Dimensión de inmersión	2					
Porcentaje de puntos próximos	44					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	35.4582	36.0690	36.5024	37.5329	36.7791	36.2544
Rend. B&H	10.6434	10.5914	10.5914	10.6087	10.6261	10.6261
Exceso de rend. sobre B&H	24.8147	25.4776	25.9111	26.9242	26.1530	25.6283
Ratio de Sharpe	0.4011	0.4074	0.4077	0.4220	0.3995	0.3932
Ratio de Sharpe B&H	0.1538	0.1498	0.1498	0.1511	0.1524	0.1524
Porcentaje de aciertos	67.0634 %	67.0634 %	67.0634 %	67.0634 %	67.0634 %	67.0634 %
Porcentaje de días en el mercado	54.3650 %	55.5555 %	55.5555 %	55.1587 %	54.7619 %	54.7619 %

Cuadro 4.51: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 9 y coste del 0 %

Dimensión de inmersión	2					
Porcentaje de puntos próximos	44					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	25.4582	26.6690	27.5024	28.9329	28.5791	28.4544
Rend. B&H	10.5347	10.4803	10.4803	10.4984	10.5166	10.5166
Exceso de rend. sobre B&H	14.9235	16.1887	17.0222	18.4345	18.0625	17.9378
Ratio de Sharpe	0.2798	0.2925	0.2995	0.3171	0.3044	0.3038
Ratio de Sharpe B&H	0.1522	0.1482	0.1482	0.1495	0.1509	0.1509
Porcentaje de aciertos	67.0634 %	67.0634 %	67.0634 %	67.0634 %	67.0634 %	67.0634 %
Porcentaje de días en el mercado	54.3650 %	55.5555 %	55.5555 %	55.1587 %	54.7619 %	54.7619 %

Cuadro 4.52: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 9 y coste del 0.10 %

Dimensión de inmersión	2					
Porcentaje de puntos próximos	44					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	25.4582	26.6690	27.5024	28.9329	28.5791	28.4544
Rend. B&H	10.5347	10.4803	10.4803	10.4984	10.5166	10.5166
Exceso de rend. sobre B&H	14.9235	16.1887	17.0222	18.4345	18.0625	17.9378
Ratio de Sharpe	0.2798	0.2925	0.2995	0.3171	0.3044	0.3038
Ratio de Sharpe B&H	0.1522	0.1482	0.1482	0.1495	0.1509	0.1509
Porcentaje de aciertos	67.0634 %	67.0634 %	67.0634 %	67.0634 %	67.0634 %	67.0634 %
Porcentaje de días en el mercado	54.3650 %	55.5555 %	55.5555 %	55.1587 %	54.7619 %	54.7619 %

Cuadro 4.53: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 9 y coste del 0.25 %

274 **4.4. Optimización de los parámetros utilizados en predicciones por  
ocurrencias análogas mediante algoritmos genéticos**

Tramo10			
Coste	0 %	0.10 %	0.25 %
Dimensión de inmersión	4	4	4
Porcentaje de puntos próximos	4.7	4.7	4.7
Ancho de banda	27	27	82
Número de puntos próximos	87	87	87

Cuadro 4.54: Parámetros obtenidos mediante AGs para las reglas basadas en predicciones NN en el tramo 10

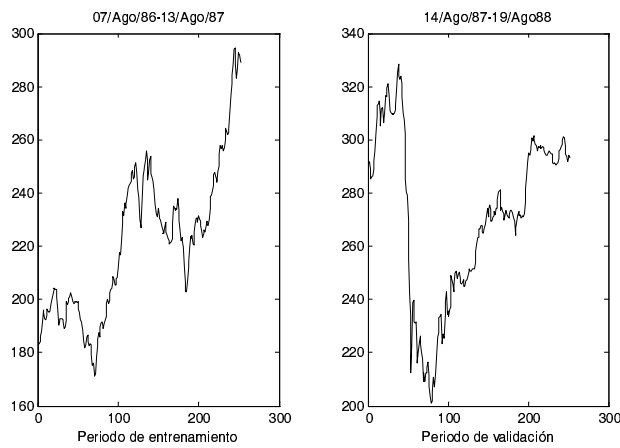


Figura 4.31: Periodo de entrenamiento y validación para el tramo 10.

Tramo10					
	Coste	Periodo de entrenamiento		Periodo de evaluación	
		Modelo NN	AR(1)	Modelo NN	AR(1)
Exceso de rendimiento ajustado en el riesgo	0 %	123.9209	100.8409	59.631	50.1306
	0.10 %	104.7583	81.1631	44.5694	35.8155
	0.25 %	79.4756	54.7318	18.8067	16.8396
Rendimiento medio anualizado	0 %	165.0877	138.5177	65.4094	55.9881
	0.10 %	145.6384	118.5615	50.1321	41.459
	0.25 %	120.2938	91.7138	23.8699	22.1628
Ratio de Sharpe	0 %	0.0238	0.0219	0.0122	0.0104
	0.10 %	0.022	0.0196	0.0098	0.0081
	0.25 %	0.0198	0.0162	0.0048	0.0047
Rendimiento B&H ajustado en el riesgo*	0 %	41.1667	37.6768	5.7784	5.8575
	0.10 %	40.8802	37.3984	5.5628	5.6435
	0.25 %	40.8182	36.9819	5.0632	5.3233
Ratio de Sharpe estrategia B&H ajustada en el riesgo*	0 %	0.0101	0.0110	0.0014	0.0016
	0.10 %	0.0101	0.0110	0.0014	0.0016
	0.25 %	0.01	0.0110	0.0013	0.0016
Porcentaje de días en el mercado	0 %	63.89 %	55.20 %	62.15 %	54.76 %
	0.10 %	63.89 %	55.20 %	62.15 %	54.76 %
	0.25 %	64.68 %	55.20 %	63.75 %	54.76 %
Días de transacción	0 %	74	86	94	97
	0.10 %	74	86	94	97
	0.25 %	56	86	78	97

Cuadro 4.55: Estadísticos obtenidos para las reglas encontradas para el tramo 10.  
 \*: A la hora de calcular estos estadísticos se ha ponderado por el porcentaje de días de cada uno de los modelos.

Dimensión de inmersión	4					
Porcentaje de puntos próximos	10					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	52.9394	56.1247	56.9345	56.6656	58.0748	58.1672
Rend. B&H	5.9975	5.9515	6.0436	6.0896	5.9515	5.9515
Exceso de rend. sobre B&H	46.9419	50.1732	50.8910	50.5761	52.1233	52.2157
Ratio de Sharpe	0.2130	0.2217	0.2251	0.2243	0.2292	0.2295
Ratio de Sharpe B&H	0.0245	0.0241	0.0248	0.0252	0.0241	0.0241
Porcentaje de aciertos	63.4920 %	63.4920 %	63.4920 %	63.4920 %	63.4920 %	63.4920 %
Porcentaje de días en el mercado	60.7142 %	61.1111 %	60.3174 %	59.9206 %	61.1111 %	61.1111 %

Cuadro 4.56: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 10 y coste del 0 %

Dimensión de inmersión	4					
	10					
Porcentaje de puntos próximos						
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	42.1394	45.5247	46.7345	46.6656	48.6748	48.9672
Rend. B&H	5.8761	5.8293	5.9229	5.9697	5.8293	5.8293
Exceso de rend. sobre B&H	36.2633	39.6954	40.8116	40.6959	42.8455	43.1379
Ratio de Sharpe	0.1692	0.1792	0.1841	0.1841	0.1915	0.1927
Ratio de Sharpe B&H	0.0240	0.0236	0.0243	0.0247	0.0236	0.0236
Porcentaje de aciertos	63.4920 %	63.4920 %	63.4920 %	63.4920 %	63.4920 %	63.4920 %
Porcentaje de días en el mercado	60.7142 %	61.1111 %	60.3174 %	59.9206 %	61.1111 %	61.1111 %

Cuadro 4.57: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 10 y coste del 0.10 %



Dimensión de inmersión	4					
Porcentaje de puntos próximos	10					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	25.9394	29.6247	31.4345	31.6656	34.5748	35.1671
Rend. B&H	5.6940	5.6460	5.7420	5.7900	5.6460	5.6460
Exceso de rend. sobre B&H	20.2454	23.9787	25.6925	25.8756	28.9288	29.5212
Ratio de Sharpe	0.1034	0.1155	0.1226	0.1238	0.1348	0.1373
Ratio de Sharpe B&H	0.0232	0.0229	0.0236	0.0239	0.0229	0.0229
Porcentaje de aciertos	63.4920 %	63.4920 %	63.4920 %	63.4920 %	63.4920 %	63.4920 %
Porcentaje de días en el mercado	60.7142 %	61.1111 %	60.3174 %	59.9206 %	61.1111 %	61.1111 %

Cuadro 4.58: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el tramo 10 y coste del 0.25 %

Los Cuadros 4.9, 4.14, 4.19, 4.24, 4.29, 4.34, 4.39, 4.44, 4.49, 4.54 muestran los parámetros seleccionados por el algoritmo genético para las reglas encontradas en cada uno de los tramos seleccionados. Se puede observar como sólo en uno de los tramos la dimensión de inmersión, al igual que el porcentaje de puntos próximos y el ancho de banda, ha aumentado cuando aumenta el coste de transacción. Esto implica que, cuando aumentan los costes, se reducen los días de transacción. Por lo tanto, el impacto que sobre el beneficio tienen los costes totales es menor ñal que hubiese aoccurrido de mantener los mismos días de transacción. Al igual que con las reglas técnicas clásicas, el algoritmo genético es capaz de modificar los parámetros de la regla en función de los costes de transacción.

Los resultados sobre la variable exceso de rendimiento respecto a la estrategia B&H, ajustada por riesgo, cuando empleamos un AG en la selección de la dimensión de inmersión, el número de puntos próximos y ancho de banda y del modelo AR(1) se resumen en los cuadros 4.59 y 4.60, respectivamente. La última fila recoge el exceso de rendimeinto sobre la estrategia B&H, ajustada por riesgo, durante el periodo completo, es decir, el conjunto de los diez subperiodos estudiados.

Podemos observar que para unos costes del 0%, el exceso de rendimientos de de nuestra regla frente a la estrategia B&H ajustada por riesgo siempre es positivo para todos los tramos. Sin embargo, para los modelos AR(1), ya existe un tramo (Tramo 6) donde los rendimientos del modelo, no superan a la de la estrategia B&H.

Para costes del 0.10% siguen existiendo más tramos con exceso de rendimientos ajustado por riesgo de la regla frente a la estrategia B&H positivos

en los modelos que emplean reglas con predicciones NN que en los modelos AR(1). Únicamente en los tramos 2 y 3 el rendimiento logrado por las reglas basadas en modelos NN no es capaz de superar a la estrategia B&H. Por otro lado, el exceso de rendimiento logrado por las predicciones de los modelos AR(1) frente a la estrategia B&H son negativos en los tramos 2,3 y 6.

Sin embargo, son con costes del 0.25 % las predicciones en modelos AR(1) demuestran su escasa habilidad para generar beneficios. Sólo en tres tramos el exceso de rendimiento es positivo frente a los 6 tramos en los modelos en donde empleamos nuestras reglas basadas en predictores NN.

Como es lógico aumentan los costes de transacción son menos los tramos donde el rendimiento de la regla supere al rendimiento de la estrategia B&H, ajustada por riesgo. Sin embargo, podemos observar que el algoritmo ajusta los parámetros de tal manera que hace disminuir los días de transacción. De esta forma, el algoritmo intenta disminuir el impacto de los costes de transacción a medida que aumentan los costes. Y esto se ve reflejado en el hecho de que en ausencia de costes, el exceso de rendimiento ajustado por riesgo de nuestras reglas es positivo para todos los tramos, mientras que con costes del 0.25 % es positivo en 6 tramos representando, aun así, más de la mitad de los tramos. La utilización de modelos que no respondan ante cambios en los costes, como pueden ser los modelos AR(1), puede provocar un comportamiento razonable en ausencia de costes de transacción (aun así, nuestra regla posee exceso de rendimientos, ajustado por riesgo, positivo en más tramos que el modelo AR(1)), pero dejan de ser efectivos cuando se le incorporan costes de transacción.

Este hecho se puede corroborar cuando comparamos, no sólo, la capacidad

Coste	0 %	0.10 %	0.25 %
Tramo 1	3.7761	0.8496	0.859
Tramo 2	3.4849	-2.432	-3.6198
Tramo 3	5.6148	-5.2656	-14.7207
Tramo 4	27.9669	17.209	1.7253
Tramo 5	35.3526	22.7318	9.0112
Tramo 6	12.5526	3.0642	-14.0113
Tramo 7	36.0343	22.862	6.0163
Tramo 8	46.3396	8.228	4.9762
Tramo 9	14.69	2.405	-15.4573
Tramo 10	59.631	44.5694	18.8067
Exceso Total	245.4428	114.2214	-6.4144

Cuadro 4.59: Exceso de rendimientos de las reglas para costes de transacción de 0 %, 0.10 % y 0.25 %

que tiene nuestro modelo de obtener un rendimiento mayor al obtenido por una estrategia de “comprar y mantener” (que era uno de los objetivos de la tesis) frente al de un modelo AR(1), sino también, cuando comparamos el exceso de rendimiento, ajustado por riesgo, logrado por nuestras reglas frente al del modelo AR(1). Con costes del 0 % y del 0.10 %, el exceso de rendimiento alcanzado por las reglas son capaces de superar en 6 de los 10 tramos al obtenido por los modelos AR(1). Mas aún, con costes del 0.25 % este porcentaje sube al 80 % de los tramos. Este hecho nos vuelve a confirmar, nuestra conclusión de que las reglas que utilizan algoritmos de optimización en cuya función objetivo se incluyan los costes de transacción se comportan mejor que el modelo alternativo AR(1).

Por último cabe destacar que si comparamos el exceso de rendimiento frente a la estrategia B&H, ajustada por riesgo, para el conjunto de los diez subperiodos se puede observar como son mayores los logrados por las reglas frente al modelo AR(1), para todos los costes de transacción.

4.4. Optimización de los parámetros utilizados en predicciones por  
 282 ocurrencias análogas mediante algoritmos genéticos

Coste	0 %	0.10 %	0.25 %
Tramo 1	18.71	3.48	-16.49
Tramo 2	18.71	-9.816	-26.9544
Tramo 3	11.1786	-1.4578	-17.8246
Tramo 4	22.708	9.8681	-6.9547
Tramo 5	19.2603	3.5653	-16.7402
Tramo 6	-11.3392	-19.92	-30.9986
Tramo 7	30.8175	16.0422	-3.1979
Tramo 8	41.7095	29.4139	13.0544
Tramo 9	36.6644	26.3226	12.2157
Tramo 10	50.1306	35.8155	16.8396
Exceso Total	238.5497	93.3138	-77.0507

Cuadro 4.60: Exceso de rendimientos del modelo AR(1) para costes de transacción de 0 %, 0.10 % y 0.25 %

#### **4.4.2. Análisis empírico para los años 1997 a 2002**

A continuación haremos un contraste empírico para los años 1997 a 2002 empleando reglas optimizadas utilizando predicciones NN y cuyos parámetros han sido optimizados mediante AGs. Por una parte, hemos aplicado una función objetivo que tiene en cuenta los costes de transacción. Por otra, hemos utilizado una función objetivo alternativa, definida en términos de la raíz cuadrada del error cuadrático medio. A su vez, se ha trabajado con un modelo AR(1) para comparar los rendimientos obtenidos a través de este modelo, uno de los más adoptados en finanzas, frente a los conseguidos por nuestras reglas.

Se ha empleado una muestra de 5665 datos comprendidos entre el 2 de enero de 1979 y 30 de diciembre de 2002. Como se ha venido repitiendo en los apartados anteriores, se evaluarán estas reglas para los años comprendidos entre 1997 y 2002. Cada año natural se corresponde con un periodo de validación y se utilizan los datos anteriores para entrenar el algoritmo.

284 **4.4. Optimización de los parámetros utilizados en predicciones por  
ocurrencias análogas mediante algoritmos genéticos**

---

Año 1997		
Coste	0.10 %	0.25 %
Dimensión de inmersión	1	1
Porcentaje de puntos próximos	2.8	3.6
Ancho de banda	4	88
Número de puntos próximos	116	150

Cuadro 4.61: Parámetros obtenidos mediante AGs para las reglas basadas en predicciones NN en el año 1997

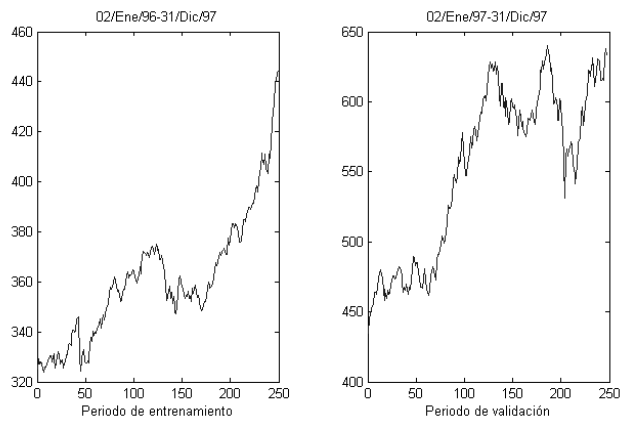


Figura 4.32: Periodo de entrenamiento y validación para el año 1997.

#### 4. Resultados empíricos: Aplicaciones a la predicción financiera del IGBM.

285

Año 1997					
	Coste	Periodo de entrenamiento		Periodo de evaluación	
		Modelo NN	AR(1)	Modelo NN	AR(1)
Exceso de rendimiento ajustado en el riesgo	0.10 %	9.7131	1.8776	5.7535	9.2902
	0.25 %	10.6000	-6.3373	-3.2900	-6.7419
Rendimiento medio anualizado	0.10 %	25.5104	34.8576	42.5922	38.6970
	0.25 %	20.5000	26.2444	35.3000	22.2774
Ratio de Sharpe	0.10 %	0.0141	0.0703	0.0076	0.0498
	0.25 %	0.0150	0.0402	0.0062	0.0241
Rendimiento B&H ajustado en el riesgo*	0.10 %	15.7973	32.9800	36.8387	29.4069
	0.25 %	9.8800	32.5817	38.6000	29.0192
Ratio de Sharpe estrategia B&H ajustada en el riesgo*	0.10 %	0.0168	0.0691	0.0074	0.0426
	0.25 %	0.0319	0.0691	0.0072	0.0426
Porcentaje de días en el mercado	0.10 %	29.72 %	88.76 %	84.21 %	55.47 %
	0.25 %	10.40 %	88.76 %	91.10 %	55.47 %
Días de transacción	0.10 %	13	44	15	84
	0.25 %	3	44	3	84

Cuadro 4.62: Estadísticos obtenidos para las reglas encontradas para el año 1997.  
\*: A la hora de calcular estos estadísticos se ha ponderado por el porcentaje de días de cada uno de los modelos



Dimensión de inmersión	6					
Porcentaje de puntos próximos	25					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	36.0619	39.1558	38.6010	38.3526	38.8825	37.4837
Rend. B&H	22.5844	22.9454	22.9454	23.3065	23.6675	23.5471
Exceso de rend. sobre B&H	13.4775	16.2103	15.6556	15.0461	15.2150	13.9365
Ratio de Sharpe	0.1479	0.1583	0.1571	0.1560	0.1573	0.1509
Ratio de Sharpe B&H	0.1230	0.1224	0.1224	0.1218	0.1213	0.1214
Porcentaje de aciertos	53.8461 %	53.8461 %	53.8461 %	53.8461 %	53.8461 %	53.8461 %
Porcentaje de días en el mercado	57.4898 %	58.7044 %	58.7044 %	59.9190 %	61.1336 %	60.7287 %

Cuadro 4.63: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el año 1997 y coste del 0 %

Dimensión de inmersión Porcentaje de puntos próximos Anchura del filtro	6 25					
	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	25.8619	29.7558	29.6010	29.7526	30.4825	29.6837
Rend. B&H	22.4694	22.8280	22.8280	23.1866	23.5452	23.4257
Exceso de rend. sobre B&H	3.3925	6.9278	6.7730	6.5659	6.9373	6.2580
Ratio de Sharpe	0.1060	0.1202	0.1203	0.1207	0.1229	0.1193
Ratio de Sharpe B&H	0.1224	0.1218	0.1218	0.1212	0.1206	0.1208
Porcentaje de aciertos	53.8461 %	53.8461 %	53.8461 %	53.8461 %	53.8461 %	53.8461 %
Porcentaje de días en el mercado	57.4898 %	58.7044 %	58.7044 %	59.9190 %	61.1336 %	60.7287 %

Cuadro 4.64: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el año 1997 y coste del 0.10 %

Dimensión de inmersión	6					
	25					
Porcentaje de puntos próximos						
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	10.5618	15.6557	16.1010	16.8525	17.8825	17.9836
Rend. B&H	22.2970	22.6519	22.6519	23.0069	23.3618	23.2435
Exceso de rend. sobre B&H	-11.7351	-6.9962	-6.5510	-6.1543	-5.4793	-5.2599
Ratio de Sharpe	0.0431	0.0629	0.0650	0.0678	0.0715	0.0718
Ratio de Sharpe B&H	0.1215	0.1209	0.1209	0.1203	0.1197	0.1199
Porcentaje de aciertos	53.8461 %	53.8461 %	53.8461 %	53.8461 %	53.8461 %	53.8461 %
Porcentaje de días en el mercado	57.4898 %	58.7044 %	58.7044 %	59.9190 %	61.1336 %	60.7287 %
Porcentaje de días en el mercado	57.4899 %	65.9919 %	80.1619 %	72.0648 %	64.3725 %	

Cuadro 4.65: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medios. Parámetros obtenidos mediante AGs para el año 1997 y coste del 0.25 %

Año 1998		
Coste	0.10 %	0.25 %
Dimensión de inmersión	3	1
Porcentaje de puntos próximos	3	2.7
Ancho de banda	11	43
Número de puntos próximos	159	119

Cuadro 4.66: Parámetros obtenidos mediante AGs para las reglas basadas en predicciones NN en el año 1998

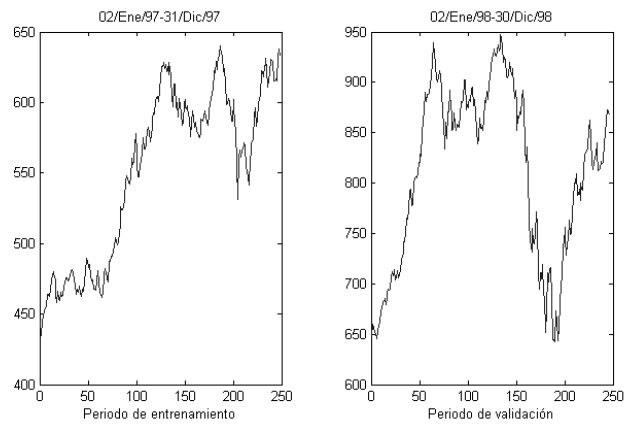


Figura 4.33: Periodo de entrenamiento y validación para el año 1998.

290 **4.4. Optimización de los parámetros utilizados en predicciones por  
ocurrencias análogas mediante algoritmos genéticos**

Año 1998					
	Coste	Periodo de entrenamiento		Periodo de evaluación	
		Modelo NN	AR(1)	Modelo NN	AR(1)
Exceso de rendimiento ajustado en el riesgo	0.10 %	15.4923	2.8862	-4.7728	5.7886
	0.25 %	11.7470	-12.1430	7.8293	-11.2192
Rendimiento medio anualizado	0.10 %	41.0328	7.1660	18.2136	9.3011
	0.25 %	47.8597	-8.1755	39.5397	-8.0167
Ratio de Sharpe	0.10 %	0.0104	0.0189	0.0032	0.0200
	0.25 %	0.0083	-0.0148	0.0050	-0.0130
Rendimiento B&H ajustado en el riesgo*	0.10 %	25.5405	4.2798	22.9864	3.5125
	0.25 %	36.1127	3.9674	31.7105	3.2025
Ratio de Sharpe estrategia B&H ajustada en el riesgo*	0.10 %	0.0076	0.0147	0.0050	0.0085
	0.25 %	0.0071	0.0147	0.0047	0.0085
Porcentaje de días en el mercado	0.10 %	59.11 %	55.47 %	60.98 %	58.54 %
	0.25 %	87.04 %	55.47 %	86.99 %	58.54 %
Días de transacción	0.10 %	97	103	15	115
	0.25 %	9	103	19	115

Cuadro 4.67: Estadísticos obtenidos para las reglas encontradas para el año 1998.  
\*: A la hora de calcular estos estadísticos se ha ponderado por el porcentaje de días de cada uno de los modelos

Dimensión de inmersión	3					
Porcentaje de puntos próximos	94					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	33.8309	34.8130	37.6813	37.6813	25.3837	27.3801
Rend. B&H	21.0902	21.4228	21.4228	21.4228	21.8661	22.0878
Exceso de rend. sobre B&H	12.7407	13.3902	16.2586	16.2586	3.5175	5.2923
Ratio de Sharpe	0.1085	0.1105	0.1220	0.1220	0.0775	0.0863
Ratio de Sharpe B&H	0.0795	0.0792	0.0792	0.0792	0.0788	0.0786
Porcentaje de aciertos	52.0325 %	52.0325 %	52.0325 %	52.0325 %	52.0325 %	52.0325 %
Porcentaje de días en el mercado	61.3821 %	62.6016 %	62.6016 %	62.6016 %	64.2276 %	65.0406 %

Cuadro 4.68: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medios. Parámetros obtenidos mediante AGs para el año 1998 y coste del 0 %

Dimensión de inmersión	3					
Porcentaje de puntos próximos	94					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	22.6309	24.4130	27.8813	27.8813	15.9837	18.7801
Rend. B&H	20.9675	21.2976	21.2976	21.2976	21.7377	21.9577
Exceso de rend. sobre B&H	1.6634	3.1154	6.5838	6.5838	-5.7540	-3.1776
Ratio de Sharpe	0.0726	0.0775	0.0902	0.0902	0.0488	0.0592
Ratio de Sharpe B&H	0.0790	0.0787	0.0787	0.0787	0.0783	0.0781
Porcentaje de aciertos	52.0325 %	52.0325 %	52.0325 %	52.0325 %	52.0325 %	52.0325 %
Porcentaje de días en el mercado	61.3821 %	62.6016 %	62.6016 %	62.6016 %	64.2276 %	65.0406 %

Cuadro 4.69: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medios. Parámetros obtenidos mediante AGs para el año 1998 y coste del 0.10 %

Dimensión de inmersión	3					
Porcentaje de puntos próximos	94					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	5.8308	8.8129	13.1813	13.1813	1.8836	5.8801
Rend. B&H	20.7833	21.1098	21.1098	21.1098	21.5450	21.7626
Exceso de rend. sobre B&H	-14.9525	-12.2968	-7.9285	-7.9285	-19.6614	-15.8825
Ratio de Sharpe	0.0187	0.0279	0.0424	0.0424	0.0057	0.0185
Ratio de Sharpe B&H	0.0784	0.0780	0.0780	0.0780	0.0776	0.0774
Porcentaje de aciertos	52.0325 %	52.0325 %	52.0325 %	52.0325 %	52.0325 %	52.0325 %
Porcentaje de días en el mercado	61.3821 %	62.6016 %	62.6016 %	62.6016 %	64.2276 %	65.0406 %

Cuadro 4.70: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el año 1998 y coste del 0.25 %



294 **4.4. Optimización de los parámetros utilizados en predicciones por  
ocurrencias análogas mediante algoritmos genéticos**

Año 1999		
Coste	0.10 %	0.25 %
Dimensión de inmersión	2	1
Porcentaje de puntos próximos	3.7	1.2
Ancho de banda	36	57
Número de puntos próximos	172	56

Cuadro 4.71: Parámetros obtenidos mediante AGs para las reglas basadas en predicciones NN en el año 1999

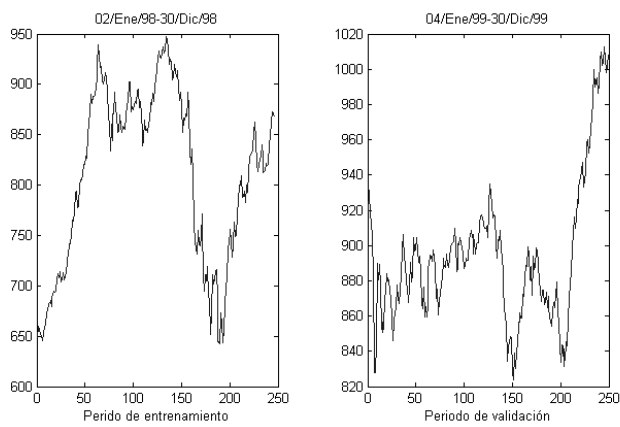


Figura 4.34: Periodo de entrenamiento y validación para el año 1999.

#### 4. Resultados empíricos: Aplicaciones a la predicción financiera del IGBM.

295

Año 1999					
	Coste	Periodo de entrenamiento		Periodo de evaluación	
		Modelo NN	AR(1)	Modelo NN	AR(1)
Exceso de rendimiento ajustado en el riesgo	0.10 %	30.0040	11.0422	-13.6230	2.7255
	0.25 %	12.49580	-6.8904	22.2847	-14.2306
Rendimiento medio anualizado	0.10 %	52.8530	15.1497	-2.0003	4.7893
	0.25 %	40.0056	-3.0948	29.0637	-12.4726
Ratio de Sharpe	0.10 %	0.0089	0.0257	-0.0006	0.0108
	0.25 %	0.0054	-0.0043	0.0082	-0.0210
Rendimiento B&H ajustado en el riesgo*	0.10 %	22.8490	4.1075	11.6227	2.0638
	0.25 %	27.5098	3.7956	6.7790	1.7580
Ratio de Sharpe estrategia B&H ajustada en el riesgo*	0.10 %	0.0050	0.0077	0.0035	0.0067
	0.25 %	0.0048	0.0077	0.0042	0.0067
Porcentaje de días en el mercado	0.10 %	60.57 %	58.54 %	69.08 %	53.01 %
	0.25 %	75.20 %	58.54 %	35.74 %	53.01 %
Días de transacción	0.10 %	98	115	103	120
	0.25 %	20	115	28	120

Cuadro 4.72: Estadísticos obtenidos para las reglas encontradas para el año 1999.

\*: A la hora de calcular estos estadísticos se ha ponderado por el porcentaje de días de cada uno de los modelos

Dimensión de inmersión	4					
Porcentaje de puntos próximos	59					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	15.5528	18.0274	17.1446	15.6548	19.0166	17.5497
Rend. B&H	9.8377	10.0850	10.0850	10.3324	10.7282	10.7282
Exceso de rend. sobre B&H	5.7151	7.9424	7.0595	5.3224	8.2884	6.8214
Ratio de Sharpe	0.0748	0.0843	0.0798	0.0725	0.0868	0.0793
Ratio de Sharpe B&H	0.0580	0.0575	0.0575	0.0570	0.0562	0.0562
Porcentaje de aciertos	51.0040 %	51.0040 %	51.0040 %	51.0040 %	51.0040 %	51.0040 %
Porcentaje de días en el mercado	57.8313 %	59.8393 %	59.8393 %	61.8473 %	65.0602 %	65.0602 %

Cuadro 4.73: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el año 1999 y coste del 0 %

Dimensión de inmersión	4					
Porcentaje de puntos próximos	59					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	3.9528	7.0274	6.5446	6.0548	10.4166	9.1497
Rend. B&H	9.7220	9.9654	9.9654	10.2087	10.5981	10.5981
Exceso de rend. sobre B&H	-5.7692	-2.9380	-3.4208	-4.1539	-0.1815	-1.4484
Ratio de Sharpe	0.0189	0.0327	0.0304	0.0280	0.0474	0.0412
Ratio de Sharpe B&H	0.0573	0.0568	0.0568	0.0563	0.0556	0.0556
Porcentaje de aciertos	51.0040 %	51.0040 %	51.0040 %	51.0040 %	51.0040 %	51.0040 %
Porcentaje de días en el mercado	57.8313 %	59.8393 %	59.8393 %	61.8473 %	65.0602 %	65.0602 %

Cuadro 4.74: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el año 1999 y coste del 0.10 %

Dimensión de inmersión	4					
Porcentaje de puntos próximos	59					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	-13.4473	-9.4727	-9.3555	-8.3452	-2.4834	-3.4504
Rend. B&H	9.5485	9.7859	9.7859	10.0232	10.4029	10.4029
Exceso de rend. sobre B&H	-22.9958	-19.2585	-19.1413	-18.3684	-12.8864	-13.8533
Ratio de Sharpe	-0.0634	-0.0435	-0.0430	-0.0382	-0.0112	-0.0154
Ratio de Sharpe B&H	0.0563	0.0558	0.0558	0.0553	0.0545	0.0545
Porcentaje de aciertos	51.0040 %	51.0040 %	51.0040 %	51.0040 %	51.0040 %	51.0040 %
Porcentaje de días en el mercado	57.8313 %	59.8393 %	59.8393 %	61.8473 %	65.0602 %	65.0602 %

Cuadro 4.75: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el año 1999 y coste del 0.25 %

Año 2000		
Coste	0.10 %	0.25 %
Dimensión de inmersión	1	1
Porcentaje de puntos próximos	2.6	2.6
Ancho de banda	70	79
Número de puntos próximos	127	127

Cuadro 4.76: Parámetros obtenidos mediante AGs para las reglas basadas en predicciones NN en el año 2000

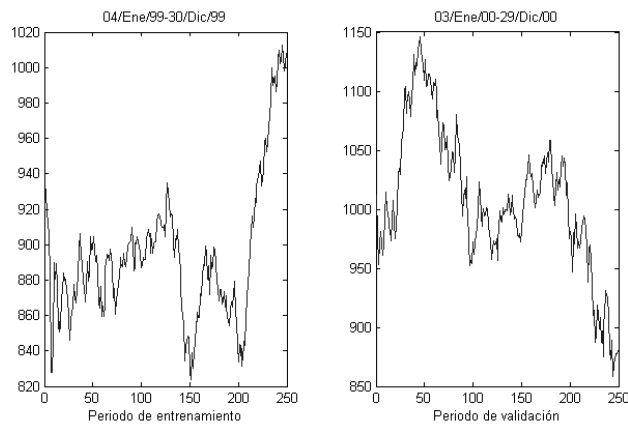


Figura 4.35: Periodo de entrenamiento y validación para el año 2000.

4.4. Optimización de los parámetros utilizados en predicciones por  
300 ocurrencias análogas mediante algoritmos genéticos

Año 2000					
	Coste	Periodo de entrenamiento		Periodo de evaluación	
		Modelo NN	AR(1)	Modelo NN	AR(1)
Exceso de rendimiento	0.10 %	38.3956	-4.7222	0.0675	-3.1066
ajustado en el riesgo	0.25 %	32.7097	-20.3733	-1.6337	-19.5785
Rendimiento medio	0.10 %	46.0842	-3.1416	-7.7446	-1.8835
anualizado	0.25 %	40.0757	-19.0971	-9.8510	-18.6586
Ratio de Sharpe	0.10 %	0.0119	-0.0118	-0.0018	-0.0072
	0.25 %	0.0106	-0.0397	-0.0023	-0.0398
Rendimiento B&H ajustado	0.10 %	7.6886	1.5806	-7.8121	1.2231
en el riesgo	0.25 %	7.3660	1.2763	-8.2173	0.9199
Ratio de Sharpe estrategia B&H	0.10 %	0.0041	0.0106	-0.0021	0.0077
ajustada en el riesgo*	0.25 %	0.0041	0.0106	-0.0021	0.0077
Porcentaje de días en el mercado*	0.10 %	60.57 %	53.60 %	69.08 %	51.22 %
	0.25 %	40.16 %	53.60 %	68.67 %	51.22 %
Días de transacción	0.10 %	28	120	18	125
	0.25 %	28	120	18	125

Cuadro 4.77: Estadísticos obtenidos para las reglas encontradas para el año 2000.

\*: A la hora de calcular estos estadísticos se ha ponderado por el porcentaje de días de cada uno de los modelos

Dimensión de inmersión	3					
Porcentaje de puntos próximos	31					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	-8.9708	-13.0915	-12.4719	-13.4779	-11.9909	-11.6969
Rend. B&H	-5.6831	-6.1087	-5.9668	-5.9668	-6.0377	-6.0377
Exceso de rend. sobre B&H	-3.2877	-6.9829	-6.5051	-7.5111	-5.9531	-5.6591
Ratio de Sharpe	-0.0351	-0.0509	-0.0480	-0.0517	-0.0463	-0.0452
Ratio de Sharpe B&H	-0.0298	-0.0307	-0.0304	-0.0304	-0.0306	-0.0306
Porcentaje de aciertos	49.3975 %	49.3975 %	49.3975 %	49.3975 %	49.3975 %	49.3975 %

Cuadro 4.78: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el año 2000 y coste del 0 %



Dimensión de inmersión	3					
Porcentaje de puntos próximos	31					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	-22.3708	-25.0916	-23.8719	-24.0779	-22.1909	-21.6969
Rend. B&H	-5.7940	-6.2243	-6.0809	-6.0809	-6.1526	-6.1526
Exceso de rend. sobre B&H	-16.5769	-18.8672	-17.7910	-17.9970	-16.0383	-15.5443
Ratio de Sharpe	-0.0878	-0.0977	-0.0921	-0.0925	-0.0859	-0.0841
Ratio de Sharpe B&H	-0.0304	-0.0313	-0.0310	-0.0310	-0.0312	-0.0312
Porcentaje de aciertos	49.3975 %	49.3975 %	49.3975 %	49.3975 %	49.3975 %	49.3975 %
Porcentaje de días en el mercado	55.4216 %	57.8313 %	57.0281 %	57.0281 %	57.4297 %	57.4297 %

Cuadro 4.79: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el año 2000 y coste del 0.10 %

Dimensión de inmersión	3					
Porcentaje de puntos próximos	31					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	-42.4709	-43.0916	-40.9720	-39.9779	-37.4909	-36.6969
Rend. B&H	-5.9602	-6.3978	-6.2520	-6.2520	-6.3249	-6.3249
Exceso de rend. sobre B&H	-36.5107	-36.6938	-34.7200	-33.7260	-31.1660	-30.3720
Ratio de Sharpe	-0.1662	-0.1674	-0.1582	-0.1532	-0.1451	-0.1422
Ratio de Sharpe B&H	-0.0313	-0.0322	-0.0319	-0.0319	-0.0320	-0.0320
Porcentaje de aciertos	49.3975 %	49.3975 %	49.3975 %	49.3975 %	49.3975 %	49.3975 %
Porcentaje de días en el mercado	55.4216 %	57.8313 %	57.0281 %	57.0281 %	57.4297 %	57.4297 %

Cuadro 4.80: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medios. Parámetros obtenidos mediante AGs para el año 2000 y coste del 0.25 %

304 **4.4. Optimización de los parámetros utilizados en predicciones por  
ocurrencias análogas mediante algoritmos genéticos**

---

Año 2001		
Coste	0.10 %	0.25 %
Dimensión de inmersión	6	1
Porcentaje de puntos próximos	0.9	2.9
Ancho de banda	18	72
Número de puntos próximos	46	149

Cuadro 4.81: Parámetros obtenidos mediante AGs para las reglas basadas en predicciones NN en el año 2001

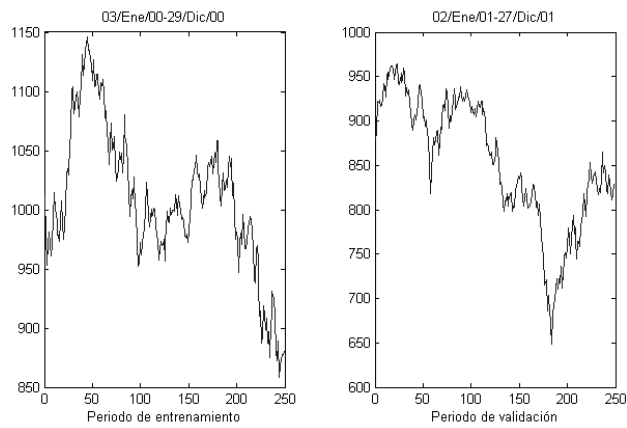


Figura 4.36: Periodo de entrenamiento y validación para el año 2001.

Año 2001					
	Coste	Periodo de entrenamiento		Periodo de evaluación	
		Modelo NN	AR(1)	Modelo NN	AR(1)
Exceso de rendimiento ajustado en el riesgo	0.10 %	15.8922	-11.7517	1.0494	-12.2796
	0.25 %	11.9346	-27.0754	-1.1524	-28.2074
Rendimiento medio anualizado	0.10 %	11.0421	-9.8962	-0.4265	-10.2433
	0.25 %	3.0105	-25.5251	-4.2877	-26.4767
Ratio de Sharpe	0.10 %	0.0027	-0.0518	-0.0001	-0.0536
	0.25 %	0.0007	-0.0592	-0.0008	-0.0620
Rendimiento B&H ajustado en el riesgo*	0.10 %	-4.8502	1.8555	-1.4760	2.0364
	0.25 %	-8.9241	1.5504	-3.1352	1.7307
Ratio de Sharpe estrategia B&H ajustada en el riesgo*	0.10 %	-0.0018	0.3033	-0.0004	0.3256
	0.25 %	-0.0022	0.3033	-0.0006	0.3256
Porcentaje de días en el mercado	0.10 %	49.80 %	48.59 %	53.63 %	47.18 %
	0.25 %	73.09 %	48.59 %	67.34 %	47.18 %
Días de transacción	0.10 %	121	127	133	133
	0.25 %	22	127	10	133

Cuadro 4.82: Estadísticos obtenidos para las reglas encontradas para el año 2001.  
\*: A la hora de calcular estos estadísticos se ha ponderado por el porcentaje de días de cada uno de los modelos.

	8					
	7					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	-1.2837	-9.1655	-6.5012	-8.9252	-13.1768	-13.3640
Rend. B&H	-0.9867	-0.9867	-1.1544	-1.3221	-1.6575	-1.6155
Exceso de rend. sobre B&H	-0.2970	-8.1787	-5.3468	-7.6031	-11.5193	-11.7485
Ratio de Sharpe	-0.0049	-0.0356	-0.0249	-0.0330	-0.0474	-0.0481
Ratio de Sharpe B&H	-0.0049	-0.0049	-0.0056	-0.0062	-0.0074	-0.0073
Porcentaje de aciertos	45.5645 %	45.5645 %	45.5645 %	45.5645 %	45.5645 %	45.5645 %
Porcentaje de días en el mercado	51.6129 %	51.6129 %	53.2258 %	54.8387 %	58.0645 %	57.6612 %

Cuadro 4.83: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medios. Parámetros obtenidos mediante AGs para el año 2001 y coste del 0 %

Dimensión de inmersión	8					
Porcentaje de puntos próximos	7					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	-15.0837	-22.5655	-18.9012	-20.7252	-23.1768	-22.7640
Rend. B&H	-1.0900	-1.0900	-1.2609	-1.4318	-1.7736	-1.7309
Exceso de rend. sobre B&H	-13.9938	-21.4755	-17.6404	-19.2934	-21.4032	-21.0332
Ratio de Sharpe	-0.0579	-0.0879	-0.0728	-0.0769	-0.0837	-0.0823
Ratio de Sharpe B&H	-0.0055	-0.0055	-0.0061	-0.0068	-0.0079	-0.0078
Porcentaje de aciertos	45.5645 %	45.5645 %	45.5645 %	45.5645 %	45.5645 %	45.5645 %
Porcentaje de días en el mercado	51.6129 %	51.6129 %	53.2258 %	54.8387 %	58.0645 %	57.6612 %

Cuadro 4.84: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el año 2001 y coste del 0.10 %

Dimensión de inmersión	8					
Porcentaje de puntos próximos	7					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	-35.7838	-42.6656	-37.5013	-38.4252	-38.1768	-36.8641
Rend. B&H	-1.2448	-1.2448	-1.4206	-1.5963	-1.9478	-1.9038
Exceso de rend. sobre B&H	-34.5390	-41.4207	-36.0807	-36.8289	-36.2290	-34.9602
Ratio de Sharpe	-0.1374	-0.1662	-0.1445	-0.1428	-0.1382	-0.1336
Ratio de Sharpe B&H	-0.0062	-0.0062	-0.0069	-0.0075	-0.0087	-0.0085
Porcentaje de aciertos	45.5645 %	45.5645 %	45.5645 %	45.5645 %	45.5645 %	45.5645 %
Porcentaje de días en el mercado	51.6129 %	51.6129 %	53.2258 %	54.8387 %	58.0645 %	57.6612 %

Cuadro 4.85: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadráticos medios. Parámetros obtenidos mediante AGs para el año 2001 y coste del 0.25 %

Año 2002		
Coste	0.10 %	0.25 %
Dimensión de inmersión	4	4
Porcentaje de puntos próximos	0.6	2.3
Ancho de banda	85	36
Número de puntos próximos	32	124

Cuadro 4.86: Parámetros obtenidos mediante AGs para las reglas basadas en predicciones NN en el año 2002

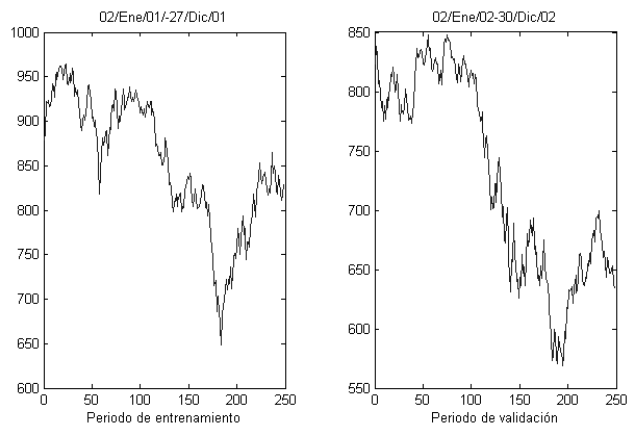


Figura 4.37: Periodo de entrenamiento y validación para el año 2002.



310 **4.4. Optimización de los parámetros utilizados en predicciones por  
ocurrencias análogas mediante algoritmos genéticos**

Año 2002					
	Coste	Periodo de entrenamiento		Periodo de evaluación	
		Modelo NN	AR(1)	Modelo NN	AR(1)
Exceso de rendimiento ajustado en el riesgo	0.10 %	47.3478	-11.7787	-10.9368	-11.7852
	0.25 %	12.4125	-27.7501	-21.5715	-27.8043
Rendimiento medio anualizado	0.10 %	45.6694	-10.0061	-21.9658	-10.3563
	0.25 %	10.6818	-26.2823	-32.6521	-26.6793
Ratio de Sharpe	0.10 %	0.0087	-0.0537	-0.0057	-0.0553
	0.25 %	0.0022	-0.0621	-0.0086	-0.0630
Rendimiento B&H ajustado en el riesgo*	0.10 %	-1.6784	1.7726	-11.0290	1.4288
	0.25 %	-1.7307	1.4677	-11.0806	1.1250
Ratio de Sharpe estrategia B&H ajustada en el riesgo*	0.10 %	-0.0004	0.1126	-0.0036	0.1001
	0.25 %	-0.00042	0.1126	-0.0036	0.1001
Porcentaje de días en el mercado	0.10 %	55.65 %	52.82 %	48.79 %	46.77 %
	0.25 %	53.23 %	52.82 %	47.98 %	46.77 %
Días de transacción	0.10 %	114	133	103	134
	0.25 %	110	133	111	134

Cuadro 4.87: Estadísticos obtenidos para las reglas encontradas para el año 2002.  
\*: A la hora de calcular estos estadísticos se ha ponderado por el porcentaje de días de cada uno de los modelos.

Dimensión de inmersión	2					
Porcentaje de puntos próximos	7					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	-27.8118	-25.6073	-24.2178	-23.9107	-24.9498	-17.6676
Rend. B&H	-9.5973	-8.9897	-8.6251	-9.1112	-8.8682	-9.3542
Exceso de rend. sobre B&H	-18.2146	-16.6176	-15.5926	-14.7995	-16.0816	-8.3133
Ratio de Sharpe	-0.1074	-0.1005	-0.0961	-0.0932	-0.0974	-0.0687
Ratio de Sharpe B&H	-0.0541	-0.0532	-0.0526	-0.0534	-0.0530	-0.0538
Porcentaje de aciertos	47.1774 %	47.1774 %	47.1774 %	47.1774 %	47.1774 %	47.1774 %
Porcentaje de días en el mercado	42.7419 %	40.7258 %	39.5161 %	41.1290 %	40.3225 %	41.9354 %

Cuadro 4.88: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el año 2002 y coste del 0 %

Dimensión de inmersión	2					
Porcentaje de puntos próximos	7					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	-41.0119	-38.0073	-35.6178	-34.7107	-35.7498	-27.4676
Rend. B&H	-9.6827	-9.0711	-8.7042	-9.1935	-8.9488	-9.4381
Exceso de rend. sobre B&H	-31.3291	-28.9362	-26.9136	-25.5172	-26.8010	-18.0295
Ratio de Sharpe	-0.1583	-0.1490	-0.1414	-0.1353	-0.1395	-0.1067
Ratio de Sharpe B&H	-0.0546	-0.0537	-0.0531	-0.0539	-0.0535	-0.0542
Porcentaje de aciertos	47.1774 %	47.1774 %	47.1774 %	47.1774 %	47.1774 %	47.1774 %
Porcentaje de días en el mercado	42.7419 %	40.7258 %	39.5161 %	41.1290 %	40.3225 %	41.9354 %

Cuadro 4.89: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el año 2002 y coste del 0.10 %

Dimensión de inmersión	2					
Porcentaje de puntos próximos	7					
Anchura del filtro	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000
Rend. Neto de la regla	-60.8119	-56.6074	-52.7178	-50.9107	-51.9499	-42.1676
Rend. B&H	-9.8110	-9.1933	-8.8227	-9.3168	-9.0698	-9.5639
Exceso de rend. sobre B&H	-51.0009	-47.4140	-43.8951	-41.5939	-42.8801	-32.6037
Ratio de Sharpe	-0.2334	-0.2205	-0.2085	-0.1977	-0.2017	-0.1632
Ratio de Sharpe B&H	-0.0553	-0.0544	-0.0538	-0.0546	-0.0542	-0.0550
Porcentaje de aciertos	47.1774 %	47.1774 %	47.1774 %	47.1774 %	47.1774 %	47.1774 %
Porcentaje de días en el mercado	42.7419 %	40.7258 %	39.5161 %	41.1290 %	40.3225 %	41.9354 %

Cuadro 4.90: Regla técnica con predicciones NN y función objetivo Raíz cuadrada del error cuadrático medio. Parámetros obtenidos mediante AGs para el año 2002 y coste del 0.25 %

En las Figuras 4.32 hasta 4.37, se muestran los periodos de entrenamiento y validación para cada uno de los años. Se puede observar como los tres primeros coinciden con periodos de alzas mientras que los tres últimos corresponden a la fase bajista correspondiente al estallido de la burbuja especulativa.

Los cuadros 4.61 hasta 4.87 muestran los parámetros obtenidos para cada uno de los periodos de entrenamiento correspondientes a cada año analizado, así como los valores de los estadísticos utilizados para medir la capacidad de generar beneficios de nuestra regla y su comparación con la estrategia B&H y el modelo AR(1).

De nuevo, los resultados confirman cómo las reglas que consideran los costes de transacción dentro de la función objetivo modifican los parámetros a optimizar de tal forma que hacen disminuir los días de transacción cuando aumentan los costes por operación. Sin embargo, esto no conduce a que los rendimientos obtenidos por estas reglas sean siempre superiores a los obtenidos a través de una estrategia B&H. Igualmente, se puede comprobar cómo en las fases bajistas, a pesar de obtener rendimientos negativos, la regla NN sí supera los rendimientos obtenidos mediante un modelo AR(1).

De forma más específica, pasaré a comentar, año por año, el comportamiento de las reglas técnicas generadas mediante la predicción NN.

El exceso de rendimientos sobre la estrategia B&H ajustada por riesgo es superior con la predicción NN que con el modelo AR(1) en la mayoría de los casos. Así durante los años 2000, 2001, y 2002 nuestra exceso de rendimientos sobre B&H supera al del modelo AR(1) para ambos costes de transacción de 0.1 % y 0.25 %. En cambio, para los años 1997,1998 y 1999 sólo supera al modelo

AR para costes de transacción del 0.25 %, aunque con rotundidad estos últimos dos periodos: un 7.8293 frente al -11.2192 en el año 1998, y un 22.2847 frente al -14.2306 para el año 1999.

Respecto a la capacidad de superar por parte de nuestra regla al propio modelo B&H hay que señalar que esto se produce durante el año 1997 (año de subidas generalizadas) con un exceso de 5.7535, para costes del 0.10 %; durante el año 1998 (año de subidas generalizadas) con un exceso de 7.8293, para costes del 0.25 %; durante el año 1999 (año de subidas generalizadas) con un exceso de 22.2847, para costes del 0.25 %; durante el año 2000 (año de descensos generalizados) con un exceso de 0.0675, para costes del 0.10 %; y durante el año 2001 (año de descensos generalizados) con un exceso de 1.0494, para costes del 0.10 %.

## **4.5. Creación de nuevas reglas técnicas mediante la programación genética**

La inmensa mayoría de los trabajos sobre reglas técnicas que aparecen en la literatura estudiaban su habilidad predictiva considerando reglas con una estructura prefijada de antemano.

Sin embargo, varios autores demostraron que estas estructuras podían provocar problemas de fisgoneo en los datos, por lo que había que utilizar técnicas de optimización que lo evitasen (Lo y MacKinley, 1990). Se han utilizado algoritmos genéticos para determinar de forma óptima los parámetros de las reglas más comunes dentro del análisis técnico (Véase Fernández Rodríguez *et al.* (2001), para un estudio sobre la relevancia del análisis técnico y Fernández Rodríguez *et al.* (2000), para una aplicación sobre la elección óptima de parámetros, ambos referidos al IGBM). En nuestro caso, las reglas estaban preestablecidas anteriormente por el analista. En este apartado de esta Tesis doctoral pretendemos encontrar nuevas técnicas de contratación mediante el uso de la programación genética. De esta forma, no sólo los parámetros, sino las mismas reglas, se seleccionan mediante un proceso de optimización.

El tipo de reglas que intentamos encontrar son aquellas que dividan los días de transacción en dos categorías: Días en el mercado (en que el inversor obtenga el rendimiento del mercado) y días fuera del mercado (donde le inversor obtenga un rendimiento libre de riesgo). Cada solución representa una regla técnica particular.

Seleccionaremos una población inicial de 500 programas cuyos nodos y terminales fueron escogidos entre funciones aritméticas (suma, resta,

multiplicación y división), valor absoluto y raíz cuadrada. Entre los terminales escogimos las constantes entre 0 y 1 y las medias móviles, precios máximos y precios mínimos con parámetros 1,2,3,4,5,10,15,20,40,60,80,100,150,200 y 250 cada uno. El algoritmo trabajaba con 45 entradas además de las constantes.

Posteriormente se seleccionan aleatoriamente cuatro individuos, se emparejan aleatoriamente y de cada par se elige aquel mejor adaptado, entendiéndolo como aquel que maximice el valor en la función objetivo. A este proceso se le denomina torneo y de aquí salen dos ganadores, uno por cada par. Después del torneo a estos ganadores se le aplican los operadores de búsqueda de cruce y mutación.

Los programas intercambian información genética a través del operador cruce. Los ganadores de cada torneo, si superan cierta probabilidad, que en nuestro caso es del 0.5, se intercambian entre ellos subárboles o bloques de instrucciones. En nuestro trabajo hemos empleado el operador cruce desarrollado por Banzhaf *et al.* (1998). Se selecciona una secuencia de bloque de instrucciones de uno de los ganadores del torneo. La posición y profundidad del subárbol se eligen de forma aleatoria. Este bloque se intercambia con un subárbol del otro ganador del torneo que esté en la misma posición y tenga la misma profundidad. Se genera de esta forma dos nuevos programas, llamados “hijos”, de la misma longitud que los padres. Se intenta de esta manera simular de una forma lo más realista la evolución natural, ya que intercambia bloques de instrucciones de grupos contiguos. Este tipo de cruce ocurre el 95 % de las veces. El otro 5 % está reservado para operar con el cruce no homólogo, que es el que más se había utilizado hasta ahora en trabajos de Programación Genética, donde el bloque de instrucciones a intercambiar no tiene porqué coincidir ni en posición ni en tamaño. En este tipo de cruce los hijos



sí pueden tener una longitud distinta a la de los padres.

La mutación causa cambios aleatorios en los programas de la población que han ganado los torneos. Si uno de los ganadores supera cierta probabilidad se genera un nuevo programa que queda anidado dentro del primero. El operador mutación se aplica probabilísticamente a todos los programas ganadores, independientemente de si éste ha sido seleccionado para el cruce.

Aunque la mayoría de los trabajos usan probabilidades de mutación baja, en nuestro trabajo hemos usado una probabilidad del 0.9, ya que experimentaciones previas sugirieron que da mejores resultados.

Finalmente, los hijos sustituyen en la población a los perdedores del torneo.

La función objetivo que se emplea en este trabajo es la de porcentaje de aciertos. Como se trata de un problema de clasificación hemos preferido usar ésta frente a la del error cuadrático medio o error lineal medio, ya que los problemas de clasificación son difíciles de resolver usando medidas basadas en el error. Debe tenerse en cuenta que no estamos buscando una función cuando ejecutamos un problema de clasificación, sino que intentamos conseguir una alta precisión en la clasificación, esto es, un alto ratio de acierto.

A continuación, las salidas de los programas que forman parte de la población existente en esa iteración se recodifican. Si la salida es igual o superior a 0.5 entonces se genera una señal que indica que tenemos que estar dentro del mercado. Por el contrario, si la salida es inferior a 0.5 se interpreta como una recomendación para estar fuera del mercado. En tal caso, se compara cuales de esas salidas proporcionadas por el programa están correctamente clasificadas. Para un torneo en concreto gana aquel programa que consigue un porcentaje de acierto mayor. En

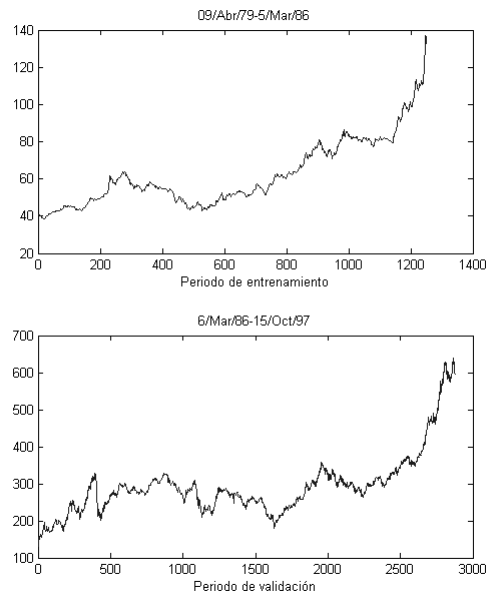


Figura 4.38: Periodo de entrenamiento y validación para el Escenario I

caso de empate, el algoritmo selecciona aquel con menor error cuadrático medio. Si aún así son iguales selecciona el ganador de forma aleatoria.

#### 4.5.1. Ejercicio experimental hasta 1997

En este apartado se ha utilizado una muestra de 4376 observaciones diarias del IGBM, al cierre, desde el desde el 2 de Enero de 1979 hasta el 15 de Noviembre de 1997. Se pierden 250 datos para la generación de las reglas que servirán de entrada para el algoritmo.

Se ha establecido tres escenarios. En el primer escenario el algoritmo se entrena con los primeros 1250 datos. La regla encontrada se valida con los 2876 datos posteriores [Figura 4.38].

En el segundo escenario se entrena con los 1251 datos posteriores al periodo de entrenamiento del escenario anterior, es decir desde el 1251 hasta el 2500, usando

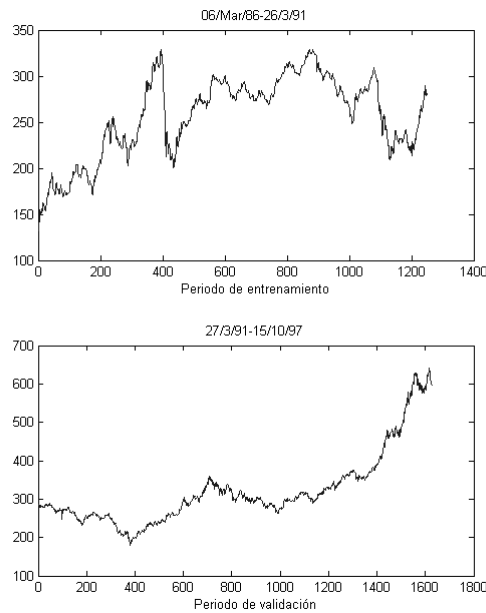


Figura 4.39: Periodo de entrenamiento y validación para el Escenario II

los datos restantes como periodo de validación [Figura 4.39].

Por último, el tercer escenario utiliza como datos de entrenamiento los precios al cierre de los 1250 días posteriores al periodo de entrenamiento del escenario anterior, dejando solamente 376 datos para validar la regla encontrada en este escenario [Figura 4.40].

Emplearemos varios estadísticos para medir la capacidad de generar beneficios que tiene la regla encontrada por el algoritmo.

- El *rendimiento neto medio anualizado* [Ecuación 4.1.6]

$$r = \sum_{t=1}^N S_{t-1} r_{m,t} + \sum_{t=1}^N (1 - S_{t-1}) r_{f,t} + d \cdot \log \frac{1 - c}{1 + c},$$

de la regla técnica de contratación consiste en su rendimiento bruto menos los costes de transacción asociados a las sucesivas entradas y salidas del

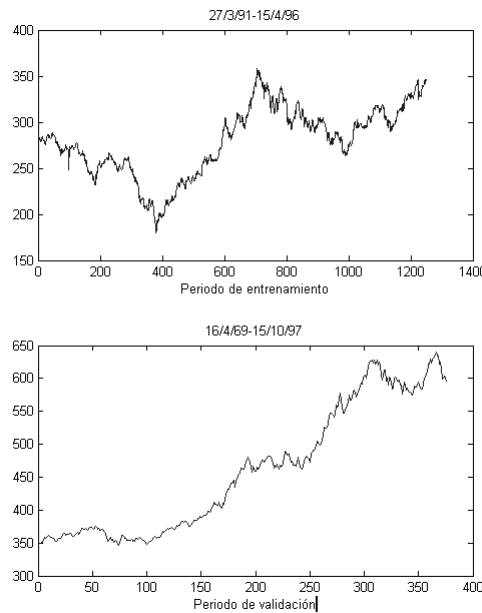


Figura 4.40: Periodo de entrenamiento y validación para el Escenario III

mercado.

- El *rendimiento neto de la estrategia B&H, ajustada por riesgo* [Ecuación 4.1.7]

$$r_{bha} = (1 - \alpha) \sum_{t=1}^N r_{f,t} + \alpha \sum_{t=1}^N r_{m,t} + \log \frac{1 - c}{1 + c}.$$

es la media ponderada del rendimiento neto de la estrategia B&H con el rendimiento obtenido por el activo libre de riesgo. Dicha ponderación se establece mediante unos coeficientes que representan el tanto por uno de los días dentro y fuera del mercado, respectivamente, que establece la regla técnica.

- El *ratio de Sharpe* [Ecuación 4.1.10] es el rendimiento medio neto, anualizado, de una estrategia de inversión por unidad de riesgo. Se obtiene a partir de la siguiente expresión:  $RS = \frac{\bar{r}}{\sigma}$  donde  $\bar{r}$  es el rendimiento medio

anualizado obtenido a través de las reglas técnicas,  $\sigma$  es la desviación típica de los rendimientos diarios.

- El *ratio de Sharpe de la estrategia BÉH ajustada en el riesgo* [Ecuación 4.1.12] que representa el rendimiento medio de la dicha estrategia por unidad de riesgo.

$$RS_{bha} = \frac{\alpha \cdot \overline{r_{m,t}} + (1 - \alpha) \cdot \overline{r_{f,t}}}{\sigma_{bh}}$$

- La variable *exceso de rendimientos sobre la estrategia BÉH, ajustada por riesgo* [Ecuación 4.1.8], que constituye un indicador clave para comprobar el comportamiento de nuestras

reglas técnicas.

- El rendimiento total estimado, que podemos definirlo como

$$\hat{R} = \sum_{t=n+1}^{n+N+1} \hat{y}_t r_{m,t}$$

donde  $N$  es el horizonte de predicción,  $r_{m,t} = \log\left(\frac{P_t}{P_{t-1}}\right)$  es el rendimiento en el periodo  $t$  e  $\hat{y}_t$  es la posición recomendada en el periodo  $t$  que se modelizará como función de los rendimientos pasados y  $n$  representa el número de observaciones.

- La *ratio de beneficio ideal* que mide el cociente entre el rendimiento total estimado por la red y un predictor perfecto (es decir, un predictor en el que se ha acertado correctamente todas las posiciones):

$$R_i = \frac{\sum_{t=n+1}^{n+N+1} \hat{y}_t r_{m,t} + T \cdot \{\ln \frac{1-c}{1+c}\}}{\sum_{t=n+1}^{n+N+1} |r_{m,t}| T \cdot \{\ln \frac{1-c}{1+c}\}} \quad (4.5.1)$$

donde  $\hat{y}_t$  representa la “posición recomendada” que nos ofrece la regla en cada instante.

- Para medir la bondad predictiva del ajuste emplearemos el contraste de Pesaran y Timmermann (1992), que evalúa la significatividad en la predicción de los signos. Bajo la hipótesis nula de independencia entre los valores reales y los predichos, el estadístico de dicho contraste se distribuye, asintóticamente como una  $\mathcal{N}(0, 1)$ , por lo que los valores críticos a niveles de significación del 1 %, 5 % y 10 % como niveles de significación serán 2.33, 1.645 y 1.82, respectivamente.
- El porcentaje en la predicción de los signos, que mide el porcentaje de veces que la predicción acierta con la posición correcta de compra o venta.

Emplearemos el algoritmo tres veces para cada uno de los escenarios. A continuación se relacionan las distintas reglas encontradas para cada uno de los escenarios.

- **Escenario I. Regla 1.**

$$P_t - w \cdot mm2 \begin{cases} \geq 0 & 1 \\ < 0 & 0 \end{cases}, \quad [I1]$$

$$w = \left[ 1 + \frac{\left( \text{Resto} \left( \frac{\left( \frac{P_t}{mm2} \right)^4 - 0,5}{\left( \left( \frac{P_t}{mm2} \right)^4 - 0,5 \right) mm60 + P_{250}^{max}} \right) \right)^2}{P_{250}^{max}} \right]^{\frac{1}{4}},$$

donde  $P_t$  constituye el valor del Índice de la Bolsa de Madrid el día  $t$ ,  $mm2$  la media móvil de orden 2,  $mm60$  la media móvil de orden 60 y  $P_{250}^{max}$  el valor máximo del Índice ente los ocurridos los 250 días anteriores.

■ **Escenario I. Regla 2.**

$$P_t - P_2^{max} \begin{cases} \geq 0 & 1 \\ < 0 & 0 \end{cases}, \quad [I2]$$

donde  $P_t$  constituye el valor del Índice de la Bolsa de Madrid el día  $t$  y  $P_2^{min}$  representa el valor mínimo del Índice ocurrido durante los 2 días anteriores.

■ **Escenario I. Regla 3**

$$\begin{aligned} &\text{Si } P_t - \left(p_2^{min} + \frac{0,4533}{mm4}\right) \geq 0, \\ &\text{entonces: } P_t - \left(p_2^{min} + 0,5^{16}mm4\right) \begin{cases} \geq 0 & 1 \\ < 0 & 0 \end{cases}, \quad [I3] \\ &\text{en otro caso: } P_t - \left(p_2^{min} + 0,5^{32}mm4\right) \begin{cases} \geq 0 & 1 \\ < 0 & 0 \end{cases} \end{aligned}$$

donde  $P_t$  constituye el valor del Índice de la Bolsa de Madrid el día  $t$ ,  $P_2^{min}$  representa el valor mínimo del Índice ocurrido durante los 2 días anteriores y  $mm4$  la media móvil de orden 4.

■ **Escenario II. Regla 1**

$$\left| P_t - 0,11 \left( \frac{P_{250}^{max}}{mm100} \right) - P_2^{min} \right| \begin{cases} \geq 0,27 & 1 \\ < 0,27 & 0 \end{cases}, \quad [II1]$$

donde  $P_t$  constituye el valor del Índice de la Bolsa de Madrid el día  $t$ ,  $P_{250}^{max}$  representa el valor máximo del Índice ocurrido durante los 250 días anteriores,  $P_{250}^{min}$  el valor mínimo del Índice ocurrido durante los 2 días anteriores y  $mm100$  la media móvil de orden 100.

■ **Escenario II. Regla 2**

$$\begin{cases} P_t - mm2 < 0 \text{ y } 2P_t - 2mm2 - mm3 + mm4 \geq 0 & 1 \\ P_t - mm2 \leq 0 \text{ y } P_t - mm2 \geq 0,23 & 0 \\ \text{en otro caso:} & 0 \end{cases}, \quad [II2]$$

donde  $P_t$  representa el valor del Índice de la Bolsa de Madrid el día  $t$  y  $mm2$ ,  $mm3$ ,  $mm4$  constituyen las medias móviles de orden 2, 3 y 4, respectivamente.

■ **Escenario II. Regla 3**

$$\begin{cases} 4P_t - 5mm2 + mm4 < 0 \text{ y} \\ P_t - mm2 - 0,03mm15 \geq 0 \text{ ó } P_t - mm2 + 0,37mm15 \leq 0 & 1 \\ P_t - mm2 - 0,03mm15 < 0 \text{ ó } P_t - mm2 + 0,37mm15 > 0 & 0 \\ \text{en otro caso:} & 0 \end{cases}, \quad [II3]$$

donde  $P_t$  representa el valor del Índice de la Bolsa de Madrid el día  $t$  y  $mm2$ ,  $mm4$ ,  $mm15$  constituyen las medias móviles de orden 2, 4 y 15, respectivamente.



■ **Escenario III. Regla 1**

$$\begin{aligned}
 a &= 2 \left[ \frac{5,61}{(1+p_{150}^{min})^2 + P_{60}^{max}} + 2 + P_{150}^{min} - 0,37 \right] P_{60}^{max} \\
 b &= \frac{2,80}{(1+p_{150}^{min})^2 + P_{60}^{max}} + 1 \\
 c &= Resto \left( \frac{a}{b} \right) , \quad [III1] \\
 c &\begin{cases} \geq -0,32 & 1 \\ < -0,32 & 0 \end{cases}
 \end{aligned}$$

donde  $P_{150}^{min}$  representa el valor mínimo del Índice de la Bolsa de Madrid ocurrido durante los 150 días anteriores y  $P_{60}^{max}$  simboliza el valor máximo del Índice ocurrido durante los 60 días anteriores.

■ **Escenario III. Regla 2**

$$\begin{aligned}
 &\text{Si } mm4 - mm2 < -0,32 \\
 &\text{entonces } P_t + mm4 - 2mm2 \begin{cases} \geq -0,32 & 1 \\ \leq -0,32 & 0 \end{cases} , \quad [III2] \\
 &\text{en otro caso } P_t - mm2 \begin{cases} \geq -0,32 & 1 \\ \leq -0,32 & 0 \end{cases}
 \end{aligned}$$

donde  $P_t$  representa el valor del Índice de la Bolsa de Madrid el día  $t$ ,  $mm2$ ,  $mm4$  constituyen las medias móviles de orden 2, y 4, respectivamente.

■ **Escenario III. Regla 3**

$$\begin{aligned}
 &\text{Si } mm4 - mm2 < -0,32 \\
 &\text{entonces } P_t + mm4 - 2mm2 \begin{cases} \geq -0,32 & 1 \\ \leq -0,32 & 0 \end{cases} , \quad [III3] \\
 &\text{en otro caso } P_t - mm2 \begin{cases} \geq -0,32 & 1 \\ \leq -0,32 & 0 \end{cases}
 \end{aligned}$$

donde  $P_t$  representa el valor del Índice de la Bolsa de Madrid el día  $t$ ,  $mm2$ ,  $mm4$  constituyen las medias móviles de orden 2, y 4, respectivamente.

La mayoría de estas reglas son de difícil interpretación financiera, a diferencia de las estudiadas en el capítulo de 2. Sin embargo, sí que se puede deducir que las reglas encontradas por el algoritmo son reglas que resultan de diferencias del precio actual con terminales de parámetros cortos, aunque ponderadas. Estas ponderaciones sí que suelen tener terminales de parámetros más largos, pero básicamente las reglas utilizan terminales no superiores a 4. Así ocurre en las reglas [I2], [I3],[III2] y [III3]. Esto sugiere, que se han encontrados reglas muy activas, que provocarán bastantes días de transacción.

En el Cuadro 4.91 se presentan los porcentajes de acierto de cada una de las reglas encontradas para cada uno de los escenarios con los que hemos trabajado. Es necesario recordar que este porcentaje de aciertos era la función de adaptación con la que trabajaba el algoritmo y por la que éste decidía con que regla quedarse al final de la ejecución. “%+” indica el porcentaje de acierto de los días que ha tenido que predecir que teníamos que estar en el mercado. “%-” indica el porcentaje de acierto de los días que ha tenido que predecir que teníamos que estar fuera del mercado. “% Total” indica el porcentaje de aciertos de forma global.

En los cuadros 4.92 y 4.93 se presentan los estadísticos que indicarían la capacidad de generar beneficios y su significatividad estadística de la capacidad predictiva que tienen estas reglas. Estos estadísticos están explicados al principio de este apartado.

En la tabla 4.94 se muestran los resultados en el periodo de entrenamiento y

Periodo de entrenamiento									
	Escenario I			Escenario II			Escenario III		
%+	67.96 %	72.88 %	67.81 %	68.40 %	67.94 %	65.04 %	62.04 %	57.26 %	57.26 %
%-	72.88 %	69.95 %	74.78 %	64.03 %	65.38 %	67.38 %	51.85 %	56.18 %	56.18 %
%Total	70.24 %	71.52 %	71.04 %	66.32 %	66.72 %	66.16 %	56.96 %	56.72 %	56.72 %
Periodo de validación									
	Escenario I			Escenario II			Escenario III		
%+	51.94 %	69.59 %	39.44 %	53.98 %	56.24 %	54.46 %	57.01 %	57.01 %	57.01 %
%-	67.03 %	48.04 %	77.83 %	57.45 %	55.67 %	57.20 %	43.21 %	58.64 %	58.64 %
%Total	59.20 %	59.25 %	57.86 %	55.66 %	55.97 %	55.78 %	51.06 %	57.72 %	57.72 %

Cuadro 4.91: Porcentaje de aciertos. %+: Porcentaje de acierto de los días en el mercado. % -: Porcentaje de acierto de los días fuera del mercado. %Total: Porcentaje de acierto total.

	Escenario I			Escenario II			Escenario III		
R	4.1425	4.4630	4.2638	4.4667	4.6773	4.5097	1.3082	1.4686	1.4686
Ri	57.92 %	62.40 %	59.62 %	42.16 %	44.15 %	42.57 %	14.67 %	16.47 %	16.47 %
Rma	84.3980	89.4100	86.8170	80.7605	84.8150	82.2099	22.1220	24.6847	24.6847
Rs	0.0270	0.0276	0.0280	0.01620	0.0172	0.0167	0.0071	0.0079	0.0079
Alfa	49.04 %	53.04 %	48.08 %	52.96 %	52.08 %	49.60 %	55.16 %	50.56 %	50.56 %
T	372	396	360	350	529	479	147	617	617
Rbha	20.6570	21.1290	20.5440	15.2849	15.2595	15.1879	7.1371	7.4309	7.4309
Rsbha	0.1200	0.0114	0.0122	0.0052	0.0053	0.0056	0.0032	0.0036	0.0036
R25	52.646	54.8940	55.594	51.3164	41.2630	42.8546	13.3350	-8.8640	-8.8640
Rs25	0.0182	0.0183	0.0192	0.0095	0.0112	0.0098	0.0045	-0.0033	-0.0033
Pt	14.3400	15.1290	15.0310	11.4740	11.7772	11.4551	4.9363	4.7513	4.7513
Ps	70.24 %	71.52 %	71.04 %	66.32 %	66.72 %	66.16 %	56.96 %	56.72 %	56.72 %

Cuadro 4.92: Periodo de entrenamiento. R: Rendimiento total. Ri: Rendimiento ideal. Rma: Rendimiento medio anualizado. Rs: Ratio de Sharpe. Alfa: Porcentaje de días en el mercado. T: Días de transacción. Rbha: Rendimiento B&H ajustado por riesgo. Rsbha: Ratio de sahrpe de la estrategia “buy and hold” ajustada por riesgo. R25: Rendimiento medio anualizado de la regla con unos costes del 0.25 %. RS25: Ratio de Sharpe de la regla con costes del 0.25 %. Pt: Test de Pesaran y Timmerman. Ps: Probabilidad en signos

	Escenario I			Escenario II			Escenario III		
R	5.8966	5.4664	5.4749	1.9537	1.9525	2.1535	0.1285	0.3421	0.3421
Ri	26.45 %	24.52 %	%	16.70 %	16.69 %	18.41 %	4.53 %	12.30 %	12.30 %
Rma	47.5706	41.9240	46.589	29.4934	29.2544	31.6262	28.4940	38.8489	38.8489
Rs	0.0132	0.0100	0.01445	0.0099	0.0097	0.0106	0.0090	0.0115	0.0115
Alfa	42.84 %	61.13 %	31.15 %	48.46 %	50.49 %	48.83 %	56.92 %	50.27 %	50.27 %
T	1152	1394	996	691	825	767	77	195	195
Rbha	12.6186	13.1170	12.3010	10.9731	11.0260	10.9828	26.268	23.7646	23.7646
Rsbha	0.0061	0.0045	0.0082	0.0055	0.0053	0.0054	0.0108	0.0111	0.0111
R25	14.3783	4.2783	17.5990	-1.2103	-6.4201	-2.5198	12.4460	-0.4474	-0.4474
Rs25	0.0045	0.0012	0.0061	-0.0005	-0.0025	-0.0010	0.00417	-0.0002	-0.0002
Pt	10.2700	9.6900	9.9882	4.6110	4.8000	4.6992	0.04252	3.0068	3.0068
Ps	59.18 %	59.25 %	57.86 %	55.66 %	55.97 %	55.78 %	51.06 %	57.71 %	57.71 %

Cuadro 4.93: Periodo de validación. R: Rendimiento total. Ri: Rendimiento ideal. Rma: Rendimiento medio anualizado. Rs: Ratio de Sharpe. Alfa: Porcentaje de días en el mercado. T: Días de transacción. Rbha: Rendimiento B&H ajustado por riesgo. Rsbha: Ratio de sahrpe de la estrategia “buy and hold” ajustada por riesgo. R25: Rendimiento medio anualizado de la regla con unos costes del 0.25 %. RS25: Ratio de Sharpe de la regla con costes del 0.25 %. Pt: Test de Pesaran y Timmerman. Ps: Probabilidad en signos

validación para la mejor regla encontrada en el periodo de entrenamiento, esto es, aquella de las tres que más porcentaje de aciertos ha tenido en este periodo.

Como se puede observar, el algoritmo seleccionó reglas muy activas que en ausencia de costes de transacción provocan altos rendimientos, que en todos los casos baten a la regla B&H ajustada por riesgo. La incorporación a las reglas de los costes de transacción, en nuestro caso del 0.25 %, produce una disminución drástica del rendimiento, aunque se siguen manteniendo por encima del B&H ajustado en el riesgo. Este hecho se acentúa aún más en el análisis de los resultados en el periodo de validación de cada uno de los escenarios. Aquí las reglas encontradas, una vez descontados los costes, no superan a la estrategia de comprar y mantener ajustadas en el riesgo, incluso muchas de ellas originan rendimientos negativos.

Hay que observar que la optimización de las reglas se afrontó como un problema de clasificación por lo que se maximizó el porcentaje de aciertos. Este objetivo se consiguió con esta técnica ya que todas las reglas superaron el 50 % de acierto en los signo, y de forma significativa como refleja el test de Pasaran y Timmerman. Por lo que el modelo goza de cierta capacidad predictiva pero, a la vista de los resultados, no garantiza que tenga capacidad de generar beneficios como demuestra el hecho de que los rendimientos una vez sustraído los costes.

Es decir, acertar en el porcentaje de acierto no garantiza la obtención de beneficios superiores a los normales en presencia de costes de transacción. Si seleccionamos reglas activas podríamos acertar un mayor número de veces, pero provocaríamos altos costes de transacción debido al número de transacciones

	Periodo de entrenamiento			Periodo de validación		
	Escenario I	Escenario II	Escenario III	Escenario I	Escenario II	Escenario III
Rma	89.41	84.8150	22.122	41.924	29.2544	28.494
Rs	0.02762	0.0172	0.0071302	0.0100	0.0097	0.0090
Alfa	53.04 %	52.08 %	55.12 %	61.13 %	50.49 %	56.92 %
T	396	529	147	1394	825	77
Rbha	21.129	15.2595	7.1371	13.117	11.0260	26.268
Rsbha	0.011361	0.0053	0.0031772	0.0045	0.0053	0.010767
R25	54.894	41.2630	13.335	4.2783	-6.4201	12.446
Rs25	0.018279	0.0095	0.0044715	0.0012	-0.0025	0.00417
Pt	15.129	11.7772	4.9363	9.69	4.80009	0.042523
Ps	71.52 %	66.72 %	56.96 %	59.25 %	55.97 %	51.06 %

Cuadro 4.94: Resultado de los estadísticos. Rma: Rendimiento medio anualizado. Rs: Ratio de Sharpe. Alfa: Porcentaje de días en el mercado. T: Días de transacción. Rbha: Rendimiento B&H ajustado por riesgo. Rsbha: Ratio de sahrpe de la estrategia “buy and hold” ajustada por riesgo. R25: Rendimiento medio anualizado de la regla con unos costes del 0.25 %. RS25: Ratio de Sharpe de la regla con costes del 0.25 %. Pt: Test de Pesaran y Timmerman. Ps: Probabilidad en signos

que ocurrirían con este tipo de reglas. Se debe crear una función que genere reglas capaces de captar y predecir las tendencias a más largo plazo, ya que de esta forma aunque perdamos por bajadas puntuales de la bolsa ganaríamos por la subida global producida en el periodo a la que se restaría unos costes de transacción bajos. En caso de predecir episodios bajistas con la regla ganaríamos el rendimiento libre de riesgo.

A modo de resumen, en este apartado hemos analizado la posibilidad de implementar en el Índice General de la Bolsa de Madrid reglas de contratación creadas mediante programación genética. Esta nueva técnica permite diseñar nuevas reglas técnicas distintas a las usadas más comúnmente en el mercado. Las reglas encontradas son capaces de superar a la estrategia “B&H” en ausencia de costes de transacción. Vemos, además, que la regla ha sido capaz de predecir correctamente más del 50% de las señales. Esto explicaría la capacidad de generar beneficios que tiene la regla encontrada en ausencia de costes. Sin embargo, no es capaz de batir a la estrategia B&H ajustada en el riesgo cuando tenemos en cuenta los costes de transacción. Esto se debe a que el algoritmo no optimiza el rendimiento medio anualizado sino el porcentaje de aciertos. Es objetivo que se plantea para futuros trabajos encontrar nuevas reglas que sean optimizadas en función del exceso de rendimiento sobre la estrategia “buy and hold” ajustada en el riesgo. No obstante, ello trae consigo el alterar los paquetes estándar de programación genética con el fin de sustituir las funciones de pérdidas estadísticas que usa habitualmente, incorporando funciones de pérdida de contenido económico y financiero.



### 4.5.2. Análisis empírico para los años 1997 a 2002

Al igual que los apartados anteriores hemos ampliado la muestra intentando medir la capacidad que tienen las reglas desarrolladas por la programación genética para generar beneficios en el IGBM durante el periodo más controvertido de la bolsa, que corresponde a la fase de burbuja especulativa y sus posteriores reajustes. Se han tomado seis años de predicción que corresponden a años naturales comprendidos entre 1997 a 2002. Las reglas han sido diseñadas atendiendo a la información proporcionada por los periodos de entrenamiento que constituyen los precios al cierre de Índice desde el 2 de enero de 1979 hasta el día anterior al comienzo del año a predecir.

Los estadísticos empleados se corresponden con los analizados para el estudio metodológico de las reglas diseñadas mediante programación genética.

Los resultados se muestran en los siguientes cuadros:

■ **Regla encontrada para el año 1997.**

$$\begin{aligned} P_{250}^{max}(P_{250}^{max} - 1) &\geq P_{10}^{max} && 1 \\ \text{en otro caso} &&& 0 \end{aligned} \quad ,$$

donde  $P_{250}^{max}$  y  $P_{10}^{max}$  constituyen los valores máximos del Índice ocurridos los 250 y 10 días anteriores, respectivamente.

■ **Regla encontrada para el año 1998.**

$$\begin{aligned} (|mm3| - mm4)(mm6 - P_t)(|mm3| - mm4) &\geq -3 && 1 \\ \text{en otro caso} &&& 0 \end{aligned} \quad ,$$

donde  $P_t$  es el valor del Índice General de la Bolsa de Madrid en el instante  $t$ ,  $mm3$ ,  $mm4$ , y  $mm6$  representan las medias móviles de órdenes 3,4 y 6, respectivamente.

- Regla encontrada para el año 1999.

$$P_t - mm2 \geq 0 ,$$

donde  $P_t$  constituye el valor del Índice General de la Bolsa de Madrid en el instante  $t$  y  $mm2$  representa la media móvil de orden 2.

- Regla encontrada para el año 2000.

$$\begin{aligned} &\text{Si } mm3 - mm20 \geq \sqrt{P_t} \\ &\text{y } mm2 + mm20 - mm3 \geq P_t - 1 , \\ &\text{en otro caso} \qquad \qquad \qquad 0 \end{aligned}$$

donde  $P_t$  constituye el valor del Índice General de la Bolsa de Madrid en el instante  $t$  y  $mm2$ ,  $mm3$  y  $mm20$  representan las medias móviles de orden 2, 3 y 20, respectivamente.

- Regla encontrada para el año 2001.

$$\begin{aligned} &1 - P_{60}^{max} - P_{100}^{min} - P_{120}^{min} - P_{250}^{max} \geq 0 - 1 , \\ &\text{en otro caso} \qquad \qquad \qquad 0 , \end{aligned}$$

donde  $P_{60}^{max}$  y  $P_{250}^{max}$  representan los valores máximos del Índice ocurridos los 60 y 250 días anteriores, respectivamente.  $P_{100}^{min}$  y  $P_{120}^{min}$  constituyen los valores mínimos del Índice General de la Bolsa de Madrid ocurridos en los 100 y 120 días anteriores, respectivamente.

- Regla encontrada para el año 2001.

$$\begin{aligned} &mm3 - mm2 \geq 0 - 1 , \\ &\text{en otro caso} \qquad \qquad \qquad 0 , \end{aligned}$$

donde  $mm2$  y  $mm3$  representan las medias móviles de orden 2 y 3 respectivamente.

Los Cuadros 4.96 y 4.97 muestran los estadísticos obtenidos para cada uno de los años estudiados. Volvemos a observar cómo un aumento de los costes disminuyen drásticamente los rendimientos a pesar de que acertamos más del 50 % de los signos. Esto indica que las reglas obtenidas mediante esta técnica son reglas que utilizan parámetros bajos y, por lo tanto, muy activas. Es por ello, que incluso en periodos de alzas generalizadas como las caracterizadas en los años 1997 a 1999 el exceso de rendimiento es negativo para costes altos. Esto se ve empeorado en periodos de caídas en bolsa. En los años 2000 a 2002 la regla no fue capaz de superar al B&H ajustado en el riesgo. Entendemos que esto podría solucionarse, o al menos mitigarse, con una función objetivo que incluyan costes de transacción.

De forma más específica, pasaré a comentar, año por año, el comportamiento de las reglas técnicas generadas mediante la programación genética durante los últimos seis años.

Durante los años 1997 y 1998, ambos de subidas generalizadas, las reglas de programación genética han obtenido un exceso de rendimientos sobre la estrategia B&H para costes de transacción del 0.1 %. Durante el año 1999, también de subidas en la bolsa de Madrid, solo se supero la estrategia B&H para costes de transacción cero. Durante los años 2000, 2001 y 2002, todos de caídas generalizadas, no fue posible superar la estrategia B&H, ni siquiera con coste de transacción cero.

Periodo de entrenamiento						
	Año 1997	Año 1998	Año 1999	Año 2000	Año 2001	Año 2002
%+	36.8 %	62.62 %	63.34 %	41.42 %	17.49 %	59.07 %
%-	80.91 %	47.89 %	58.10 %	52.00 %	79.48 %	55.42 %
%Total	57.76 %	55.60 %	60.85 %	46.46 %	47.17 %	57.33 %
Periodo de validación						
%+	56.20 %	71.53 %	59.83 %	54.74 %	32.85 %	56.93 %
%-	65.45 %	41.82 %	47.27 %	40.91 %	80.91 %	38.18 %
%Total	60.32 %	58.30 %	54.25 %	48.58 %	54.25 %	48.58 %

Cuadro 4.95: Porcentaje de aciertos. %+: Porcentaje de acierto de los días en el mercado. % -: Porcentaje de acierto de los días fuera del mercado. %Total: Porcentaje de acierto total.

	Coste	Exceso de rendimiento ajustado por riesgo	Rendimiento medio anualizado	Ratio de Sharpe	Rendimiento B&H ajustado por riesgo	Ratio de Sharpe estrategia B&H ajustada por riesgo	Porcentaje de días en el mercado	Dt
Año 1997	0 %	42.6266	57.7042	0.015	15.0776	0.0059	0.5853	1676
	0.10 %	26.4436	41.5063	0.0114	15.0627	0.0059	0.5853	1676
	0.25 %	5.2344	20.2748	0.0059	15.0404	0.0059	0.5853	1676
Año 1998	0 %	30.0214	44.2249	0.0179	14.2035	0.0101	0.3166	1120
	0.10 %	20.5353	34.7249	0.0146	14.1896	0.0101	0.3166	1120
	0.25 %	7.4664	21.6352	0.0095	14.1688	0.0101	0.3166	1120
Año 1999	0 %	11.7604	25.1529	0.0095	13.3925	0.0148	0.1941	1204
	0.10 %	3.4161	16.7956	0.0066	13.3795	0.0148	0.1941	1204
	0.25 %	-8.0657	5.2942	0.0022	13.36	0.0148	0.1941	1204
Año 2000	0 %	-9.0558	5.2221	0.0019	14.2779	0.008	0.3805	261
	0.10 %	-10.5265	3.7389	0.0014	14.2655	0.008	0.3805	261
	0.25 %	-12.6935	1.5533	0.0006	14.2468	0.008	0.3805	261
Año 2001	0 %	-14.0606	-1.0902	-0.0004	12.9704	0.0086	0.3175	544
	0.10 %	-16.7887	-3.8299	-0.0013	12.9588	0.0086	0.3175	544
	0.25 %	-20.7392	-7.7979	-0.0026	12.9413	0.0086	0.3175	544
Año 2002	0 %	22.1443	34.4324	0.0136	12.2881	0.0084	0.3029	1756
	0.10 %	11.0377	23.3148	0.0096	12.2771	0.0084	0.3029	1756
	0.25 %	-3.9224	8.3382	0.0036	12.2605	0.0084	0.3029	1756

Cuadro 4.96: Resultado de los estadísticos para el periodo de entrenamiento. Dt: Días de transacción

	Coste	Exceso de rendimiento ajustado por riesgo	Rendimiento medio anualizado	Ratio de Sharpe	Rendimiento B&H ajustado por riesgo	Ratio de Sharpe estrategia B&H ajustada por riesgo	Porcentaje de días en el mercado	Dt
Año 1997	0 %	21.6965	49.1027	0.0104	27.4062	0.0076	0.6113	117
	0.10 %	5.0561	32.2005	0.0073	27.1444	0.0076	0.6113	117
	0.25 %	-16.3819	10.3708	0.0025	26.7527	0.0076	0.6113	117
Año 1998	0 %	11.4205	19.8223	0.0094	8.4018	0.0085	0.1341	49
	0.10 %	5.7327	13.9109	0.0068	8.1782	0.0085	0.1341	49
	0.25 %	-2.2576	5.5861	0.0028	7.8436	0.0085	0.1341	49
Año 1999	0 %	0.6385	7.4515	0.0024	6.8129	0.0045	0.3092	119
	0.10 %	-11.4267	-4.8314	-0.0016	6.5952	0.0045	0.3092	119
	0.25 %	-26.9436	-20.6741	-0.0076	6.2695	0.0045	0.3092	119
Año 2000	0 %	-1.9463	-3.2921	-0.001	-1.3458	-0.0008	0.3012	21
	0.10 %	-3.7948	-5.3417	-0.0017	-1.5469	-0.0008	0.3012	21
	0.25 %	-6.4872	-8.335	-0.0026	-1.8477	-0.0008	0.3012	21
Año 2001	0 %	-15.5394	-16.6473	-0.004	-1.1079	-0.0002	0.4677	42
	0.10 %	-10.2882	-11.0917	-0.0026	-0.8035	-0.0002	0.4677	42
	0.25 %	-15.5394	-16.6473	-0.004	-1.1079	-0.0002	0.4677	42
Año 2002	0 %	-1.3256	-2.4261	-0.0012	-1.1004	-0.0012	0.1411	67
	0.10 %	-7.5943	-8.8971	-0.0045	-1.3028	-0.0012	0.1411	67
	0.25 %	-16.2024	-17.808	-0.0093	-1.6056	-0.0012	0.1411	67

Cuadro 4.97: Resultado de los estadísticos para el periodo de validación. Dt: Días de transacción



## Capítulo 5

# Conclusiones y desarrollos futuros

Pese a que por mucho tiempo los economistas han considerado los mercados bursátiles como impredecibles, justificando al análisis técnico como un conjunto de reglas carentes de toda capacidad predictiva (Veánse Samuelson, 1965; Malkiel, 1992, como trabajos pioneros en este sentido), durante los últimos quince años han aparecido numerosos indicios que confirman la posibilidad de predicción de un número elevado de activos financieros. En particular, el trabajo de Brock *et al.* (1992) ha supuesto todo un renacimiento del análisis técnico, poniendo de manifiesto que las conclusiones sobre su falta de poder predictivo eran, cuanto menos, precipitadas. En este sentido, se están produciendo numerosas investigaciones en muy diversos campos. Por una parte, debido a que la eficiencia e integración puede variar considerablemente de unos mercados a otros, tiene elevado valor empírico los trabajos aplicados a los mercados bursátiles de tamaño mediano como ocurre con el caso de la Bolsa de Madrid. Por otra parte, se ha tratado de desarrollar nuevas metodologías predictivas capaces tanto de mejorar las reglas técnicas de contratación ya existentes como de producir otras nuevas.



Esta Tesis presenta diversas aportaciones sobre la capacidad de generar beneficios de determinadas metodologías para crear reglas técnicas sobre el IGBM, que dan pie al siguiente primer grupo de conclusiones.

## **5.1. Conclusiones específicas sobre las metodologías predictivas aportadas a la serie IGBM**

### **5.1.1. Reglas técnicas obtenidas mediante redes neuronales artificiales**

Se ha examinado la capacidad de obtener beneficios de una simple regla técnica de contratación basada en redes neuronales aplicada sobre el IGBM. Los resultados sugieren que, en ausencia de costes de transacción, la regla técnica de contratación es siempre superior a la estrategia de “comprar y mantener”, tanto para episodios de bajas generalizadas en el precio de los activos (“bear” markets) como para episodios estables.

Por otra parte, se ha encontrado que la estrategia de “comprar y mantener” genera rendimientos más altos que la regla técnica basada en redes neuronales artificiales para subperiodos que presentan tendencias al alza (“bull” markets), lo cual parece razonable teniendo en cuenta las importantes revalorizaciones registradas en la Bolsa de Madrid durante los años noventa.

Tales conclusiones están en línea con las presentadas en Fernández Rodríguez *et al.* (1999c) en que se aplicaron predicciones no lineales en el índice Nikkei.

### **5.1.2. Optimización de reglas técnicas clásicas mediante algoritmos genéticos**

Se ha investigado la capacidad de obtener beneficios de las reglas técnicas de contratación más comúnmente usadas por los inversores como son las de cruce de medias móviles o la de ruptura de rango, pero realizando una aportación metodológica en el sentido de optimizar los valores de los parámetros que aparecen en tales reglas mediante un algoritmo genético.

Los resultados muestran que si se emplea como función de pérdida el rendimiento neto (una vez descontados los costes de transacción), al aumentar dichos costes las estrategias técnicas se hacen menos intensas, reduciéndose considerablemente el número de entradas y salidas del mercado.

Cuando se emplean costes de transacción razonables, similares a los que actualmente registra el mercado, las reglas técnicas optimizadas mediante algoritmos genéticos obtienen rendimientos superiores a los de la estrategia de “comprar y mantener” ajustada por riesgo.

### **5.1.3. Creación de nuevas reglas técnicas mediante el uso de la programación genética**

Se ha estudiado la posibilidad de implementar en el IGBM nuevas reglas técnicas de contratación creadas mediante la programación genética. Se trata de una nueva técnica de optimización que, gracias a su gran flexibilidad, permite diseñar reglas técnicas óptimas distintas a las más comúnmente usadas en el mercado, tales como las de cruce de medias móviles o las de ruptura de rango.

La nueva regla técnica encontrada mediante programación genética es capaz de predecir correctamente el signo de más del 50 % de las señales y, en ausencia

de costes de transacción, es capaz de superar a la estrategia de “comprar y mantener”. No obstante, si se consideran costes de transacción, la nueva regla técnica no es capaz de superar a la estrategia de “comprar y mantener”. Ello puede explicarse porque el proceso de optimización empleado en el programa de programación genética (Francone, 2000) utiliza como función de pérdida el porcentaje de aciertos predictivos y no el rendimiento tras considerar los costes de transacción. Las grandes dificultades que entraña la elaboración propia de un programa de programación genética ha obligado a acudir a uno de los pocos disponibles en el mercado. No obstante, en futuros trabajos se plantea el objetivo de encontrar nuevas reglas técnicas obtenidas mediante programación genética en el que la función de pérdidas emplee los rendimientos tras descontar los costes de transacción.

#### **5.1.4. Optimización de reglas técnicas mediante el uso de predictores por ocurrencias análogas**

Se ha diseñado un procedimiento para ajustar, mediante validación cruzada, los parámetros claves de una regla técnica simple basada en las predicciones por ocurrencias análogas. En este sentido, tales parámetros (como son la dimensión de inmersión, el número de entornos próximos y la anchura del filtro empleado para desechar las señales engañosas), se obtienen mediante optimización con un algoritmo genético de la función de pérdida dada por el exceso de rendimientos sobre la estrategia de “comprar y mantener” ajustada por riesgo.

Para niveles de coste de transacción razonables la regla técnica obtenida mediante las predicciones NN supera, en la mayoría de los periodos considerados, tanto a la estrategia de comprar y mantener como a una estrategia similar basada

en las predicciones del mejor modelo ARIMA.

## 5.2. Sobre la hipótesis de mercados eficientes en el IGBM

Si una determinada regla técnica produce, de forma consistente, beneficios positivos después de descontar los costes de transacción y de ajustarla por riesgo, ello puede considerarse como una evidencia contra la hipótesis de mercados eficientes.

En este sentido algunas de las metodologías empleadas en esta Tesis Doctoral parecen apuntar hacia la existencia de ineficiencia en la Bolsa de Madrid pues son capaces de mejorar la estrategia de “comprar y mantener” después de tener en cuenta los costes de transacción y de considerar alguna forma de prima de riesgo. No obstante tales afirmaciones deben ser tomadas con cautela porque se encuentran estrechamente relacionadas con el nivel de costes de transacción empleados e el trabajo y la determinación de la prima de riesgo asociada a las estrategias técnicas empleadas.

El problema de determinar una adecuada prima de riesgo es complejo, particularmente en un índice bursátil donde, como señalaremos posteriormente, no es posible una aplicación de la teoría CAPM y existe la posibilidad de un riesgo dinámico asociado a la propia estrategia.

En relación a los costes de transacción hay que señalar que en estos momentos los costes de transacción en los mercados bursátiles españoles rondan el 0.25 % gracias a las nuevas tecnologías de la información y la comunicación. Hace algunos años los costes eran varias veces superiores y descendieron muy considerablemente

durante los últimos años. Lo único que es capaz de señalar este trabajo es que durante los últimos años, donde los costes de transacción eran similares a los actuales, aparecen indicios de ineficiencia por los excesos de rendimientos obtenidos por nuestras reglas. No obstante, sólo si una regla técnica es capaz de obtener beneficios sobre un extenso periodo de tiempo la hipótesis de mercados eficientes puede ser rechazada. En este sentido será interesante averiguar si dentro de los próximos años este tipo de metodologías continúan produciendo beneficios generales.

### **5.3. Conclusiones generales**

Más allá de estas conclusiones específicas que se refieren a determinadas metodologías predictivas de obtención de reglas técnicas en la serie del IGBM, el trabajo realizado invita a la consideración de determinadas conclusiones más generales sobre la propia naturaleza del análisis técnico y sus características

#### **5.3.1. Sobre la necesidad de introducir modelos no lineales de predicción**

Los resultados desarrollados a lo largo de esta Tesis Doctoral apuntan hacia la importancia de utilizar modelos no lineales en la predicción de series financieras. Como queda patente a lo largo de todo el trabajo, existen importantes posibilidades de predicción si se va mas allá de los simples modelos lineales ARIMA tanto en su versión univariante como multivariante y se emplean modelos no lineales. No obstante, pese a la complejidad que entrañan algunos de los modelos no lineales no resulta trivial mejorar ostensiblemente los modelos ARIMA, resultando estos métodos lineales como una referencia obligatoria a la

que hay que batir.

### 5.3.2. Nueva visión de la Hipótesis de Mercados Eficientes

A comienzos de los años setenta era ampliamente aceptado por los investigadores en finanzas la validez de la hipótesis de mercados eficientes que se formulaba sobre la base del modelo de paseo aleatorio y que podía enunciarse diciendo que los precios de los activos financieros son una martingala. En aquella época no se encontraron reglas técnicas beneficiosas en las series examinadas de datos diarios o mensuales.

Fruto de esta actividad, se ha venido documentando sobradamente la habilidad predictiva de los mercados financieros, si bien no existe aún un consenso sobre el origen de dicha predicibilidad Bekaert (2001). Algunos autores señalan que la habilidad predictiva detectada en los mercados financieros podría estar reflejando primas de riesgo cambiantes en el tiempo. Otros han sugiriendo la existencia de fricciones (tales como costes de transacción, restricciones a la ventas a corto plazo, problemas de agencia, etc.) que podrían explicar la persistencia de anomalías en los precios de los activos financieros, incluida su predecibilidad. Unos cuantos sostienen que se debe a la pobre inferencia estadística. Por último, un nutrido grupo de autores proponentes de la denominada "finanzas del comportamiento" (behavioural finance) indica que dicha predicibilidad se debe al comportamiento irracional por parte de los participantes en el mercado (Véase Shleifer, 2001). Está por ver aún qué enfoque (o que combinación de enfoques) es el más relevante. En este sentido, Pennock *et al.* (2001) han abierto una vía interesante de investigación al explorar la eficiencia y la capacidad predictiva utilizando mercados artificiales.

Desde los años ochenta hasta nuestros días se ha producido toda una explosión de nueva actividad predictiva en los mercados financieros usando métodos más sofisticados de predicción, nuevos conjuntos de datos con series más largas y diferentes frecuencias y nuevas variables explicativas.

En este contexto y siguiendo a Granger (1992) otra nueva formulación de la Hipótesis de Mercados Eficientes se abre paso. Esta nueva hipótesis se expresaría diciendo que en un mercado eficiente no es posible obtener beneficios extraordinarios después de descontar los costes de transacción y una prima por el riesgo asumido por participar en él. En ese sentido los beneficios de un activo financiero ajustados por riesgo  $\pi$  se definirían como

$$\pi = R_t - r_t - \beta(\text{rendimientos del mercado}) - c, \quad (5.3.1)$$

donde  $R_t$  es el rendimiento del mercado,  $r_t$  es el rendimiento de un activo libre de riesgo,  $\beta$  es el nivel de riesgo del stock y  $c$  los costes de transacción. El nivel de riesgo del activo examinado suele medirse por medio del modelo CAPM (Capital Asset Pricing Model):

$$R_t - r_t = \beta(\text{exceso de rendimientos del mercado}) + \epsilon_t, \quad (5.3.2)$$

donde los rendimientos del mercado se refieren a una medida global del mercado en su conjunto como suele ser un índice, en nuestro caso particular el IGBM, representando  $\beta$  el riesgo no diversificable del activo.

No obstante, el problema de la obtención de la prima de riesgo de una determinada regla técnica puede ser algo más complejo que la obtención de  $\beta$  tal como formula la teoría CAPM. Por una parte, hay que tener en cuenta que cuando empleamos un índice como es el caso del IGBM sería mucho más complicado

encontrar el riesgo no diversificable del índice, teniendo que recurrir para ello a algún tipo de comparación con carteras diversificadas internacionalmente. Por otra parte, es preciso tener en cuenta el carácter estático del riesgo medido por  $\beta$ , mientras que una estrategia técnica de contratación puede incurrir en algún tipo de riesgo dinámico cuya estimación puede resultar compleja. En esta Tesis hemos considerado que el riesgo dinámico de la estrategia técnica está directamente asociado a los días que la estrategia permanece dentro del mercado.

### **5.3.3. La influencia de los costes de transacción en la búsqueda de reglas técnicas óptimas**

El disponer de un sistema predictivo capaz de obtener un alto porcentaje de aciertos en los cambios de signo de los rendimientos estadísticamente significativo no garantiza que la estrategia técnica de contratación que se derive de dicha estrategia sea más rentable que la de comprar y mantener.

Puede ocurrir que una buena estrategia predictiva desde el punto de vista estadístico sea muy activa e incurra en numerosos costes de transacción. En este sentido, los resultados de este trabajo coinciden con los de Satchell y Timmermann (1995) quienes consideran que, en la predicción no lineal financiera, las medidas clásicas de evaluación de las predicciones, como la Raíz Cuadrada del Error Cuadrático Medio (RCEMC), no son siempre adecuadas para evaluar la bondad de las predicciones. Satchell y Timmermann señalan que la significación estadística de patrones predecibles no implica que las predicciones sean económicamente significativas, aunque represente mejoras sustanciales en precisión predictiva en comparación con los modelos lineales estándar.

Así pues, los costes de transacción constituyen una variable clave en las



estrategias de contratación basadas en modelos no lineales de predicción, produciéndose un curioso dilema entre la frecuencia de los datos y los costes de transacción. Por un lado para la predicción de componentes no lineales en una serie temporal se necesitan datos de alta frecuencia con el fin de que el tamaño muestral sea lo suficientemente grande. Por otra parte, una estrategia técnica de contratación basada en datos de alta frecuencia puede incurrir fácilmente en el problema de que, si es muy activa, los costes de transacción pueden eliminar los beneficios que se obtenían bajo el supuesto de costes de transacción nulos. A lo largo de esta Tesis se ha comprobado, con frecuencia, que es fácil conseguir estrategias técnicas de contratación muy activas que superan estadísticamente al modelo de “comprar y mantener”. No obstante, tales estrategias dejan de ser rentables en el momento que se consideran los costes de transacción.

Por todo ello, de cara a la optimización de cualquier tipo de regla técnica de contratación es muy conveniente que la función de pérdida empleada sea el beneficio neto tras descontar los costes de transacción. Este tipo de funciones de pérdida darán lugar a estrategias poco activas que reducirán el impacto derivado los costes de transacción. Este problema se pone de manifiesto cuando se emplean en la construcción de reglas técnicas de contratación en paquetes informáticos estándar de redes neuronales o de programación genética donde la función de pérdida sólo considera medidas de bondad estadística y no financiera. Todo ello matiza diversas conclusiones obtenidas en recientes trabajos sobre la predicción de los rendimientos de los activos como es el caso de Min y Maddala (1999) donde se afirma que los modelos de redes neuronales pueden mejorar los modelos lineales de regresión en términos de habilidad predictiva, pero no en términos de su

capacidad para generar beneficios. Los resultados obtenidos en esta Tesis apuntan a que las afirmaciones de Min y Maddala (1999) pueden estar relacionadas con que las estrategias técnicas basadas en las redes neuronales, sin ningún tipo de filtro previo, pueden ser más activas que las basadas en los modelos lineales, hecho que puede repercutir notablemente en su rentabilidad.

#### **5.3.4. Sobre el fisgoneo de datos y la sobreparametrización**

Para que un método de predicción pueda ser considerado como aceptable no es suficiente que produzca buenas predicciones dentro de la muestra que empleamos como conjunto de entrenamiento del modelo. Antes al contrario, es necesario demostrar el poder predictivo del modelo fuera de la muestra, como única forma de evaluación relevante del modelo. La sobreparametrización de un modelo y el fisgoneo de datos dentro del conjunto de entrenamiento puede conducir a ajustes extraordinariamente buenos, cuya capacidad predictiva desaparece cuando se consideran nuevos datos. Este fenómeno puede ser especialmente significativo cuando la naturaleza del modelo es no lineal. Por esta razón, la optimización de reglas técnicas debe estar presidida por el principio de validación cruzada, de modo que los datos de un subconjunto de entrenamiento sean los encargados de evaluar los parámetros de la regla técnica óptima, mientras que será en otro subconjunto diferente denominado de validación o predicción, donde se considere la capacidad de generar beneficios económicos de dicha regla.

### 5.3.5. Evaluación del riesgo de las predicciones financieras

Como ya se ha señalado, los meros contrastes estadísticos sobre predecibilidad no son suficientes para evaluar el poder predictivo de una regla técnica. Como señalan Satchell y Timmermann Satchell y Timmermann (1995), la única forma adecuada de evaluar una regla técnica es por medio de los beneficios económicos que se obtienen mediante su aplicación.

No obstante, en cualquier toma de posición financiera los rendimientos no pueden ser evaluados al margen del nivel de riesgo en que se incurre. Por tanto, las estrategias técnicas de contratación necesitan incorporar tanto los costes de transacción como los niveles de riesgo de dichas estrategias. El riesgo soportado por las diferentes estrategias es una variable económica relevante que debe estar contrapesado con su capacidad de realizar beneficios. Existen diversas formas de evaluar el riesgo de una estrategia técnica empleada tradicionalmente tales como el ratio de Sharpe o  $\beta$ , cuando nos referimos a un determinado título y queremos referirnos a su riesgo específico.

En este trabajo hemos considerado como medida más idónea de riesgo la propuesta por Allen y Karjalainen (1999) que se basa en considerar la proporción del número de días en los que la regla está fuera del mercado, parámetro que permite definir como variable económica clave el exceso de rendimientos de la estrategia “comprar y mantener” ajustada por riesgo. De manera adicional, hemos considerado también la corrección de los rendimientos ajustados por riesgo proporcionada por el ratio de Sharpe, convenientemente modificado para incorporar esa posibilidad de mantener activos financieros alternativos durante el periodo de negociación considerado. La evaluación económica de la rentabilidad

de tales estrategias resulta incompleta si no se considera la variable riesgo y se compara con el riesgo asumido por la estrategia de comprar y mantener.

### 5.3.6. Sobre el carácter de las reglas técnicas

Existe un tema delicado planteado por diversos autores tales como Bowden (1990), o Granger (1992), que es el de la eventualidad de las reglas técnicas. Como señala Granger (1992), existen muchas posibles reglas técnicas que para las que es realista suponer que los inversores no hayan intentado, especialmente aquellas que necesiten para su detección y descubrimiento computación costosa y técnicas estadísticas sofisticadas.

A medida que se va ampliando el conocimiento y utilización de determinadas reglas técnicas, cabría esperar que el comportamiento de los especuladores elimine y suprima su capacidad de generar beneficios, a menos que existan otras reglas técnicas que los especuladores consideren superiores y sean empleadas, inicialmente, con más profusión.

En un sentido similar Bowden (1990) considera que la efectividad de las reglas chartistas tiene un carácter efímero y eventual. Aunque existe una multitud de reglas del tipo de expectativas extrapolativas que pueden ser empleadas con éxito de forma muy localizada en el tiempo, en la medida en que determinada regla tenga éxito y se generaliza su uso por parte de los agentes, los patrones de comportamiento (origen de dicha regla) evolucionen hasta convertirse en nuevos patrones que sólo serán predecibles por nuevas reglas.

Por estos motivos parece razonable el emplear sistemas muy flexibles de obtención de reglas técnicas como son los derivados de la regresión no paramétrica, las redes neuronales y la programación genética.

No obstante, pese a que la concepción evolutiva de las reglas técnicas apuntada por Bowden y Granger parece razonable, subsiste el problema de cómo explicar los excesos de rendimientos obtenidos mediante las reglas técnicas clásicas como las de cruce de medias móviles o ruptura de rango. Como se ha señalado en el capítulo 2 de esta Tesis, tales reglas optimizadas mediante algoritmos genéticos, pese a que son conocidas y empleadas por los analistas de mercado desde hace un siglo, son capaces de generar en el IGBM exceso de rendimientos sobre la estrategia de “comprar y mantener”.

Otros trabajos como Brock *et al.* (1992) o Fernández Rodríguez *et al.* (2001) apuntan en el mismo sentido, el último de ellos dentro del propio IGBM. La literatura financiera no da grandes respuestas a este problema de capacidad de generar beneficios por parte de las reglas técnicas, considerando en general que su carácter predictivo descansa en el se trata del reflejo de “profecías autocomplacientes”. Lo cierto es que en este momento no existe todavía ningún modelo capaz de ayudar a entender la capacidad de generar beneficios de las reglas técnicas y su evolución a lo largo del tiempo.

Quizás desde el prometedor campo de la psicología económica pueda llegar, en un futuro próximo, nuevos enfoques que permitan avanzar en este sentido (Rabin, 2002; Tirole, 2002; Chang y Majnoni, 2002).

## 5.4. Desarrollos futuros

Existen varias direcciones posibles a la hora de realizar futuras extensiones del trabajo de investigación realizado en esta Tesis Doctoral. En particular, podemos señalar las siguientes líneas de actuación:

- 
- En primer lugar, una ampliación natural de la Tesis consiste en explorar las consecuencias que sobre la rentabilidad puede tener el diseño de nuevas técnicas mediante programación genética cuando la función objetivo es el rendimiento neto, y no el porcentaje de aciertos. Somos conscientes de que ello supone un esfuerzo notable de programación, pero las potenciales ganancias en términos de rentabilidad podrían justificar este esfuerzo.
  - En segundo lugar, y en línea con lo anterior, cabría la posibilidad de crear nuevas reglas de negociación más elaboradas empleando sistemas híbridos de algoritmos genéticos y programación genética en la línea de los propuestos por Rayan y Raywaer-Smith (1998).
  - En tercer lugar, dado que reglas técnicas analizadas se comportan relativamente mejor durante episodios alcistas mientras que otras tienen una cierta ventaja en episodios bajista, y dado que tales episodios se suceden continuamente en los mercados, cabe también la posibilidad de combinar las recomendaciones realizadas por cada una de las distintas reglas de forma que se obtenga la mayor rentabilidad promedio, independientemente de la fase en la que se encuentre el mercado. En este sentido, se podría explorar el efecto de implementar, en este área de investigación, las ideas desarrolladas por Li y Tkacz (2001) y Elliot y Timmermann (2002) sobre combinaciones óptimas de predictores.
  - En cuarto lugar, se podría aplicar el enfoque no paramétrico de Brandt (1999) para determinar la selección de la cartera óptima de un inversor adverso al riesgo en el mercado cambiario, suponiendo que éste utiliza las

señales generadas por las reglas técnicas examinadas en esta tesis como instrumentos de información para realizar sus inversiones. Este enfoque nos permitiría evaluar el valor económico de dichas reglas técnicas a partir de las preferencias frente al riesgo del inversor.

- Por último, y desde el punto de vista empírico, sería interesante extender el análisis a otros mercados financieros (bursátiles, cambiarios, etcétera), con el fin de evaluar el comportamiento de las reglas técnicas examinadas en otros contextos institucionales.

Dados los resultados relativamente favorables obtenidos en esta Tesis Doctoral, cabe albergar un cierto optimismo sobre los futuros rendimientos derivados de la realización de estas extensiones.

## Referencias bibliográficas

- Abarbanel, H. D. I. (1996), *Analysis of Observed Chaotic Data*, New York: Springer-Verlag.
- Akaike, H. (1974), “A new look at the statistical model identification,” en *IEEE Transaction on Automatic Control*, tomo AC-19, 716–723.
- Alexander, S. (1961), “Price movements in speculative markets: trends or random walks?” *Industrial Management Review*, **2**, 7–26.
- Allen, F. y Karjalainen, R. (1999), “Using genetic algorithms to find technical trading rules,” *Journal Finance Economics*, **51**, 245–271.
- Altenberg, L. (1995), *Genome Growth and the Evolution of the Genotype-Phenotype Map*, Berlin: Springer-Verlag.
- Bajo Rubio, O., Fernández Rodríguez, F. y Sosvilla Rivero, S. (1992a), “Chaotic behaviour in exchange-rate series: First results for the peseta-u.s. dollar case,” *Economic Letters*, **39**, 207–211.
- Bajo Rubio, O., Fernández Rodríguez, F. y Sosvilla Rivero, S. (1992b), “Volatilidad y predecibilidad en las series de tipo de cambio peseta-dólar: Un enfoque basado en el caos determinista,” *Revista Española de Economía*, **9**, 91–109.
- Banzhaf, W., Francone, F. y Nordin, P. (1996), “The effect of extensive use of the mutation operator on generalization in genetic programming using sparse



- data,” en Voigt, H., Ebeling, W., Recenberg, I. y Schwefwl, H. (Eds.), *Parallel Problem Solving From Nature IV. Proceedings of the International Conference on Evolutionary Computations*, tomo 1141 de *Lectures Notes in Computer Science*, Berlin: Springer-Verlag, 300–310.
- Banzhaf, W., Nordin, P., Keller, R. y Francone, F. (1998), *Genetic Programming. An Introduction on the Automatic Evolution of Computer Programs and Its Applications*, California: Morgan Kaufmann Publisher.
- Bauer, R. J. (1994), *Genetic Algorithms and Investment Strategies*, New York: Wiley Finance Editors, John Wiley and Sons.
- Beja, A. y Goldman, M. B. (1980), “On the dynamic behavior of prices in disequilibrium,” *Journal of Finance*, **34**, 235–247.
- Bekaert, G. (2001), “On the predictability of asset returns,” *Journal of Empirical Finance*, **8**, 451–457.
- Benachenhou, D. (1994), “Smart trading with FRET,” *Genetic and Fuzzy Systems for Chaotic Financial Markets*.
- Bessembinder, H. y Chan, K. (1995), “The profitability of technical trading rules in the asian stock markets,” *Pacific-Basic Finance Journal*, **3**, 257–284.
- Bowden, R. J. (1990), “Predictive disequilibria and the short run dynamics of asset prices,” *Australian Journal of Management*, **15**, 65–87.
- Box, G. E. P. y Jenkins, G. M. (1976), *Time Series Análisis, Forecasting and Control*, San Francisco: Holden Day.
- Brandt, M. (1999), “Estimating portfolio and consumption choice: A conditional euler equations approach,” *Journal of Finance*, **54**, 1609–1645.

- Brock, W. y Hommes, C. (1997), “A rational route to randomness,” *Econometrica*, **65**, 1059–1096.
- Brock, W., Lakonishhock, J. y LeBaron, B. (1992), “Simple technical trading rules and the stochastic of stock returns,” *Journal of Finance*, **47**, 1731–1764.
- Brock, W. A. y Hommes, C. H. (1998), “Heterogeneous beliefs and routes to chaos in a simple asset pricing model,” *Journal of Economic Dynamics and Control*, **22**, 1235–1274.
- Brock, W. A., Hsieh, D. y LeBaron, B. (1991), *Nonlinear Dynamics, Chaos and Instability*, Cambridge, MA: MIT Press.
- Brock, W. A. y LeBaron, B. (1996), “A dynamic structural model for stock return volatility and trading volume,” *Review of Economics and Statistics*, **78**, 94–110.
- Broomhead, D. S. y King, G. P. (1986), “Extracting qualitative dynamics from experimental data,” *Physica D*, **20**, 217–231.
- Brown, S., Goetzmann, W. y Kumar, A. (1998), “The down theory: William peter hamilton’s track record reconsidered,” Working Paper, Stern school of business. New York University.
- Campbell, J. Y., Lo, A. W. y MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton, New Jersey: Princeton University Press.
- Canova, F. (1993), “Modelling and forecasting exchange rates with bayesian time-varying coefficient model,” *Journal of Economics Dynamics and Control*, **17**, 233–261.
- Cao, L. (1997), “Practical method for determining the minimum embedding dimension for a scalar time series,” *Physica D*, **110**, 43–50.

- Cao, L. y Soofi, A. S. (1999), “Nonlinear deterministic forecasting of daily dollar exchange rates,” *International Journal of Forecasting*, **15**, 421–430.
- Casdagli, M. (1989), “Nonlinear prediction of chaotic time series,” *Physica D*, **35**, 335–356.
- Casdagli, M. (1992a), *Applied Chaos*, capítulo Nonlinear Modeling of Chaotic Time Series: Theory and Applications, New York: John Wiley and Sons.
- Casdagli, M. (1992b), “Chaos and determinitics versus stochastic nonlinear modeling,” *Journal of the Royal Statistical Society*, **54**, 303–328.
- Casdagli, M. y Weigend, A. S. (1994), *Exploring the Continuum Between Deterministic and Stochastic Modelling*, Massachussets: Addison Wesley.
- Cecchetti, S. G., Lam, P. S. y Mark, N. C. (1990), “Mean reversion in equilibrium asset price,” *American Economic Review*, **80**, 398–418.
- Chang, R. y Majnonib, G. (2002), “Fundamentals, beliefs, and financial contagion,” *European Economic Review*, **46**, 801–808.
- Chiarella, C. (1992), “The dynamics of speculative behaviour,” *Annal of Operations Research*, **37**, 101–124.
- Cleveland, W. (1979), “Robust locally weighted regression and smoothing scatterplots,” *Journal of the American Statistical Association*, **74**, 829–836.
- Cleveland, W. S. y Devlin, S. J. (1988), “Locally weighted regression: An approach to regression analysis by local fitting,” *Journal of the American Statistical Association*, **83**, 596–610.
- Corrado, C. y Lee, S. (1992), “Filter rule test of the economic significance of serial dependence in daily stock returns,” *Journal of Financial Research*, **15**, 369–387.

- Curtis, H. y Barnes, N. S. (1989), *Biology*, Worth Publishers.
- Darwin, C. (1859), *On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*, London: Murray.
- Daubechies, I. (1990), "The wavelet transform, time frequency localization and signal analysis," *IEEE Trans. Inform. Theory*, **36**, 961–1004.
- Dawes, R., Fildes, R., Lawrence, M. y Ord, K. (1994), "The past and the future of forecasting research," *International Journal of Forecasting*, **10**, 151–159.
- Day, R. y Huang, W. (1990), "Bulls, bears and market sheep," *Journal of Economic Behavior and Organization*, **14**, 299–329.
- Deboeck, G. (1994), *Trading on the Edge. Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets*, New York: John Wiley & Sons.
- D'haeseleer, P. (1994), "Context preserving crossover in genetic programming," en *Proceedings of the 1994 IEEE World Congress on Computational Intelligence*, tomo 1, New York: FL. IEEE Press, 256–261.
- Díaz, A., Glover, F., Ghaziri, H., González, J., Laguna, M., Moscato, P. y Tseng, F. (1996), *Optimización Heurística Y Redes Neuronales En Dirección de Operaciones e Ingeniería*, Madrid: Editorial Paraninfo.
- Diebold, F. y Nason, J. A. (1990), "Nonparametric exchange rate prediction?" *Journal of International Economics*, **28**, 315–332.
- Dorsey, R. D. y Mayer, J. M. (1995), "Genetic algorithms for estimation problems with multiple optima, nondifferentiability, and other irregular features," *Journal of Business and Economics Statistics*, **13**, 53–66.
- Downsland, K. y Downsland, W. (1992), "Packing problems," *European Journal of Operational Research*, **56**, 2–14.

- Efrom, B. (1983), "Estimating the error rate of a prediction rule. improvement on cross-validation," *Journal of American Statistical Association*, **78**, 316–331.
- Elliot, G. y Timmermann, A. (2002), "Optimal forecast combinations under general loss functions and forecast error distributions," Economics Discussion Paper 2002-08, Unniversity of California.
- Engle, R., Granger, C. y Hallman, J. (1989), "Merging short- and long-run forecasts: An application of seasonal cointegration to monthly electricity sales forecasts," *Journal of Econometrics*, **40**, 45–62.
- Engle, R. F. (1982), "Autoregressive conditional heterocedasticity with estimates of the variance of U. K. inflation," *Econometrica*, **50**, 987–1008.
- Eubank, E. (1999), *Nonparametric Regression and Spline Smoothing*, New York: Marcel Dekker, Inc.
- Fama, E. (1965), "The behavior of stock-market prices," *Journal of Business*, **38**, 34–105.
- Fama, E. y Blume, M. (1966), "Filter rules and stock market trading," *Journal of Business*, **39**, 226–241.
- Fama, E. F. (1970), "Efficient capital market: A review of theory and empirical work," *Journal of Finance*, **25**, 383–417.
- Fama, E. F. (1991), "Efficient capital markets: II," *Journal of Finance*, **46**, 575–1617.
- Fama, E. F. y French, K. R. (1988), "Permanent and temporary components of stock prices," *Journal of Political Economy*, **96**, 246–273.
- Farmer, J. D. y Sidorowich, J. (1987), "Predicting chaotic time series," *Physical Review Letters*, **59**, 845–848.

- Fernández Rodríguez, F., García Artiles, M. y Martín, J. (2002), “A model of speculative behaviour with a strange attractor,” *Applied Mathematical Finance*, **9**, 1–19.
- Fernández Rodríguez, F., González Martel, C. y Sosvilla Rivero, S. (2001), “Optimization of technical rules by genetic algorithms: Evidence from madrid stock market,” Documento de trabajo 2001-14, Fundación de Estudios de Economía Aplicada.
- Fernández Rodríguez, F. y Sosvilla Rivero, S. (1998), “Testing nonlinear forecastability in time serie: Theory and evidence from the EMS,” *Economic Letters*, **59**, 49–63.
- Fernández Rodríguez, F., Sosvilla Rivero, S. y Andrada Félix, J. (1997), “Combining information in exchange rate forecasting: Evidence from the EMS,” *Applied Economics Letters*, **4**, 441–444.
- Fernández Rodríguez, F., Sosvilla Rivero, S. y Andrada Félix, J. (1999a), “Análisis técnico en la bolsa de madrid,” *Moneda y Crédito*, **213**, 11–37.
- Fernández Rodríguez, F., Sosvilla Rivero, S. y Andrada Félix, J. (1999b), “Exchange-rate forecasts with simultaneous nearest-neighbour methods: Evidence from the EMS,” *International Journal of Forecasting*, **15**, 383–392.
- Fernández Rodríguez, F., Sosvilla Rivero, S. y García Artiles, D. (1997), “Using nearest neighbour predictors to forecast the spanish stock market,” *Investigaciones Económicas*, **21**, 75–91.
- Fernández Rodríguez, F., Sosvilla Rivero, S. y García Artiles, M. D. (1999c), “Dancing with bulls and bears: Nearest-neighbour forecasts for the nikkei index,” *Japan and the World Economy*, **11**, 395– 413.

- Fernández Rodríguez, F., Sosvilla Rivero, S. y González Martel, C. (2000), “On the profitability of technical trading rules based on artificial neural networks: Evidence from the madrid stock market,” *Economic Letters*, **69**, 89–94.
- Fogel, I., Owens, A. y Walsh, M. (1966), *Artificial intelligence through simulated evolution*, New York: Wiley.
- Frances, P. H. y Dijk, D. (2000), *Non-Linear Time Series Models in Empirical Finance*, New York: Cambridge University Press.
- Francone, F. D. (2000), *Discipulus. Owner’s Manual*, register Machione Learning Technologies, Inc.
- Fraser, A. M. y Swinney, H. L. (1986), “Independent coordinates for strange attractors from mutual information,” *Physical Review*, **33**, 1134–1140.
- French, K. y Roll, R. (1986), “Stock return variances: The arrival of information and the reaction of trades,” *Journal of Financial Economics*, **17**, 5–26.
- Friedberg, R., Dunham, B. y Y North, J. (1959), “A learning machine, part II,” *IBM Journal of Research and Development*, **3**, 282–287.
- Futuyma, D. J. (1986), *Evolutionary Biology*, Sinauer Ass.
- Gençay, R. (1996), “Non-linear prediction of security returns with moving average rules,” *Journal of Forecasting*, **15**, 165–194.
- Gençay, R. (1998), “Optimization of technical trading strategies and the profitability in security markets,” *Economic Letters*, **59**, 249–254.
- Goldberg, D. (1989), *Genetic Algorithms in Search, Optimization, and Machine Learning*, Mass: Addison Wesley.

- Granger, C. W. J. (1992), "Forecasting stock markets prices: Lessons for forecasters," *International Journal of Forecasting*, **8**, 3–13.
- Grassberger, P. y Procaccia, I. (1983), "Characterization of strange attractors," *Phys. Rev. Lett.*, **50**, 346–349.
- Grossman, S. y Stiglitz, J. E. (1980), "On the impossibility of informationally efficient markets," *American Economic Review*, **70**, 393–408.
- Guillaume, D. M., Dacorogna, M. M., Dave, R. D., A., M. U., Olsen, R. B. y Pictet, O. V. (1995), "From the birds eye to the microscope: A survey of a new stylized facts of the intra-daily foreign exchange markets," Internal document DMG. 1994-04-06, Olsen and Associates.
- Hansen, B. E. (1999), "Discussion of "data mining reconsidered"," *Econometrics Journal*, **2**, 192–201.
- Härdle, W. (1992), *Applied Nonparametric Regression*, Cambridge: Cambridge University Press.
- Hastie, T., Tibshirani, R. y Friedman, J. (2001), *The Elements of Statistical Learning*, New York: Springer Series in Statistics. Springer.
- Haupt, R. L. y Haupt, S. E. (1998), *Practical Genetic Algorithms*, New York: John Wiley and Sons.
- Hebb, D. (1949), *The Organization of Behavior*, New York: John Wiley and Sons.
- Hendry, D. y Krolzig, H. (1999), "Improving on "data mining reconsidered" by k. d. hoover and s. j. perez," *Econometrics Journal*, **2**, 167–191.
- Hinchliffe, M., Hiden, H., McKay, B., Willis, M., Tham, M. y Barton, G. (1996), *Modelling Chemical Process Using a Multi-Gene Genetic Programming Algorithm*, Standford: Standford Bookstore, 56–65.



- Holland, J. (1962), "Outline for a logical theory of adaptive systems," *Journal of the Association for Computing Machinery*, **3**, 297–314.
- Holland, J. (1975), *Adaptation in natural and artificial systems*, Michigan: University of Michigan Press, Ann Arbor.
- Holland, J. (1980), "Adaptive algorithms for discovering and using general patterns in growing knowledge-bases," *International Journal of Policy Analysis an Information System*, **4**, 217–240.
- Holland, J. H. (1976), "Adaptation," *Progress in Theoretical Biology*, **4**, 263–293.
- Holly, S. y Hughes-Hallet, A. (1989), *Optimal Control, Expectations and Uncertainty*, Cambridge: Cambridge University Press.
- Hoover, K. y Pérez, S. (1999), "Data mining reconsidered: Encompassing and the general-to-specific approach to specification search," *Econometrics Journal*, **2**, 1–25.
- Hsieh, D. (1991), "Chaos and nonlinear dynamics: Applications to financial markets," *Journal of Finance*, **47**, 1145–1189.
- Hudson, R., Dempsey, M. y Keasey, K. (1996), "A note on the weak form of efficiency of capital markets: The application of simple technical trading rules to the u.k. stock markets-1935-1994," *Journal of Banking and Finance*, **20**, 1121–1132.
- Hutchinson, J., Lo, A. y Poggio, T. (1994), "A nonparametric approach to the pricing and hedging of derivative securities via learning networks," *Journal of Finance*, **49**, 851–889.
- Iba, H. y de Garis, H. (1996), *Extending Genetic Programming with Recombinative Guidance*, MA.: MIT Press.

- Jaditz, T. y Sayers, C. L. (1995), *Chaos and Nonlinear Dynamics in the Financial Markets*, capítulo Nonlinearity in the Interest Rate Risk Premium, New York: Irwin, 335–361.
- Jaditz, T. y Sayers, C. L. (1998), “Out of sample forecast performance as a test for nonlinearity in time series,” *Journal of Business & Economic Statistics*, **16**, 110–117.
- Keim, D. B. (1986), “The CAPM and equity return regularities,” *Financial Analysts Journal*, **42**, 19–34.
- Kennel, M., Brown, R. y Abarbanel, H. (1992), “Determining embedding dimension for phase-space reconstruction using a geometrical construction,” *Physical Review*, **45**, 3403–3411.
- Kirkpatrick, S., Gelatt, C. y Vecchi, M. (1983), “Optimization by simulated annealing,” *Science*, **220**, 671–680.
- Koza, J. (1992), *Genetic Programming: On the Programming of Computers By the Means of Natural Selection*, Cambridge, Mass: MIT Press.
- Lang, K. J. (1995), “Hill climbing beats genetic search on a boolean circuit synthesis of Koza’s,” en *Proceedings of International Conference on Machine Learning*, San Francisco: Morgan Kaufmann.
- Levich, R. y Thomas, L. (1993), “The significance of technical trading rule profits in the foreign exchange market: A bootstrap approach,” *Journal of International Money and Finance*, **12**, 451–474.
- Li, F. y Tkacz, G. (2001), “Evaluating linear and non-linear time-varying forecast-combination methods,” Documento de trabajo 2001-12, Bank of Canada.

- Linsay, P. (1991), “An efficient method of forecasting chaotic time series using linear interpolation,” *Physics Letters A*, **153**, 353–356.
- Lisi, F. y Medio, A. (1997), “Is a random walk the best exchange rate predictor?” *International Journal of Forecasting*, **13**, 255–267.
- Lo, A. y MacKinley (1990), “Data snooping biases in test of financial asset pricing models,” *The Review of Financial Studies*, **3**, 431–467.
- Lo, A. W. y MacKinlay, A. C. (1988), “Stock market price do not follow random walks: Evidenc from a simple specification test,” *The Review of Financial Studies*, **1**, 41–66.
- Loftsgaarden, D. O. y Quesenberry, G. P. (1965), “A nonparametric estimate of a multivariate density function,” *Annals of Mathematical Statistics*, **36**, 1049–1051.
- Lorenz, H. W. (1963), “Deterministic nonperiod flows,” *Journal of atmosphere science*, **20**, 130–141.
- Lovell, M. C. (1983), “Data mining,” *Review of Economics and Statistics*, **65**, 1–12.
- Lux, T. (1995), “Herd behaviour, bubbles and crashes,” *Economic Journal*, **105**, 881–896.
- MacDonald, R. y Taylor, M. P. (1994), “The monetary model of the exchange rate: Lon-run relationship, short-run dynamica and how to beat a random walk,” *Journal of International Money and Finance*, **3**, 276–290.
- Malkiel, B. (1992), *Un Paseo Aleatorio Por Wall Street*, Madrid: Alianza Editorial.

- Manynard-Smith, J. (1994), *Evolutionary Genetics*, Oxford: Oxford University Press.
- McCulloch, W. S. y Pitts, W. (1943), "A logical calculus of the ideas immanent in nervous activity," *Bouulletin of Mathematics and Biophysics*, **5**, 115–133.
- Meese, R. y Rose, A. (1990), "Nonlinear, nonparametric exchange rate estimation," *American Economic Review*, **80**, 192–196.
- Mendel, G. (1865), *Versuche über Planzanhhybriden*, tomo 4, Brünn.
- Michalewicz, Z. (1992), *Genetic Algorithms + Data Structures = Evolution Programs*, Berlin: Springer-Verlag.
- Mills, T. (1993), *The Econometric Modelling of Financial Time Series*, Cambridge: Cambridge University Press.
- Mills, T. (1996), "Technical analysis and the london stock exchange: Testing trading rules using the FT30," Economic Research Paper 96/1, Department of Economics, Loughborough of Technology.
- Min, Q. y Maddala, G. S. (1999), "Economic factors and the stock market: A new perspective," *Journal of Forecasting*, **18**, 151–166.
- Minsky, M. y Papert, S. (1969), *Perceptrons*, Mass: MIT Press.
- Mitchell, T. (1996), *Machine Learning*, New York: McGraw-Hill.
- Mizrach, B. (1992), "Forecast comparations in  $l^2$ ," Documento de Trabajo 1995-24, Rutgers University.
- Montana, D. J. y Davis, L. (1989), "Training feedfoward neural networks using genetic algorithms," en *Proceeding of Th Eleventh International Joint Conference on Artificial Intelligence*, Detroit, 762–767.

- Mulhern, F. J. y Caprara, R. J. (1994), "A nearest neighbor model for forecasting market response," *International Journal of Forecasting*, **10**, 191–207.
- Nadaraya, E. A. (1964), "On estimating regression," *Theory Probability and its Applications*, **10**, 186–190.
- Neely, C., Weler, P. y Dittmar, R. (1997), "Is technical analysis in the foreign exchange market profitable? a genetic programming approach," *Journal of Financial Quantitative Analysis*, **32**, 405–426.
- Nordin, P., Francone, F. y Banzhaf, W. (1996), *Explicitly Defined Introns and Destructive Crossover in Genetic Programming*, capítulo 6, MA.: MIT Press, 111–134.
- Olmeda, I. y Pérez, J. (1995), "Non-linear dynamics and chaos in the spanish stock market," *Investigaciones Económicas*, **19**, 217–248.
- Osborne, M. M. (1959), "Brownian motion in the stock market," *Operations Research*, **7**, 145–175.
- Osler, C. L. y Chang, K. P. H. (1995), "Head and shoulders: Not just a flaky pattern," Staff Report 4, Federal Reserve Bank of New York.
- Parker, D. (1982), "Learning logic," Invention Report S81-64, Office of Technology Licensing. Stanford University.
- Pennock, D. M., Lawrence, S., Giles, G. L. y Nielsen, F. A. (2001), "The power of play: Efficiency and forecast accuracy in web market games," *Science*, **291**, 987–988.
- Pesaran, M. y Timmermann, A. (1992), "A simple non parametric test of predictive performance," *Journal of Business and Economic Statistics*, **10**, 461–465.

- Peters, E. E. (1994), *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*, New York: John Wiley & Sons.
- Plaut, D., Nowlan, S. y Hinton, G. (1986), “Experiments on learning by back propagation,” Documento de Trabajo CMU-CS-86-126, Department of Computer Science, Carnegie Mellon University.
- Plummer, T. (1989), *Forecasting Financial Markets: The Truth Behind Technical Analysis*, London: Kogan Page.
- Poterba, J. M. y Summers, L. H. (1987), “Mean reversion in stock prices: Evidence and implications,” Documento de Trabajo 2343, NBER.
- Pring, M. (1991), *Technical Analysis Explained: The Successful Investor’s Guide to Spotting Investment Trends and Turning Points*, New York: MacGraw-Hill.
- Rabin, M. (2002), “A perspective on psychology and economics,” *European Economic Review*, **46**, 657–685.
- Rayan, M. D. y Raywaer-Smith, V. J. (1998), “The evolution of decision trees,” en *Genetic Programming 1998: Proceedings of the Third Annual Conference*, San Francisco: Morgan Kaufmann Publishers, 350–358.
- Rechenberg, I. (1973), *Evolutionsstrategie: Optimierung Technischer Systeme nach Prinzipien der Biologischen Evolution*, Stuttgart: Frommann-Holzboog Verlag.
- Refenes, P. (Ed.) (1995), *Nural Networks in the Capital Markets*, New York: John Wiley.
- Roberts, H. V. (1959), “Stock-market “patterns” and financial analysis: Methodological suggestions,” *Journal of Finance*, **14**, 1–10.

- Rosenblatt, F. (1958), "The perceptron: A probabilistic model for informatyion storage and organization in the brain," *Psychological Review*, **65**, 386–408.
- Rumelhart, D., Hinton, G. y Williams, R. (1986), "Learning representations by back-propagation errors," *Nature*, **323**, 533–536.
- Ryden, T., Teräsvirta, T. y Asbrink, S. (1998), "Stylized facts of daily return series and the hidden markov model," *Journal of Applied Econometrics*, **13**, 217–244.
- Samuel, A. (1963), *Some Studies in Machine Learning Using the Game of Checkers*, New York: McGraw-Hill.
- Samuelson, P. A. (1965), "Proof that prperly anticipated prices fluctuate randomly," *Industrial Managment Review*, **6**, 41–49.
- Saratis, N. (2001), "Nonlinearities, cyclical behaviour and predictability in stock markets: International evidence," *International Journal of Forecasting*, **17**, 459–482.
- Satchell, S. y Timmermann, A. (1995), "An assesment of economic value of non-linear foreign exchange rate forecast," *Journal of Forecasting*, **14**, 477–479.
- Sawa, T. (1978), "Information criteria for discriminating among alternative regression models," *Econometrica*, **46**, 1273–1292.
- Schwarz, G. (1978), "Estimating the dimension of a model," *The Annals of Statistics*, **6**, 461–464.
- Schwefel, H. (1981), *Numerical optimization of computer models*, New York: Wiley.
- Shiller, R. J. (1989), *Market Volatility*, Cambridge: M.I.T. Press.

- Shleifer, A. (2001), *Inefficient Markets: An Introduction to Behavioural Finance*, Oxford: Oxford University Press.
- Soofi, A. S. y Cao, L. (1999), "Nonlinear deterministic forecasting of daily peseta-dollar exchange rate," *Economic Letters*, **62**, 175–180.
- Stone, J. (1997), "Consistent nonparametric regression," *The Annals of Statistics*, **5**, 595–645.
- Suárez, A. S. S. (1996), *Decisiones Óptimas de Inversión Y Financiación de la Empresa*, Madrid: Pirámide.
- Sugihara, G. y May, R. (1990), "Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series," *Nature*, **344**, 734–741.
- Sullivan, R., Timmermann, A. y White, H. (1997), "Data-snooping, technical trading rule performance, and the bootstrap," Documento de Trabajo 97-31, University of California at San Diego, Department of Economics.
- Tackett, W. A. (1994), *Recombination, Selection and the Genetic Construcción of Computer Programs*, Tesis Doctoral, University of Southern California, Department of Electrical Engineering Systems.
- Takens, F. (1981), *Lectures Notes Math*, capítulo Detecting Strange Attractors in Turbulence, 898, New York: Springer-Verlag.
- Taylor, M. y Allen, H. (1992), "The use of technical analysis in the foreign exchange market," *Journal of Money and Finance*, **11**, 304–314.
- Taylor, S. J. (1994), "Trading futures using a channel rule: A study of the predictive power of technical analysis with currency examples," *The Journal of Future Markets*, **14**, 215–235.



- Teller, A. y Veloso, M. (1995), "Learning tree structured algorithms for orchestration into an object recognition system," Documento de Trabajo CMU-CS-95-101, Department of Computer Science, Carnegie Mellon University, Pittsburgh.
- Teller, A. y Veloso, M. (1996), *A New Learning Architecture for Object Recognition*, Oxford: Oxford University Press.
- Tirole, J. (2002), "Rational irrationality: Some economics of self-management," *European Economic Review*, **46**, 633–655.
- Trippi, R. R. y Lee, J. K. (1992), *Artificial Intelligence in Finance & Investing: State-Of-The-Art Technologies for Securities Selection and Portfolio Management*, New York: McGraw-Hill Trade.
- Van Eyden, R. J. (1996), *The Applications of Neural Networks in the Forecasting of Share Prices*, New York: Finance & Tecnology Publishing.
- Venogopal, V. y Narendan, T. (1992), "A genetic algorithm approach to the machine component grouping problem with multiple objectives," *Computer & Industrial Engineering*, **21**, 469–480.
- Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. y Weiner, A. M. (1987), *Molecular Biology of the Gene*, California: Benjamin-Cummings.
- Wayland, R., Pickett, D., Bromley, D. y Passamante, A. (1994), "Measuring spatial spreading in recurrent time series," *Physica D*, **79**, 320–334.
- Werbos, P. (1974), *Beyond Regression: New Tools for Prediction and Analisis in the Behavioral Sciences*, Tesis Doctoral, Harvard University.
- West, K. D. (1988), "Bubbles, fads and stock price volatility: A partial evaluation," *Journal of Finance*, **43**, 639–660.

Whitley, D. (1989), “The GENITOR algorithm and selection pressure: Why rank-based allocation of reproductive trial is best,” en Schaffer, D. J. (Ed.), *Proceedings of Third International Conference on Genetic Algorithms*, California: Morgan Kaufman, 116–121.

Widrow, B. y Hoff, M. (1960), “Adaptative switching circuits,” en *IREWESCON Convention Record*, tomo 4, 96–104.

Zannoni, E. y Reynolds, R. (1996), *Extracting Design Knowledge from Genetic Programs Using Cultural Algorithms*, MA.: MIT Press.