



**Universidad Nacional Autónoma de México**  
**Programa de Posgrado en Ciencias de la Administración**

**Anomalías de mercado para la estimación del rendimiento de emisoras de mayor y menor bursatilidad de la BMV a través del aprendizaje automático: 2005-2020.**

**T e s i s**

Que para optar por el grado de:

**Doctor en Ciencias de la Administración**

Presenta:

**Carlos Omar Chávez Sánchez**

Comité Tutor

Tutor principal:

**Dr. Arturo Morales Castro**

**Facultad de Contaduría y Administración, U.N.A.M.**

**Dr. Arturo Rodríguez Castellanos**

**Facultad de Ciencias Econ. y Empresariales, U.P.V.**

**Dr. Oswaldo García Salgado**

**Facultad de Economía, U.A.E.M.**

**Ciudad de México, febrero de 2024.**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## Dedicatoria

*Queridos padres, ese “ingeniero” serio y creativo con la facilidad para resolver los problemas cotidianos y esa “administradora” alegre y determinada con el don de la gestión de los recursos, ambos sin títulos, quienes provocaron en mí una enorme curiosidad y motivación para entender el mundo, quienes con amor y sacrificio sembraron las raíces de mi educación.*

*A mi amada familia, fuente inagotable de apoyo y comprensión, esta tesis es el fruto de nuestros valores compartidos y de la fe inquebrantable que han depositado en mí.*

*A mis hijas, Valen y Carmis, mi mayor inspiración, este logro es también suyo. Cada página escrita es un paso más hacia un futuro que espero que esté lleno de oportunidades y orgullo.*

*A Paola por haberme permitido formar un hogar.*

*A mi familia extendida, mis profesores y amigos, cuyas risas y alientos han iluminado este arduo camino, les dedico con gratitud cada palabra y descubrimiento plasmado en estas páginas.*

*En este momento de culminación académica, celebro la fortaleza de los lazos familiares que han tejido un soporte inquebrantable. Gracias por ser el faro que me guía en la oscuridad y por ser la razón detrás de cada logro. Este logro es nuestro, y esta tesis lleva impresa la esencia de quienes me han sostenido a lo largo de este viaje.*

## Agradecimientos

A lo largo de este arduo pero enriquecedor camino académico, he tenido el privilegio de contar con el apoyo y la colaboración de diversas personas e instituciones, a quienes deseo expresar mi sincero agradecimiento:

A mis directores de tesis:

- Dr. Arturo Morales por su guía experta, paciencia infinita y dedicación incansable. Su mentoría ha sido fundamental en mi desarrollo académico y profesional. Quien nunca ha perdido la fe en mi y me ha motivado a continuar en esta hermosa carrera ¡Gracias por creer en mi doc!
- Dr. Arturo Rodríguez por calidez, amabilidad y gran sabiduría para mostrarme siempre un camino. Su presencia y palabras fueron de gran apoyo en los momentos cuando parecía que no había más rumbo ¡Agradezco su acompañamiento desde la distancia doc!
- Dr. Oswaldo García por su perspicacia, decisión y pasión por lo que hace. Con tu incorporación le diste frescura al proyecto. Tus comentarios y observaciones me dieron el rumbo ¡Gracias por tu amistad incondicional Oz!

A mis revisores de tesis por sus valiosas contribuciones, críticas constructivas y por compartir su sabiduría en cada etapa de esta investigación:

- Dr. Henry Bricard por su sentido estricto y experiencia. Su profesionalismo y forma de ver las cosas me hizo apreciar otros panoramas.
- Dr. Hideo Taniguchi por su meticulosidad, ojo microscópico y dedicación. Quien invirtió un enorme esfuerzo y tiempo invertidos en precisar mi documento y me inculcó fuertemente la idea de la calidad en las cosas que hacemos.

Agradezco al Posgrado de la Facultad de Contaduría y Administración por brindarme los recursos y el ambiente propicio para llevar a cabo esta investigación. A la U.N.A.M., mi alma

*mater*, todos estos años cobijándome en sus aulas y por la enorme calidad de la formación recibida que ha sido clave en mi crecimiento académico y personal.

Expreso mi gratitud a mis colegas y compañeros de investigación, por las discusiones enriquecedoras, el intercambio de ideas y el espíritu colaborativo que ha caracterizado nuestro trabajo conjunto.

A mi familia, por su amor incondicional, comprensión y apoyo constante. Han sido mi roca en los momentos difíciles y mi fuente de alegría en los triunfos.

A mis amigos y seres queridos, quienes han compartido este viaje conmigo, brindándome aliento y motivación. Su presencia ha hecho este camino más ameno y significativo.

A todos aquellos cuya contribución no puede ser enumerada aquí pero que han dejado una huella positiva en mi vida, gracias por ser parte de este viaje.

Este logro no es solo mío, sino de todos ustedes. Aprecio profundamente el papel que cada uno ha desempeñado en mi trayectoria académica. ¡Gracias!

## Resumen

El presente trabajo tiene la finalidad de evaluar la presencia de anomalías en el mercado accionario mexicano, específicamente se concentra en determinar el desempeño del *momentum*, la reversión a la media en precio y en volatilidad, efecto enero, efecto de fin de mes y de fin de semana. El objetivo es hacer una revisión de los efectos de éstas en términos de los beneficios que pudieran haber generado su identificación, en contraste con el rendimiento del mercado para el periodo comprendido de 2005 a 2020. Las series analizadas corresponden a los precios mensuales, semanales y diarios de las dos empresas con la mayor y menor bursatilidad del S&P/BMV IPC, Cemex S.A. de C.V. e Industrias Peñoles S.A.B. de C.V. respectivamente, y de dos portafolios de las empresas mexicanas de mayor y menor bursatilidad de la BMV. Se proponen algoritmos de cálculo de las anomalías usando el precio de las cotizaciones y los resultados se presentan en términos de rendimientos ajustados al mercado, representado por el Índice de Precios y Cotizaciones mexicano o IPC. Las anomalías son exploradas de manera independiente y su presencia es probada con cuatro metodologías de aprendizaje automático: *regresión logística*, *perceptrón multicapa*, *máquinas de soporte vectorial* y *bosques aleatorios*, cuyos modelos son evaluados con el conjunto completo y validación cruzada. Los resultados muestran que el *momentum* tiene mayor presencia dado el comportamiento del índice, sin gran frecuencia de cambios u oscilaciones. Le sigue la reversión a la media y la volatilidad con presencia media. Mientras que el efecto enero logra mejores porcentajes de rendimientos que el efecto fin de mes y fin de semana. Además, se identifican que los primeros cuatro periodos posteriores a la señal de las anomalías son los más rentables.

**Palabra clave:** Anomalías de mercado, rendimientos ajustados al riesgo, aprendizaje automático, mercado mexicano, rendimiento de IPC.

## Abstract

The present work has the purpose of evaluating the presence of anomalies in the Mexican stock market, specifically it focuses on determining the performance of the *momentum*, the reversion to the mean in price and in volatility, January effect, end-of-month and end-of-week effect. The objective is to review the effects of these in terms of the benefits that their identification could have generated, in contrast to the market performance for the period from 2005 to 2020. The series analyzed correspond to the monthly, weekly and daily of the two companies with the highest and lowest marketability, Cemex S.A. de C.V. and Industrias Peñoles S.A.B. de C.V. respectively, and two portfolios of Mexican companies with the highest and lowest marketability on the BMV. Algorithms for calculating the anomalies using the price of the quotes are proposed and the results are presented in terms of returns adjusted to the market, represented by the Mexican Price and Quotes Index or IPC. The anomalies are explored independently and their presence is tested with four machine learning methodologies: logistic regression, multilayer perceptron, support vector machines, and random forests, whose models are evaluated with full set and cross validation. The results show that the *momentum* has a greater presence given the behavior of the index, without great frequency of changes or oscillations. It is followed by reversion to the mean and volatility with a medium presence. While the January effect achieves better yield percentages than the end of the month and weekend effect. In addition, it is identified that the first four periods after the signal of the anomalies are the most profitable.

**Keyword:** Market anomalies, return forecast, machine learning, Mexican market, IPC performance.

# Índice

<i>Dedicatoria</i> .....	<i>ii</i>
<i>Agradecimientos</i> .....	<i>iii</i>
<i>Resumen</i> .....	<i>v</i>
<i>Abstract</i> .....	<i>vi</i>
<i>Índice</i> .....	<i>iv</i>
<i>Índice de tablas</i> .....	<i>vi</i>
<i>Índice de figuras</i> .....	<i>vii</i>
<b>PROYECTO DE INVESTIGACIÓN Y SU PLANTEAMIENTO</b> .....	<b>1</b>
<i>Introducción</i> .....	<i>1</i>
<i>Objetivo general</i> .....	<i>3</i>
<i>Objetivos específicos</i> .....	<i>3</i>
<i>Preguntas de investigación</i> .....	<i>4</i>
<i>Hipótesis</i> .....	<i>4</i>
<i>Matriz de Congruencia</i> .....	<i>5</i>
<i>Antecedente</i> .....	<i>5</i>
<i>Justificación</i> .....	<i>8</i>
<i>Datos</i> .....	<i>11</i>
<i>Metodología</i> .....	<i>12</i>
<i>Estructura de la Investigación</i> .....	<i>14</i>
<i>Resumen</i> .....	<i>15</i>
<b>Capítulo 1 MARCO TEÓRICO</b> .....	<b>16</b>
1.1 <i>Hipótesis de los Mercados Eficientes, HME</i> .....	<i>17</i>
1.1.1 <i>Antecedentes</i> .....	<i>17</i>
1.1.2 <i>Niveles de eficiencia de los mercados</i> .....	<i>23</i>
1.2 <i>Anomalías de Mercado</i> .....	<i>26</i>
1.2.1 <i>Anomalías de calendario</i> .....	<i>27</i>
1.2.2 <i>Anomalías fundamentales</i> .....	<i>28</i>
1.2.3 <i>Anomalías técnicas</i> .....	<i>29</i>
1.2.4 <i>Anomalías como factores financieros</i> .....	<i>30</i>
1.3 <i>El Mercado mexicano</i> .....	<i>33</i>
1.3.1 <i>Historia de la BMV</i> .....	<i>38</i>
1.3.2 <i>Índice de Precios y Cotizaciones, IPC</i> .....	<i>39</i>
1.3.3 <i>Otros índices</i> .....	<i>43</i>
1.3.4 <i>Regulación</i> .....	<i>44</i>
1.3.5 <i>Análisis y estudios empíricos sobre comportamiento de los precios en la BMV</i> .....	<i>46</i>
<i>Resumen</i> .....	<i>50</i>
<b>Capítulo 2 HERRAMIENTAS DE ANÁLISIS</b> .....	<b>51</b>
2.1 <i>Introducción</i> .....	<i>51</i>
2.2 <i>Valuación de activos</i> .....	<i>54</i>
2.3 <i>El Modelo de valoración de activos financieros, CAPM</i> .....	<i>55</i>
2.4 <i>Modelos multifactoriales</i> .....	<i>58</i>
2.4.1 <i>Teoría de fijación de precios por arbitraje, APT</i> .....	<i>59</i>

2.4.2	<i>Modelo Fama-MacBeth</i> .....	61
2.4.3	<i>Modelos de análisis multivariado</i> .....	64
2.5	<i>Métodos de análisis no lineales</i> .....	67
2.5.1	<i>Modelos ARCH y GARCH</i> .....	70
2.5.2	<i>Aprendizaje automático</i> .....	71
2.5.2.1	<i>Perceptrón multicapa</i> .....	73
2.5.2.2	<i>Regresión Logística</i> .....	76
2.5.2.3	<i>Máquinas de soporte vectorial</i> .....	79
2.5.2.4	<i>Bosque aleatorio</i> .....	81
2.6	<i>Sobre ajuste y sesgo de datos</i> .....	84
2.7	<i>Evaluación de desempeño</i> .....	86
	<i>Resumen</i> .....	88
<b>Capítulo 3 METODOLOGÍA PROPUESTA</b> .....		89
3.1	<i>Métodos de clasificación</i> .....	89
3.2	<i>Fuente de datos</i> .....	91
3.3	<i>Cálculo de anomalías</i> .....	95
3.4	<i>Procedimiento general</i> .....	102
3.5	<i>Evaluación de las anomalías</i> .....	103
3.6	<i>Análisis de las anomalías</i> .....	104
	<i>Resumen</i> .....	106
<b>Capítulo 4 RESULTADOS Y DISCUSIONES</b> .....		108
4.1	<i>Ejecución de pruebas</i> .....	108
4.1.1	<i>Desempeño del modelo</i> .....	108
4.1.2	<i>Análisis de las anomalías</i> .....	111
4.1.3	<i>Rendimientos ajustados</i> .....	119
4.1.4	<i>Discusión</i> .....	123
	<i>Resumen</i> .....	127
<b>CONCLUSIONES</b> .....		128
	<i>Trabajo futuro</i> .....	130
<b>REFERENCIAS</b> .....		132
<b>Anexo A PROCEDIMIENTO DE ANÁLISIS</b> .....		144
	<i>Modelo de análisis</i> .....	144
	<i>Automatización del análisis</i> .....	153
<b>Anexo B RESULTADOS EXTENDIDOS</b> .....		164

## Índice de tablas

<i>Tabla 1.1. Anomalías de calendario.</i>	28
<i>Tabla 1.2. Anomalías fundamentales.</i>	28
<i>Tabla 1.3. Anomalías técnicas.</i>	29
<i>Tabla 1.4. Tipos de índices comercializados en la BMV.</i>	43
<i>Tabla 1.5. Entidades supervisoras de las regulaciones.</i>	44
<i>Tabla 3.1. Empresas del IPC ordenadas de menor a mayor con operaciones en cada año dentro del periodo 2005 a 2020.</i>	92
<i>Tabla 3.2. Portafolio de empresas mexicanas de la BMV con la mayor bursatilidad.</i>	93
<i>Tabla 3.3. Portafolio de las empresas mexicanas de la BMV con la menor bursatilidad.</i>	93
<i>Tabla B.1. Porcentajes de ajuste de los modelos para Peñoles en tendencia positiva: resultados de la clasificación por anomalía.</i>	169
<i>Tabla B.2. Porcentajes de ajuste de los modelos para Peñoles en tendencia negativa: resultados de la clasificación por anomalía.</i>	170
<i>Tabla B.3. Porcentajes de ajuste de los modelos para Cemex en tendencia positiva: resultados de la clasificación por anomalía.</i>	173
<i>Tabla B.4. Porcentajes de ajuste de los modelos para Cemex en tendencia negativa: resultados de la clasificación por anomalía.</i>	174
<i>Tabla B.5. Porcentajes de ajuste de los modelos para el portafolio de empresas con menor bursatilidad en tendencia positiva.</i>	183
<i>Tabla B.6. Porcentajes de ajuste de los modelos para el portafolio de empresas con menor bursatilidad en tendencia negativa.</i>	184
<i>Tabla B.7. Porcentajes de ajuste de los modelos para el portafolio de empresas de mayor bursatilidad en tendencia positiva.</i>	187
<i>Tabla B.8. Porcentajes de ajuste de los modelos para el portafolio de empresas de mayor bursatilidad en tendencia negativa.</i>	188

## Índice de figuras

<i>Figura 1. Precios anuales del IPC de México de 2000 a 2020.....</i>	<i>9</i>
<i>Figura 1.1. Niveles de eficiencia de los mercados.....</i>	<i>23</i>
<i>Figura 1.2. Métodos tradicionales para probar la forma débil de eficiencia de mercado.....</i>	<i>25</i>
<i>Figura 1.3. Índice EM MSCI (Morgan Stanley, 2021) que sigue el comportamiento de los mercados emergentes más importantes de 2003 a 2020.....</i>	<i>37</i>
<i>Figura 1.4. Línea de tiempo de progreso de la BMV.....</i>	<i>40</i>
<i>Figura 2.1. Modelo de umbral binario .....</i>	<i>73</i>
<i>Figura 2.2. Gráficas de función de activación: a) escalón y b) logística .....</i>	<i>74</i>
<i>Figura 2.3 Perceptrón multicapa con una única capa oculta.....</i>	<i>75</i>
<i>Figura 2.4. Regresión logística: a) la transformada logit y b) ejemplo de función logística .....</i>	<i>78</i>
<i>Figura 2.5. Hiper plano de separación óptima.....</i>	<i>80</i>
<i>Figura 2.6. Representación gráfica de un árbol de decisión.....</i>	<i>82</i>
<i>Figura 2.8. Proceso de ejecución de validación cruzada.....</i>	<i>87</i>
<i>Figura 3.1. La serie mensual de América Móvil muestra que en periodos previos al 2008 no había o había muy bajo volumen de operaciones. ....</i>	<i>94</i>
<i>Figura 3.2. La serie mensual de Cemex muestra volumen de operaciones para todos periodos analizados.....</i>	<i>94</i>
<i>Figura 3.3. Representación de la señal de momentum calculado sobre la serie de mensual del IPC y la aproximación de su correspondencia en la serie de precios.....</i>	<i>96</i>
<i>Figura 3.4. Representación de la reversión a la media en la serie de precios mensuales del IPC. .</i>	<i>99</i>
<i>Figura 3.5. Representación de la anomalía de volatilidad calculada sobre la serie mensual del IPC, junto con su correspondencia aproximada. ....</i>	<i>100</i>
<i>Figura 3.6. Representación de la señal del efecto enero en la serie de precios mensuales del IPC. ....</i>	<i>101</i>
<i>Figura 3.7. Diagrama de flujo del procedimiento de análisis. ....</i>	<i>106</i>
<i>Figura 4.1. Porcentajes de ajuste promedio de los clasificadores evaluados con validación cruzada para Peñoles y Cemex. ....</i>	<i>109</i>
<i>Figura 4.2. Porcentajes de ajuste promedio de los clasificadores evaluados con validación cruzada para los portafolios de las empresas de mayor y menor bursatilidad. ....</i>	<i>109</i>
<i>Figura 4.3. Porcentajes de ajuste promedio de los clasificadores evaluados con el conjunto de datos completo para Peñoles y Cemex. ....</i>	<i>110</i>
<i>Figura 4.4. Porcentajes de ajuste promedio de los clasificadores evaluados con el conjunto de datos completo para los portafolios de empresas. ....</i>	<i>110</i>
<i>Figura 4.5 Presencia de la anomalía de momentum en Peñoles mediante validación cruzada. ....</i>	<i>112</i>
<i>Figura 4.6 Presencia de la anomalía de reversión a la media en Peñoles mediante validación cruzada. ....</i>	<i>112</i>
<i>Figura 4.7 Presencia de la anomalía de volatilidad en Peñoles mediante validación cruzada. ....</i>	<i>112</i>
<i>Figura 4.8 Presencia de anomalías de calendario en Peñoles mediante validación cruzada.....</i>	<i>113</i>
<i>Figura 4.9 Presencia de la anomalía de momentum en Cemex mediante validación cruzada.....</i>	<i>113</i>
<i>Figura 4.10 Presencia de la anomalía de reversión a la media en Cemex mediante validación cruzada. ....</i>	<i>114</i>
<i>Figura 4.11 Presencia de anomalía de volatilidad en Cemex mediante validación cruzada. ....</i>	<i>114</i>
<i>Figura 4.12 Gráfica de la presencia de calendario en Cemex evaluando con validación cruzada. ....</i>	<i>114</i>

<i>Figura 4.13 Presencia de la anomalía de momentum en el portafolio de empresas de menor bursatilidad mediante validación cruzada.</i>	115
<i>Figura 4.14 Presencia de la anomalía de reversión a la media en el portafolio de empresas de menor bursatilidad mediante validación cruzada.</i>	116
<i>Figura 4.15 Presencia de la anomalía de volatilidad en el portafolio de empresas de menor bursatilidad mediante validación cruzada.</i>	116
<i>Figura 4.16 Presencia de las anomalías de calendario en el portafolio de empresas de menor bursatilidad mediante validación cruzada.</i>	116
<i>Figura 4.17 Presencia de la anomalía momentum en el portafolio de empresas de mayor bursatilidad mediante validación cruzada.</i>	117
<i>Figura 4.18 Presencia de la anomalía de reversión a la media en el portafolio de empresas de mayor bursatilidad mediante validación cruzada.</i>	117
<i>Figura 4.19 Presencia de la anomalía de volatilidad en el portafolio de empresas de mayor bursatilidad mediante validación cruzada.</i>	118
<i>Figura 4.20 Presencia de las anomalías de calendario en el portafolio de empresas de mayor bursatilidad mediante validación cruzada.</i>	118
<i>Figura 4.21 Rendimiento promedio de la serie mensual de Cemex e IPC para periodos posteriores a la señal de reversión a la media.</i>	120
<i>Figura 4.22 Rendimientos promedio, máximos y mínimos para la anomalía de momentum en la serie de precios diarios de Peñoles.</i>	120
<i>Figura 4.23 Rendimientos promedio de las series mensual, semanal y diaria de Peñoles para los diferentes periodos rezagados de la anomalía momentum.</i>	121
<i>Figura 4.24 Rendimientos promedio, máximos y mínimos de la serie diaria de Cemex posteriores a la presencia de la anomalía momentum.</i>	122
<i>Figura 4.25 Rendimientos promedio de las series mensual, semanal y diaria de Cemex para los diferentes periodos rezagados de la anomalía momentum.</i>	122
<i>Figura 4.26 Proyección de los rendimientos para periodos rezagados largos en la serie mensual de Cemex.</i>	123
<i>Figura 4.27 Rendimientos promedio de la anomalía momentum en la serie mensual de Cemex.</i>	123
<i>Figura A.1. Proceso de ejecución del estudio.</i>	145
<i>Figura A.2. Pantalla principal de Weka en ejecución de aplicación de escritorio.</i>	147
<i>Figura A.3. Preparación de prueba.</i>	147
<i>Figura A.4. Creación de archivo de configuración de datos arff.</i>	148
<i>Figura A.5. Descripción de los datos y su especificación.</i>	149
<i>Figura A.6. Selección y filtrado de variables del modelo.</i>	150
<i>Figura A.7. Configuración de modelo de análisis.</i>	151
<i>Figura A.8. Selección de clasificador.</i>	151
<i>Figura A.9. Selección de método de evaluación de modelo.</i>	152
<i>Figura A.10. Ejecución de análisis de un modelo.</i>	153
<i>Figura A.11. Salida de la ejecución de una batería de pruebas automatizadas en lenguaje java.</i>	159
<i>Figura A.12. Resultado de tabla en Excel de importar una batería de prueba en formato csv.</i>	159
<i>Figura B.1. Ejemplo de las señales en la serie mensual de Peñoles: momentum y reversión a la media.</i>	165
<i>Figura B.2. Ejemplo de las señales en la serie mensual de Peñoles: volatilidad y efecto de calendario.</i>	166
<i>Figura B.3. Ejemplo de las señales en la serie mensual de Cemex: momentum y reversión a la media.</i>	167
<i>Figura B.4. Ejemplo de las señales en la serie mensual de Cemex: volatilidad y efecto de calendario.</i>	168
<i>Figura B.5. Gráficas comparativas del desempeño promedio de los clasificadores con las series de Peñoles por anomalía en tendencia positiva.</i>	171

<i>Figura B.6. Gráficas comparativas del desempeño promedio de los clasificadores con las series de Peñoles por anomalía en tendencia negativa.....</i>	<i>172</i>
<i>Figura B.7. Gráficas comparativas del desempeño promedio de los clasificadores con las series de Cemex por anomalía en tendencia positiva.....</i>	<i>175</i>
<i>Figura B.8. Gráficas comparativas del desempeño promedio de los clasificadores con las series de Cemex por anomalía en tendencia negativa.....</i>	<i>176</i>
<i>Figura B.9. Porcentajes de exactitud obtenido con el modelo de Máquina de soporte vectorial en las series de Peñoles.....</i>	<i>177</i>
<i>Figura B.10. Porcentajes de exactitud obtenido con el modelo de Máquina de soporte vectorial en las series de Cemex. ....</i>	<i>178</i>
<i>Figura B.11. Rendimientos lineales ajustados al riesgo de Peñoles en tendencia positiva.....</i>	<i>179</i>
<i>Figura B.12. Rendimientos lineales ajustados al riesgo de Peñoles en tendencia negativa. ....</i>	<i>180</i>
<i>Figura B.13. Rendimientos lineales ajustados al riesgo de Cemex en tendencia positiva. ....</i>	<i>181</i>
<i>Figura B.14. Rendimientos lineales ajustados al riesgo de Cemex en tendencia negativa. ....</i>	<i>182</i>
<i>Figura B.15. Desempeño promedio de los clasificadores para el portafolio de empresas con menor bursatilidad en tendencia positiva.....</i>	<i>185</i>
<i>Figura B.16. Desempeño promedio de los clasificadores para el portafolio empresas con menor bursatilidad en tendencia negativa.....</i>	<i>186</i>
<i>Figura B.17. Desempeño promedio de los clasificadores para el portafolio de empresas con mayor bursatilidad en tendencia positiva.....</i>	<i>189</i>
<i>Figura B.18. Desempeño promedio de los clasificadores para el portafolio de empresas con mayor bursatilidad en tendencia negativa.....</i>	<i>190</i>
<i>Figura B.19. Porcentajes de exactitud obtenido con el modelo de Máquina de soporte vectorial para el portafolio de empresas con menor bursatilidad. ....</i>	<i>191</i>
<i>Figura B.20. Porcentajes de exactitud obtenido con el modelo de Máquina de soporte vectorial para el portafolio de empresas con mayor bursatilidad. ....</i>	<i>192</i>

# PROYECTO DE INVESTIGACIÓN Y SU PLANTEAMIENTO

## *Introducción*

Desde su introducción, el Modelo de apreciación de activos de capital<sup>1</sup> (CAPM, por las siglas en inglés: *Capital Asset Pricing Model*) ha sido el principal referente teórico en el ámbito económico-financiero para la valoración de activos y éste ha servido como base de comparación de modelos posteriores, como el modelo de valuación de opciones (Black et al., 1972), el modelo basado en la teoría de arbitraje de precios (Ross, 1976), modelo de tres factores (Fama y French, 1992), modelo de cuatro factores (Carhart, 1997), Martingalas (Shreve, 1992) y otros. Sin embargo, a pesar de los esfuerzos por desarrollar modelos más robustos para obtener estimaciones más precisas, la alta sensibilidad durante periodos de tensión en los mercados sigue siendo un desafío común que afecta su fiabilidad, lo que sigue siendo motivo de debate en la actualidad (Cervantes Zepeda, 1999) (Rubistein, 2001) (Damodaran, 2012).

Hoy en día, gracias a la multidisciplinariedad existente entre las áreas científicas y la expansión del conocimiento, especialmente en términos de métodos de análisis de información para su procesamiento y estudio, se ha hecho posible disponer de herramientas adicionales para poder poner a prueba las hipótesis prevalecientes en cada área. Esto ha llevado a la obtención de resultados variables y, en muchos casos, a conclusiones nuevas y mejor fundamentadas. Entre los paradigmas más comúnmente utilizados, que han mostrado un rendimiento sobresaliente, se encuentran aquellos relacionados a las ciencias de la computación, como el aprendizaje automático y la inteligencia artificial (Montalván et al.,

---

<sup>1</sup> El CAPM es un modelo de valuación de activos que explica los rendimientos esperados de un activo en términos de la correlación de los rendimientos de otro activo que representa al mercado. El resultado es un factor al que se le denomina  $\beta$  (beta) y al que se le refiere como riesgo sistemático (Sharpe, 1964).

2018) (Fernández et al., 2021) (Hinojosa et al., 2021). Estos modelos han sido probados en el área económico-financiera en tareas de predicción y clasificación (activos financieros, quiebra de empresas, selección de carteras, análisis de riesgos) con resultados prometedores (Damodaran, 2012) (García et al., 2013) (Semaan et al., 2014) (Wong, 2020) (Chopra y Sharma, 2021). No obstante, a pesar de la existencia de cuerpos teóricos sólidos en el ámbito financiero, cuyos principios siguen generando escepticismo en cuanto a la predicción de los precios y los rendimientos de los activos (Fama, 1970) (Fama y French, 1992) (Barberis y Thaler, 2003), han surgido enfoques alternativos para analizar los fenómenos financieros. Estos enfoques incluyen la teoría de mercados ineficientes (Ashadi et al., 2012), que se centra en el análisis de los eventos esporádicos que impactan los mercados, las finanzas conductuales (Ferreira et al., 2021) (Weiwei, 2021) (Ansah et al., 2022), que considera el flujo de información desde la perspectiva de la orientación de los participantes, y más recientemente, la teoría de los mercados adaptativos (Lo, 2005) (Khursheed et al., 2020) (Munir et al., 2022). Esta última se basa en la creación de modelos responsivos que se ajustan a medida que se dispone de nueva información. Esto plantea la necesidad de reconsiderar la efectividad de los modelos existentes y de desafiarlos con nuevas y mejores hipótesis.

Asimismo, diversos trabajos sobre análisis de activos financieros se han centrado en el estudio de las discrepancias de los precios y rendimientos de los activos. Se han enfocado en identificar los elementos o factores que impactan de manera extraordinaria, con la esperanza de aprovechar sus beneficios. Los resultados han mostrado evidencia empírica que valida la presencia de estas discrepancias, al menos en mercados de Estados Unidos (Basu, 1977) (Ariel, 1987) (Brock et al., 1992) (Jegadeesh y Titman, 1993) (Frankfurter y McGoun, 2001) (Balvers y Wu, 2006) (Barberis y Huang, 2009) (Baker et al., 2011) (Green et al., 2012) (Harvey et al., 2016). Estos factores suelen corresponder a anomalías en el comportamiento de los activos financieros, como introducen Kahneman y Tversky (1984), lo que los hace oscilar entre la eficiencia y la ineficiencia (Akwolaga, 2020) (Nti et al., 2020) (Ansah et al., 2022) (Argade et al., 2022).

Por último, el mercado mexicano ha experimentado un desarrollado significativo en el último siglo y, con la última estructuración realizada en la década de 1970, éste ha alcanzado su

estado actual. Aunque se considera que el mercado mexicano es relativamente joven y ha sido poco explotado, existe un creciente interés en comprender su comportamiento, los precios de los activos, sus rendimientos y los ajustes de riesgo. Esto enriquece la literatura existente sobre las características importantes que influyen en los participantes del mercado. Por lo tanto, esta investigación se centra en analizar el mercado mexicano en busca de la presencia de algunas de las anomalías comúnmente encontradas en otros mercados como el *momentum*, reversión a la media en el precio y la volatilidad, así como los efectos enero, fin de mes y fin de semana utilizando aprendizaje automático<sup>2</sup>. El periodo de análisis abarca desde 2005 hasta 2020.

### ***Objetivo general***

Evaluar la influencia de las anomalías de mercado *momentum*, reversión a la media, volatilidad, el efecto enero, fin de semana y fin de mes en los rendimientos de los activos financieros del mercado mexicano durante el periodo 2005 a 2020.

### ***Objetivos específicos***

1. Analizar el impacto de las anomalías de mercado en los rendimientos ajustados de los activos del mercado mexicano en el periodo 2005 a 2020.
2. Emplear técnicas no lineales de aprendizaje automático como mecanismos de evaluación para mejorar el ajuste de los modelos elaborados.

---

<sup>2</sup> Aprendizaje automático, o *Machine Learning* (ML) en inglés, es una extensión de la Inteligencia Artificial, AI por las siglas en inglés de *Artificial Intelligence*, que permite aprender y mejorar automáticamente sin ser explícitamente programado. Utiliza algoritmos de paradigmas variados para generar reglas sobre la información que utiliza para auto aprendizaje. Se divide en aprendizaje supervisado, no supervisado y de reforzamiento. El ML extrae conocimiento de los datos el cual puede ser usado para predecir y generar nueva información. La eficiencia de los modelos en analizar, clasificar y predecir eventos en complicados y vastos cúmulos de datos ha hecho la ML una herramienta confiable. En el sector financiero ha sido utilizado con excelentes resultados para el análisis y predicción de precios, así como en la gestión de riesgos.

En un sistema autónomo, el ML es usado para la creación del modelo de reglas que identifican y discriminan la información entrante y la IA es el sistema integral que toma decisiones a partir de dichas reglas, incluyendo si estas deben ser retroalimentadas o reajustadas. En este trabajo se utiliza ML para referirse a los métodos y técnicas de representación de la información que conducen a las reglas de aprendizaje que modelan un determinado problema con fines de clasificación.

3. Identificar los periodos de mayor influencia de las anomalías en los rendimientos ajustados.

### ***Preguntas de investigación***

¿Qué influencia tienen las anomalías de mercado *momentum*, reversión a la media, volatilidad, efecto enero, efecto fin de semana y efecto fin de mes en los rendimientos de los activos financieros del mercado mexicano en el periodo 2005 y 2020?

1. ¿Cuál es el impacto de las anomalías de mercado en los rendimientos ajustados de los activos financieros en el mercado mexicano?
2. ¿Puede mejorarse el ajuste de los modelos elaborados al utilizar técnicas no lineales de aprendizaje automático como mecanismo de evaluación?
3. ¿Qué periodos son los de mayor influencia de las anomalías en los rendimientos ajustados?

### ***Hipótesis***

Las anomalías de mercado *momentum*, reversión a la media, volatilidad, efecto enero, efecto fin de semana y efecto fin de mes tienen influencia significativa en los rendimientos ajustados de los activos financieros del mercado mexicano en el periodo 2005 a 2020.

- a) Las anomalías de mercado impactan en los rendimientos ajustados de los activos financieros del mercado mexicano.
- b) Las técnicas no lineales de aprendizaje automático mejoran el ajuste de los modelos al utilizarse como mecanismos de evaluación.
- c) Los primeros periodos son los de mayor influencia de las anomalías en los rendimientos ajustados.

## ***Matriz de Congruencia***

<b>OBJETIVO</b>	<b>PREGUNTA DE INVESTIGACIÓN</b>	<b>HIPÓTESIS</b>
<b>Evaluar la influencia de las anomalías de mercado <i>momentum</i>, reversión a la media, volatilidad, el efecto enero, fin de semana y fin de mes en los rendimientos de los activos financieros del mercado mexicano durante el periodo 2005 a 2020.</b>	¿Qué influencia tienen las anomalías de mercado <i>momentum</i> , reversión a la media, volatilidad, efecto enero, efecto fin de semana y efecto fin de mes en los rendimientos de los activos financieros del mercado mexicano en el periodo 2005 y 2020?	Las anomalías de mercado <i>momentum</i> , reversión a la media, volatilidad, efecto enero, efecto fin de semana y efecto fin de mes tienen influencia significativa en los rendimientos ajustados de los activos financieros del mercado mexicano en el periodo 2005 a 2020.
<b>Analizar el impacto de las anomalías de mercado en los rendimientos ajustados de los activos del mercado mexicano en el periodo 2005 a 2020.</b>	¿Cuál es el impacto de las anomalías de mercado en los rendimientos ajustados de los activos financieros en el mercado mexicano?	Las anomalías de mercado impactan positivamente en los rendimientos ajustados de los activos financieros del mercado mexicano.
<b>Emplear técnicas no lineales de aprendizaje automático como mecanismos de evaluación para mejorar el ajuste de los modelos elaborados.</b>	¿Puede mejorarse el ajuste de los modelos elaborados al utilizar técnicas no lineales de aprendizaje automático como mecanismo de evaluación?	Las técnicas no lineales de aprendizaje automático mejoran el ajuste de los modelos al utilizarse como mecanismos de evaluación.
<b>Identificar los periodos de mayor influencia de las anomalías en los rendimientos ajustados.</b>	¿Qué periodos son los de mayor influencia de las anomalías en los rendimientos ajustados?	Los primeros periodos son los de mayor influencia de las anomalías en los rendimientos ajustados.

## ***Antecedente***

En el ámbito financiero y económico, la Hipótesis de Mercados Eficientes (HME), también conocida como *Efficient Market Hypothesis* (EMH), constituye uno de los pilares teóricos más robustos (Cervantes Zepeda, 1999) (Barberis y Thaler, 2003) (Campbell et al., 2012) (Damodaran, 2012) (Delcey, 2019) (Díaz Ruiz et al., 2020) (Chopra y Sharma, 2021) (Gajapathy y Sathyanarayana, 2021) (Ansah et al., 2022) (Argade et al., 2022). Esta hipótesis postula la racionalidad de los mercados, afirmando que los precios de las acciones reflejan plenamente toda la información disponible. Según este principio, los precios de los activos se ajustan de manera inmediata a nueva información, dejando a los inversores incapaces de

superar al mercado, ya que cualquier evento relevante o noticia que pueda conferirles una ventaja se refleja rápidamente en los precios, o estos se ajustan en tiempo real (Fama, 1970) (Damodaran, 2012). Sin embargo, varias investigaciones y análisis empíricos, como los de Fama y French (1993) o Jegadeh y Titman (1993), han revelado mercados cuyo funcionamiento se desvía de los principios de la HME, suscitando interrogantes sobre su validez.

Fama y French (1992) introducen un modelo compuesto por tres factores para la estimación futura del precio de acciones: valor de capital en libros, apalancamiento y ganancia por precio. Este modelo, ampliamente citado y utilizado, se ha complementado con nuevos factores y criterios propuestos en diversos estudios, basados en asimetrías identificadas en la conducta de diferentes activos: razón precio-ganancia, P/E (Basu, 1977), reversión a la media (Fama y French, 1988) (Poterba y Summers, 1988), cruce de medias móviles (Brock et al., 1992), *momentum* (Carhart, 1997), reversión a la media en la volatilidad (Kuhe y Audu, 2019), etc.

El término de “anomalía” es acuñado por Kahneman y Tversky (1984) y desde entonces se asocia con cualquier irregularidad en los precios que difiere de la premisa de la HME. Tversky y Thaler (1990) definen las anomalías como hechos o características de los rendimientos en los que un grupo de activos presenta rendimientos mayores que otros y no pueden explicarse mediante modelos de valuación simples como el CAPM, o similares. Estas anomalías pueden ser eventos únicos o recurrentes (Latif et al., 2011).

En cuanto a la percepción de los precios de las acciones por parte de los inversores, se ha clasificado a los "inversionistas ruidosos" como irracionales y a los "arbitrajistas" como racionales. Se sugiere que los inversores racionales influyen en la corrección de las desviaciones de precios causadas por los irracionales. A pesar de las diferencias entre los precios de las acciones y su valor intrínseco, muchos investigadores sostienen que estos precios deben ser correctos y no ofrecen oportunidades de inversión excepcionales ni que éstos estén libres de riesgos (Friedman, 1953) (Barberis y Thaler, 2003). Este debate mantiene abierta la discusión sobre la eficiencia del mercado, donde todavía no se ha llegado

a un consenso sobre si los gestores de dinero y profesionales en finanzas pueden superar consistentemente al mercado, respaldando así la HME (Rubinstein, 2001).

A pesar de las evidencias empíricas, las anomalías no han sido consideradas eventos lo suficientemente disruptivos como para descartar completamente la HME. Además, no hay pruebas concluyentes de que, una vez detectada una anomalía, esta persista o sea de utilidad para los inversores.

Investigaciones recientes discuten sobre una suavización o degradación en el desempeño de las anomalías una vez que éstas se han hecho públicas (McLean et al., 2016). También se ha cuestionado la autenticidad y eficacia de varias de las anomalías reportadas en la literatura, ya que algunos podrían no ser estrategias significativas o no se replican en el tiempo. Otros autores sugieren que la liquidez de los mercados, impulsada por la participación de nuevos actores (Green et al., 2012), la actividad de *trading*, las innovaciones tecnológicas y el mejoramiento en la capacidad de cómputo, ha facilitado la corrección de desviaciones en los precios, fomentando así la eficiencia del mercado y reduciendo los efectos de anomalía, al menos en el contexto de los Estados Unidos (Fama y French, 1988) (Chordia et al., 2014).

Esta investigación se centra en enriquecer la literatura sobre el mercado mexicano y la influencia de las anomalías que ponen en entredicho la HME. Entre los cuestionamientos abordados están el de determinar si estas anomalías son específicas de ciertos mercados o si son fenómenos más amplios y comunes a todos los mercados. El análisis se enfoca en evaluar la influencia de estas anomalías en los rendimientos de los activos del mercado mexicano y si estos rendimientos presentan algún impacto, sea positivo o negativo, en comparación con los rendimientos de mercado, representado por el IPC mexicano. Los resultados se validan utilizando diversos modelos de aprendizaje automático con datos tanto dentro como fuera de la muestra. Es importante señalar que el propósito de este estudio no radica en desarrollar un modelo predictivo de precios o rendimientos, sino más bien explorar estas anomalías de manera independiente, evitando convertirse en un modelo de valoración de activos.

## ***Justificación***

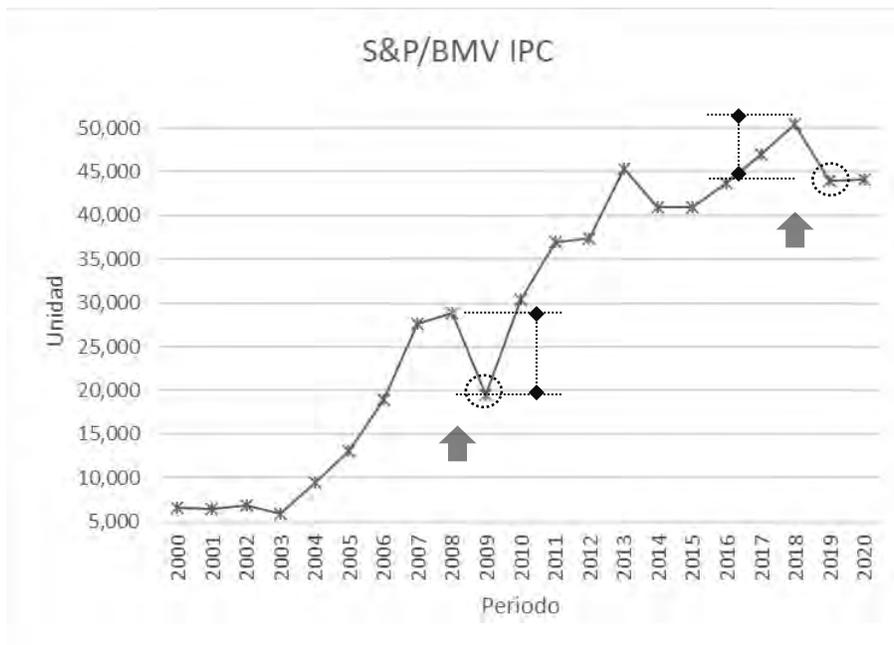
En las últimas décadas, numerosas investigaciones han revelado la existencia de irregularidades en los precios y rendimientos de activos en diversos mercados. Entre estas anomalías, las relacionadas con aspectos de calendario han sido objeto de un estudio más profundo y se han demostrado más robustas que otras. Esto ha sido especialmente evidente cuando se han investigado de forma individual.

En este estudio, se analiza el mercado mexicano para evaluar la presencia de anomalías específicas, incluyendo el *momentum*, la reversión a la media en los precios y la volatilidad, así como los efectos relacionados con el calendario, como el efecto enero, el efecto de fin de mes y el efecto de fin de semana. El enfoque se centra en contribuir a la literatura existente sobre anomalías de mercado al analizar el mercado.

Hasta el momento, se ha hablado de estas imperfecciones y asimetrías en los precios y rendimientos de los activos. Sin embargo, más allá de analizar su presencia, es importante discernir si estas anomalías pueden proporcionar evidencia que ayude a los diversos participantes del mercado en la toma de decisiones de inversión. Esto incluye a inversores minoristas, profesionales financieros reguladores entre otros.

En la figura 1 se muestra el comportamiento del IPC (Índice de Precios y Cotizaciones) de la Bolsa Mexicana de Valores en el periodo comprendido entre enero de 2000 y 2020. Se puede observar un crecimiento significativo, ya que en el año 2000 el IPC tenía un valor de aproximadamente 5,000 unidades, y a mediados de 2018, superó las 50,000 unidades, lo que representa un aumento de alrededor de diez veces su valor (un incremento del 1,000% en términos de beneficio simple en un período de inversión de 20 años). Lo más interesante de esta figura son los dos mercados declives que presenta, uno en 2007 y otro en 2020. En ambos casos, el IPC experimentó una caída de alrededor de 10,000 unidades, lo que representó aproximadamente una disminución del 30% y el 25% de su valor en cada período, respectivamente. Por ejemplo, y aunque no se refleja en la figura dado que esta muestra precios anuales, el IPC llegó a las 32,000 unidades en precios intradía, lo que equivale a una caída aproximada del 38%.

Figura 1. Precios anuales del IPC de México de 2000 a 2020.



Fuente: Elaboración propia.

Es importante destacar que, si bien una inversión en el periodo mencionado resultaría atractiva desde una perspectiva de largo plazo, los inversores que participaron en el mercado unos años antes de estas caídas experimentarían una considerable disminución de su capital, con importantes pérdidas. En cualquier caso, contar con herramientas adecuadas y probadas en el mercado mexicano puede ser beneficioso para enfrentar los movimientos del mercado, ya sea como una protección preventiva o, incluso, de manera predictiva. Esto es lo que se propone en este estudio.

La contribución de esta investigación se divide en dos rubros: en primer lugar, se relaciona con la estimación de la sensibilidad de las anomalías de manera individual como indicadores de rendimientos ajustados al riesgo de los activos; y, en segundo lugar, se identifican los periodos en los cuales los rendimientos promedio ajustados al riesgo son los más beneficiosos. El grado o nivel de confianza de la presencia de las anomalías es evaluado utilizando técnicas no lineales de aprendizaje automático: *regresión logística*, *perceptrón multicapa*, *máquinas de soporte vectorial* y *bosques aleatorios*. Estas son usadas como clasificadores y el porcentaje de exactitud de los modelos construidos es utilizado junto con

un criterio de categorización para indicar el grado de presencia de las anomalías. Así, las principales aportaciones son las siguientes:

- **Avance teórico.** Los criterios de construcción del modelo y las propuestas de cálculo de las anomalías, que en los estudios del estado del arte las confirman a partir de sus correlaciones, en este trabajo se realiza con algoritmos basado en los precios y sus rendimientos. Asimismo, el modelo desarrollado para la estimación de las anomalías es representado como clasificador el cual utiliza metodologías de aprendizaje automático a diferencia del análisis tradicional de series de tiempo con base estadística.
- **Contribución metodológica.** Se propone un criterio de categorización de la presencia y el procedimiento y flujo detallado de construcción de los modelos para el análisis de las anomalías que está basado en un sistema de reconocimiento de patrones de aprendizaje supervisado. Además, el código fuente que corresponde a la automatización de la batería de pruebas, que integra la plataforma proveedora de los algoritmos, permite su replicación a otros mercados y activos.
- **Relevancia práctica.** Se muestra evidencia empírica del análisis y conclusión de las anomalías, así como de los periodos para los que el rendimiento promedio ajustado al riesgo es mayor. Esto permite a los inversionistas tener idea del número de periodos que debería mantener una posición para obtener los mejores rendimientos de acuerdo a cada anomalía.

En resumen, este trabajo aporta significativamente en términos teóricos al proponer un enfoque para calcular y analizar anomalías en los mercados financieros. Además, ofrece una metodología sólida y la posibilidad de replicación en diferentes contextos, lo que hace que sus resultados tengan relevancia tanto para la comunidad académica como para los profesionales del mercado.

## ***Datos***

El enfoque de análisis utilizado para llevar a cabo este estudio comparativo implica el uso de datos de precios del IPC (Índice de Precios y Cotizaciones), que representa el mercado financiero mexicano. Se exploran dos casos distintos:

Caso 1: Se examinan las cotizaciones diarias, semanales y mensuales de dos acciones específicas que cotizan en la Bolsa Mexicana de Valores: Cemex y Peñoles.

Caso 2: Se construyen dos portafolios. El primer portafolio se forma utilizando las empresas con mayor bursatilidad de la Bolsa Mexicana de Valores, mientras que el segundo portafolio se compone de las empresas con menor bursatilidad en dicha bolsa.

Este análisis se realiza en un horizonte de tiempo que abarca 15 años, desde 2005 hasta 2020, considerando cotizaciones diarias, semanales y mensuales. En conjunto, estos casos y el periodo de estudio proporcionan una amplia perspectiva de la evolución de los precios y rendimientos en el mercado financiero mexicano, lo que permite evaluar con mayor profundidad la presencia de anomalías y su impacto en diferentes contextos y horizontes temporales.

Para el cálculo de los rendimientos se utilizan los precios al cierre en su forma continua obtenidos con la siguiente expresión:

$$R_{log} = \ln\left(\frac{V_t}{V_{t-1}}\right) \quad (1.1)$$

Donde:

- $R_{log}$  representa el rendimiento logarítmico o compuesto continuamente.
- $t$  es el periodo de tiempo.
- $V_t$  es el precio de cierre en tiempo o periodo  $t$ .

Se consideran en la muestra dos empresas tomadas de las 35 empresas que conforman el IPC, Cemex y Peñoles, que cumplen los siguientes criterios:

- Se han mantenido vigentes y cuya serie de observaciones histórica cubre los 15 años de estudio.

- Los precios están presentes en al menos el 98%, es decir, no existen faltantes, corresponden a las fechas observadas y tienen valor numérico.
- Existen observaciones para las series evaluadas: mensuales, semanales y diarias.
- Las empresas corresponden a las de mayor y menor bursatilidad.

Los datos son tomados de la base de datos dada por el servicio Infosel Financiero<sup>3</sup> (2020). Las variables dependientes están conformadas por segmentos de las series de rendimientos de los activos y una variable de *proxy* formada indicando la activación de la señal de la anomalía.

## ***Metodología***

El enfoque metodológico de esta investigación se sustenta en dos tipos de técnicas:

- **Técnica Documental:** Se lleva a cabo una recopilación exhaustiva de información relevante a partir de observaciones de las acciones que componen el IPC de México. Además, se analizan investigaciones previas y estudios relacionados que proporcionan un marco de trabajo. Esta técnica documental es esencial para la comprensión y contextualización de las anomalías estudiadas, tanto en el mercado mexicano como en otros mercados internacionales.
- **Técnica de Campo:** Una vez recopilada y revisada la información documental, se procede a la construcción de modelos destinados a probar la presencia de las anomalías. Estos modelos se aplican a los datos relativos a los precios de las acciones, y se utilizan criterios de selección para identificar aquellas acciones que exhiben una mayor acentuación de estas anomalías. Posteriormente, los modelos se someten a un análisis detallado basado en criterios específicos que permiten categorizar el grado de presencia de las anomalías en cada caso.

Esta combinación de técnicas documentales y de campo respalda la base teórica y empírica de la investigación, lo que contribuye a la comprensión de las anomalías en el mercado financiero mexicano y su posible impacto en la toma de decisiones de inversión.

---

<sup>3</sup> Es un servicio facilitado por la BMX con costo bajo esquema de suscripción (Infosel Financiero, 2020).

El enfoque metodológico de esta investigación se clasifica como correlacional, ya que su objetivo principal es establecer relaciones y dimensionar el comportamiento de los rendimientos de las acciones y del IPC en el mercado mexicano, particularmente a través del análisis de anomalías. Dentro de este contexto, el método de investigación se compone de tres enfoques interrelacionados:

- **Método Histórico:** El análisis se basa en la recopilación y evaluación de observaciones de precios y rendimientos que abarcan el periodo entre 2005 y 2020. Estos datos históricos se aplican al IPC, las acciones individuales (Cemex y Peñoles) y a los portafolios conformados por las empresas mexicanas con la mayor y menor bursatilidad en la Bolsa Mexicana de Valores. Este enfoque histórico proporciona una visión retrospectiva del comportamiento del mercado a lo largo de los últimos 15 años.
- **Método Deductivo:** Se utiliza el razonamiento deductivo al extrapolar y adaptar conceptos y hallazgos de estudios y análisis previos aplicados en otros mercados y condiciones distintas. La idea subyacente es aplicar el conocimiento existente sobre la identificación de anomalías en el mercado financiero a la realidad del mercado mexicano. Este método deduce una base de conocimiento que se adapta a las circunstancias específicas del mercado mexicano.
- **Método Sintético:** Los resultados obtenidos a partir del análisis permiten validar la presencia de anomalías en el mercado mexicano y evaluar su impacto en los rendimientos promedio ajustados. Esta síntesis de resultados confirma la existencia de ciertas anomalías y demuestra su relevancia como factores de influencia en las decisiones de inversión de los participantes del mercado. En esencia, se busca sintetizar el conocimiento obtenido en una comprensión más completa de las anomalías en el mercado financiero mexicano.

En conjunto, estos métodos histórico, deductivo y sintético proporcionan una sólida base para el análisis de las anomalías en el mercado mexicano y su potencial impacto en las decisiones de los distintos participantes del mercado.

### ***Estructura de la Investigación***

El trabajo de investigación inicia con una revisión de la literatura en la cual se presentan trabajos de relevancia alrededor de la teoría de mercados eficientes y sobre todo de sus anomalías. Se hace hincapié en las anomalías que analizan y se explica la metodología para su identificación. Posteriormente, se muestran los resultados y, finalmente, se concluye con el breve razonamiento motivo de los resultados presentados.

A continuación, se describe el detalle de los apartados:

- El primer apartado se conforma, por una parte, de la introducción de la investigación, esto es, el proyecto de investigación, y por otra del primer capítulo. Se detalla el planteamiento de la investigación, definiendo los antecedentes, el problema, los objetivos, las hipótesis y la justificación de la misma. Asimismo, integra el marco conceptual y el planteamiento teórico general, así como la consideración del mercado mexicano de valores.
- El segundo apartado está integrado por el segundo capítulo, referente a la teoría relacionada las técnicas de análisis de anomalías y de métodos con fundamento estadístico para estimación, típicos del área económico-financiera. En general, se muestra la compilación de información relevante para que el lector conozca los aspectos teóricos alrededor de los cuales se sustenta el aporte de la investigación.
- En el tercer apartado, que corresponde a los capítulos tercero y cuarto, se detalla la investigación: selección de datos, procesamiento, clasificación, criterios de cálculo y obtención de anomalías, elaboración de modelos para la estimación y su validación; así como el diseño del estudio comparativo para la estimación de la rentabilidad de los modelos. Es aquí donde se localiza el aporte y definición de la propuesta de investigación y análisis empírico. Asimismo, se exponen y discuten los resultados de dicho análisis.
- Finalmente, en el último apartado se presentan las conclusiones del estudio y el trabajo de investigación futuro.

## ***Resumen***

En este capítulo, se han presentado las bases fundamentales de la investigación. Se han enunciado las generalidades teóricas para destacar la oportunidad que dio origen a este trabajo. Además, se ha realizado una revisión de las investigaciones más relevantes a nivel internacional en relación con las anomalías de mercado que se desvían de la HME, las cuales, más allá de la discusión sobre el cumplimiento o no de la hipótesis, obtienen resultados relevantes, que abren nuevas puertas para la investigación. También se ha delineado la metodología utilizada en la investigación y se ha destacado la importancia y los beneficios que los resultados ofrecen al mercado financiero mexicano, enriqueciendo así la literatura existente y brindando oportunidades para investigaciones futuras.

El siguiente capítulo establecerá los fundamentos teóricos relacionados con los mercados financieros, centrándose en la HME y sus anomalías, con un enfoque especial en investigaciones previas sobre su aplicabilidad en el mercado financiero mexicano y los resultados obtenidos.

# Capítulo 1

## MARCO TEÓRICO

Este capítulo tiene como objetivo resumir las ideas clave y los análisis más destacados en relación con los estudios sobre la Hipótesis del Mercado Eficiente (HME) y las investigaciones sobre anomalías que han afectado a diversos mercados financieros. En primer lugar, se presentan las generalidades y se destacan los resultados de investigaciones realizadas en mercados de todo el mundo. Luego, se realiza una recopilación de investigaciones que se han centrado en el mercado financiero mexicano y en América Latina, así como en otros mercados similares en cuanto a desarrollo.

La HME ha sido un tema recurrente en la literatura financiera, y a pesar de las diversas corrientes que han intentado cuestionarla o proporcionar enfoques alternativos, la HME ha mantenido su posición predominante. Aunque se han señalado divergencias en sus postulados y condiciones subyacentes, hasta el momento, estas discrepancias no han sido lo suficientemente contundentes como para reemplazar o invalidar por completo la HME. Las investigaciones que han identificado asimetrías en varios activos y productos financieros no han logrado demostrar la consistencia de estas anomalías a lo largo del tiempo.

La HME sigue siendo un referente clave en las discusiones sobre los mercados financieros, especialmente en lo que respecta a la noción de que los rendimientos de los precios son independientes. A continuación, se describen algunos conceptos fundamentales que son relevantes para abordar este tema, seguidos por una revisión de la literatura que destaca investigaciones relacionadas con el análisis de rendimientos de precios.

## ***1.1 Hipótesis de los Mercados Eficientes, HME***

La HME es uno de los términos básicos de las finanzas corporativas y una de las teorías más utilizadas para estudiar el comportamiento de los precios de los mercados financieros. Tiene su origen en la teoría del paseo aleatorio y parte de que las cotizaciones bursátiles son totalmente impredecibles; de modo que cualquier intento de estimación futura está destinado al fracaso (Dimson y Marsh, 1984) (Gutiérrez, 2016).

La HME postula, de manera general, que los inversores, como personas racionales que son, valoran todos sus activos y títulos por su valor intrínseco u objetivo, en el sentido del precio de equilibrio. Sobre esta base, dicho valor intrínseco se verá modificado por la nueva información que llegue al mercado, que, por definición, será impredecible (Gutiérrez, 2016).

### ***1.1.1 Antecedentes***

Las bases financieras para la formulación de esta hipótesis, tal como las conocemos, derivan en gran parte de la tesis doctoral del matemático francés Louis Bachelier en 1900, titulada “Teoría de la Especulación”, que describe un modelo sobre el comportamiento aleatorio de los precios de las acciones de la bolsa de París (Samuelson, 1965) (Cheong y Isa, 2007) (Delcey, 2019). En ella, es introducido el concepto de “movimiento browniano” para describir el comportamiento de los precios en los mercados además de obtener la primera formulación acerca de las caminatas aleatorias. Entre las premisas básicas de dicho trabajo se mantiene que los cambios en los precios son independientes unos de otros y que pueden ser representados por una distribución Gaussiana, es decir, deduce que la esperanza matemática de las ganancias de un inversionista es cero (Campbell et al., 2012).

En (1965) Samuelson aporta un importante argumento a favor de la eficiencia de los mercados. Desde la perspectiva del proceso de información en un mercado eficiente, concluye que los precios deben ser impredecibles. En otras palabras, las tasas de retorno de los activos, denotadas como  $r_t$ , deben comportarse como un juego justo, lo que significa que la predicción del proceso estocástico  $r_t$  debe ser igual a cero para cualquier conjunto de información posible  $\Omega_t$  ( $E[r_t + 1 | \Omega_t] = 0$ ) (Campbell et al., 2012).

Esta definición implica que los retornos  $r_t$  se comportan como un juego justo cuando los precios de los activos siguen un patrón conocido como martingala, lo que se traduce en que las expectativas futuras de los precios, dados los conocimientos actuales, son iguales a los precios actuales ( $E[P_t + 1 \mid \Omega_t] = P_t$ ). En otras palabras, en un mercado eficiente, no existe una diferencia sistemática entre el rendimiento actual del juego (los precios) y el rendimiento esperado antes del juego (las expectativas).

Este concepto subraya la noción fundamental de que, en un mercado eficiente, la información disponible se refleja rápidamente en los precios de los activos, haciendo que sea difícil predecir futuros movimientos de precios con base en información pasada (Cheong y Isa, 2007).

Fama (1970) define un mercado eficiente como un mercado con alta proporción de individuos racionales que maximizan activamente los rendimientos de sus inversiones, compitiendo unos con otros y haciendo intentos para predecir los valores de mercado de precios de acciones, con la premisa de que toda la información posible está disponible para todos. En este sentido, en el caso de un evento impredecible, si un mercado es eficiente, los precios de las acciones inmediatamente reflejan el efecto del evento; y estos tomarán tiempo en ajustarse a la nueva información si el mercado es ineficiente. Por el contrario, si el evento es predecible, y antes que este pase, los precios de las acciones se incrementarán y rápidamente se ajustarán en la fecha del evento, si el mercado es eficiente; pero si el mercado es ineficiente, los precios se moverán muy lentamente en la dirección del ajuste. Esta definición se deriva de un modelo de equilibrio, como el Modelo de Fijación de Precios de Activos de Capital (CAPM por sus siglas en inglés), desarrollado por Sharpe (1964), Lintner (1965) y Mossin (1966).

Siguiendo la línea de pensamiento de Samuelson (1965), Fama (1970) expresa las condiciones de equilibrio del mercado en términos de tasas de retorno esperadas, las cuales se calculan en función del conjunto de información disponible  $\Omega_t$ . Fama identifica un mercado eficiente con la suposición de que las tasas de retorno de los activos, denotadas

como  $r_t$ , se comportan como un juego justo y hace referencia al comportamiento generalizado de esta conjetura a otros modelos, incluyendo los modelos de martingala y el modelo de camino aleatorio. Estos modelos fundamentales en la teoría de eficiencia de mercado ayudan a comprender cómo la información y las expectativas afectan la formación de precios en los mercados financieros. Jensen (1978) se declara a favor de la definición HME como uno de los paradigmas más importantes de la teoría financiera tradicional e indica que no hay teoría económica más sólida con mayor evidencia empírica.

Malkiel (2003) argumenta que los mercados financieros son en gran medida eficientes, lo que significa que la información pública se refleja rápidamente en los precios de los activos, y es difícil o imposible obtener consistentemente ganancias extraordinarias a través del análisis de valores o mediante una selección de activos. Uno de los conceptos clave de Malkiel es la noción del "paseo aleatorio", que sugiere que los precios de los activos financieros siguen un patrón aleatorio y que los intentos de predecir los movimientos futuros de los precios son esencialmente inútiles. Esto se alinea con la HME de Fama, que postula que es imposible superar consistentemente el mercado y obtener ganancias anormales debido a la eficiencia con la que los precios incorporan la información disponible. Malkiel también aboga por la inversión pasiva, como la inversión en fondos indexados que siguen un índice de mercado, en lugar de la inversión activa que implica la selección de activos individuales. Argumenta que, en promedio, los inversores activos no superan a los índices de mercado a largo plazo después de tener en cuenta las tarifas y los costos asociados.

La conclusión de Malkiel es que la eficiencia pudiera corresponder solo a un conjunto de información, por lo que quizá en ese sentido existirían datos dentro de un subconjunto u otro para el que no aplicara. Con esto respalda la noción de que los mercados son en gran medida eficientes y que los inversores pueden beneficiarse de estrategias de inversión pasiva y de largo plazo, en lugar de tratar de vencer al mercado a través de la selección de activos o del análisis técnico.

La información relevante incluye información pasada, información públicamente disponible e información privada. Sobre esta base un mercado eficiente puede ser de tres tipos: forma

débil, semi fuerte y fuerte (Roberts, 1967) (Fama, 1970). En su forma débil toda la información pasada, incluyendo precios y rendimientos pasados, ya está reflejada en los precios actuales de las acciones. Esta suposición es consistente con la hipótesis de caminata aleatoria que dice que los precios de las acciones se mueven aleatoriamente y que sus cambios son independientes unos de otros. Así que nadie puede predecir el futuro a partir de la información pasada. De ahí que no sea posible obtener rendimientos anormales bajo este esquema. Por lo tanto, el análisis técnico, que consiste en la clasificación de gráficas e identificación de figuras en los movimientos pasados de los precios para predecir sus cambios futuros, es inútil. Sin embargo, sería posible con análisis fundamental o con información privada.

En la forma semi fuerte los precios de las acciones reflejan toda la información pública disponible, así como la pasada. Así que nadie puede hacer ganancias adicionales sobre la base de análisis fundamental. Sin embargo, se podría obtener rendimientos anormales con información privada. En su forma fuerte, toda la información relevante, pasada, pública y privada, ya está reflejada en los precios actuales de las acciones. Así que, si esta forma persiste no es posible obtener rendimientos anormales de ninguna manera, ni siquiera manejando datos privados.

Según Fama (1970), las siguientes serían las condiciones suficientes, pero no necesarias, para la eficiencia de un mercado de capitales (Campbell et al., 2012):

- i. No deben existir costos de negociación de los valores.
- ii. Toda la información disponible debe ser disponible sin costo para todos los participantes del mercado.
- iii. Todos deben estar de acuerdo en las implicaciones de la actual información para el precio actual y en las distribuciones de los precios futuros de cada valor. Esto es, todos los participantes del mercado deben interpretar de igual forma la información, o, de otro modo, debe existir homogeneidad de expectativas.

Así mismo, Damodaran (2012) adiciona condiciones necesarias no sólo para eliminar la ineficiencia del mercado, sino también para hacer que este último opere como un aparato

eficiente informacionalmente; es decir, que dicho autor parte de una situación particular en la que asume que en el mercado existen ineficiencias, por lo cual, aporta las pautas requeridas para confeccionar la transición desde la ineficiencia a la eficiencia de los mercados financieros (Ramírez Morales et al., 2015):

- Las ineficiencias del mercado deben proveer las pistas o bases para armar un modelo que derrote al mercado y permita ganar excesos de retornos.
- El activo o los activos que originan la ineficiencia tienen que ser transados.
- Los costos de transacción por emplear el modelo que vence al mercado, tienen que ser menores que los beneficios generados por la utilización de dicho modelo.
- Debe haber inversores que buscan maximizar el beneficio que:
  - a. Reconozcan el potencial de los excesos de retornos.
  - b. Puedan replicar el modelo que vence al mercado, y que además permite generar excesos de rentabilidad.
  - c. Tengan los recursos para transar en valores financieros, hasta que la ineficiencia desaparezca.

Además, es pertinente agregar, que las decisiones del mercado en su conjunto van a determinar el precio que refleja la información disponible, y, por ende, dicho precio es la mejor estimación del valor de los activos intercambiados (Timmermann y Granger, 2004) (Ramírez Morales et al., 2015).

La eficiencia de mercado no requiere que el precio de mercado tenga que ser igual al valor correcto para cualquier momento del tiempo. Lo que requiere es que los errores en el precio de mercado sean imparciales, es decir, que los precios puedan ser mayores o menores que el valor verdadero siempre y cuando dichas desviaciones sean aleatorias (Ramírez Morales et al., 2015).

Un elemento importante de la HME es que los agentes del mercado estén constantemente interesados en buscar y analizar nueva información, para así crear valores esperados acerca

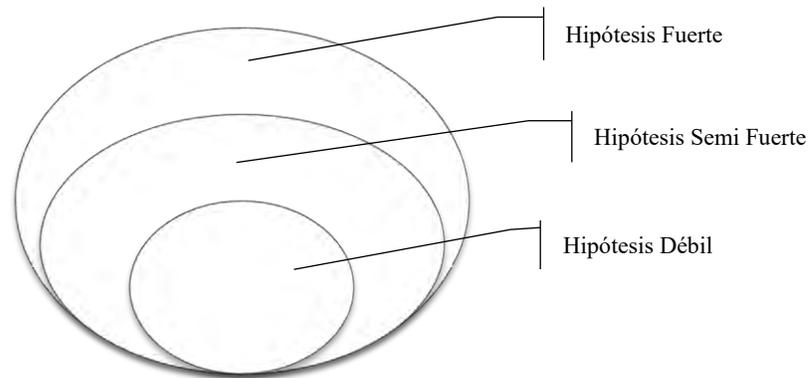
de los precios futuros y, que cuando se vean enfrentados a esta nueva información, vayan redefiniendo las expectativas.

Cuestionamientos preliminares sobre el comportamiento de los mercados en cuanto a la evolución de distintas variables económicas y la capacidad de anticipar los ciclos alcistas y bajistas son introducidos por Kendall (1953). En este trabajo se estudian patrones de comportamiento de las acciones con la premisa inicial de que los precios de éstas, para una determinada empresa, deben reflejar las perspectivas de beneficios, el impacto de los ciclos económicos y toda aquella información que pudiera ser relevante para dicha empresa. El autor concluye que los precios parecen seguir un cierto patrón aleatorio, que suben y bajan sin tener en cuenta ninguna pauta previa. De ahí que, no sean predecibles. Establece también la diferencia entre aleatorio e irracional y que el comportamiento aparente de los precios de las acciones significa que el mercado es eficiente, por lo tanto, que funciona correctamente y, por el contrario, que el mercado sea irracional significa que el mercado no es eficiente (Brun y Moreno, 2008).

La situación en la que los precios de una acción son consecuencia de toda la información existente en el mercado se encuentra en la HME (Fama, 1970) donde se plantea que los precios de un instrumento financiero reflejan ya toda la información disponible y relevante, lo que no permitiría obtener rendimientos de manera consistente por encima del promedio del mercado. Adicionalmente, reconsidera y defiende el entendimiento de los mercados eficientes en función de distintos aspectos, propuesto previamente por Roberts (1967) y que desprende tres niveles de eficiencia del mercado en función del nivel de información disponible y su impacto en las cotizaciones.

### 1.1.2 Niveles de eficiencia de los mercados

Figura 1.1. Niveles de eficiencia de los mercados.



Fuente: Elaboración propia.

#### *Hipótesis débil*

Los precios incorporan la información que se deriva de la evolución histórica de las cotizaciones y volúmenes de negociación. Por lo tanto, analizando las pautas seguidas por las cotizaciones en el pasado, no se puede adivinar ninguna regla que permita obtener beneficios extraordinarios (Brun y Moreno, 2008).

Esta hipótesis implica que las series históricas de las acciones son públicas y no tienen costo. Por lo tanto, todos los inversores tendrán los mismos datos. Si los analistas han aprendido a detectar comportamientos futuros de las acciones estudiando las series históricas, entonces cualquier signo que determine un comportamiento será inmediatamente detectado, y el precio de la acción aumentará o disminuirá automáticamente puesto que todos los participantes del mercado habrán explotado dichas señales ajustando así su valor (figura 1.1).

#### *Hipótesis semi-fuerte*

Los precios incorporan toda la información pública disponible. Es decir, los precios no incluyen sólo la información pasada (la que hace referencia a los volúmenes y precios) sino también la información actual pública, como los resultados obtenidos, los dividendos a pagar,

en resumen, la referente a sus fundamentales económicos (crecimiento en resultados, situación financiera, situación competitiva, etc.). Este es el concepto cercano al utilizado por el análisis fundamental (Brun y Moreno, 2008).

### *Hipótesis fuerte*

Los precios incorporan toda la información referente a una empresa, incluso la no pública o privilegiada. Esta hipótesis va más allá de las demás y asume que existen inversores que utilizan información privada de su empresa para obtener un beneficio propio adicional (Brun y Moreno, 2008).

Entonces, un mercado es eficiente debido a la competencia entre los distintos participantes del mercado que, guiados por el principio de maximización del beneficio, buscarán oportunidades de arbitraje en los mercados, lo que conducirá a una situación de equilibrio en la que los precios de las acciones en el mercado es la mejor referencia de su valor intrínseco.

De acuerdo con Lo (Lo, 2005) (Hull y McGroarty, 2014), el concepto de la HME puede percibirse como: entre mayor eficiencia en el mercado, mayor aleatoriedad debe haber en la secuencia de cambios en los precios, por lo que el mercado más eficiente es uno en el cual, el cambio en los precios sea completamente aleatorio e impredecible. Esto no es naturalmente accidental, sino que es el resultado directo de muchos participantes activos del mercado que tratan de sacar provecho de su información.

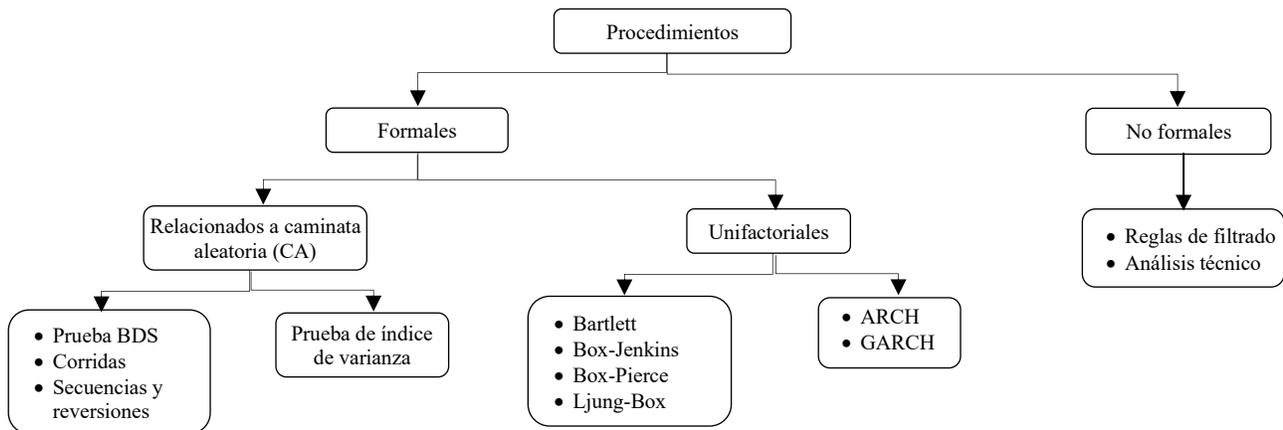
Los métodos tradicionales utilizados para contrastar la forma débil de eficiencia pueden ser clasificados dentro de dos principales grupos, dependiendo del nivel de uso de inferencia estadística como se muestra en la figura 1.2 (Roldán y García, 2022).

La literatura financiera muestra que los métodos descritos anteriormente tradicionalmente se han aplicado para probar la hipótesis de eficiencia débil en los mercados financieros.

Las pruebas de correlación para determinar la eficiencia de un mercado se utilizaron por primera vez cuando Fama (1965) y Samuelson (1965) sentaron las bases de la teoría de

mercado eficiente. A partir de estos comienzos, destacan los trabajos desarrollados por Moore (1964), Theil y Leenders (1965), Fama (1965) y Fisher (1966).

Figura 1.2. Métodos tradicionales para probar la forma débil de eficiencia de mercado.



Fuente: Elaboración propia con datos de (Roldán y García, 2022).

Estas pruebas se usaron como la única herramienta para analizar la eficiencia de un mercado hasta que, en la década de 1970, los efectos estacionales y las anomalías del calendario se volvieron relevantes para el análisis. Entonces surgieron nuevas metodologías que incorporaban estos efectos, como las pruebas de estacionalidad aplicadas por Roseff y Kinney (1976), French (1980) y Gultekin y Gultekin (1983).

En la década de 1990, comienzan a aparecer estudios que analizan la HME utilizando métodos tradicionales. Esta práctica ha continuado hasta el día de hoy, como lo demuestran los trabajos empíricos más destacados sobre eficiencia financiera en los últimos años (Gimba, 2012) (Akwolaga, 2020) (Gajapathy y Sathyanarayana, 2021) (Roldán y García, 2022).

Por su parte, el análisis técnico como herramienta para probar la HME ha sido también utilizada para incorporar factores creados a partir de las observaciones y conjeturas realizadas sobre los gráficos de precios de los activos. Entre algunos estudios se encuentra Potvin et al.

(2004), Marshall et al. (2006), Chen et al. (2009), Alexeev y Tapon (2011), Shynkevich (2012), Ho et al. (2012), Leković (2018), Picasso et al. (2019) y Nti et al. (2020).

En resumen, en la mayoría de los estudios empíricos existentes los reportes de los resultados coinciden que no se puede rechazar por completo la existencia en los mercados de la HME en su forma débil. Si bien, varios autores muestran que los mercados tienen cierto nivel de ineficiencia, aunque las pruebas no son contundentes para poder establecer predictibilidad en los movimientos de los precios, y otros muestran evidencia de rechazo de la hipótesis para ciertos mercados, especialmente para aquellos en desarrollo.

Desde que Fama (1970) formaliza las discusiones sobre la caminata aleatoria y establece la versión de la HME, los fundamentos de la misma han sido motivo de estudio, análisis y cuestionamiento en innumerables investigaciones desde su aparición, entre ellas Brock, Lakonishok y LeBaron (1992); Carhart (1997); Bhattacharya, Daouk, Jorgenson y Kehr (2000); Frankfurter y McGoun (2001); Green, Hand y Zhang (2012); Semaan, Harb y Kassem (2014); Harvey, Liu y Zhu (2016) y Rossi y Gunardi (2018). Esto da lugar al concepto de “anomalía” y su posterior estudio.

## ***1.2 Anomalías de Mercado***

El significado literal de una “anomalía” es ser una ocurrencia inusual. La palabra anomalía es definida por Frankfurter y McGoun (2001) como una irregularidad o desviación del orden común o natural o una condición excepcional. “Anomalía” es un término que es genérico por naturaleza y aplica a cualquier fenómeno fundamentalmente novedoso en hechos, nuevo e inesperado o a una sorpresa con relación a cualquier teoría, modelo o hipótesis.

Las anomalías son indicadores de ineficiencias de mercado, algunas pasan una única vez y desaparecen mientras que otras lo hacen frecuentemente o periódicamente (Tversky y Thaler, 1990). Kahneman y Tversky (1984) definen “anomalía de mercado” como una desviación de los paradigmas en el presente aceptados, que es tan evidente como para no ser ignorada, tan

sistemática como para no ser rechazada como un error aleatorio, y tan fundamental como para necesitar ser acomodada ajustando el sistema normativo (1984, p. 257).

En la teoría financiera estándar, una anomalía del mercado financiero significa una situación en la cual un desempeño de una acción o un grupo de ellas se desvían de los supuestos de la hipótesis de los mercados eficientes (Frankfurter y McGoun, 2001). Dichos movimientos o eventos que no pueden ser explicados utilizando la hipótesis de los mercados eficientes son llamados “anomalías de los mercados financieros”.

Harvey, Liu y Zhu (2016) documentan alrededor de 400 diferentes anomalías sobre los rendimientos de acciones en el mercado de los Estados Unidos. En general, la mayoría de los trabajos que documentan algún tipo de anomalía se enfocan en aquellas que pueden ser evaluadas localmente, esperando que la relación de interés se mantenga vigente y con cierto grado de predicción. Sin embargo, McLean, Pontiff y Pontiff (2016) muestran resultados bajo los cuales las ocurrencias de algunas anomalías son reducidas o dejan de presentarse una vez que son compartidas en la literatura. Estos autores estudian comparaciones de rendimientos con datos dentro de la muestra, posteriores a la muestra y posteriores a su publicación, reportando que, en promedio, los rendimientos son reducidos en un 56%, el 25% por estimación estadísticamente tendenciosa de los inversionistas y el 31% por el efecto de la publicación.

A continuación, se presentan algunas de las anomalías por categoría (Latif et al., 2011):

### ***1.2.1 Anomalías de calendario***

Éstas anomalías están relacionadas con comportamientos recurrentes en ciertos periodos de tiempo y estación.

Tabla 1.1. Anomalías de calendario.

ANOMALÍA	DESCRIPCIÓN
Efecto de fin de semana	Los precios de las acciones con probabilidad de caída los lunes, significando que el precio de apertura de la semana es menor que el del cierre en viernes anterior.
Efecto de fin de mes	Los precios de las acciones tienen probabilidad de incremento al cierre del mes y/o primeros días del mes siguiente.
Efecto enero	Describe incremento de los precios de las acciones y el volumen de estas en la última semana del mes de diciembre y mitad de enero.

Fuente: Elaboración propia.

### 1.2.2 Anomalías fundamentales

Estas anomalías se relacionan con conductas propias de los factores y características internas de las empresas. Éstas incluyen indicadores unitarios y combinaciones lineales entre estos.

Tabla 1.2. Anomalías fundamentales.

ANOMALÍA	DESCRIPCIÓN
Anomalía de valor	La anomalía de valor ocurre debido a una falsa predicción de los inversionistas. Estos sobre estiman la ganancia y crecimiento futuro de las compañías o subestiman los rendimientos y ganancias futuras de valor de estas.
Precio bajo en libro	La acción con razón baja de precio de mercado sobre precio en libros genera más rendimiento que las acciones con razones altas.
Producto de dividendos altos	Las acciones con alto dividendo superan el mercado y generan mayor rendimiento.
Precio bajo en ganancias (P/E)	Las empresas con razón P/E baja son propensas a que generen mayores rendimientos y superen al mercado mientras que las de una razón P/E alta tiendan a hacerlo menor.

Fuente: Elaboración propia.

### 1.2.3 Anomalías técnicas

Estas anomalías incluyen aquéllas que producen ganancias anormales en el mercado de valores cuando se emplea cualquier técnica de análisis de precios utilizadas para pronosticar precios futuros de las acciones sobre la base de precios pasados e información pasada relevante. El análisis técnico comúnmente usa juegos y combinaciones lineales de origen estadístico, figuras y formaciones en los precios.

Tabla 1.3. Anomalías técnicas.

ANOMALÍA	DESCRIPCIÓN
Promedios móviles	Técnica de análisis en la cual las señales de compra o de venta de acciones son generadas por periodos cortos y largos de promedios de precios. Específicamente, éstas son señaladas por los cruces de las líneas.
Rompimiento de rangos	Esta técnica de análisis está basada en los niveles llamados “de resistencia” y “de soporte”, donde “descansan” y “se topan” los precios. A pesar de que esta técnica es bastante intuitiva su implementación como estrategia sistematizada es difícil de lograr.
<i>Momentum</i>	Esta anomalía se refiere a que activos con precios o rendimientos positivos en el pasado continuarán siendo positivos en el futuro. Hace alusión al movimiento en física que supone su fuerza mantenida en la misma dirección
Reversión a la media	Supone que después de movimientos largos en la misma dirección le suceden movimientos cortos en contra, para después continuar en la misma dirección anterior (muy relacionado con el <i>momentum</i> ), o movimientos largos mantenidos, que cambian totalmente su dirección (se le conoce como <i>contrarian strategy</i> ). Esta anomalía puede aplicarse a precios, rendimientos o a la volatilidad reflejada por la desviación estándar.
Volatilidad	Se relaciona también con cambios en la varianza de los precios, ya si esta se mantiene dentro de rangos con umbrales específicos o si rompe esos umbrales.

Fuente: Elaboración propia.

### ***1.2.4 Anomalías como factores financieros***

Reconsiderando el antecedente de la HME y sus anomalías, Uribe y Ulloa (2011) interpretan la definición de HME como:

*“Aunque la información que ocurre en los mercados financieros es extensiva, esta es absorbida y reflejada en todo momento por los precios de mercado de los activos. La razón de que futura información no sea reflejada aun en los precios del mercado es simplemente porque es aún desconocida por los participantes y su ocurrencia depende de los procesos aleatorios. De ahí que, si los mercados financieros son eficientes es imposible sistemáticamente obtener ganancias excesivas a partir de la predicción de los precios de mercado, sin importar el tipo de información o técnica de predicción utilizada”.*

De acuerdo a Lo (2005) una anomalía puede ser definida como un patrón regular en los rendimientos de un activo el cual es consistente, ampliamente conocido e inexplicable. Como se ha comentado hasta ahora existen numerosas anomalías o factores financieros reportados en la literatura, y cuyo interés aun es vigente no sólo para los investigadores, sino también para los diversos participantes de los mercados. Por lo que encontrar medios que permitan aproximar mejor las estimaciones de los rendimientos futuros de los precios de los activos sigue siendo un área de investigación que se aborda desde distintas aristas.

A continuación, se detallan las anomalías que consideramos más populares, y que por ello son usadas como factores en los modelos analizados en este documento.

#### ***1.2.4.a Momentum***

Es la anomalía en la que el efecto de los rendimientos que presentan tendencia positiva o negativa en el pasado persiste en el futuro durante un largo periodo. Jegadeesh y Titman (1993) documentan un análisis de estrategias evidenciando que la compra de activos con desempeño bueno o positivo en el pasado y la venta de los de peor desempeño presentan rendimiento significativo manteniendo las posiciones de 3 a 12 periodos.

Se atribuye este comportamiento a una sobre reacción de los participantes en el mercado, exhibida en los precios, o por posible atraso de la reacción de los precios ante factores comunes del mercado (Chan et al., 1996). Esta anomalía encuentra relación con la idea general de que a largos intervalos en una dirección, ya sea precios o rendimientos, le siguen intervalos cortos en dirección opuesta para posteriormente continuar con el movimiento en la misma dirección, del intervalo largo inicial (Scholes y Williams, 1977) (Hong, 2014).

#### ***1.2.4.b Reversión a la media***

Esta anomalía se refiere a la conducta de los precios de los activos en la que estos tienden a retornar a la línea de tendencia. Está muy relacionada a la anomalía de *momentum*, en la que, como se ha indicado, se mantiene que los individuos sobre reaccionan a la información; esta anomalía sería, pues, una extensión directa de la anterior, en la que se sugiere que los precios también sobre reaccionan a la información comportándose de manera contraria, es decir, que la compra de los activos perdedores y la venta de los ganadores en el pasado logran obtener rendimientos anormales (De Bondt y Thaler, 1985) (Fama y French, 1992) (Fama y French, 1996) (Balvers y Wu, 2006). Hay evidencia que seguido de movimientos largos se presentan retornos en los precios con duración corta (Jegadeesh N. , 1990) (Lehmann, 1990), y se ha encontrado correlación negativa para ciertos horizontes la cual es normalmente referida como reversión a la media (Cutler et al., 1991). Esta no es contundente y no hay una explicación clara de esos retornos excesivos, aunque se le atribuye a la falta de liquidez del mercado, desconfianza de los inversionistas para continuar en el mismo sentido o únicamente por retraso en los efectos de información en los precios. Wu y Balvers (2000) sugieren que pudiera ser únicamente producto de la minería de datos más que algo relacionado a la teoría o con soporte fundamental. Este tipo de anomalía puede ser calculada sobre los precios, los rendimientos o en la volatilidad, expresada por la desviación estándar (Kuhe y Audu, 2019).

#### ***1.2.4.c Volatilidad***

En cuanto a esta anomalía se encuentran distintos conceptos, y cada uno con su forma de cálculo que bien podrían ser variantes o distintas anomalías bajo la categoría “Volatilidad”.

Se presenta como “anomalía de baja volatilidad” y se refiere a que los activos o portafolios con baja volatilidad de precios o rendimientos tienden a producir más utilidad en términos de rendimientos que aquellos que presentan alta volatilidad. También se identifica como “volatilidad idiosincrática” o “sesgada” ya que se tiene la creencia de que a periodos largos de muy baja volatilidad, con movimientos de precios casi totalmente horizontales, le siguen movimientos radicales y prolongados. Los inversionistas tienen a preferir este tipo de activos, al que también se le conoce como “tipo lotería”, dado que hay poca posibilidad de multiplicar sustanciosamente el valor en el corto plazo y aún existe mayor posibilidad de que su valor disminuya en el largo plazo (Baker et al., 2011), aunque la evidencia empírica apunta a que posiblemente este riesgo es el que genera los rendimientos excesivos (Barberis y Huang, 2009).

#### ***1.2.4.d Efecto fin de semana***

Esta anomalía se refiere a la baja de los rendimientos al acercarse el fin de semana, alcanzando los mínimos el viernes, para incrementar o crecer a inicios de ésta, en especial el lunes. Tiene presencia cuando llegando fin de semana los rendimientos promedio son consistentemente más bajos y la volatilidad es sistemáticamente más grande que los otros días (Cross, 1973) (French, 1980) (Gibbons y Hess, 1981) (Lakonishok y Levi, 1982). Por una parte, se sugiere que esta conducta en los precios o sus rendimientos es ocasionada por la información negativa que las compañías tienden a difundir al final de la semana o durante el fin de semana. Por otra parte, esta conducta se atribuye simplemente a que los inversionistas cierran sus posiciones para tener tranquilidad y poder disfrutar su fin de semana.

#### ***1.2.4.e Efecto fin de mes***

También llamada de “cambio de mes”, se refiere a los rendimientos a la baja al final de mes y su incremento durante los primeros días del siguiente. En los primeros días del mes los rendimientos son superiores al resto (Ariel, 1987). Se reporta esta conducta como motivo de que las compañías normalmente declaran ganancias o dan a conocer nuevos reportes durante

el primer día del mes o como resultado de nuevas posiciones de inversionistas con poder adquisitivo debido a la paga de salarios y estímulos.

#### ***1.2.4.f Efecto enero***

Conocido también como “efecto fin de año”, se relaciona con rendimientos superiores a los del resto de los meses. En especial, el mes de enero se ha reportado con rendimientos mayores. Ritter (1988) define el efecto enero como el fenómeno de retorno donde las compañías tienen rendimientos anormales durante el periodo que comienza en el último día de diciembre y continúa durante el siguiente mes de enero. La explicación de este efecto se relaciona a una supuesta estrategia fiscal de parte de los inversionistas en la que se asume que venden sus activos con pérdida al final del año para minimizar impuestos; retomando posiciones a inicios del año siguiente (Roseff y Kinney, 1976). También se le atribuye este efecto a la presencia de liquidez en el mercado que caracteriza el cierre de año por concepto de pago de primas, aguinaldos y bonos.

### ***1.3 El Mercado mexicano***

La Bolsa de Valores de México es la principal bolsa de valores de México y la segunda mayor de América Latina, después de Brasil, con un total de capitalización de mercado de 530 miles de millones de dólares. En el año 2020, debido al confinamiento poblacional y la escasa actividad económica por la pandemia del Covid-19, el PIB<sup>4</sup> mexicano tuvo una contracción de 8.5%. Durante este periodo el IPC alcanzó 44,066 unidades al cierre de año y la acción de la BMV, BOLSAA, tuvo un cierre de \$47.66 pesos por acción. En ese mismo año el tipo de cambio FIX<sup>5</sup> alcanzó un máximo de \$25.12 pesos por dólar y cerró con \$19.95 el año. La

---

<sup>4</sup> PIB, de Producto Interno Bruto.

<sup>5</sup> Por sus siglas en inglés de *Fixed Exchange Rate*, y significa tipo de cambio “fijo” o “fijado”. El tipo de cambio (FIX) es determinado por el Banco de México con base en un promedio de cotizaciones del mercado de cambios al mayoreo para operaciones liquidables el segundo día hábil bancario siguiente, que son obtenidas de plataformas de transacción cambiaria y otros medios electrónicos con representatividad en el mercado de cambios. El Banco de México da a conocer el FIX a partir de las 12:00 horas de todos los días hábiles bancarios, se publica en el Diario Oficial de la Federación (DOF) un día hábil bancario después de la fecha de determinación y es utilizado para solventar obligaciones denominadas en dólares liquidables en la República Mexicana al día siguiente de la publicación en el DOF. Por tanto, el tipo de cambio FIX es usado como una referencia fundamental de la relación entre el peso mexicano y el dólar estadounidense.

tasa de inflación se informó de 3.15% y se fijó una tasa objetivo de interés en 4.35% (Grupo BMV, 2020).

Al cierre del 2020 la BMV se conforma de 140 emisoras mexicanas de acciones y 5 extranjeras. El principal índice, como ya se ha indicado, es el Índice de Precios y Cotizaciones (IPC), que está compuesto por las cotizaciones de las acciones de las 35 empresas con mayor bursatilidad y cuyo valor de capitalización representa el 81% del total del mercado accionario. En ese mismo año, y en cuanto a las iniciativas de desarrollo de mercados sostenibles, se realizaron 5 listados de bonos temáticos, 4 bonos sustentables y un bono verde, representando un monto de 17,000 millones de pesos. Con 333 valores extranjeros listados en el SIC<sup>6</sup>, de los cuales 221 son empresas y 112 son ETF<sup>7</sup>, suman 2,576 instrumentos extranjeros listados en el SIC. El mercado accionario tuvo un volumen de negociación de 77,431 millones de acciones con un valor de \$4,044,572 millones de pesos, lo que es un incremento del 22% en importe y del 29% en volumen. El valor de capitalización total de las emisoras nacionales equivale al 46.9% del PIB, con el IPC representando el 80.96% del valor total (Grupo BMV, 2020).

Lo siguientes son los tipos de valores operados:

- Acciones:  
Acciones comunes, acciones de voto limitado, acciones preferentes, certificados de participación ordinarios sobre acciones, obligaciones convertibles y warrants domésticos.
- Instrumentos de deuda domésticos.

---

<sup>6</sup> El Sistema Internacional de Cotizaciones (SIC) es una plataforma donde se pueden negociar acciones o fondos cotizados (ETF) listados en otras partes del mundo.

<sup>7</sup> Los ETF, de las siglas de *Exchange Traded Funds*, es un fondo de índice cotizado. Básicamente es un fondo de inversión cuya compra de participaciones supone la adquisición de un cesto de valores, aunque tal cesto se conforme de un único activo subyacente, pero que cotiza en bolsa como si fuera una acción y puede ser comercializado de la misma forma.

Instrumentos de deuda de corto y largo plazo emitidos por entidades gubernamentales mexicanas, por compañías privadas y fideicomisos constituidos en México. La negociación secundaria se suscita en plataformas OTC<sup>8</sup>.

- TRACs.

Emisiones de fideicomisos privados constituidos en México, varios de los cuales siguen índices locales accionarios.

- CKDs.

Títulos fiduciarios lanzados en 2009 dedicados a la inversión en acciones, partes sectoriales o el financiamiento de sociedades mexicanas, directa o indirectamente, a través de vehículos de inversión. Están diseñados para impulsar proyectos de infraestructura, inmobiliarios, minería, empresas en general y proyectos para desarrollo de tecnología. Los flujos a recibir son variables y dependen de los resultados de cada proyecto.

- FIBRAS.

Lanzadas en 2010 son valores emitidos por fideicomisos dedicados a la adquisición o desarrollo de bienes inmuebles que se destinen al arrendamiento o a la adquisición del derecho a percibir ingresos provenientes del arrendamiento de tales bienes, así como otorgar financiamiento para esos fines.

- FIBRAS E.

Certificados bursátiles para la inversión en sociedades, proyectos o activos de energía o infraestructura maduros, en donde ya existían flujos probados. El instrumento puede ser adquirido por el gran público inversionista por lo que se incorporan requisitos de revelación y gobierno corporativo.

- CERPIs.

Certificados bursátiles que se podrán emitir mediante oferta pública restringida, dirigidos exclusivamente a inversionistas institucionales y calificados. A través de estos instrumentos, se promoverá el financiamiento de proyectos en etapa de desarrollo, bajo marco normativo flexible que mejore su administración.

---

<sup>8</sup> Por sus siglas *Over The Counter*, son un tipo de contrato que se negocia entre particulares y en el que los términos y restricciones son acordadas entre ambas partes. Estos mercados son más riesgosos que los mercados tradicionales dado que tienen riesgo de contrapartida, además que no hay una cámara de compensación.

- Títulos opcionales o *warrants*.

Documentos que otorgan a sus tenedores, a cambio del pago de una prima de emisión, el derecho de comprar o de vender al emisor un determinado número de acciones a las que se encuentren referidos o bien, a un grupo o canasta de acciones, o bien de recibir del emisor una determinada suma de dinero resultante de la variación de un índice de precios a un cierto precio (precio de ejercicio) y durante un periodo y fecha determinada.

- SPAC.

Lanzado en 2017, es una oferta pública inicial de acciones, con la característica de que los recursos que se obtengan se destinan solamente a la compra de una o varias compañías. El dinero que se levanta de mercado se coloca en una cuenta en custodia en el exterior para transparentar que el uso del recurso se destinará a las compras; en tanto no se use, el dinero se invierte en valores gubernamentales de Estados Unidos.

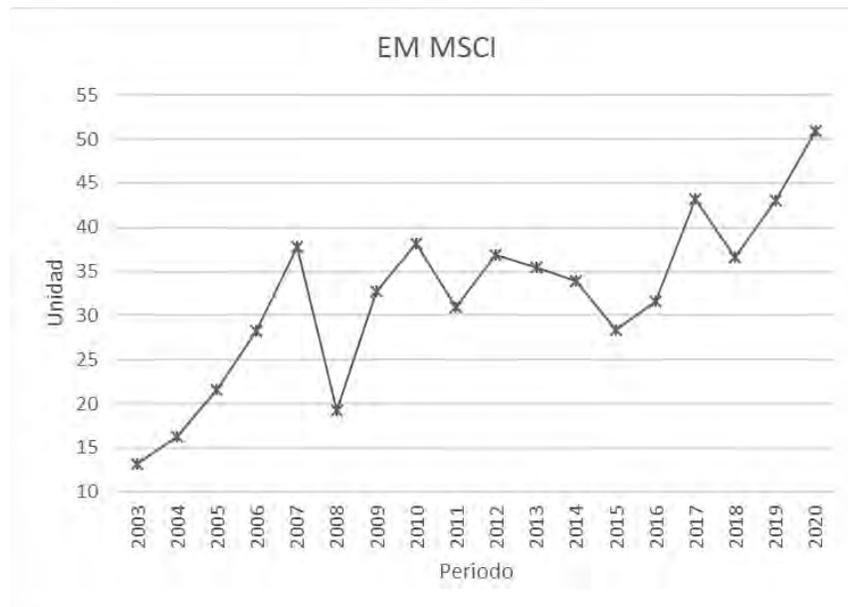
El incremento notorio de flujos de capital hacia los mercados emergentes, como el mexicano, en las últimas décadas, los hacen ser una gran opción para los inversionistas y creadores de portafolios. Como ejemplo del crecimiento de este tipo de economías, en la figura 1.3 se muestra el comportamiento del índice EM MSCI<sup>9</sup> el cual replica un portafolio de economías en desarrollo, que incluye la de México, y se conforma por empresas de diferentes sectores, capitalización y bursatilidad.

Sin embargo, a pesar del gran interés en este tipo de economías, poco es conocido acerca de los procesos de determinación de precios en estos mercados y cómo estos son comparables con aquellos de mercados más minuciosamente estudiados en economías desarrolladas, con los que existen diferencias de rendimientos y volatilidad (Harvey, 1995).

---

<sup>9</sup> El índice *MSCI Emerging Markets*, de las siglas en inglés de Morgan Stanley Capital International, es un índice utilizado para medir el desempeño global de los mercados emergentes. Fue creado en 1988 y está conformado por 24 economías en desarrollo incluyendo la de México. El índice es de capitalización de mercado de ajuste flotante y representa el 13% de la capitalización del mercado (Morgan Stanley, 2021).

Figura 1.3. Índice EM MSCI (*Morgan Stanley, 2021*) que sigue el comportamiento de los mercados emergentes más importantes de 2003 a 2020.



Fuente: Elaboración propia.

La búsqueda de anomalías de mercado es también motivada por la preocupación de que ciertas características institucionales puedan inducir la conducta de los rendimientos que se desvían de la conducta esperada (por ejemplo, caminata aleatoria u otra forma de martingalas) (Harvey, 1995), por ejemplo, cambios en políticas económicas para promover interés en los diferentes participantes del mercado conducen a reducción de la presencia de anomalías, como es deseable, no pueden ser resultados aislados.

Las anomalías no necesariamente indican ineficiencias de mercado, sino que pudieran simplemente reflejar ciertas mejoras institucionales de los mercados (o ciertos factores de riesgo para los cuales el modelo de precio de activos utilizado pudiera no ser correcto), por ejemplo, una reestructuración de impuestos pudiera provocar una corrección de efecto de impuestos anual (Harvey, 1995).

### ***1.3.1 Historia de la BMV***

La bolsa mexicana de valores es una empresa privada del sector financiero y bursátil cuya operación está bajo la concesión de la Secretaría de Hacienda y Crédito Público, SHCP. Fue creada en la Ciudad de México hace 125 años. En 1850, dado el éxito de las empresas mineras las negociaciones de compra y venta de acciones motivan el surgimiento de recintos privados de operación, con lo que en 1867 se promulga la Ley Reglamentaria de corretaje de valores. Sin embargo, no es hasta 1895 que aparece formalmente el primer mercado bursátil organizado del país y se inaugura La Bolsa de México. Entre los primeros títulos negociados y que instauraron las bases del mercado de valores fueron las empresas cerveceras.

Aunque la BMV sufrió inicialmente interrupciones debido a la revolución mexicana y a la primera guerra mundial, sus operaciones han continuado de manera constante, ofreciendo un recinto de intercambio y proveyendo de las facilidades para el comercio de valores. El sistema de valores y financiero mexicano comenzó a tomar forma cuando en 1933, en uno de sus momentos más significativos, se promulga la Ley Reglamentaria de Bolsas y se constituye la Bolsa de Valores de México bajo la supervisión de la hoy Comisión Nacional Bancaria y de Valores, CNBV.

En 1975 la bolsa cambia su denominación a Bolsa Mexicana de Valores, BMV, con la entrada en vigor de la Ley del Mercado de Valores, y consolida actividades integrando las bolsas que operaban en Guadalajara y Monterrey. La BMV se da a la tarea de enlazar el mercado de valores con el crecimiento económico de México, propiciando así que las empresas financiaran su desarrollo mediante la emisión de acciones.

En 1978 el gobierno creó el Instituto para el Depósito de Valores (Indeval) y en 1980 surge la Asociación Mexicana de Instituciones Bursátiles, AMIB. En 1990 las autoridades reinstituyen lo que llaman “régimen mixto”, estableciendo las bases para la reprivatización de los bancos y la formación de grupos financieros bajo la Ley de Grupos Financieros. Actualizaciones posteriores motivan la internacionalización de la actividad de valores nacionales, liberación de la comercialización para el inicio de la auto regulación, simplificación de procedimientos y la formación de grupos financieros no bancarios.

En 1996, por primera vez los valores listados para comercialización son negociados a través del sistema electrónico BVM-SENTRA Capitales. En 1997, la CNBV emite las reglas para la creación de un mercado de futuros y opciones. Por último, en 1997 el gobierno federal mexicano comienza la emisión de instrumentos de deuda para su comercialización y el Sistema Internacional de Cotizaciones, SIC, comienza sus operaciones.

La jornada más activa en la historia de la BMV se registró el 17 de mayo de 2001, estableciendo niveles récord de operatividad, ascendiendo el número de operaciones a 11,031, cifra superior en 516 operaciones (+4.91%) respecto al nivel máximo anterior registrado el día tres de marzo del 2000.

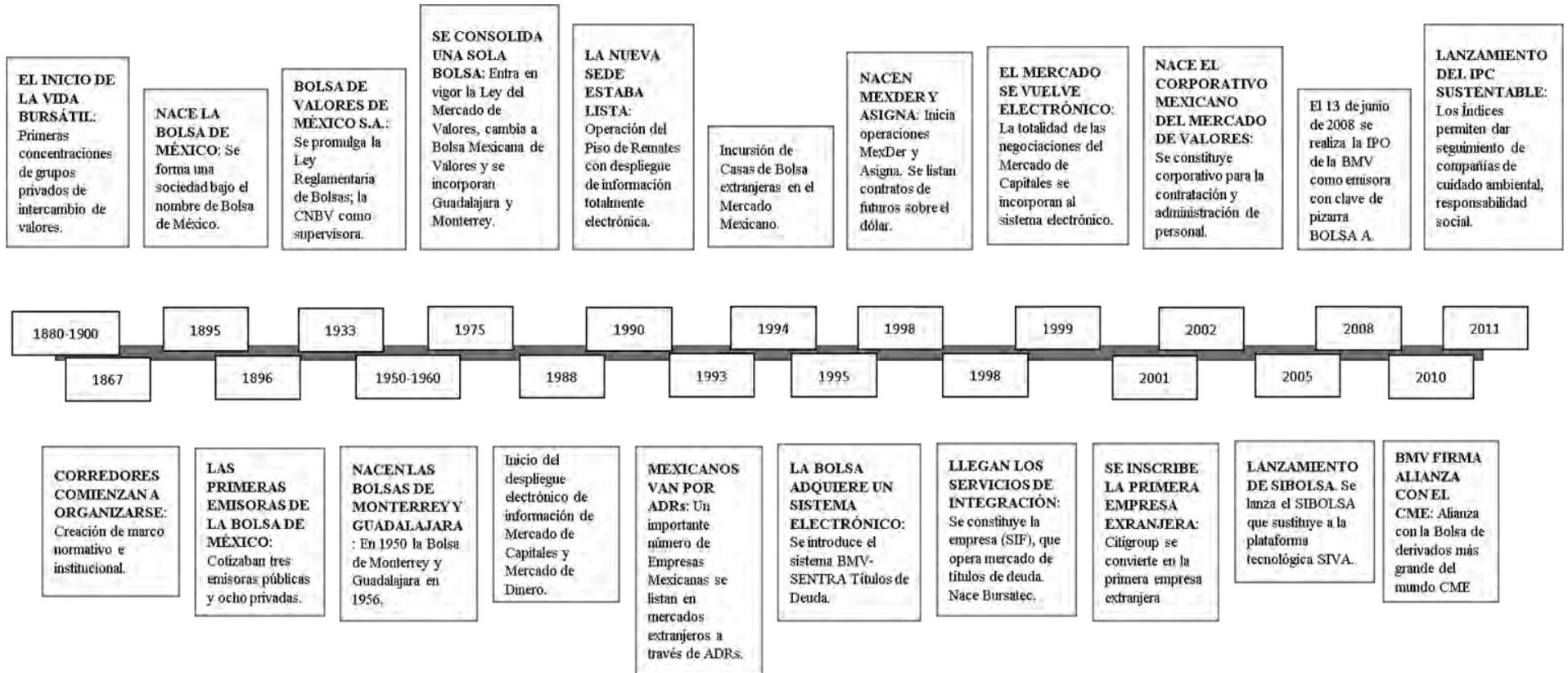
En 2002 se constituye la empresa de servicios Corporativo Mexicano del Mercado de Valores, S.A. de C.V. para la contratación, administración y control del personal de la Bolsa y de las demás instituciones financieras del Centro Bursátil que se sumaron a este proceso, y en 2008 la Bolsa Mexicana se convierte en empresa cuyas acciones son susceptibles de negociarse en el mercado de valores bursátil, llevando a cabo el 13 de junio de 2008 la Oferta Pública Inicial de sus acciones representativas de su capital social con clave de pizarra: BOLSAA y razón social: Bolsa Mexicana de Valores, S.A.B. de C.V. (figura 1.4).

### ***1.3.2 Índice de Precios y Cotizaciones, IPC***

En noviembre de 1978 se comienza a calcular el Índice de Precios y Cotizaciones, IPC, de la BMV. En un inicio se conformaba de las 42 empresas más bursátiles del mercado accionario. Este índice es de principal relevancia, ya que ha funcionado como indicador de la economía mexicana. Su comportamiento es observado por los participantes del mercado dado que es reflejo del mercado mexicano, así como de la evolución del mercado accionario.

Entre los eventos más relevantes alrededor del índice está el del año 1987, donde sufre una caída del 75% debido al exceso de operaciones apalancadas y el préstamo temporal de acciones (Grupo BMV, 2018). La bolsa, estando estrechamente vinculada con la economía del país, se vio sujeta a fluctuaciones con las devaluaciones del peso y los conflictos políticos.

Figura 1.4. Línea de tiempo de progreso de la BMV.



Fuente: Elaboración propia.

El IPC también ha experimentado una serie de cambios antes de llegar a ser lo que es ahora: era identificado como Promedio de Hechos entre 1900 y 1957; utilizando como el Promedio de Cotizaciones de Acciones entre 1958 y 1965, y Promedio de Precios y Cotizaciones en 1978, convirtiéndose en este último año y hasta la fecha como el principal indicador de rendimiento del mercado accionario mexicano (Grupo BMV, 2018).

El Índice de Precios y Cotizaciones, IPC, es el indicador oficial de la evolución del mercado accionario, ya que busca medir el rendimiento de las acciones de mayor tamaño y liquidez listadas en la Bolsa Mexicana de Valores.

El IPC es calculado diariamente por Standard and Poor's (S&P), ya que desde 2015 la BMV y S&P cuentan con una alianza para el cálculo y metodología de todos los índices de la Bolsa. Dicho cálculo se realiza con base en los resultados de la sesión cotidiana de remates y toma como referencia 35 emisoras de distintos sectores de la economía (BMV, 2019).

El IPC está basado en una muestra representativa de acciones con diferentes pesos. Esta se selecciona semestralmente (marzo y septiembre) en un proceso llamado *rebalanceo*, el cual, toma en cuenta variables como: número de operaciones, importe negociado, días operados y razón entre el monto operado y monto suscrito. De la ponderación de las 35 emisoras, ninguna emisora o componente deberá rebasar una ponderación que exceda del 25% y la ponderación total de las 5 acciones de mayor tamaño no podrán sobrepasar el 60% (BMV, 2019).

Dicha muestra es revisada una vez cada dos meses. Se calcula como sigue:

$$I_t = I_{t-1} \frac{\sum_{i=1}^N P_{it} Q_{it}}{\sum_{i=1}^N P_{i,t-1} Q_{i,t-1} F_{it}} \quad (1.2)$$

Donde:

- $I_t$  IPC en el día  $t$
- $P_{it}$  precio del emisor en día  $t$
- $Q_{it}$  Acciones del emisor  $i$  listadas en la BMV en día  $t$

- $F_{it}$  factor de ajuste por derechos sobre el activo  $i$  en el día  $t$
- $t-1$  día laboral anterior
- $i = 1, 2, 3, \dots, N$
- $N$  número de emisoras en la muestra

El  $P_{it}$  es usado para determinar el valor del índice de las 14:50 a las 15:00 horas, es equivalente al promedio de volumen del  $P_{it}$  de los primeros 10 minutos. Esta variable es calculada cada vez que una negociación es registrada. El propósito es determinar un precio representativo de los últimos diez minutos de negociación para calcular el valor de cierre del índice. La ecuación 1.1 indica que la suma de la capitalización de mercado de la muestra en día laboral, ajustada, determina el cambio en el IPC.

Cualquier cambio en el número de activos listados modifica la estructura del índice dado que el precio y la capitalización de mercado son usados como factor de peso. Esto debido a que el valor de las emisiones que declaran dividendos debe ser ajustado aplicando el factor de ajuste del día previo de capitalización de mercado. Si no existe ajuste, el factor será igual a uno. Existen posturas de inconformidad de los inversionistas, quienes no están a favor de esta valuación, dado que algunas acciones reciben gran peso; aun así, resulta un buen indicador del mercado.

Una característica importante del IPC es que es un índice invertible, es decir, que existen instrumentos que replican el comportamiento del índice, como por ejemplo los TRAC's o ETF's. El TRAC más representativo es el NAFTRAC<sup>10</sup>, en el cual existen inversiones por más de 60 mil millones de pesos (BMV, 2019).

Además del IPC, la BMV publica otros varios índices que siguen los diferentes sectores, los cuales miden el desempeño de ciertas áreas o grupos de actividad económica. El método de su cálculo es el mismo que el descrito anteriormente, solo que se aplica a muestras de activos relacionados por giro económico.

---

<sup>10</sup> El NAFTRAC es un ETF de tipo TRAC emitido por Nacional Financiera, Nafinsa, en abril 2002, que replica el comportamiento del IPC de la Bolsa Mexicana de Valores.

### 1.3.3 Otros índices

Además del IPC, la BMV ofrece otros índices que reflejan el comportamiento del mercado mexicano desde diferentes vistas (Tabla 1.4).

Tabla 1.4. Tipos de índices comercializados en la BMV.

#	ÍNDICE	DESCRIPCIÓN
1	Bursa Óptimo	Parametrizar el desempeño de las compañías más líquidas por su valor de capitalización de mercado.
2	Rentable	Parametrizar el desempeño de las compañías más líquidas y grandes que pagan dividendos.
3	Habita	Reflejar el desempeño del mercado de vivienda.
4	Sectoriales Analíticos	Parametrizar el comportamiento de las acciones listadas según la cobertura de siete sectores económicos (servicios de consumo no básico, productos de consumo frecuente, servicios financieros, salud, industrial, materiales y servicios de telecomunicación).
5	Actividad Económica	Medir el rendimiento de las acciones cubriendo varios sectores económicos del mercado de capitales.
6	Sectoriales Invertibles	Medir el desempeño de las acciones negociadas cubriendo cuatro sectores del mercado de capitales (consumo frecuente, financiero, industrial y materiales).
7	IPC CompMx	Indicador de referencia representativo del mercado de capitales y refleja el comportamiento de las emisoras más grandes y líquidas.
8	Formador de Mercado	Medir el desempeño de las acciones negociadas las cuáles mantienen un contrato activo con formadores de mercado para incentivar su liquidez y diversificación.
9	INMEX	Parametrizar el comportamiento de las 20 emisoras más grandes y líquidas del IPC.
10	IPC Sustentable	Medir el desempeño de empresas líderes en México que han adoptado políticas sobre los temas económicos, medioambientales y sociales.
11	Calidad, Valor y Rendimiento	Medir el desempeño de las emisoras del IPC con el multifactor más alto calculado a partir de la combinación de sus puntuaciones por calidad, valor y crecimiento.
12	Unifactoriales	Medir el desempeño de las estrategias que seleccionan el 30% superior de las acciones ordenadas de mayor a menor en función de sus factores respectivos dentro del IPC.

Fuente: Elaboración propia.

### 1.3.4 Regulación

Para el correcto funcionamiento y proveeduría de los diversos servicios que conforman el intercambio y comercialización de valores en la BMV, se han creado organismos e instituciones que dan seguimiento y permiten mantener dentro de las reglas y los procedimientos instruidos el comportamiento de las operaciones para garantizar el buen funcionamiento de la operatividad de los participantes (tabla 1.5).

Tabla 1.5. Entidades supervisoras de las regulaciones.

#	SIGLAS	INSTITUCIÓN	FUNCIÓN
1	SHCP	Secretaría de Hacienda y Crédito Público	Planear, coordinar, evaluar y vigilar el sistema bancario.
2	BANXICO	Banco de México	Proveer a la economía del país y procurar la estabilidad de precios, es decir, mantener la inflación baja
3	CNBV	Comisión Nacional Bancaria y de Valores	Vigilar y regular a las entidades que forman parte del sistema financiero mexicano
4	IPAB	Instituto para la Protección al Ahorro Bancario	Garantizar los depósitos que los ahorradores han realizado en los bancos
5	CNSF	Comisión Nacional de Seguros y Fianzas	Supervisar que las empresas de seguros y fianzas cumplan con el marco normativo que las regula, para que estas empresas mantengan su solvencia y estabilidad
6	CONSAR	Comisión Nacional del Sistema de Ahorro para el Retiro	Regular las aportaciones que hacen los trabajadores y las empresas, durante la vida laboral de los trabajadores
7	CONDUSEF	Comisión Nacional para la Protección y Defensa de los Usuarios de Servicios Financieros	Informar, orientar y promover la educación financiera, así como atender y resolver las quejas y reclamaciones

Fuente: Elaboración propia.

Las leyes aplicables con su normatividad para regular el mercado de valores se clasifican en: especiales, generales y disposiciones secundarias.

Especiales:

- *Ley del Mercado de Valores.*  
Regula los valores, su oferta pública e intermediación, las actividades de las sociedades anónimas que intervienen en el mercado de valores, el Registro Nacional de Valores y las autoridades y servicio en materia de mercado de valores.
- *Ley de la Comisión Nacional Bancaria y de Valores.*  
Define la naturaleza, objeto y facultades de la CNBV. Tiene como finalidad mejorar la calidad de la supervisión por parte del gobierno a las entidades prestatarias de servicios financieros que operan en el país.
- *Ley de Protección y Defensa al Usuario de Servicios Financieros.*  
Su objeto es la defensa y protección de los usuarios de servicios financieros prestados por las instituciones públicas y privadas, autorizadas por la SHCP. Esta ley regula la organización, procedimientos y funcionamiento de la CONDUCEF.

Generales:

- *Ley de Títulos y Operaciones de Crédito.*  
Reglamenta los actos y operaciones relativos a la emisión, expedición, endoso, aval, aceptación y otras actividades que se realicen con los títulos de crédito.
- *Ley General de Sociedades Mercantiles.*  
Regula la constitución, organización y objetos sociales de las sociedades mercantiles.

Disposiciones secundarias:

- Circulares de la CNBV.  
En las diferentes leyes del sistema financiero se faculta a la CNBV para interpretar a efectos administrativos los preceptos de estas y para facilitar su adecuada aplicación:
  - a) Circulares expedidas por la CNBV
  - b) Circular única para casas de bolsa
  - c) Circular única para emisoras
  - d) Circular única para sociedades de inversión

- Circulares del Banco de México.

Recomendaciones a las disposiciones de carácter general en las que se establece la interpretación que el Banco de México realiza sobre las leyes aplicables a instituciones de crédito, casas de bolsa, sociedades de inversión y otras entidades financieras.

### ***1.3.5 Análisis y estudios empíricos sobre comportamiento de los precios en la BMV***

Varios autores describen que la cantidad de investigaciones relacionadas con las anomalías en el contexto del mercado mexicano es escasa (Cervantes Zepeda, 1999) (Cabello y Ortiz, 2003) (Cruz, 2007) (Rodríguez y Morales, 2009) (López y Rodríguez, 2010) (García Blandón, 2011) (Dorantes Dosamantes, 2013) (Kristjanpoller y Arenas Yanez, 2015) (Duarte Duarte et al., 2016) (Garay Alvarado y Demmler, 2019) (Díaz Ruiz et al., 2020). Se observa una carencia de estudios exhaustivos que se centren en identificar patrones o indicios que puedan estar presentes en el comportamiento del mercado, así como análisis detallados sobre cómo determinadas políticas y reformas institucionales han repercutido o han tenido impacto en el desarrollo de la bolsa de valores.

En su análisis del periodo de 1986 a 2001 en el mercado accionario mexicano, Cabello y Ortiz (2003) examinan el impacto del mes de enero y del día de la semana en los rendimientos. Los hallazgos revelan que el lunes se presenta como el día de la semana con los peores rendimientos, mientras que el martes muestra un desempeño significativamente positivo. Los autores reportan que los patrones identificados se mantienen incluso al ajustar los rendimientos a la inflación y al tipo de cambio. Respecto a los meses, los resultados señalan la existencia del conocido “efecto enero”, aunque los autores indican que, al realizar el ajuste al riesgo, los rendimientos suelen ser negativos. En relación con este fenómeno a diferencia de los mercados estadounidenses donde se ha asociado con consideraciones fiscales, la anomalía en México no puede ser explicada por teorías impositivas, ya que no existen impuestos sobre las ganancias de capital. A pesar de este contexto, la anomalía

persiste en los resultados presentados en ese trabajo. Además, los autores destacan que, de agosto a febrero, se registran rendimientos negativos de manera consistente.

En el análisis realizado por Rodríguez Benavides y Morales Castro (2009), se investigan los efectos enero y de fin de semana en los rendimientos de 23 acciones de la Bolsa Mexicana de Valores en el periodo de 1992 a 2009, utilizando precios diarios y mensuales. Se emplearon modelos ARCH, GARCH y Mínimos Cuadrados Ordinarios (OLS, por sus siglas en inglés) para las evaluaciones. Los resultados ofrecen una imagen mixta respecto a la presencia de las anomalías estudiadas: siete acciones revelaron la existencia del efecto de fin de semana (Cemex, ICA, Ind, Modelo, Simec, Telmex y Wal-Mart), mientras que diez mostraron evidencia del efecto enero (Autlan, Bimbo, Cemex, Elektra, Gmex, ICA, Ind, Kimberly, Peñoles y Telmex).

En un estudio posterior, López Herrera y Rodríguez Benavidez (2010) replican el análisis del efecto enero y de fin de semana de Rodríguez Benavides y Morales Castro (2009) para las bolsas de América Latina, encontrando la presencia del efecto enero únicamente en Argentina, Brasil, Colombia, Chile y México, cuando se consideran rendimientos en monedas locales, y solo en Argentina y Chile cuando se miden en dólares estadounidenses. Mientras que para el efecto de fin de semana no se reporta presencia para estos países. No obstante, no se detecta la presencia de ninguna de las anomalías analizadas en Perú.

Asimismo, García Blandón (2011) lleva a cabo un análisis en los mercados de Argentina, Brasil, México y Chile, evaluando el efecto de fin de semana, fin de mes, fin de año y días feriados. Los resultados indican rendimientos negativos los lunes en todas las bolsas, excepto en México. En Argentina, se evidencia el efecto de fin de año, pero no se encuentra relación con el cambio de mes y días feriados, mientras que, en Brasil y México, se encuentra relación solamente con el cambio de mes y días feriados.

Kristjanpoller Rodríguez y Arenas Yáñez (2015) se enfocan en el efecto de fin de mes en seis países latinoamericanos (Brasil, México, Chile, Colombia, Perú y Argentina) en el periodo de 1993 a 2011. Sus hallazgos coinciden con investigaciones anteriores, mostrando

rendimientos positivos anormales en la mayoría de los países analizados, lo que confirma la presencia de esta anomalía.

En su análisis del mercado mexicano, Dorantes Dosamantes (2013) se enfoca en examinar el comportamiento de los rendimientos mediante el uso de factores basados en anomalías fundamentales. Su estudio se basa en datos trimestrales de 196 acciones de la Bolsa Mexicana de Valores durante el periodo de 1991 a 2011. Los resultados de su investigación indican que las anomalías relacionadas con las ganancias (*earnings*), la relación libro-valor (*book-to-market*) y el tamaño (*size*) son características relevantes en el portafolio analizado.

Duarte, Sierra y Garcés (2016) se centran en analizar periodos de tendencia posteriores a crisis y épocas de crecimiento, evaluando empíricamente la presencia de anomalías, el efecto fin de mes, fin de semana y día posferiado en la BMV. El estudio considera el IPC y tres acciones de mayor peso en el índice, que componen el 50% del mismo: Telcel, Walmart y Femsa. Los autores utilizan modelos binarios de regresión multivariantes durante el periodo desde enero de 2002 hasta abril de 2013. Los resultados de la investigación no muestran evidencia del efecto de fin de semana. En cuanto al efecto posferiado y de fin de mes, los autores reportan evidencia de una disminución en las anomalías mencionadas. Los autores concluyen que, durante periodos alcistas, las anomalías se hacen más evidentes, sugiriendo que esta conducta podría deberse a la confianza de los inversionistas ante situaciones de optimismo generalizado en el contexto de una aparente bonanza del país debido a la estabilidad política y económica. Además, para periodos más cortos de la muestra, se exhiben resultados positivos con mayores rendimientos para los días posteriores a los feriados y para el primer día de cada mes.

Garay Alvarado y Demmler (2019) emplean el análisis econométrico GARCH para identificar la presencia del efecto enero en el IPC y en los índices sectoriales de la BMV. El estudio considera precios diarios para el periodo comprendido entre enero de 2010 y diciembre de 2018. Los investigadores concluyen que los resultados obtenidos no proporcionan evidencia significativa respecto a la anomalía analizada en ninguna de las series probadas.

Díaz, Herrerías y Vásquez (2020) exploran la relación entre lo que los autores denominan las 19 anomalías más comunes informadas para el mercado de Estados Unidos y los rendimientos de acciones en el mercado mexicano. El estudio abarca el período de 1994 a 2017, utilizando datos mensuales, y revela que solo 5 de las 19 anomalías consideradas están presentes en el mercado mexicano: tamaño, *momentum*, volatilidad idiosincrática, efecto lotería y rentabilidad. La metodología empleada incluye regresiones univariadas Fama-MacBeth (1973) para evaluar cada anomalía. Según los resultados, se observa una capacidad predictiva en los precios a un período (un mes) para tamaño, *momentum* y rentabilidad. En contraste, se identifica una relación negativa para volatilidad idiosincrática y el efecto lotería. Para tamaño y *momentum*, los resultados sugieren horizontes más extensos de 3 y 6 meses.

Villarreal-Samaniego y Santillán (2023) examinan la anomalía del efecto de fin de semana desde la perspectiva de la Hipótesis de Mercados Adaptativos (HMA) en los índices bursátiles de Argentina, Brasil, Chile, Colombia, México y Perú. El periodo completo de análisis abarca desde 1995 hasta 2000, y se incluyen subperiodos de 5 años. Los investigadores emplean especificaciones de Promedio móvil autoregresivo (ARMA), Heteroscedasticidad condicional autorregresiva generalizada (GARCH), y pruebas de Kruskal-Wallis. En relación con el IPC mexicano los resultados indican, por un lado, la presencia de la mencionada anomalía solo durante el primer subperiodo, de 1995 a 2000. Este hallazgo, junto con los resultados obtenidos para los índices de los otros países, respalda la idea de que a medida que aumenta la capitalización de mercado, la presencia de la anomalía disminuye y, por ende, la eficiencia del mercado aumenta en términos débiles. Por otro lado, durante condiciones bajistas, la prueba de Kruskal-Wallis señala la presencia de la anomalía del lunes. Una explicación plausible que los autores ofrecen para este resultado es que, en condiciones bajistas, los inversores están más atentos a los anuncios de políticas gubernamentales durante los fines de semana. Otra posibilidad es la "hipótesis del estacionamiento de los ingresos" que sugiere que los inversores, después de sufrir pérdidas, retienen temporalmente sus fondos antes de reinvertirlos. Ésa es una estrategia que podría considerarse especialmente sólida en condiciones de mercado en declive.

## ***Resumen***

Como se ha destacado en este capítulo, el mercado mexicano, al igual que otros mercados emergentes, ha sido objeto de una exploración superficial, lo que implica que los estudios relacionados con el análisis de sus activos y comportamiento sean escasos. Dado que México ocupa la posición de segundo mercado más importante en América Latina, poseer mayor información sobre el mercado resulta atractivo para inversionistas, tanto nacionales como extranjeros, y actualmente es un área de interés para todos los participantes: analistas profesionales, inversionistas corporativos y minoristas, por mencionar solo algunos. Las evidencias presentadas acerca de la presencia de ciertas anomalías, especialmente las de calendario que han sido las más investigadas, no son concluyentes y en algunos casos exhiben resultados con opiniones contradictorias. Debido a la amplia variedad de anomalías y la diversidad de sus combinaciones, este tema sigue siendo motivo de discusión y constituye un área de investigación significativa en la actualidad.

En este trabajo, se aborda un subconjunto específico de anomalías presentes en el mercado mexicano, tomando como referencia aquellas que han demostrado relevancia en otros mercados. La propuesta consiste en evaluarlas mediante modelos construidos con métodos no lineales provenientes del ámbito de aprendizaje automático, que, como se ha evidenciado, han recibido una atención creciente en comparación con los métodos tradicionales. Los próximos capítulos se centrarán en presentar estos análisis detallados.

## Capítulo 2

### HERRAMIENTAS DE ANÁLISIS

En este capítulo se presentan las herramientas técnicas y analíticas que son utilizadas para estudiar las anomalías en el mercado mexicano. Primeramente, se aborda los métodos tradicionales de valuación de activos financieros, posteriormente se introducen las anomalías como factores y finalmente se concluye el análisis con las técnicas de aprendizaje automático y las métricas de evaluación de modelos.

#### *2.1 Introducción*

En la literatura existe gran variedad de investigaciones relacionadas con la explicación de los precios de los activos financieros y de sus rendimientos esperados. Para la valuación de activos, el modelo principal y más mencionado es el CAPM (*Capital Asset Pricing Model*). En éste, se utiliza de referencia un activo base, llamado libre de riesgo, y se busca la obtención del factor beta. Dicho factor beta corresponde al valor explicativo de los rendimientos de un activo financiero determinado del cual se desea conocer el desempeño respecto de los rendimientos del activo base. Asimismo, en la literatura del análisis multifactorial, Fama y MacBeth (1973) introducen un modelo de regresión para el cálculo de las betas de mercado y obtención de coeficientes de los factores analizados basados en el t-estadístico con el objeto de determinar precios de un activo.

Los modelos CAPM y Fama-MacBeth son considerados herramientas de valuación de activos y estos se utilizan para análisis de pronósticos de rendimientos de activos a partir de factores o para la obtención de los coeficientes y pesos de las variables explicativas. Entre estos factores, se utilizan las anomalías documentadas en la literatura y algunas derivaciones a partir de sus combinaciones lineales. Cabe mencionar que dichas investigaciones

generalizan el impacto de los factores que utilizan, que estos están presentes y aplican de igual forma en todos los mercados. Por lo que los resultados con diferentes condiciones pudieran variar notablemente.

En esta investigación se analiza el desempeño de las anomalías *momentum*, reversión a la media, volatilidad y los efectos enero, fin de mes y fin de semana de manera aislada e independiente dentro del ámbito del mercado mexicano, con el objeto de analizar empíricamente su presencia y, sobre todo, su utilidad.

El mercado mexicano ha sido poco estudiado debido a los siguientes motivos:

1. El mercado poco desarrollado conlleva a tener limitaciones en términos de disponibilidad de la información histórica y contar con bajo número de empresas comparados con los mercados desarrollados.
2. La información de las cotizaciones no es pública y la que está disponible es de dudosa calidad.
3. La insuficiencia de registros de reportes de las empresas con historia de datos.
4. Los datos de las cotizaciones existentes no están homologados en cálculo, periodos y tiempo.

El estudio de las anomalías de mercado no es un tema nuevo y existe gran variedad de análisis e investigaciones sobre el tema. Las anomalías documentadas y sus combinaciones han sido utilizadas de igual forma para predicción de activos, pronósticos de rendimientos y formación de portafolios. Muchas de estas anomalías han sido replicadas en diferentes mercados, activos y productos financieros, en su mayoría confirmando su presencia (Moore, 2021). A pesar que los primeros factores presentados como anomalías de mercado fueron reportados hace más de 50 años, el detalle de su existencia y sobre todo el impacto que tienen en los mercados sigue siendo de gran interés debido al área de oportunidad que ofrecen a los inversionistas. Por lo tanto, conocer más sobre el comportamiento de las anomalías es aún un tema vigente y éstas continúan siendo un motivo de investigación.

Entre las metodologías de evaluación de modelos utilizados para estudiar las anomalías sobresalen las econométricas y de análisis de series de tiempo, las cuales tienen fundamento estadístico, por lo que estas son consideradas las herramientas paramétricas clásicas de análisis. Aunque cada vez es más frecuente encontrar nuevas propuestas de modelos de análisis provenientes de otras áreas como la inteligencia artificial y minería de datos de las ciencias de la computación, de física o incluso de modelos híbridos construidos de sus combinaciones. Estos modelos paramétricos y no paramétricos resultantes ofrecen alternativas diferentes de estudio y nuevas áreas de oportunidad (Cruz, 2007) (García et al., 2013) (Rodríguez et al., 2015).

En este trabajo se analiza la presencia de anomalías en el mercado mexicano. Las anomalías se obtienen sobre los rendimientos de los activos y esos son comparados con los del mercado, que en este trabajo son los del IPC. Las metodologías de aprendizaje automático se utilizan para la evaluación del grado de presencia de las anomalías. Por un lado, estas metodologías son conocidas por su utilidad para descubrir relaciones o patrones no lineales entre los factores, una limitante de las metodologías de análisis tradicionales. Por otro lado, cuando estas metodologías se utilizan como clasificadores, esto es que las variables dependientes están categorizadas, el porcentaje de exactitud resultante de la evaluación del modelo es bastante intuitivo y legible para reportar su desempeño (García et al., 2013).

Las técnicas de aprendizaje automático pueden ser usadas para la predicción o estimación de los rendimientos. En el presente trabajo, estas técnicas de aprendizaje automático se utilizan como evaluadores del grado de presencia de las anomalías mediante la categorización del porcentaje de exactitud resultante de la validación de los modelos. Cabe mencionar que el análisis a realizar consiste en probar las anomalías en su forma simple, esto es, no se intentan optimizar o localizar los mejores periodos, número de periodos, mes, semana o día para el cual los modelos creados presentan el mejor ajuste en el mercado mexicano. Tampoco está dentro del alcance combinar las anomalías descritas para probar su desempeño en conjunto ni que estas formen parte de un modelo multifactorial de valuación de activos.

## 2.2 *Valuación de activos*

En esta sección se presentan algunas de las teorías para la valuación de activos financieros comúnmente usadas por profesionales y académicos cuyo objetivo es determinar el valor fundamental de los activos y su relación con sus rendimientos.

La valuación de activos financieros se refiere al proceso de determinar el valor intrínseco de diferentes tipos de activos que tienen un componente financiero, como acciones, bonos, opciones, divisas y otros instrumentos financieros. Este proceso busca estimar cuánto vale un activo en función de diversos factores, incluyendo su flujo de efectivo esperado, el riesgo asociado y las tasas de interés prevalecientes en el mercado.

La valuación de activos financieros es esencial en el mundo de las inversiones y las finanzas, ya que ésta proporciona una base para la toma de decisiones de inversión. Los inversionistas utilizan diversas técnicas y modelos para valorar los activos financieros con el objetivo de determinar si un activo está subvaluado o sobrevalorado en relación con su precio de mercado actual.

Algunas de las técnicas de valuación de activos financieros comunes incluyen el ya mencionado CAPM, el análisis fundamental, el análisis técnico, el valor presente neto (VPN), el método de comparables y otros enfoques específicos según el tipo de activo en consideración. Estas técnicas pueden variar en complejidad y enfoque, pero todas tienen como objetivo proporcionar una estimación del valor de un activo en el mercado financiero.

En la teoría de valoración de activos financieros, el enfoque se centra en encontrar métodos para determinar los precios de estos activos con el propósito de comprender la relación entre el valor de los factores fundamentales de un activo y su retorno apropiado. Este supuesto se basa en la idea de que el valor fundamental de un activo explica su precio de mercado. Adam Smith se refiere al valor fundamental como el "precio natural" de los activos, que incorpora elementos como el costo de la mano de obra y los conceptos de rentabilidad que se derivan de la posesión del activo, y que proporciona al propietario rendimientos adecuados (Hirsch, 2021).

Además, el precio de mercado de un activo está determinado por las fuerzas de la oferta y la demanda, y aunque puede desviarse temporalmente de su valor fundamental, a largo plazo tiende a converger hacia este último. Por lo tanto, esta valoración desempeña un papel crucial en la asignación de recursos financieros para todos los participantes del mercado, ya sea en la toma de decisiones de inversión, el consumo de recursos económicos, proyecciones financieras, entre otros.

A continuación, se procede a examinar algunos de los modelos más destacados en el ámbito de la estimación y valoración de activos financieros. Estos modelos desempeñan un papel fundamental en la determinación de los precios de los activos en los mercados financieros, brindando una comprensión más profunda de cómo se evalúan los activos y se toman decisiones de inversión. Se exploran estos modelos con el fin de comprender mejor su funcionamiento y sus aplicaciones en el entorno financiero.

### ***2.3 El Modelo de valoración de activos financieros, CAPM***

En (1964), William F. Sharpe inició una discusión fundamental sobre las expectativas de rendimiento de los inversionistas, un paso esencial para alcanzar el equilibrio en los mercados financieros. Sharpe exploró cómo el precio de los activos financieros se ve influenciado tanto por las preferencias de los inversionistas como por los atributos intrínsecos de los activos. Paralelamente, Lintner (1965) y Mossin (1966), utilizando la teoría de carteras de manera independiente, llegaron a resultados similares. Estos resultados dieron lugar al desarrollo del Modelo de Activos de Capital (CAPM, por sus siglas en inglés de *Capital Asset Pricing Model*), que refleja las preferencias de los inversionistas en términos de la relación entre riesgo y rendimiento de los activos, el cual ha tenido una influencia significativa tanto en la teoría como en la práctica financieras.

El modelo propone que el rendimiento esperado de un activo financiero debe estar por encima de la tasa libre de riesgo más una prima proporcional al riesgo sistémico que ofrece el mercado. El riesgo sistémico no es diversificable y se mide por la covarianza entre el

rendimiento de un activo y el rendimiento de los activos de un portafolio compuesto de todos los activos existentes, llamado en la literatura financiera “portafolio de mercado”. Además del riesgo sistémico, los activos financieros presentan un riesgo intrínseco diversificable; debido precisamente a que puede ser eliminado por diversificación, en equilibrio el mercado no paga una prima por este riesgo (Campbell et al., 2012).

El modelo se expresa como sigue:

$$E(R_{it}) = R_{ft} + \beta_i [E(R_{mt}) - R_{ft}] \quad (2.1)$$

Donde  $E(R_{it})$  es el rendimiento esperado del  $i$ -ésimo activo,  $R_{ft}$  es la tasa libre de riesgo,  $E(R_{mt})$  es el rendimiento esperado del portafolio de mercado y  $\beta_i$ , denominado también “coeficiente beta” o “coeficiente de volatilidad”, es la medida de riesgo sistémico o parámetro de sensibilidad del mercado. En otras palabras,  $\beta_i$  mide la sensibilidad del rendimiento del activo a la variación en el rendimiento del mercado.

La siguiente expresión identifica los rendimientos en exceso:

$$E(R_{it} - R_{ft}) = \beta_i [E(R_{mt}) - R_{ft}] \quad (2.2)$$

El rendimiento en exceso sobre el activo libre de riesgo del  $i$ -ésimo activo es:

$$r_{it} = R_{it} - R_{ft} \quad (2.3)$$

La prima por el riesgo de mercado es:

$$r_{mt} = R_{mt} - R_{ft} \quad (2.4)$$

La ecuación considerando los rendimientos excesivos:

$$E(r_{it}) = \beta_i E(r_{mt}) \quad (2.5)$$

$E(r_{it})$  es el rendimiento esperado del excedente del activo  $i$ ,  $E(r_{mt})$  es la rentabilidad esperada del excedente en el portafolio de mercado y  $\beta_i$  es la medida de riesgo sistemático, la cual es la misma que en ecuación (2.1). La medida de riesgo sistemático  $\beta_i$  también puede definirse como:

$$\beta_i = \frac{cov(r_{it}, r_{mt})}{var(r_{mt})} \quad (2.6)$$

La beta del modelo CAPM es un concepto conocido ampliamente en la valuación de activos relacionados con los rendimientos de las acciones respecto a los rendimientos del mercado. Por ejemplo, en el mercado mexicano, Herrera y Lockwood (1994) reportan una relación positiva entre los rendimientos promedio y la beta del mercado con una muestra entre 1987 y 1992.

La necesidad de la existencia de un activo libre de riesgo es la principal limitación del CAPM. Black (1972) deriva una versión menos restrictiva del modelo en la que utiliza el hecho de que todos los conjuntos de portafolios eficientes pueden ser formados como una combinación lineal de portafolios eficientes. Si bien esta versión del CAPM no es dependiente de la existencia de un activo libre de riesgo, aún ésta tiene como supuesto base que todos los inversionistas tengan una distribución de probabilidad común sobre los rendimientos de los activos disponibles.

Como se ha indicado en el capítulo anterior, el hallazgo de anomalías en series de tiempo de precios y rendimientos financieros, respecto de lo previsto por el CAPM, ha sido objeto de numerosas investigaciones empíricas. Estas anomalías, que se refieren a patrones persistentes que no pueden explicarse adecuadamente mediante este modelo, cuestionan su significancia, ya que éstas ponen en entredicho la hipótesis conjunta de eficiencia del mercado y validez del CAPM (Cervantes Zepeda, 1999). En este contexto, es importante destacar que concluir simplemente que el mercado es ineficiente basándose en la presencia de anomalías sería un enfoque erróneo. La razón detrás del rechazo de la hipótesis nula de eficiencia del mercado podría deberse, en parte, a que el modelo de valoración de activos no captura de manera adecuada la complejidad del mercado real. Esto sugiere que algunas anomalías que persisten a lo largo del tiempo podrían indicar limitaciones en el modelo de equilibrio utilizado (Hawawini y Keim, 1995).

Por lo tanto, la presencia de anomalías en los mercados financieros plantea interrogantes sobre la eficiencia del mercado y la capacidad de los modelos teóricos para explicar el comportamiento real de los activos. Esto ha llevado a la exploración de modelos y teorías alternativas que puedan ofrecer una mejor comprensión de los movimientos de precios y rendimientos en los mercados financieros.

## **2.4 Modelos multifactoriales**

Las investigaciones han revelado la existencia de múltiples factores que influyen en los rendimientos esperados de los activos, más allá de lo que captura el CAPM. Estos factores adicionales pueden incluir elementos como el tamaño de la empresa, el valor contable, el *momentum* de los precios y otros. Dado que el mercado real es más complejo de lo que el CAPM puede reflejar con su única beta del mercado, ha surgido la necesidad de desarrollar modelos de valoración de activos que utilicen múltiples factores.

Si bien el CAPM ha sido una herramienta valiosa en la valoración de activos y la toma de decisiones de inversión, la evidencia empírica, como hemos visto, ha mostrado que este modelo no es suficiente para explicar completamente los rendimientos esperados de los activos financieros en la práctica.

Los modelos multifactoriales reconocen que los rendimientos esperados de los activos pueden estar influenciados por una variedad de factores, y no solo por el riesgo sistémico respecto del mercado en su conjunto. Estos modelos incorporan múltiples variables independientes, con betas adicionales, para capturar mejor la complejidad del mercado financiero. Al hacerlo, dichos modelos proporcionan una representación más precisa y completa de los rendimientos esperados de los activos y en consecuencia estos ayudan a los inversores a tomar decisiones más informadas y precisas (Campbell et al., 2012).

A continuación, vamos a considerar los tres modelos multifactoriales más conocidos, en concreto la teoría de fijación de precios por arbitraje (APT), el modelo Fama-MacBeth y el conjunto de modelos de análisis multivariado.

### ***2.4.1 Teoría de fijación de precios por arbitraje, APT***

La “teoría de fijación de precios por arbitraje” o APT del inglés *Arbitrage Pricing Theory*, es un modelo teórico utilizado en finanzas para estimar el precio o el rendimiento esperado de un activo financiero, como una acción o un bono. Fue desarrollado por el economista Stephen Ross (1976) como una alternativa al CAPM.

El APT se basa en la idea de que el precio o el rendimiento de un activo financiero se puede explicar en función de varios factores de riesgo o variables macroeconómicas. Estos factores pueden incluir tasas de interés, tasas de inflación, tasas de crecimiento económico u otros indicadores económicos relevantes. El APT asume que los inversores son racionales y buscan maximizar su riqueza, por lo que los precios de los activos deben reflejar el equilibrio entre el riesgo y el rendimiento esperado.

El APT es más general que el CAPM en que permite la integración de múltiples factores de riesgo. A diferencia del CAPM, el APT no requiere de la identificación de un portafolio de mercado. El APT provee una relación aproximada de los rendimientos esperados de los activos con un número desconocido de factores no identificados. Esto es, mientras el CAPM se basa en un único factor beta como riesgo de mercado, el APT considera un conjunto de factores con los que captura el riesgo del mercado. A este nivel, el rechazo de la teoría es imposible (a menos que existan oportunidades de arbitraje) y como consecuencia la prueba del modelo depende de la introducción de supuestos adicionales (Campbell et al., 2012).

El APT asume que el mercado es perfectamente competitivo y sin fricciones y que los inversionistas prefieren tener mayor beneficio asumiendo mayor riesgo en comparación con tener una menor utilidad, pero de manera segura, con menor riesgo.

El proceso estocástico de generación de rendimientos de activos es expresado como una función lineal de un conjunto de  $K$  factores de riesgo. El APT, al igual que el CAPM, asume que los “efectos específicos” o “riesgo específico” de los activos ( $\varepsilon_i$ ) son independientes y pueden ser diversificados dentro de un largo portafolio. De acuerdo al APT, el rendimiento esperado de un activo puede ser dado por:

$$E(R_i) = \lambda_0 + \lambda_1 b_{i1} + \lambda_2 b_{i2} + \dots + \lambda_k b_{ik} \quad (2.7)$$

Donde  $E(R_i)$  es el rendimiento esperado sobre cualquier activo,  $\lambda_0$  es el retorno esperado sobre un activo con cero riesgos sistemáticos;  $\lambda_k$  es la prima de riesgo relacionada al  $k$ -ésimo factor de riesgo común, y  $b_{ik}$  es la relación de precio entre la prima de riesgo y el activo el cual es conocido como “factor beta” o “factor de cargas”.

El APT asume que los inversionistas creen que el vector columna ( $N \times 1$ )  $r$  de un único periodo de rendimientos aleatorios sobre activos de capital satisface el modelo:

$$r = \mu + \beta f + e \quad (2.8)$$

Donde  $e$  es un vector columna ( $n \times 1$ ) de variables aleatorias,  $f$  es un vector columna ( $k \times 1$ ) de los diferentes factores de riesgo,  $\mu$  es un vector columna ( $n \times 1$ ) del rendimiento esperado y  $\beta$  es una matriz ( $n \times k$ ) que denota la influencia de cada factor sobre el rendimiento. Sin pérdida de generalización, se normaliza la ecuación con los supuestos  $E(f) = 0$  y  $E(e) = 0$ , donde  $E(\ )$  denota la expectativa  $r$  y  $0$  la matriz de ceros con la dimensión requerida. El modelo de factores de la ecuación 2.8 implica que  $E(r) = \mu$ .

La demostración del APT requiere restricciones sobre  $\beta$  y la matriz de covarianza  $\Omega = E(ee')$ . Un supuesto adicional es que  $E(e | f) = 0$ , siendo esto último no necesario en algunos de los desarrollos del modelo.

El número de activos  $n$ , se asume ser más grande que el número de factores  $k$ . En algunos modelos,  $n$  es infinita o muy cercano a infinito. En tal caso, la representación de ecuación 2.8 aplica a una secuencia de mercado de capitales; los primeros  $n$  activos en el mercado  $n + 1$  son los mismos que aquellos en el mercado  $n$ -ésimo y las primeras  $n$  filas de la matriz  $\beta$  en los primeros  $n + 1$  mercados constituyen la matriz  $\beta$  en el  $n$ -ésimo mercado.

El APT confirma la existencia de una constante  $a$  tal que, para cada  $n$ , la desigualdad:

$$(\mu - X\lambda)Z^{-1}(\mu - X\lambda) \leq a \quad (2.9)$$

Se mantiene para un vector columna  $\lambda$  de  $(k + 1) \times 1$ , y una matriz positiva  $Z$  definida de  $n \times n$ . De ahí que,  $X = (i, \beta)$ , en el cual  $i$  es un vector columna  $n \times 1$  de unos. Sea  $\lambda_0$  el primer componente de  $\lambda$  y  $\lambda_1$  consiste en el resto de componentes. Si algún portafolio de los activos es libre de riesgo, entonces  $\lambda_0$  es el rendimiento del portafolio libre de riesgo. La matriz positiva  $Z$  definida es frecuentemente la matriz de covarianza  $E(ee')$ . La valuación exacta del arbitraje se obtiene si ecuación 2.8 es reemplazada por:

$$\mu = X\lambda = i\lambda_0 + \beta\lambda_1 \quad (2.10)$$

El vector  $\lambda_1$  es referido como la prima de riesgo y la matriz  $\beta$  como la beta o carga del factor de riesgo.

La interpretación de ecuación 2.8 es que cada componente de  $\mu$  depende aproximadamente de la linealidad de la correspondiente fila de  $\beta$ . Esta relación lineal es la misma sobre todos los activos. La estimación es mejor cuanto más pequeña sea la constante  $a$ ; si  $a = 0$  la relación lineal es exacta y se obtiene la ecuación 2.10.

La limitación del APT es que la teoría no sugiere factores para un activo en particular. Los inversionistas deben percibir las fuentes de riesgo y estimar la sensibilidad de los factores. En la práctica, un activo debería ser más sensible a un factor que a otro. En consecuencia, el reto real para un inversionista yace en la determinación de cada factor que influya a un cierto activo, los rendimientos ajustados esperados y la sensibilidad del activo ante dicha exposición.

### ***2.4.2 Modelo Fama-MacBeth***

Las teorías de valuación de activos financieros frecuentemente utilizan diversos factores de riesgo como descriptores de los rendimientos de los activos. El modelo Fama-MacBeth es un enfoque estadístico utilizado en el campo de las finanzas para evaluar la relación entre los rendimientos de los activos financieros y múltiples factores de riesgo. Este modelo lleva los nombres de sus creadores, Eugene Fama y James MacBeth en (1973), y se desarrolló como una extensión del CAPM y del APT.

El objetivo principal del modelo Fama-MacBeth es analizar cómo los factores de riesgo afectan a los rendimientos esperados de los activos financieros, como acciones o bonos. En lugar de depender de un único factor de riesgo, como en el CAPM, el modelo Fama-MacBeth permite la inclusión de varios factores, lo que lo convierte en un modelo multifactorial. Estos factores pueden incluir tasas de interés, tasas de inflación, tasas de crecimiento económico, indicadores macroeconómicos u otros indicadores financieros relevantes.

Ahora bien, debe tenerse en cuenta que, a diferencia del CAPM y la APT que son “modelos de equilibrio” donde las relaciones 2.1 o 2.7 implican la existencia de equilibrio en los mercados financieros, el modelo Fama-MacBeth es un “modelo de comportamiento”, esto es, no supone que los mercados están en equilibrio, sino que se busca la relación que mejor se ajuste a un conjunto de datos sobre rendimientos de activos financieros.

En resumen, el modelo Fama-MacBeth es un enfoque estadístico que permite analizar la relación entre los rendimientos de los activos financieros y múltiples factores de riesgo, lo que lo convierte en una herramienta importante en la valoración de activos y la gestión de inversiones.

El método consiste en proyectar los rendimientos sobre las betas y entonces agregar las estimaciones en la dimensión del tiempo. Asumiendo que las betas son conocidas, el modelo de regresión del  $t$ -ésimo estimador de  $N$  activos es:

$$Z_t = \gamma_{0t}L + \gamma_{1t}\beta_m + \eta_t \quad (2.11)$$

Donde  $Z_t$  es el vector columna ( $N \times 1$ ) de los rendimientos excesivos del activo para la serie de tiempo  $t$ ,  $L$  es un vector columna ( $N \times 1$ ) de unos y  $\beta_m$  es el vector columna ( $N \times 1$ ) de las betas del modelo CAPM.

La implementación del estimador Fama-MacBeth involucra dos pasos: en el primero, cada rendimiento del portafolio es regresado contra cada una de las series de tiempo de los factores para determinar la exposición a cada uno. En el segundo paso, los rendimientos del portafolio

son regresados contra la exposición de la etapa anterior, para cada unidad de tiempo, y obtener una serie de tiempo de los coeficientes de prima de riesgo de cada factor. La lógica es promediar dichos coeficientes, una vez por cada factor, para obtener la prima esperada por unidad expuesta a cada factor de riesgo en el tiempo.

Para el primer paso, dados  $T$  periodos de datos, la ecuación 2.11 es estimada utilizando mínimos cuadrados ordinarios, OLS del inglés *Ordinary Least Squares*, para cada  $t$ ,  $t = 1, \dots, T$ , dando las  $T$  estimaciones de  $\gamma_{0t}$  y  $\gamma_{1t}$ . Entonces, en el segundo paso, las series de tiempo de  $\hat{\gamma}_{0t}$ 's y  $\hat{\gamma}_{1t}$ 's son analizadas. Definiendo  $\gamma_0 = E[\gamma_{0t}]$  y  $\gamma_1 = E[\gamma_{1t}]$ , las implicaciones de modelo CAPM son  $\gamma_0 = 0$  (intercepto) y  $\gamma_1 > 0$  (prima de riesgo de mercado positiva). Dado que los rendimientos son normalmente distribuidas y temporalmente independientes e idénticamente distribuidos, las gammas serán también normalmente distribuidas, independientes e idénticamente distribuidas. Asimismo, dadas las series de tiempo  $\gamma_{0t}$  y  $\gamma_{1t}$ ,  $t = 1, \dots, T$ , se puede probar las implicaciones utilizando la típica t-test. Definiendo  $\omega(\hat{\gamma}_j)$  como la estadística  $t$ , se tiene:

$$\omega(\hat{\gamma}_j) = \frac{\hat{\gamma}_j}{\hat{\sigma}_{\gamma_j}} \quad (2.12)$$

Donde

$$\hat{\gamma}_j = \frac{1}{T} \sum_{t=1}^T \hat{\gamma}_{jt} \quad (2.13)$$

$$\hat{\sigma}_{\gamma_j}^2 = \frac{1}{T(T-1)} \sum_{t=1}^T (\hat{\gamma}_{jt} - \hat{\gamma}_j)^2 \quad (2.14)$$

La distribución de  $\omega(\hat{\gamma}_j)$  es la  $t$ -student con  $(T - 1)$  grados de libertad y asintóticamente es normal estándar. Dadas las pruebas estadísticas, las inferencias pueden ser realizadas de la manera típica.

El estimador Fama-MacBeth es particularmente útil porque puede ser modificado para acomodar medidas de riesgo adicionales más allá de las betas del CAPM. Al agregar medidas de riesgo adicionales al estimador se puede examinar la hipótesis de que la beta

completamente describe la variación en los rendimientos esperados. Por ejemplo, se puede considerar el factor “tamaño de la empresa” como poder implicatorio de los retornos esperados, donde este factor está definido como el logaritmo del valor de mercado del capital. Si se define  $\zeta_t$  como el vector columna ( $N \times 1$ ) con elementos correspondientes al “tamaño de la empresa” en el periodo  $t$ , se puede extender la ecuación 2.11 para investigar si ese factor tiene poder explicativo no capturado por la beta del mercado:

$$Z_t = \gamma_{0t}L + \gamma_{1t}\beta_m + \gamma_{2t}\zeta_t + \eta_t \quad (2.15)$$

Usando la forma  $\hat{\gamma}_{2t}$  de la ecuación 2.15, se puede probar la hipótesis de que el factor “tamaño de la empresa” no tiene poder explicativo más allá de la beta, es decir, haciendo  $\gamma_2 = 0$  mediante la asignación de  $j = 2$  en las ecuaciones 2.12 y 2.14

La metodología Fama-MacBeth, si bien es de utilidad, presenta algunos problemas, propios de su carácter de “modelo de comportamiento”: no puede ser aplicada directamente, dado que las betas de mercado no son conocidas, y las regresiones son realizadas utilizando betas estimadas de los datos, las cuales introducen errores. Estos errores pueden ser tratados de dos formas: primera, y es la que adopta el método, consiste en minimizar los errores mediante la agrupación de los activos en portafolios e incrementar la precisión de las estimaciones de las betas; y segunda, ajustar explícitamente los errores estándar para corregir el sesgo introducido.

### ***2.4.3 Modelos de análisis multivariado***

Los métodos basados en análisis multivariado como alternativas para explicar los rendimientos de los activos han sido motivados dados los antecedentes de las discrepancias reportadas a la HME y al CAPM. Los factores probados varían y pueden ser desde variables macroeconómicas, características internas y externas de la empresa o cualquier otra anomalía reportada o combinaciones de estas. De nuevo nos encontramos con procedimientos “de comportamiento” dirigidos a encontrar el modelo que mejor se ajuste a los datos disponibles (Campbell et al., 2012) (Damodaran, 2012) (García et al., 2013) (Akwolaga, 2020).

Este tipo de modelos multifactoriales, o llamados portafolios de factores “cero betas”, están basados en un portafolio que no contiene ningún factor de sensibilidad del mercado; los rendimientos excesivos están linealmente relacionados con la sensibilidad inherente de los rendimientos en la matriz.

Sea  $R_t$  un vector columna ( $N \times 1$ ) de rendimientos reales para  $N$  activos (o portafolio de activos), el modelo sin restricciones lineal de  $K$  factores está dado por:

$$R_t = a + BR_{Kt} + \epsilon_t \quad (2.16)$$

$$E[\epsilon_t] = 0 \quad (2.17)$$

$$E[\epsilon_t \epsilon_t'] = \Sigma \quad (2.18)$$

$$E[R_{Kt}] = \mu_K, E[(R_{Kt} - \mu_K)(R_{Kt} - \mu_K)'] = \Omega_K \quad (2.19)$$

$$Cov[R_{Kt}, \epsilon_t'] = 0 \quad (2.20)$$

Donde  $B$  es la matriz ( $N \times K$ ) de sensibilidad de los factores,  $R_{Kt}$  es el vector columna ( $K \times 1$ ) de rendimientos reales de los factores del portafolio, y  $a$  y  $\epsilon_t$  son vectores columna ( $N \times 1$ ) de los interceptos de los rendimientos y perturbaciones respectivamente.  $O$  es una matriz ( $K \times N$ ) de ceros y  $\Sigma$  es la matriz de covarianza de las perturbaciones mientras que  $\Omega_K$  es la matriz de varianza-covarianza de los factores de rendimientos excesivos del portafolio.

Se asume que el término  $e$  de perturbación está no correlacionado con los rendimientos de los activos. Esto implica que en un mercado eficiente:

$$\mu = i\lambda_0 + B\lambda_K \quad (2.21)$$

Donde  $\mu$  es el vector columna ( $N \times 1$ ) de  $N$  rendimientos esperados de los activos,  $i$  es el vector columna ( $N \times 1$ ) de unos,  $\lambda_0$  es el parámetro libre de riesgo o beta cero  $\lambda_K$  es el vector columna ( $K \times 1$ ) de la prima de factores de riesgo.

Los modelos revisados en esta sección suelen ser la base para el análisis de activos financieros y usados no solo por académicos para la realización de estudios empíricos e investigación sino también por profesionales y analistas de mercado para aplicaciones prácticas sobre la valoración de precios. Sin embargo, estos métodos de valoración de activos financieros, conocidos como tradicionales, presentan diversos inconvenientes y limitaciones

que han sido objeto de críticas y discusiones en la literatura financiera. A continuación, se resumen algunos de los principales problemas asociados con estos métodos:

- **Suposiciones simplificadas:** Muchos métodos tradicionales, como el CAPM, se basan en suposiciones simplificadas que pueden no reflejar adecuadamente la complejidad del comportamiento real del mercado.
- **Expectativas racionales:** Algunos métodos presuponen que los participantes del mercado siempre toman decisiones de inversión basadas en expectativas racionales, lo cual puede no ser realista, ya que los inversores pueden estar influenciados por emociones, sesgos cognitivos y comportamientos irracionales.
- **Modelos estáticos:** Muchos modelos tradicionales consideran factores estáticos y no tienen en cuenta la dinámica cambiante de los mercados financieros. Esto puede llevar a una falta de capacidad para capturar eventos inesperados o cambios significativos en las condiciones del mercado.
- **Dependencia de datos históricos:** Algunos métodos dependen en gran medida de datos históricos, lo que puede no ser indicativo de futuros comportamientos del mercado, especialmente en entornos económicos cambiantes.
- **Igualdad de acceso a la información:** Los modelos tradicionales a menudo asumen que todos los participantes del mercado tienen acceso a la misma información, lo cual no refleja la realidad, ya que algunos inversores pueden tener acceso privilegiado o información exclusiva.
- **No linealidades y complejidades ignoradas:** Los modelos tradicionales asumen relaciones lineales entre variables y pueden no capturar de manera efectiva no linealidades o complejidades en la estructura del mercado.
- **No consideración de eventos extremos:** Los métodos tradicionales a veces no están diseñados para manejar eventos extremos o crisis financieras, lo que puede llevar a evaluaciones inexactas en situaciones de volatilidad extrema.
- **Falta de consideración de factores externos:** Algunos modelos no tienen en cuenta factores externos como cambios regulatorios, eventos geopolíticos o crisis económicas, lo que puede afectar significativamente el valor de los activos.

En respuesta a estos inconvenientes, ha habido un creciente interés en enfoques más avanzados, como métodos basados en aprendizaje automático y modelos estocásticos, que buscan abordar algunas de estas limitaciones y mejorar la precisión en la estimación y análisis de activos financieros.

### ***2.5 Métodos de análisis no lineales***

Los métodos econométricos están casi todos diseñados para detectar la estructura lineal de los datos financieros: análisis de series de tiempo, comprobaciones, datos, correlaciones, etc. Incluso los modelos de valuación de activos están basados en modelos lineales. Este énfasis no es sorprendente, puesto que el fundamento de la teoría sobre la que yacen está basado en modelos estadísticos lineales.

Sin embargo, muchos de los aspectos de la conducta económica pueden no ser lineales. La evidencia empírica y la introspección casual sugieren que las actitudes de los inversionistas hacia el riesgo y la expectativa de rendimientos son no lineales. Por ejemplo, los términos de varios productos financieros como las opciones y otros derivados son no lineales. Asimismo, las interacciones estratégicas entre los participantes del mercado, los procesos por los cuales la información es incorporada a los precios de los activos, y las dinámicas de las fluctuaciones, son todas inherentemente no lineales. Por lo tanto, la frontera natural para la econometría financiera es el modelado de fenómenos no lineales (Campbell et al., 2012). Esto es una gran ventana de oportunidades, dado que el conjunto de los modelos no lineales es mucho mayor que los lineales. Ahora bien, los modelos no lineales son mucho más difíciles de analizar y estos, rara vez, producen formas cerradas de expresiones que permitan su fácil manipulación y su implementación empírica (Campbell et al., 2012). Comúnmente, su único modo de análisis es computacional, lo que resulta complejo desde el punto de vista analítico, intuitivo y lineal. Entre estos modelos están la teoría de sistemas dinámicos, el análisis de series de tiempo no lineales, los modelos de volatilidad estocástica, la estadística no paramétrica, las redes neuronales artificiales, el aprendizaje, el cómputo evolutivo y los algoritmos bio inspirados, y la economía inspirada en la física; estos últimos cuatro rubros, correspondientes al campo denominado “Inteligencia Artificial” (IA), han resultado de gran

interés en los últimos años para la exploración de datos financieros (Gómez Martínez et al., 2020) (Sarin y Sharma, 2023).

A pesar de que muchas de las técnicas tienen la capacidad de descubrir una gran variedad de relaciones no lineales entre las variables, se les atribuye incurrir frecuentemente a una alta dependencia de los datos, lo que se conoce como “sobreajuste” y “sesgo a la información”. Además, las técnicas son conocidas por presentar costos computacionales altos, por la gran cantidad de iteraciones debido a los criterios y heurísticas de las que normalmente éstas hacen uso para la selección de las “condiciones iniciales” con el objeto de localizar la mejor solución. Cabe mencionar que estos problemas no son únicos de estos modelos, sino que se presentan también en los modelos lineales, no en el mismo grado, dada su naturaleza.

Entre los modelos usados más frecuentemente para el análisis y estudios en el mercado mexicano figuran el ARCH y el GARCH. Este tipo de modelos han probado capturar de manera adecuada el comportamiento de los rendimientos esperados en diversos mercados.

Un modelo de series de tiempo típico relaciona una serie de tiempo observada  $x_t$  a una secuencia de eventos  $\epsilon_t$ . En un análisis de series de tiempo lineal, se asume que los eventos no están correlacionados, pero no son necesariamente independientes e idénticamente distribuidos. Por el teorema de descomposición de Wold<sup>11</sup>, cualquier serie de tiempo puede ser escrita como una media móvil lineal de orden infinito, esta representación resume la varianza y la autocorrelación incondicional de las series (Campbell et al., 2012).

Una representación general es:

$$x_t = f(\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \dots) \quad (2.23)$$

---

<sup>11</sup> El teorema de descomposición de Wold (Wold, 1938) se ha usado para justificar el uso generalizado de los modelos ARIMA para series de tiempo estacionarias. Este teorema enuncia que un proceso aleatorio general puede ser descrito como la suma de dos procesos  $x(n) = xp(n) + xr(n)$ .

Donde los eventos se asumen ser media cero y varianza unitaria, y  $f(\cdot)$  es alguna función desconocida. La generalización de la representación lo hace difícil de trabajar. Un modelo más restrictivo puede escribirse como:

$$x_t = g(\epsilon_{t-1}, \epsilon_{t-2}, \dots) + \epsilon_t h(\epsilon_{t-1}, \epsilon_{t-2}, \dots) \quad (2.23)$$

La función  $g(\cdot)$  representa la media de  $x_t$  condicional sobre la información pasada, dado  $E_{t-1}(x_t) = g(\epsilon_{t-1}, \epsilon_{t-2}, \dots)$ . La mejora en  $x_t$  es proporcional al evento  $\epsilon_t$ , donde el coeficiente de proporcionalidad es la función  $h(\cdot)$ . El cuadrado de la función es la varianza de  $x_t$  condicional sobre la información pasada dado  $E_{t-1}[(x_t - E_{t-1}[x_t])^2] = h(\epsilon_{t-1}, \epsilon_{t-2}, \dots)^2$ . Los modelos con no linealidad  $g(\cdot)$  son llamados “no lineales en la media”, mientras que los modelos con no linealidad en  $h(\cdot)^2$  son llamados “no lineales en la varianza”.

Muchos modelos de series de tiempo se concentran en una forma de no linealidad u otra. Un modelo de media móvil no lineal sería de la forma:

$$x_t = \epsilon_t + \alpha \epsilon_{t-1}^2 \quad (2.24)$$

Donde  $g(\cdot) = \alpha \epsilon_{t-1}^2$  y  $h(\cdot) = 1$ . Este modelo es no lineal en la media, pero no en varianza. Por ejemplo, el modelo autorregresivo condicionalmente heterocedástico ARCH de primer orden toma la forma:

$$x_t = \epsilon_t \sqrt{\alpha \epsilon_{t-1}^2} \quad (2.25)$$

Donde  $g(\cdot) = 0$  y  $h(\cdot) = \sqrt{\alpha \epsilon_{t-1}^2}$ . Este modelo es no lineal en la varianza.

Una manera de entender la distinción entre la no linealidad de la media y la varianza es considerar los momentos del proceso  $x_t$ . Los modelos no lineales pueden construirse tal que los segundos momentos, la autoconvarianza,  $E[x_t x_{t-1}]$  todas sean cero para  $i > 0$ . En las dos expresiones anteriores ecuación 2.24 y ecuación 2.25 se puede confirmar que  $\epsilon_t$  es sistemáticamente distribuida, su tercer momento es cero. Para el caso de la media móvil no lineal se tendría  $E[x_t x_{t-1}] = E[(\epsilon_t + \alpha \epsilon_{t-1}^2)(\epsilon_{t-1} + \alpha \epsilon_{t-2}^2)] = \alpha E[\epsilon_{t-1}^3] = 0$  cuando  $E[\epsilon_{t-1}^3] = 0$ .

En la siguiente subsección se presentan los dos modelos no lineales de series de tiempo más conocidos: el ARCH y el GARCH.

### 2.5.1 Modelos ARCH y GARCH

Una observación básica acerca de los datos de rendimientos de los activos es que rendimientos grandes (de cualquier signo) tienden a ser seguidos por más rendimientos grandes (en cualquier signo), y lo mismo sucede con rendimientos pequeños. En otras palabras, la volatilidad de los rendimientos de los activos puede ser correlacionada en serie.

Para capturar la correlación serial de la volatilidad, Engle (1982) propone el modelo ARCH, el cual describe la varianza condicional como un rezago distribuido de eventos cuadrados:

$$\sigma_t^2 = \omega + \alpha(L)\eta_t^2 \quad (2.26)$$

Donde  $\alpha(L)$  es un polinomio en el operador de rezago. Para mantener la varianza positiva,  $\omega$  y los coeficientes en  $\alpha(L)$  deben ser no negativos.

Como una forma para modelar movimientos persistentes en volatilidad sin estimar un muy largo número de coeficientes en un polinomio de alto orden  $\alpha(L)$ , Bollerslev (1986) sugirió el modelo autorregresivo condicionalmente heterocedástico generalizado, o GARCH del inglés *Generalized Autorregresive Conditionally Heterocedastic*:

$$\sigma_t^2 = \omega + \beta(L)\sigma_{t-1}^2 + \alpha(L)\eta_t^2 \quad (2.27)$$

Donde  $\beta(L)$  es también el polinomio en el operador de rezago. Por analogía con modelos ARMA, lo anterior es conocido como un modelo GARCH( $p, q$ ) cuando el orden del polinomio  $\beta(L)$  es  $p$  y el orden del polinomio  $\alpha(L)$  es  $q$ . El modelo más comúnmente usado en la clase GARCH es el GARCH(1, 1) y puede ser escrito como:

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\eta_t^2 \quad (2.28)$$

$$= \omega + (\alpha + \beta)\sigma_{t-1}^2 + \alpha(\eta_t^2 - \sigma_{t-1}^2) \quad (2.29)$$

$$= \omega + (\alpha + \beta)\sigma_{t-1}^2 + \alpha\sigma_{t-1}^2(\epsilon_t^2 - 1) \quad (2.30)$$

En la segunda igualdad, ecuación 2.29, el término  $(\eta_t^2 - \sigma_{t-1}^2)$  significa cero, el condicional en tiempo  $t - 1$  de información, y puede ser tomado como el evento de volatilidad. El coeficiente  $\alpha$  mide la extensión a la cual un evento de volatilidad en este momento se alimenta dentro del siguiente periodo de volatilidad, mientras que  $(\alpha + \beta)$  mide la razón a la que este efecto muere en el tiempo. La tercera igualdad, ecuación 2.30, reescribe el suceso de volatilidad como  $\sigma_{t-1}^2(\epsilon_t^2 - 1)$ , el cuadrado de la desviación estándar normal menos su media, es decir, una variable aleatoria  $\chi^2(1)$  degradada, multiplicada por la volatilidad pasada  $\sigma_{t-1}^2$ .

El modelo GARCH(1, 1) puede ser escrito también en términos de sus implicaciones para los sucesos cuadrados  $\eta_{t+1}^2$ :

$$\eta_{t+1}^2 = \omega + (\alpha + \beta)\eta_t^2 + (\eta_{t+1}^2 - \sigma_t^2) - \beta(\eta_t^2 - \sigma_{t-1}^2) \quad (2.31)$$

Esta representación permite observar que el modelo GARCH(1, 1) es un modelo ARMA(1, 1) para eventos cuadrados; pero el modelo ARMA(1, 1) tiene sucesos heterocedásticos, mientras los eventos  $(\eta_{t+1}^2 - \sigma_t^2)$  son por sí mismos heterocedásticos.

Finalmente, la expresión 2.28, representando un modelo GARCH univariado, puede ser extendida para modelar una configuración multivariada permitiendo la agregación de una matriz de producto cruz o cartesiano, como:

$$\sigma_{ij,t}^2 = \omega_{ij} + \beta_{ij}\sigma_{ij,t-1} + \alpha_{ij}\eta_{it}\eta_{jt} \quad (2.32)$$

### 2.5.2 Aprendizaje automático

Sumado a las técnicas econométricas tradicionales se tienen los modelos de análisis de datos basados en el amplio abanico de técnicas agrupadas bajo la denominación de “inteligencia artificial” (IA) entre las que se encuentra el “aprendizaje automático” (*Machine Learning*, ML).

La inteligencia artificial puede definirse, de forma no estrictamente rigurosa, como el conjunto de tareas no preprogramadas de tipo cognitivo –percepción, razonamiento,

aprendizaje, resolución de problemas–, tradicionalmente asociadas a las funciones propias de las mentes humanas, pero que son efectuadas por sistemas computacionales artificiales.

Las tecnologías de IA más destacadas suelen estar incluidas dentro del denominando “aprendizaje automático”. Se trata de sistemas basados en algoritmos que se entrenan con datos. De esta forma pueden detectar patrones y aprender a hacer predicciones y recomendaciones mediante el procesamiento de datos, en lugar de recibir instrucciones de programación explícitas. También se adaptan en respuesta a nuevos datos y experiencias, mejorando así su eficacia con el tiempo. El volumen y la complejidad de los datos que se generan actualmente, imposibles de ser razonablemente procesados por los seres humanos, han aumentado tanto el potencial como la necesidad de esta tecnología.

El aprendizaje automático sirve como una aproximación adicional que, junto a los análisis técnico y fundamental, ofrece a los investigadores una amplia variedad de opciones para poder explotar la información financiera y localizar nuevas relaciones en esta (Sonkavde et al., 2023) (Sood et al., 2023). El aprendizaje automático puede proveer soluciones a problemas tales como predicción y clasificación de precios de acciones, gestión de portafolios, *trading* algorítmico, análisis de sentimiento de mercado, evaluación de riesgo, etc. (Argade et al., 2022).

Actualmente, dentro del aprendizaje automático existen en la literatura numerosos técnicas, métodos y combinaciones entre estos que ofrecen diversidad de algoritmos para la predicción y clasificación de datos. Algunos de los algoritmos más populares, y que se consideran dentro del alcance de este trabajo, están el bosque aleatorio, la regresión logística, las máquinas de soporte vectorial y el perceptrón multicapa, este último perteneciente a las redes neuronales artificiales. Debido al éxito de la inteligencia artificial y el aprendizaje automático, la cantidad de publicaciones referentes al análisis de los mercados financieros mediante el empleo de estas técnicas se ha incrementado considerablemente en el campo de las finanzas (Loke, 2017) (Sonkavde et al., 2023).

### 2.5.2.1 Perceptrón multicapa

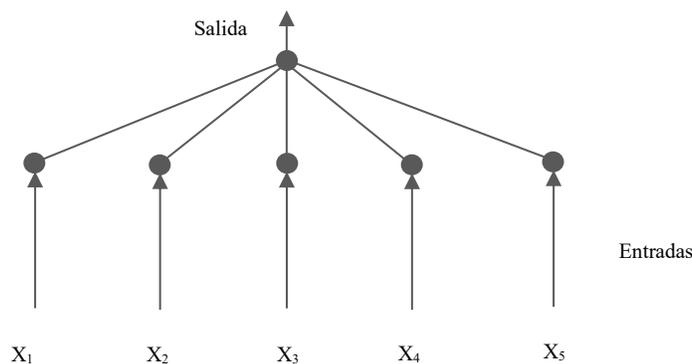
El Perceptrón Multicapa, MLP por sus siglas en inglés *Multi Layer Peceptron*, es uno de los tipos de redes neuronales más populares en la literatura, dado que este es el ejemplo más simple de una red neuronal (Campbell et al., 2012) (García et al., 2013) (Ansah et al., 2022). Es el modelo de umbral binario en el que la variable de salida  $Y$ , tomando únicamente valores de 0 y 1, tiene relación no lineal con una colección de  $J$  variables de entrada  $X_j, j = 1 \dots J$ , de la siguiente manera:

$$Y = g\left(\sum_{j=1}^J \beta_j X_j - \mu\right) \quad (2.33)$$

$$g(u) = \begin{cases} 1 & \text{Si } u \geq 0 \\ 0 & \text{Si } u < 0 \end{cases} \quad (2.34)$$

Cada  $X_j$  tiene peso  $\beta_j$ , llamado “fuerza de conexión”, y estos son sumados en todas sus entradas. Si la sumatoria excede el umbral  $\mu$ , entonces la neurona artificial es activada de acuerdo a la función  $g()$ , de otra forma se mantiene intacta. Esta simple neurona se representa como en la figura 2.1, en la que se observa la capa de entrada conectada a la capa de salida.

Figura 2.1. Modelo de umbral binario

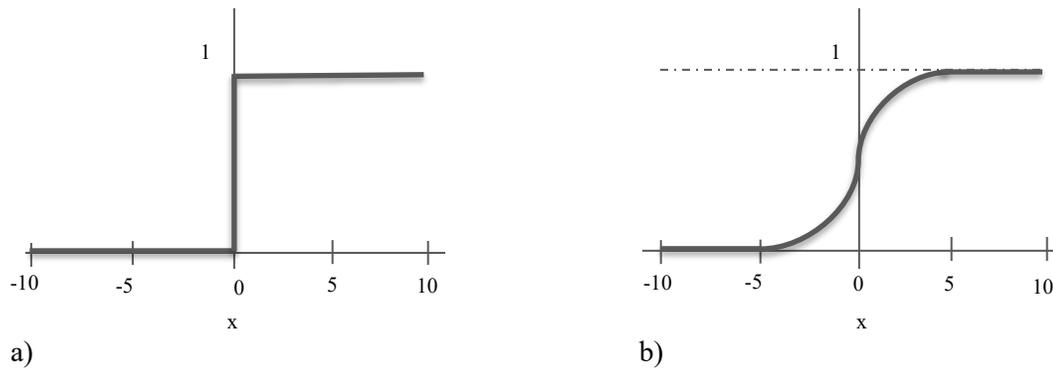


Fuente: Elaboración propia.

Diferentes modelos de redes neuronales parten de esta generalización. En particular, se recurre a cambiar la función de activación para permitir salidas con valores continuos. Por ejemplo, utilizar la función logística (figura 2.2.a y 2.2.b):

$$g(u) = \frac{1}{1 + e^{-u}} \quad (2.35)$$

Figura 2.2. Gráficas de función de activación: a) escalón y b) logística



Fuente: Elaboración propia.

De la misma forma y sin perder la generalización, se puede asignar  $\mu$  a cero, dado que es siempre posible modelar un nivel de activación distinto de cero mediante la definición de la primera entrada  $X_j = 1$ ; este caso, el negativo de dicha entrada de fuerza de conexión,  $-\beta_j$ , llega a ser el nivel de activación.

La más importante extensión de modelo de umbral binario es la introducción de la “capa oculta” entre la capa de entradas y la capa de las salidas (figura 2.3):

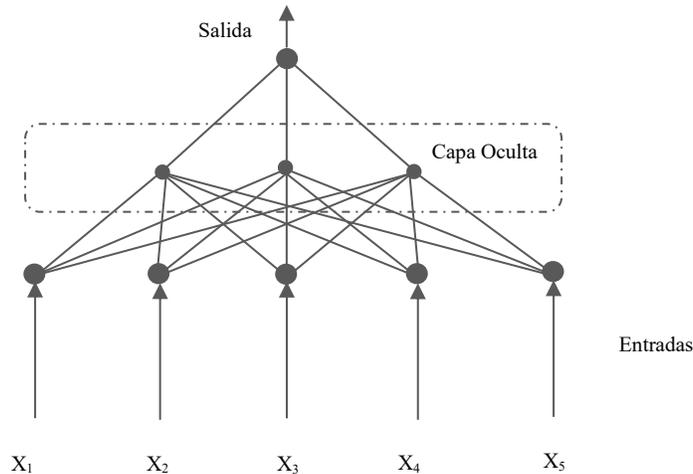
$$Y = h\left(\sum_{k=1}^K \alpha_k g(\beta_j X_j)\right) \quad (2.36)$$

$$\beta_k = [\beta_{k1} \quad \beta_{k2} \quad \dots \quad \beta_{kJ}]', X = [X_1 \quad X_2 \quad \dots \quad X_J]' \quad (2.37)$$

Donde  $h( )$  es otra función de activación arbitraria no lineal. En este caso las entradas son conectadas a múltiples unidades ocultas y en cada unidad oculta estas tienen pesos diferentes,

aunque son transformadas por la misma función  $g(\cdot)$ . La salida de cada unidad oculta es entonces pesada nuevamente y ahora por  $\alpha_k$ 's, y sumadas y transformadas por la segunda función de activación  $h(\cdot)$ . Esta configuración, corresponde a un perceptrón multicapa de una capa oculta, ha sido la más común y ampliamente utilizada en las aplicaciones y cuya topología más compleja se observa en la figura 2.3. Esta configuración puede ser generalizada agregándose más capas ocultas, de ahí que adquiera el término de “perceptrón multicapa”.

Figura 2.3 Perceptrón multicapa con una única capa oculta



Fuente: Elaboración propia.

Para un conjunto de entradas y salidas  $\{X_t, Y_t\}$ , el resultado del MLP equivale a la estimación de los parámetros de la red del MLP, los vectores  $\beta_k$  y los escalares  $\alpha_k, k = 1, \dots, K$ , típicamente minimizando la sumatoria de las desviaciones cuadradas entre la salida y la red:

$$\sum_y [Y_t - \sum_k \alpha_k g(\beta_k' X)]^2 \quad (2.38)$$

Al proceso de estimación de parámetros se llama “fase” o “etapa” de “entrenamiento” de la red, también se le conoce como “etapa de aprendizaje”. Incluso con una única capa oculta, el MLP posee una característica de estimación universal: el MLP puede estimar casi cualquier función no lineal a un arbitrario grado de precisión con un adecuado número de unidades ocultas. Este tipo de característica de estimación puede ser encontrada en muchas de las técnicas de estimadores no lineales. Esto no significa que deban tener alto desempeño para

todos los problemas, sino que para un determinado conjunto de datos es posible que alguna de las técnicas domine a las otras en términos de exactitud, es decir, del desempeño.

Una de las principales ventajas del MLP es que puede estimar problemas complejos usando una red de funciones simples: funciones de base radial, polinomiales, gaussianas, por mencionar algunas.

Aunque este tipo de técnicas de representación de la información poseen grandes atributos para la estimación de funciones no lineales, también presentan varias desventajas: particularmente, no hay procedimientos que sean ampliamente aceptados para determinar la arquitectura para una aplicación específica como las capas ocultas, la función de activación, etc. que en su mayoría se basan en heurísticas y observaciones empíricas.

### ***2.5.2.2 Regresión Logística***

La regresión logística es una técnica de aprendizaje estadístico categorizada en los métodos de clasificación supervisada de aprendizaje automático, que ha ganado reputación las últimas décadas especialmente en el sector financiero. Entre las razones que se atribuyen a este modelo para su utilización destacan su no linealidad y su baja relación con supuestos estadísticos estrictos, a diferencia de lo que sucede con la regresión lineal. Este tipo de regresión estima la probabilidad de que un evento ocurra para una variable dicotómica dependiente (Hua et al., 2007) (Campbell et al., 2012).

La regresión logística es preferible a otras técnicas en finanzas por varias razones:

1. La regresión lineal multivariada es inapropiada para modelado de valores dependientes dicotómicas ya que la distribución de errores no es normal y los valores pronosticados no son interpretados como probabilidades, entre 0 y 1.
2. Los análisis, como el discriminante, requieren que las variables independientes sean normalmente distribuidas y que las matrices de covarianza de los grupos estudiados sean iguales.

3. La regresión logística expresa la ecuación de la regresión lineal en términos logarítmicos y el modelo es estimado usando el método de máxima probabilidad. De esta manera la regresión localiza las estimaciones de parámetros mediante la función de máxima probabilidad la cual expresa la probabilidad de los datos observados como funciones de parámetros desconocidos.

El procedimiento es usado de manera iterativa hasta encontrar todos los parámetros. Supóngase que hay únicamente dos clases. La regresión logística reemplaza la variable objetivo original:

$$Pr[1|a_1, a_2, \dots, a_k] \quad (2.39)$$

La cual no puede ser estimada exactamente utilizando una función lineal, con:

$$\frac{\log(Pr[1|a_1, a_2, \dots, a_k])}{(1 - Pr[1|a_1, a_2, \dots, a_k])} \quad (2.40)$$

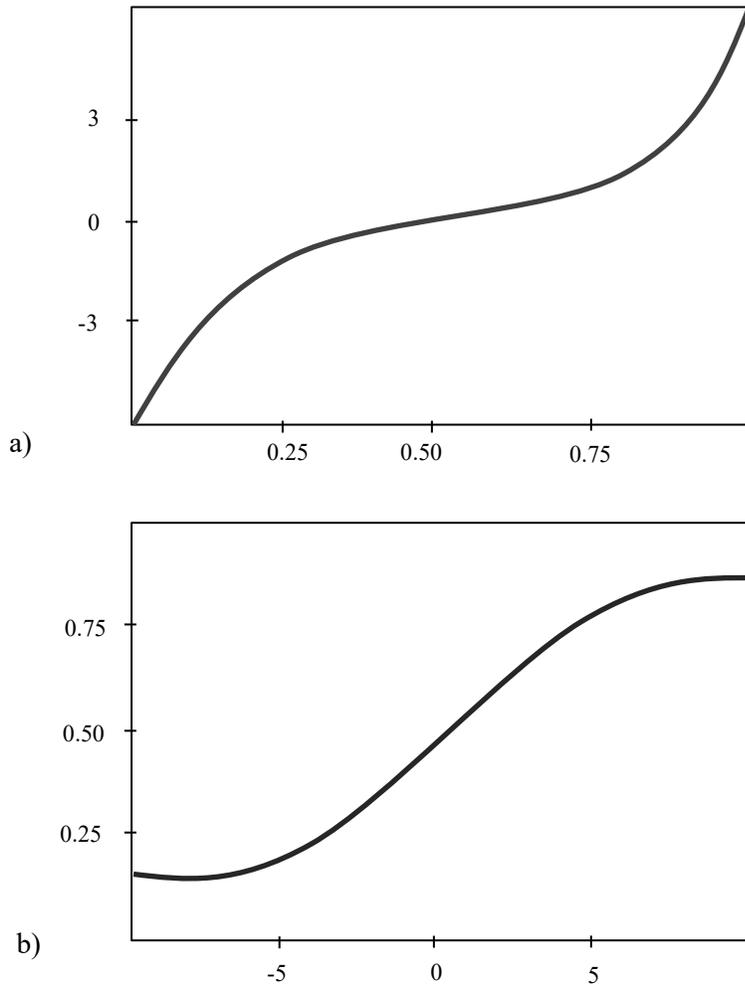
Los valores resultantes no están restringidos al intervalo de 0 a 1, es decir pueden caer entre el infinito negativo y el positivo. La figura 2.4 muestra la función de transformación, la cual es frecuentemente llamada “transformación *Logit*”.

La variable transformada es estimada usando una función lineal en la misma manera como la generada por regresión lineal. El modelo resultante con pesos  $w$ , es:

$$Pr[1|a_1, a_2, \dots, a_k] = \frac{1}{(1 + \exp[-w_0 - w_1 a_1 - \dots - w_k a_k])} \quad (2.41)$$

La figura 2.4.b muestra en ejemplo de esta función en una dimensión con dos pesos  $w_0 = 0.5$  y  $w_1 = 1$ .

Figura 2.4. Regresión logística: a) la transformada *logit* y b) ejemplo de función logística



Fuente: Elaboración propia.

Igual que como una regresión lineal, los pesos son aquellos de mejor ajuste arrojados en la fase de entrenamiento de datos. La regresión lineal mide el ajuste usando el error cuadrático. En la regresión logística se utiliza el modelo de probabilidad logarítmica, que está dado por:

$$\sum_{i=1}^n (1 - x^{(i)}) \log(1 - \text{Pr}[1|a_1^{(i)}, a_2^{(i)}, \dots, a_k^{(i)}]) + x^{(i)} \log(\text{Pr}[1|a_1^{(i)}, a_2^{(i)}, \dots, a_k^{(i)}]) \quad (2.42)$$

Donde  $x^{(i)}$  es 0 o 1. Los pesos  $w_i$  son seleccionados por el máximo de la probabilidad logarítmica; hay varios métodos para resolver el problema de maximización. El método más simple es resolver iterativamente una secuencia de problemas de regresión de mínimos cuadrados hasta que converja con el máximo.

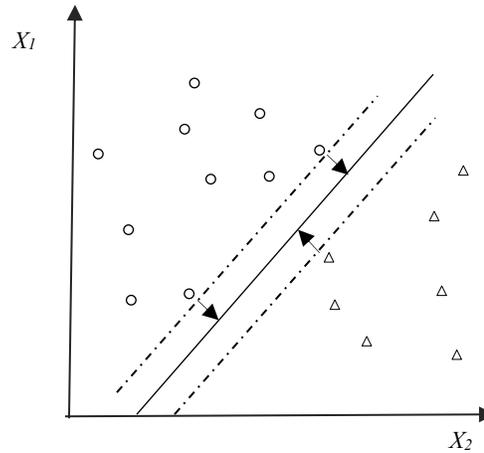
Para generalizar la regresión logística a varias clases, se resuelve independientemente la regresión logística para cada clase. Esto puede conducir a que las probabilidades resultantes no sumen 1. Para obtener la probabilidad correcta es necesario acoplar los modelos para cada clase. Esto produce un problema de optimización de su integración, pero existen métodos para resolver esto.

Una manera muy general y conceptualmente sencilla de resolver problemas de multi clasificación es conocida como “clasificación de paridad”. Un clasificador es construido para cada par de clases utilizando únicamente instancias de esas dos clases. La salida sobre un patrón desconocido de entrada se basa en la clase que recibe más votos. Este método generalmente produce resultados más exactos en términos de error de clasificación. Puede ser usado para producir estimaciones aplicando el método de paridad acoplada, el cual calibra la estimación de probabilidad individual de los diferentes clasificadores.

### ***2.5.2.3 Máquinas de soporte vectorial***

Las máquinas de soporte vectorial, o SVM, del inglés *Support Vector Machines*, son una técnica bastante útil para la clasificación de datos. El objetivo de este método es el de identificar un hiper plano en un espacio  $n$  dimensional ( $n$  número de factores o atributos) que discrimine o diferencie los puntos (elementos) de datos; básicamente, resuelven el problema de manera espacial. Así es posible que varios hiper planos conformen o sean identificados para resolver un determinado problema (figura 2.5). La métrica utilizada para medir la separación entre clases es la de la máxima distancia (Hua et al., 2007) (Liu et al., 2021).

Figura 2.5. Hiper plano de separación óptima



Fuente: Elaboración propia.

Sean dos clases, dado un conjunto de puntos de datos  $G = \{(x_i, d_i)\}_i^n$ ,  $x_i$  como vector de entrada,  $d_i$  el valor deseado  $d \in \{0, 1\}$  es conocido como el objetivo binario y  $n$  el total de patrones de datos, el SVM usa la siguiente función de estimación:

$$y = f(x) = w\Phi(x) + b \quad (2.43)$$

Donde  $\Phi(x)$  es el espacio característico de alta dimensión el cual es no linealmente mapeado del espacio de entrada  $x$ . Los coeficientes  $w$  y  $b$  son estimados minimizando:

$$R_{SVM_s} = C \frac{1}{n} \sum_{i=1}^n L_\varepsilon(d_i y_i) + \frac{1}{2} \|w\|^2 \quad (2.44)$$

$$L_\varepsilon(d, y) = \begin{cases} |d - y| - \varepsilon, & |d - y| > \varepsilon \\ 0, & |d - y| \leq \varepsilon \end{cases} \quad \text{otherwise } \varepsilon \quad (2.45)$$

Donde:

- $C \frac{1}{n} \sum L_\varepsilon(d_i y_i)$ , el primer término, es conocido como el error empírico (riesgo) el cual es medido por la función insensible al ruido  $\varepsilon$ .
- $\frac{1}{2} \|w\|^2$ , el segundo término, es llamado el termino regularizado.
- $\varepsilon$  es el tamaño de SVM.

- $C$  es la constante de regularización determinando el intercambio entre el error empírico y el término regularizado.

Los dos términos mencionados son parámetros prescritos por el usuario y son seleccionados empíricamente.

El SVM construye un modelo lineal para implementar fronteras de clase no lineales a través de la transformación de las entradas en el espacio característico de alta dimensión. Para el caso no linealmente separable, una versión de alta dimensión es representado como:

$$Y = b + \sum \alpha_i y_i K(x_i, x_j) \quad (2.46)$$

La función  $K(x_i, x_j)$  es definida como la “función kernel”. El valor es igual al producto interno de dos vectores  $x_i$  y  $x_j$  en el espacio característico  $\Phi(x_i)$  y  $\Phi(x_j)$ , esto es,  $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ . Hay algunos kernels diferentes para generación de productos internos para construir máquinas con tipos diferentes de superficies de decisión no lineales dentro del espacio de entrada. Entre los diferentes kernels del modelo que minimiza la estimación, se selecciona el mejor modelo. Algunos ejemplos comunes de función kernel son los polinomiales  $K(x, y) = (xy + 1)^d$  y la función de base radial gaussiana  $K(x, y) = \exp(-1/\delta^2(x - y)^2)$ , donde  $d$  es el grado del kernel polinomial y  $\delta^2$  es el ancho de banda de la función kernel con base radial gaussiana (Kumar y Thenmozhi, 2006).

La clasificación basada en la óptima separación de hiperplanos puede llegar a ser altamente eficiente; sin embargo, el inconveniente es que sea muy sensible a o dependiente de observaciones individuales.

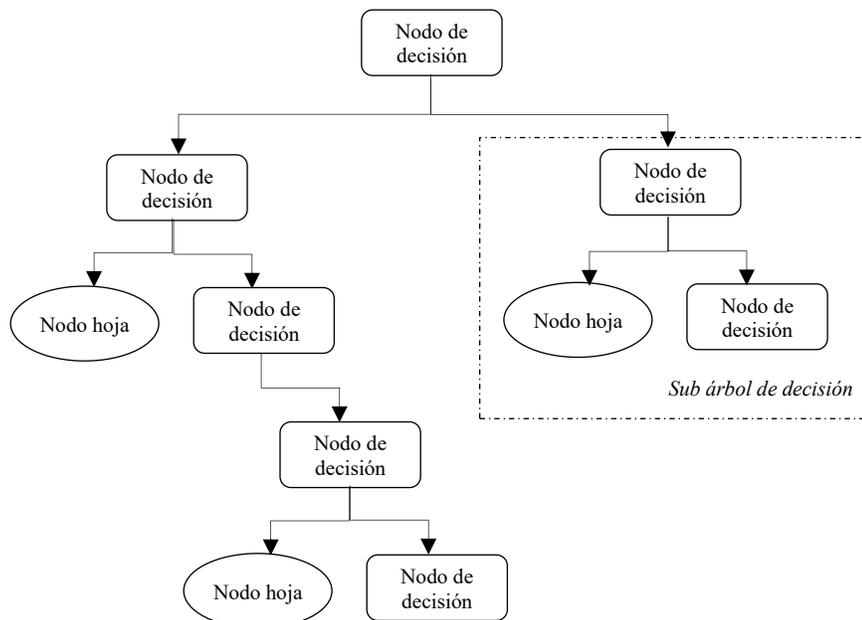
#### **2.5.2.4 Bosque aleatorio**

Los Bosques Aleatorios, o *Random Forests* en inglés, son un algoritmo de aprendizaje automático que se utiliza tanto en tareas de clasificación como en regresión. Este algoritmo es una técnica que combina múltiples árboles de decisión y es conocido por su simplicidad y

gran efectividad. Este método es usado comúnmente como clasificador y ha sido probado en problemas de detección de intrusos, filtrado de contenido de información, análisis de sentimiento y como variedad de aplicaciones de procesamiento de imágenes. En el área financiera se ha probado para predicción de precios, riesgos y quebranto de empresas (Breiman, 2001) (Kumar y Thenmozhi, 2006) (Loke, 2017).

Un árbol de decisión es un tipo de método de clasificación el cual construye un modelo en la forma de un árbol. Esta técnica consiste en dividir la información en subconjuntos más pequeños, por lo que el árbol de decisión es creado a partir de la sucesión iterativa de esta descomposición. De esta manera el árbol está compuesto por “ramas” y “nodos”, siendo los nodos de las decisiones de división en subconjuntos más pequeños, las ramas el resultado de esas divisiones, esto es, ruta de acción o consecuencias de decisiones, y van de unos nodos a otros; los nodos terminales, o nodos “hoja”, son los nodos más alejados, que no dan lugar a nuevas particiones, y por lo tanto representan los resultados (figura 2.6).

Figura 2.6. Representación gráfica de un árbol de decisión.



Fuente: Elaboración propia.

Los árboles de decisión son escalables y de los algoritmos de aprendizaje automático más fáciles de usar, dado que ofrecen resultados muy buenos sin la necesidad de emplear gran esfuerzo en la optimización de parámetros.

El bosque aleatorio es un algoritmo de clasificación de los conocidos como “ensamblados”, que usa una colección de combinaciones de árboles de decisión, el cual fue introducido por Breiman (2001). El método usa una selección aleatoria de atributos para realizar las bifurcaciones o ramificaciones de cada nodo de los árboles de decisión. Este factor aleatorio permite que los árboles individuales no estén correlacionados, lo cual hace que el bosque sea robusto al ruido y resistente a la sobreclasificación. Cada uno de los árboles creados al final del árbol transversal será considerado un voto para la clasificación de la clase de entrada, y la suma del total de votos que se obtenga será el resultado de la clasificación. Un único árbol aleatorio pudiera ser ligeramente mejor que una clasificación aleatoria, pero al utilizarse dentro de una combinación de varios de estos dentro de un ensamble se obtiene una mejor precisión (Loke, 2017).

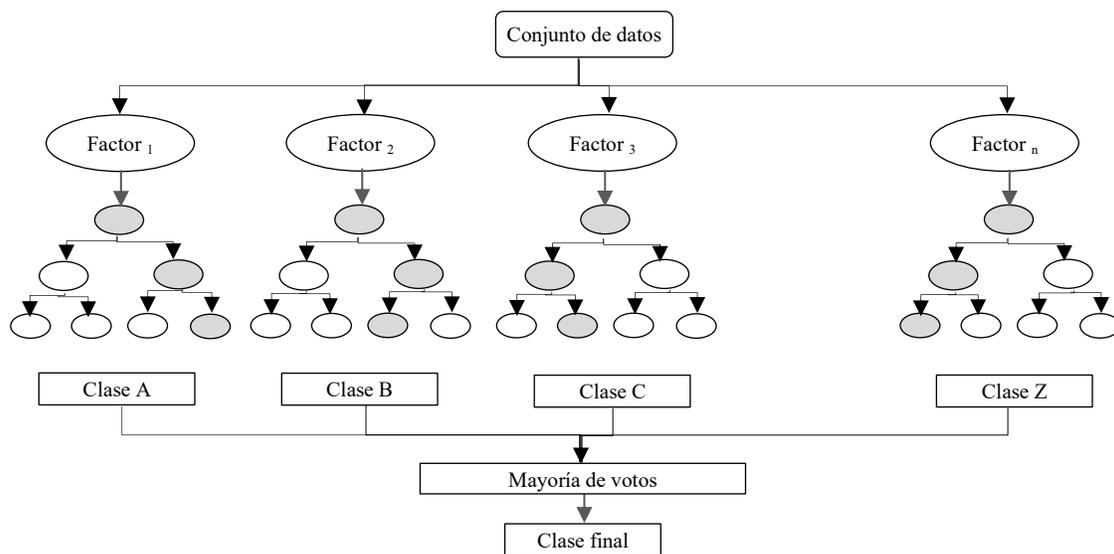
El proceso de generación de un bosque aleatorio es el siguiente (figura 2.7) (Sonkavde et al., 2023):

1. *Creación de los árboles de decisión*: El primer paso en la construcción de un Bosque Aleatorio implica la creación de múltiples árboles de decisión. Cada árbol se entrena de manera independiente utilizando una porción aleatoria del conjunto de datos de entrenamiento. Esta selección aleatoria se realiza tanto en las observaciones (filas) como en las características (columnas). Esto introduce variabilidad en el proceso de entrenamiento.
2. *Votación o promedio*: Una vez que se han entrenado todos los árboles, se utiliza un proceso de votación (en el caso de clasificación) o promedio (en el caso de regresión) para combinar las predicciones de cada árbol y llegar a una predicción final. En clasificación, la clase que obtiene la mayoría de votos se elige como la predicción final. En regresión, las predicciones de todos los árboles se promedian.
3. *Reducción del sobreajuste*: El proceso de entrenamiento de múltiples árboles con selecciones aleatorias ayuda a reducir el sobreajuste. Cada árbol podría sobre

ajustarse ligeramente a diferentes partes del conjunto de datos, pero al combinar sus predicciones, se obtiene un modelo más robusto y generalizado.

4. *Importancia de características*: Los Bosques Aleatorios también proporcionan una medida de la importancia de cada característica en la tarea de predicción. Esto puede ser útil para comprender qué características son más relevantes en un problema dado.

Figura 2.7. Representación gráfica del proceso de generación de un bosque aleatorio.



Fuente: Elaboración propia.

Los Bosques Aleatorios son conocidos por su rendimiento robusto y su capacidad para manejar conjuntos de datos grandes y complejos. Estos se utilizan en una variedad de aplicaciones, como clasificación de imágenes, predicción del mercado de valores, diagnóstico médico y más, debido a su versatilidad y eficacia.

## 2.6 *Sobreajuste y sesgo de datos*

Estas características que se le adjudican a cualquier técnica de estimación afectan por igual a lineales y no lineales. La sobre clasificación o sobreajuste (*overfitting* en inglés) se produce cuando un modelo ajusta demasiado a los datos: ruido blanco y relaciones lineales y no lineales. Entre las causas principales están (Campbell et al., 2012):

- a) Número reducido de factores, o muchos, pero con alta similitud.
- b) Desempeño alto del modelo en pruebas dentro de la muestra, pero bajo en pruebas fuera de la muestra.
- c) Poca cantidad de registros en la muestra o con variedad reducida.

El sesgo de datos, o *data snooping* en inglés, está relacionado con un alto desempeño del modelo en pruebas fuera de la muestra, pero con significado dudoso. Comúnmente sucede este sesgo cuando se analiza demasiado la especificación inicial para obtener un modelo final que ajusta a los datos. Incluso si el modelo es incorrecto, con la búsqueda exhaustiva de conjuntos de datos y parámetros a utilizar, es posible encontrar una combinación con la que se obtenga un ajuste aceptable. Se dice entonces que el ajuste es espurio y rebuscado. Otra causa bastante común es la consideración de factores o de combinaciones de estos utilizando registros adelantados, es decir, que motivados por tener al alcance el conjunto total de datos se incurra, por descuido, en perder la temporalidad de las dependencias con respecto a los valores estimados.

No hay maneras de detectar estos problemas dado que se emplean los mismos procedimientos de análisis que para construir modelos legítimos. Sin embargo, la poca experiencia de los analistas para trabajar con experimentos controlados y la alta dependencia en la inferencia estadística, en el intento de tener la comprensión de los datos, está muy relacionadas, y pueden conducir al sesgo.

Entre las formas de mitigar estos problemas se mencionan las siguientes (Campbell et al., 2012):

- a) El análisis de la especificación sistémica, que sugiere el uso de la distribución estadística adecuada para inferencia con lo que se intenta encontrar resultados con determinada significatividad.
- b) La evaluación adecuada del desempeño de los modelos con datos fuera de la muestra, con suficiente cantidad de elementos y conjuntos variados.
- c) El uso de diferentes métodos de evaluación y de errores: estadísticos, como el error cuadrático y cuadrático medio; y validación cruzada.

- d) La que se considera como la más efectiva es la búsqueda de consideraciones teóricas y de fundamentos sustentados de la especificación del modelo que pueden ser basados en elementos matemáticos, bases económicas o financieras, conductuales, psicológicas o de algún tipo de evidencia sobre el fenómeno de estudio.

## 2.7 Evaluación de desempeño

Para evaluar un modelo de clasificación entre las métricas preferidas están la exactitud, sensibilidad (o *recall* en inglés), precisión y el factor F (*F-score* factor en inglés) y para regresiones o predicciones se tiene el error cuadrático medio (o RMSE por sus siglas en inglés *root mean square error*) y el porcentaje de error medio absoluto (MAPE por sus siglas en inglés *mean absolute porcentaje error*) (Sonkavde et al., 2023).

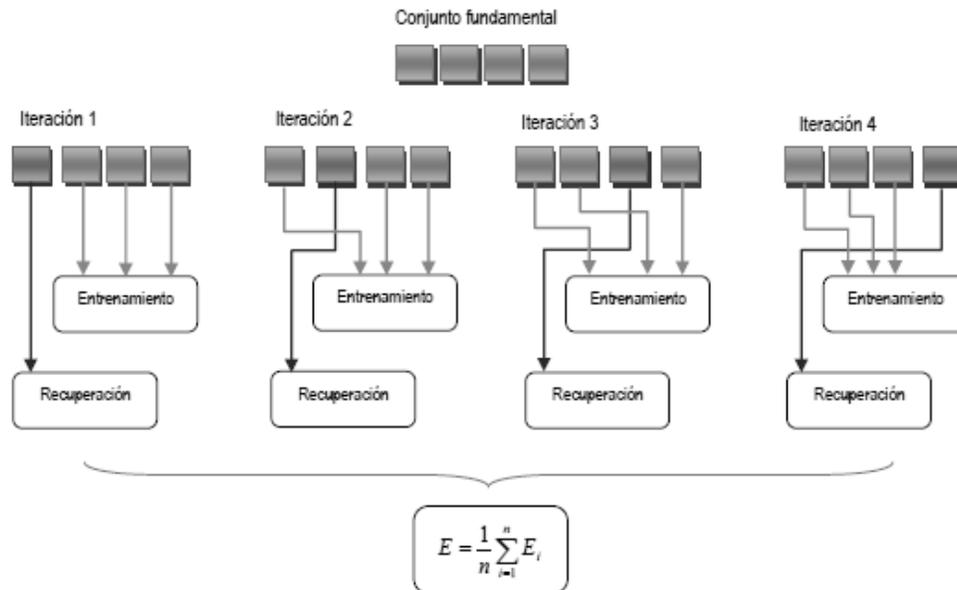
En este trabajo se usa el porcentaje de exactitud como métrica de evaluación del desempeño, también conocido como “porcentaje de recuperación” del clasificador. Para esta evaluación, el clasificador evalúa cada elemento aprendido con clase conocida, y el resultado es etiquetado como correcto si la clase pronosticada corresponde a la asignada, o erróneo, en el caso contrario. De esta forma, el factor de recuperación es la proporción de aciertos realizados sobre un conjunto completo de instancias y es utilizado para medir de forma estadística el desempeño de un clasificador (Chávez Sánchez, 2012, p. 29).

La manera más directa de evaluar un clasificador es utilizando todos los datos disponibles para primeramente entrenar el modelo y posteriormente utilizar todos los datos para probarlo. El inconveniente de este procedimiento yace en que el modelo presente alto ajuste a los datos, pero incurriendo en sobre ajuste de tal forma que sea muy sensible a datos desconocidos.

Otra alternativa de medición es con validación cruzada, o *cross validation* en inglés, el cual es uno de los métodos más utilizados para la medir el desempeño de un clasificador. Esta técnica de evaluación consiste en dividir el conjunto de datos en un determinado número  $k$  de partes, tal que una de las partes sea tomada como conjunto de prueba, y las restantes  $k - 1$  se usen como datos de entrenamiento de modelo. Este procedimiento es repetido para cada

uno de los subconjuntos  $k$ . El porcentaje de exactitud es obtenido al aplicar la media aritmética sobre los resultados de cada una de las  $k$  iteraciones como se muestra en la figura 2.8 (Chávez Sánchez, 2012).

Figura 2.8. Proceso de ejecución de validación cruzada.



Fuente: (Chávez Sánchez, 2012, p. 30).

La validación cruzada deriva del método de retención, o *holdout* en inglés, el cual consiste en dividir el conjunto de datos en dos partes, tal que una se emplee en la fase de entrenamiento y la otra en la fase de recuperación; una alternativa común es dividir en tercios, de los cuales dos son para la primera etapa y el restante para la segunda, respectivamente. En este método hay una única iteración y el porcentaje de exactitud es el factor de resultado. No se considera un método confiable, ya que la muestra considerada para realizar la fase de entrenamiento podría no ser la de mejor calidad de las del conjunto en cuanto a la utilidad de la información provista por sus elementos.

El método “deja uno afuera”, o *leave one out* en inglés, es una variante del método de validación cruzada que consiste en realizar la partición del conjunto en muestras del tamaño del conjunto a evaluar, es decir, que se hacen tantas muestras como haya elementos en el conjunto, tal que la cardinalidad de los subconjuntos es 1; al cual se asigna únicamente un

elemento para la etapa de prueba y el  $n - 1$  restante para la de entrenamiento, de ahí el nombre del método. Cada elemento o patrón en el conjunto es considerado en ambas etapas, entrenamiento y recuperación, por lo menos una vez. El proceso es repetido  $n$  veces y el resultado es obtenido al aplicar la media aritmética de todos los resultados de las iteraciones. Es un método muy preciso, ya que todos los elementos son considerados en ambas etapas lo cual robustece al modelo, pero de ejecución lenta (Chávez Sánchez, 2012, p. 30).

### ***Resumen***

En este capítulo, dadas las anomalías introducidas en el capítulo anterior, en primer lugar, se han presentado modelos tradicionales de valuación de activos, como el CAPM y el APT, así como otros modelos multivariantes lineales y no lineales. De los no lineales se han considerado tanto los basados en modelos econométricos de series temporales como aquellos que emplean técnicas de inteligencia artificial, en concreto aprendizaje automático. No hay un método de análisis de anomalías por excelencia. Si bien los modelos econométricos conocidos ofrecen resultados eficaces, los métodos basados en aprendizaje automático han resultado tener excelente ajuste en muchos casos donde las relaciones entre variables no son claras o evidentes y, como se ha mencionado, han mostrado desempeño sobresaliente cuando han sido probados en problemas complejos, incluyendo el área financiera y riesgo. Se han detallado las metodologías pertenecientes al área de aprendizaje automático más representativas, por ser las utilizadas en el presente trabajo. Asimismo, se han señalado también dos puntos de los que adolecen los modelos, el sesgo de datos y la sobreclasificación, y la manera de poder mitigar y prevenir su ocurrencia, al menos en cuanto al proceso utilizado de modelado. Por último, se han indicado las técnicas a aplicar para evaluar el desempeño de los modelos de clasificación, escogiendo para este trabajo el porcentaje de exactitud o porcentaje de recuperación del clasificador como métrica más adecuada.

En el siguiente capítulo se mostrará cómo las generalidades mencionadas de modelado de datos y métodos de análisis son utilizadas en conjunto para la validación de la presencia de las anomalías de mercado empleando a los clasificadores como método de ajuste.

## Capítulo 3

### METODOLOGÍA PROPUESTA

En el capítulo anterior se sentaron las bases teóricas necesarias para comprender el presente apartado. Se introdujeron los conceptos clave de análisis de datos, junto con algunas de las corrientes y paradigmas metodológicos de aprendizaje automático. Además, se abordaron algunos conceptos sobre los problemas comunes al trabajar con datos y se revisaron los métodos de evaluación de modelos. Todos estos elementos resultan de gran utilidad para comprender de mejor forma el marco de trabajo del modelo de análisis de anomalías desarrollado, el cual tiene como objetivo validar la hipótesis planteada.

En el presente capítulo se explica el procedimiento detallado para obtener los modelos que son el eje principal de este trabajo. Se presentan los criterios propuestos para la determinación de las anomalías, el flujo para la creación de los modelos y su evaluación.

A continuación, se muestran los detalles que dan fundamento al trabajo y se realiza un breve compendio de los cálculos de las anomalías y el procedimiento de análisis.

#### ***3.1 Métodos de clasificación***

Como se ha mencionado previamente, las herramientas utilizadas para el análisis de datos financieros se basan comúnmente en métodos estadísticos paramétricos y lineales. Si bien estas técnicas han demostrado ser efectivas en la solución de problemas, aún existen otros enfoques que pueden aprovechar las relaciones indirectas y ocultas entre los datos, mejorando la precisión de las estimaciones. En el área económico-financiera, una de estas áreas emergentes es el Aprendizaje Automático, que pertenece a las Ciencias de la Computación y

cuyas metodologías han demostrado altos niveles de predicción en problemas de distintas áreas.

Como también se ha indicado, los modelos desarrollados en este trabajo son evaluados utilizando cuatro clasificadores de los más utilizados en el estado del arte. El porcentaje de exactitud de ajuste de los modelos se utiliza como métrica, junto con el criterio de categorización de las anomalías, para determinar su grado de presencia. Es importante destacar que en este estudio, los clasificadores se utilizan principalmente como mecanismos de evaluación de las anomalías, en lugar de como predictores de precios o rendimientos, o de las propias anomalías.

Asimismo, y como se indicó, los siguientes métodos de clasificación han sido seleccionados de la literatura debido a su utilización en problemas similares en el área económico-financiera y de riesgos:

1. Regresión logística.
2. Perceptrón multicapa.
3. Máquinas de soporte vectorial.
4. Bosques aleatorios.

La Regresión logística es utilizada como punto de referencia debido a su base paramétrica y funcionamiento similar a los procesos estadísticos tradicionales de análisis. Los resultados obtenidos con esta metodología son utilizados como control y comparación para los otros métodos empleados en este estudio: Perceptrón multicapa, Máquinas de soporte vectorial y Bosques aleatorios.

Las implementaciones de los clasificadores utilizados en este estudio se tomaron de Weka (Holmes et al., 1994) (Hall et al., 2009). Weka es una plataforma informática que proporciona un ambiente integrado de fácil acceso que incluye una variedad de técnicas de análisis, pre y post procesamiento. En este trabajo, Weka se utiliza como proveedor de la biblioteca de código para los algoritmos de clasificación utilizados y de herramientas de filtrado para la estratificación de datos y la ejecución automática de los conjuntos de pruebas.

### 3.2 Fuente de datos

Los rendimientos de las series se calculan utilizando los precios de cierre ajustados de las cotizaciones, lo que significa que se aplican los descuentos correspondientes por concepto de dividendos y *splits*<sup>12</sup>. Para obtener los rendimientos en su forma continua, se utiliza la siguiente expresión:

$$R_{log} = \ln\left(\frac{V_t}{V_{t-1}}\right) \quad (3.1)$$

Donde:

- $R_{log}$  representa el rendimiento logarítmico o compuesto continuamente.
- $t$  es el periodo de tiempo actual.
- $V_t$  es el precio de cierre en tiempo o periodo  $t$ .
- $V_{t-1}$  es el precio de cierre en tiempo o periodo anterior a  $t$ .

El IPC es utilizado como activo de referencia del mercado y como punto de comparación en este estudio. Se analizan los precios durante un periodo de 15 años, comprendido entre 2005 al 2020. En cuanto a la selección de las empresas, se consideran dos casos: probar las anomalías en dos empresas que conforman el IPC y en dos portafolios conformados por empresas de la BMV.

En el caso de las empresas del IPC, se seleccionaron dos empresas representando la mayor y la menor bursatilidad<sup>13</sup> y con operaciones durante el periodo de estudio. En la Tabla 3.1 se muestran las empresas que cumplen con el criterio de selección, ordenadas por volumen y se indican sus claves de pizarra. Cemex S. A. B. de C. V.<sup>14</sup> es la empresa seleccionada con la

---

<sup>12</sup> Palabra del inglés que significa dividir o repartir. En el área financiera, un *Split* se refiere a la operación sobre acciones en la que se aumenta o reduce el número de emisiones circulantes en una proporción establecida. Con esta operación, el capital social de la empresa no varía.

<sup>13</sup> La bursatilidad está relacionada con actividad de negociación de una acción en un mercado, es decir, que existe un intercambio frecuente de las acciones entre los distintos participantes y, por lo tanto, hay registros históricos de su cotización. Por esta razón, una de las características de selección de las acciones para realizar análisis basados en sus precios es que tengan suficiente información anual durante el periodo de estudio.

<sup>14</sup> **Cemex S. A. B. de C. V.** es una empresa multinacional mexicana dedicada a la industria de la construcción, que ofrece productos y servicio a clientes y comunidades en más de 50 países en el mundo. En la lista Forbes Global 2000 del año 2021, Cemex fue clasificada como la 1178a empresa privada con información pública más grande del mundo, con ventas anuales de casi 13,000 millones de dólares. La compañía mexicana ocupa el tercer lugar mundial en ventas de cemento con una capacidad de producción anual de 92 millones de toneladas. Es la principal empresa productora de concreto premezclado, con una capacidad de producción de

mayor bursatilidad y que ha tenido cotizaciones para para todos los años del periodo de estudio, mientras que Industrias Peñoles S. A. B. de C. V.<sup>15</sup> es la empresa con la menor bursatilidad que cumple con dicha condición.

Tabla 3.1. Empresas del IPC ordenadas de menor a mayor con operaciones en cada año dentro del periodo 2005 a 2020.

EMPRESA	CLAVE	VOLUMEN ANUAL PROMEDIO
Cemex S.A.B. de C.V.	CEMEXCPO.MX	147,354,862
Grupo México S. A. B. de C. V.	GMEXICOB.MX	59,481,950
Alfa S. A. B. de C. V.	ALFAA.MX	53,857,553
Grupo Financiero Banorte S. A. B. de C. V.	GFNORTEO.MX	26,302,863
Kimberly-Clark de México S. A. B. de C. V.	KIMBERA.MX	15,081,401
Fomento Económico Mexicano S. A. B. de C. V.	FEMSAUBD.MX	14,257,020
Grupo Financiero Inbursa S. A. B. de C. V.	GFINBURO.MX	11,788,803
Orbia Advance Corporation S. A. B. de C. V.	ORBIA.MX	10,992,488
Grupo BIMBO S. A. B. de C. V.	BIMBOA.MX	10,744,999
Alsea S. A. B. de C. V.	ALSEA.MX	6,398,212
GRUMA S. A. B. de C. V.	GRUMAB.MX	2,863,479
Promotora y operadora de infraestructura S. A. B. de C. V.	PINFRA.MX	1,847,604
El puerto de Liverpool S. A. B. de C. V.	LIVEPOLC-1.MX	1,559,128
Industria Peñoles S. A. B. de C. V.	PE&OLES.MX	996,335

Fuente: Elaboración propia.

Note que, en este primer caso, es importante que las cotizaciones presenten un volumen significativo de operaciones. Dado que los cálculos se aplican sobre los precios de las series, si los movimientos se presentan de manera horizontal o con poca operación, estos podrían no mostrar las anomalías y, por tanto, no habría suficiente información para alimentar los modelos y obtener resultados confiables.

aproximadamente 92 millones de toneladas anuales, atendiendo así los mercados de América, Europa, Asia, África y Medio Oriente.

<sup>15</sup> **Industrias Peñoles S. A. B. de C. V.** es una empresa subsidiaria propiedad del Grupo BAL. Peñoles es la segunda minera mexicana más grande de México, con sede en Torreón, Coahuila, México. Es el primer productor mexicano de oro, zinc y plomo y el líder mundial en producción de plata. Peñoles es una empresa con minas activas en México y con algunos proyectos de prospección en Sudamérica. Las tenencias incluyen la Mina de Plata Fresnillo / Mina Proaño, el complejo metalúrgico Met-Mex Peñoles y Química del Rey; una planta química y tres operaciones. Peñoles produce alrededor de 80,500,000 onzas troy (2,500,000 kg) de plata y 756,100 onzas troy (23,520 kg) de oro al año. Otros metales que produce la empresa son zinc, plomo, cobre, bismuto y cadmio. Se constituyó el 1 de marzo de 1887 en la ciudad de Torreón por iniciativa de un grupo de empresarios mexicanos con el fin de explotar varias minas de la región.

En el segundo caso, se construyen dos portafolios: el primero con las empresas ubicadas en el decil nueve ( $D_9$ ) o alto y, el segundo, con las empresas en el decil uno ( $D_1$ ) o bajo, empresas de mayor y menor bursatilidad, respectivamente, de las empresas de la BMV. La selección se compone de todas las empresas mexicanas listadas dentro del periodo de 2005 a 2020, ordenadas por volumen de operación. En las tablas 3.2 y 3.3 se muestran los dos portafolios, cada uno integrado por siete empresas.

Tabla 3.2. Portafolio de empresas mexicanas de la BMV con la mayor bursatilidad.

EMPRESA	CLAVE	VOLUMEN ANUAL PROMEDIO
América Móvil S. A. B. de C. V.	AMXL.MX	54,645,645
Cemex S.A.B. de C.V.	CEMEXCPO.MX	35,750,330
Walmart de México S. A. B. de C. V.	WALMEX.MX	17,981,665
Grupo México S. A. B. de C. V.	GMEXICOB.MX	13,527,217
Alfa S. A. B. de C. V.	ALFAA.MX	9,173,562
Grupo Financiero Banorte S. A. B. de C. V.	GFNORTEO.MX	5,945,057
Desarrolladora Homex S. A. B. de C. V.	HOMEX.MX	4,113,942

Fuente: Elaboración propia.

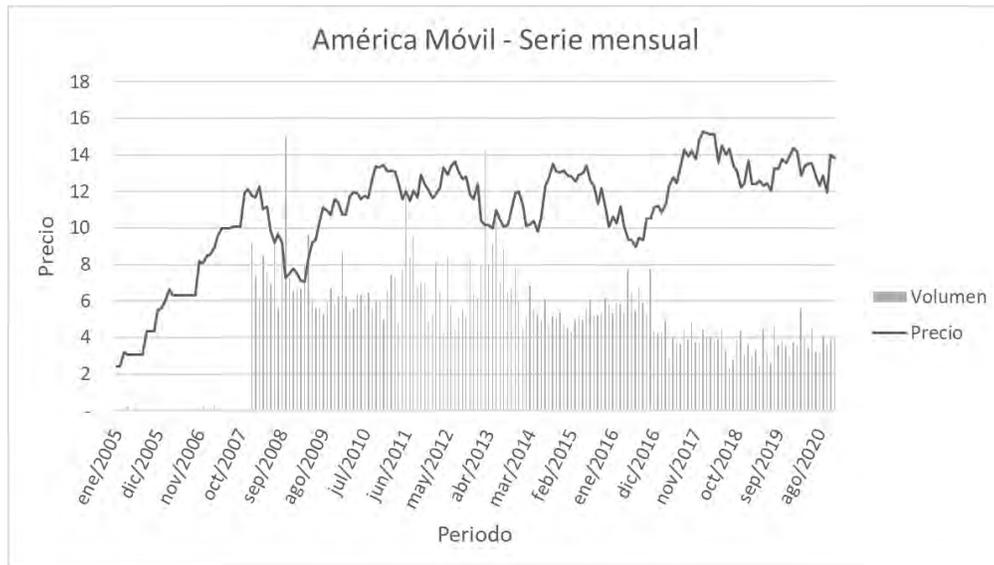
Tabla 3.3. Portafolio de las empresas mexicanas de la BMV con la menor bursatilidad.

EMPRESA	CLAVE	VOLUMEN ANUAL PROMEDIO
Grupo Profuturo S. A. B. de C. V.	GPROFUT.MX	3,043
Grupo Comercial S. A. B. de C. V.	GOMO.MX	2,160
Farmacias Benavides S. A. B. de C. V.	BEVIDESB.MX	1,397
Urvi Desarrollos urbanos S. A. B. de C. V.	URBI.MX	946
Empresas Cablevisión S. A. B. de C. V.	CABLECPO.MX	743
Edoardos Martin S. A. B. de C. V.	EDOARDOB.MX	727
Industria Automotriz S. A. B. de C. V.	IASASA.MX	282

Fuente: Elaboración propia.

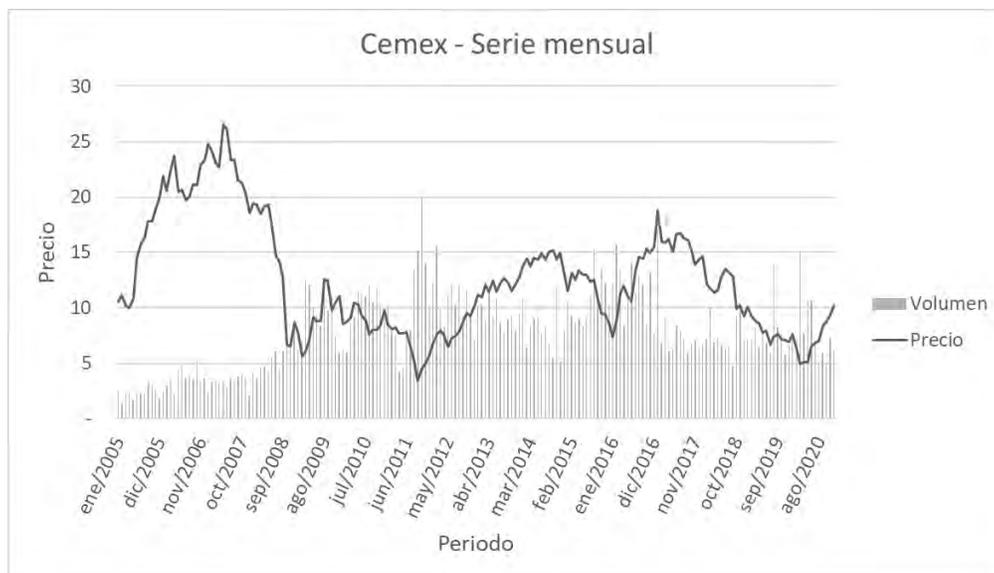
Es importante señalar que, en el caso de las empresas individuales (Peñoles y Cemex), se trata de empresas que forman parte del IPC, mientras que, en el caso de los portafolios, se toman empresas que simplemente cotizan en la BMV y no necesariamente forman parte del IPC.

Figura 3.1. La serie mensual de América Móvil muestra que en periodos previos al 2008 no había o había muy bajo volumen de operaciones.



Fuente: Elaboración propia.

Figura 3.2. La serie mensual de Cemex muestra volumen de operaciones para todos periodos analizados.



Fuente: Elaboración propia.

Cabe destacar que en el portafolio de las empresas de mayor bursatilidad de la tabla 3.2, la empresa de mayor bursatilidad es América Móvil S.A.B. de C.V. y ésta no corresponde a la

empresa de mayor bursatilidad del IPC mostrada en la tabla 3.1, que fue Cemex S.A.B. de C.V. Esto se debe a que, en el primer caso, una de las condiciones de selección de las empresas es que presenten cotizaciones durante todos los periodos analizados, para evitar falta de información (figuras 3.1 y 3.2). En contraste, en el segundo caso, los portafolios se constituyen con la junta de los datos de sus señales, y las cantidades de éstas se complementan y presentan información suficiente para la ejecución de los análisis, por lo que la condición del mínimo de cotizaciones anuales es descartada. En ambos casos, las series son obtenidas de la base de datos Infosel Financiero (2020).

### 3.3 *Cálculo de anomalías*

En este apartado se detallan los cálculos de las anomalías<sup>16</sup> y se analizan teniendo en cuenta las siguientes consideraciones:

- El *momentum* se calcula utilizando la diferencia entre seis periodos, tal como se ha informado en la literatura, ya que se ha mostrado que tiene un buen desempeño con ese rango (Díaz Ruiz et al., 2020).
- La reversión a la media se calcula sobre los precios y se obtiene utilizando un algoritmo propuesto.
- La volatilidad, que se calcula como una reversión a la media en la desviación estándar de una ventana móvil, también se obtiene utilizando un algoritmo propuesto.
- Para las anomalías de calendario, se evalúa el efecto enero sobre las series mensuales, el efecto fin de mes para las series semanales y el efecto de fin de semana para las series diarias.

A continuación, se especifica el cálculo de cada una de las anomalías:

a) El *momentum* se calcula como la diferencia entre dos precios, en la que el segundo precio tiene una separación de cierto número de periodos. Esta separación indica la velocidad de los movimientos y se asocia comúnmente con cambios de tendencia.

---

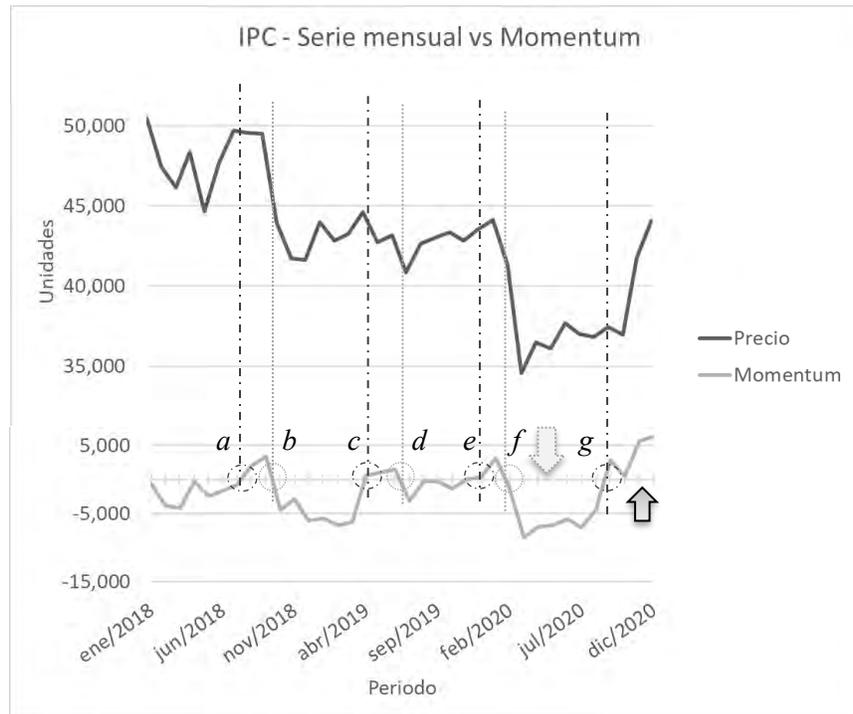
<sup>16</sup> Algunas anomalías, particularmente las técnicas, se aplican a los precios, rendimientos y desviación estándar, incluyendo las series suavizadas, como las medias móviles. En este trabajo, el *momentum* y la reversión a la media se calculan sobre los precios, y la volatilidad sobre la desviación estándar de una ventana móvil de precios.

Sea  $x_t$  la serie de precios de una acción en el periodo  $t = 0, 1, 2, \dots, N$ . Se define el *momentum*, MOM, como:

$$MOM_t = (x_t - x_{t-n}) \quad (3.2)$$

Donde  $t$  es el periodo de análisis,  $x$  representa el precio de cierre de una cotización dentro del periodo de análisis  $N$  y  $n$  es el rango de periodos para el cual se evalúa el *momentum*. Como se indicó previamente, en este estudio  $n=6$ . Los valores resultantes pueden ser positivos o negativos, y la señal se considera activada al cruzar el cero, de abajo hacia arriba si la tendencia es positiva, y de arriba hacia abajo si la tendencia es negativa (figura 3.3).

Figura 3.3. Representación de la señal de *momentum* calculado sobre la serie de mensual del IPC y la aproximación de su correspondencia en la serie de precios.



Fuente: Elaboración propia.

La expresión para obtener la señal en tendencia positiva, está dada por:

$$signal_{MOM} = \begin{cases} 1 & \text{Si } MOM_{t-1} < 0 \text{ y } MOM_t > 0, \text{ con } t \in N \\ 0 & \text{cualquier otro caso} \end{cases} \quad (3.3)$$

Para obtener la señal en tendencia negativa lo signos se invierten de la siguiente forma:

$$signal_{-MOM} = \begin{cases} 1 & \text{Si } MOM_{t-1} > 0 \text{ y } MOM_t < 0, \text{ con } t \in N \\ 0 & \text{cualquier otro caso} \end{cases} \quad (3.4)$$

En la parte inferior de la gráfica de la figura 3.3, se puede observar el cálculo del *momentum* sobre la serie mensual del IPC, el cual oscila alrededor del cero con valores positivos y negativos. También se muestra una aproximación de la correspondencia entre la anomalía del *momentum* y la serie de precios del índice. Los círculos en *a*, *c*, *e* y *g* señalan el cambio del *momentum* de negativo a positivo dando origen a la señal, mientras que los círculos en *b*, *d*, y *f* indican el cambio de positivo a negativo.

**b) La reversión a la media, *REV***, se calcula sobre los precios y se detecta cuando a una serie de variaciones negativas en los precios, con precios de cierre cada vez más bajos, es seguida por una serie de variaciones positivas en los precios, con precios de cierre cada vez mayores que los anteriores, en el caso de una tendencia alcista o positiva. En una tendencia bajista o negativa, los términos se invierten y la anomalía se detecta cuando una serie de variaciones positivas en los precios, con precios de cierre cada vez más altos, es seguida por una serie de variaciones negativas en los precios, con precios de cierre cada vez más bajos que los anteriores.

Sea  $x_t$  la serie de precios de una acción en el periodo  $t = 0, 1, 2, \dots, N$ . Se define la reversión a la media como una serie de  $p$  y  $n$  precios que pertenecen a  $x_t$  y se cumple  $x_{p-0} < x_{p-1} < x_{p-2} < \dots < x_{p-p}$  para un  $p$  dado seguida de la serie de precios  $n$  que cumplen  $x_{n+(p-1)-1} > x_{n+(p-1)-2} > \dots > x_{n+(p-1)-n}$  para un  $n$  dado.

Donde  $t$  es el periodo de análisis,  $x$  representa el precio de cierre de una cotización dentro del periodo de análisis  $N$ ,  $n$  es la extensión de la serie de precios con variaciones negativas y  $p$  la extensión de precios con variaciones positivas.

La expresión general para obtener la señal de la *REV* para tendencia positiva es la siguiente:

$$f_{Down} = \begin{cases} 1 & \text{Si } x_{t-(n+p-1-i)} - x_{t-(n+p-1-i-1)} > 0 \text{ para } \forall i = 1, 2, \dots, n-1, \text{ con } n, i \in N \\ 0 & \text{cualquier otro caso} \end{cases} \quad (3.5)$$

$$f_{Up} = \begin{cases} 1 & \text{Si } x_{t-(p-i)} - x_{t-(p-i-1)} < 0 \text{ para } \forall i = 1, 2, \dots, p-1, \text{ con } p, i \in N \\ 0 & \text{cualquier otro caso} \end{cases} \quad (3.6)$$

$$signal_{REV} = \begin{cases} 1 & \text{Si } f_{Down} = 1 \text{ y } f_{Up} = 1 \\ 0 & \text{cualquier otro caso} \end{cases} \quad (3.7)$$

Para la serie de variaciones negativas se cumple  $x_{t-(n+p-1-i)} > x_{t-(n+p-1-i-1)}$  para  $i=1, 2, \dots, n-1$ . Mientras que para la serie de variaciones positivas se cumple  $x_{t-(p-i)} < x_{t-(p-i-1)}$  para  $i=1, 2, \dots, p-1$ . En este caso,  $n$  representa la longitud del movimiento negativo y  $p$  representa la longitud del movimiento positivo (figura 3.2).

En el caso de la tendencia negativa se tiene:

$$f_{Up} = \begin{cases} 1 & \text{Si } x_{t-(n+p-1-i)} - x_{t-(n+p-1-i-1)} < 0 \text{ para } \forall i = 1, 2, \dots, n-1, \text{ con } n, i \in N \\ 0 & \text{cualquier otro caso} \end{cases} \quad (3.8)$$

$$f_{Down} = \begin{cases} 1 & \text{Si } x_{t-(p-i)} - x_{t-(p-i-1)} > 0 \text{ para } \forall i = 1, 2, \dots, p-1, \text{ con } p, i \in N \\ 0 & \text{cualquier otro caso} \end{cases} \quad (3.9)$$

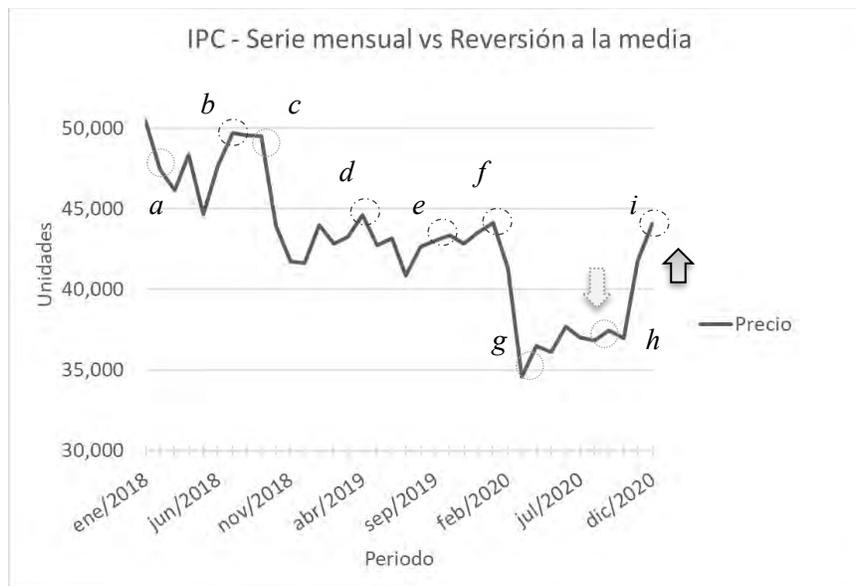
$$signal_{REV} = \begin{cases} 1 & \text{Si } f_{Up} = 1 \text{ y } f_{Down} = 1 \\ 0 & \text{cualquier otro caso} \end{cases} \quad (3.10)$$

Es importante destacar que las fórmulas para el cálculo de ambas tendencias son idénticas. La diferencia radica en el orden en que se evalúan las direcciones de las series de variaciones o periodos de movimiento de los precios. Además, es posible componer la señal utilizando más de dos series de variaciones para validar movimientos complejos o combinados.

Las expresiones son generales y se pueden asignar valores específicos a  $p$  y  $n$  según la longitud de la serie de variaciones que se desee evaluar. En este estudio, se prefiere que el tamaño de la serie de variaciones negativas sea menor que el de la serie de variaciones positivas cuando se evalúa la tendencia positiva, y mayor que el de la serie de variaciones positivas cuando se evalúa la tendencia negativa. Por lo tanto, los valores utilizados en los cálculos son  $n=2$  y  $p=3$ .

En la figura 3.4 se presentan ejemplos de señales de la anomalía sobre la serie de precios mensuales del IPC. Los puntos *b, d, e, f e i* señalan la anomalía para el caso de una tendencia positiva mientras que lo puntos *a, c, g y h* la señalan para la tendencia negativa. La interpretación de este criterio sugiere que, si se produce un cambio en la dirección de los precios, puede deberse a un ajuste temporal para continuar con la tendencia previa, o, si ocurre en dirección opuesta a la tendencia principal, podría indicar una divergencia o cambio de tendencia.

Figura 3.4. Representación de la reversión a la media en la serie de precios mensuales del IPC.

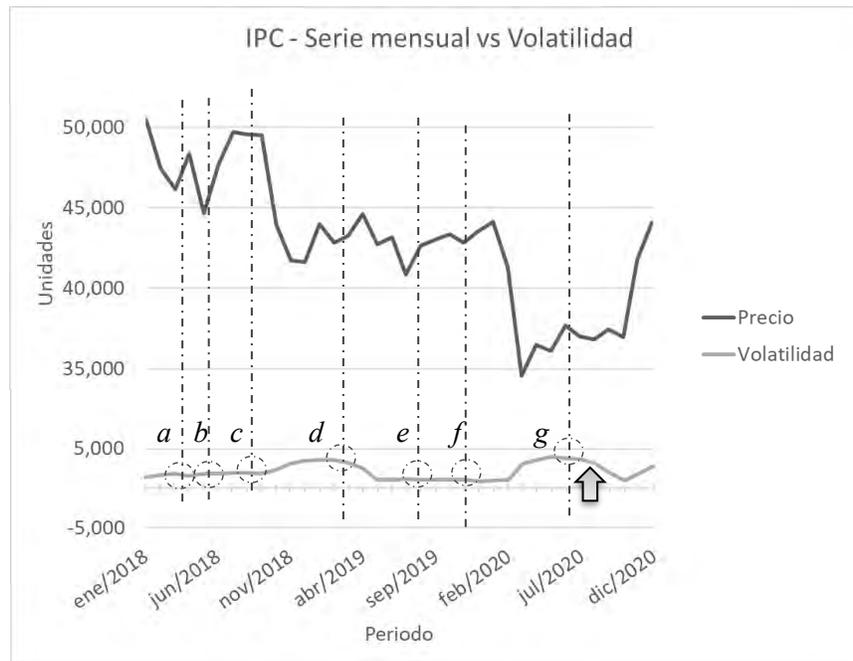


Fuente: Elaboración propia.

**c) La reversión a la media en la volatilidad**, o simplemente **volatilidad**, *VOL*, se calcula utilizando ventanas móviles de la desviación estándar de los precios al cierre. Tanto para la tendencia positiva como para la tendencia negativa, se evalúa de la misma manera debido a que la desviación estándar se comporta de la misma forma, independientemente de la dirección de los precios. La desviación estándar se acerca y se aleja de cero, y muestra picos cuando los valores de los precios cambian rápidamente, y valles cuando los valores se mantienen más estables dentro de un rango estrecho.

La interpretación del cálculo de esta anomalía es que después de un movimiento brusco, es decir, que los valores de la desviación estándar de los precios presentan variaciones mayores a los anteriores, le sigue un movimiento estable, es decir, que los valores de la desviación estándar de los precios no varían o presentan poca variación, que continua en la misma dirección, lo que podría indicar una tendencia (figura 3.5).

Figura 3.5. Representación de la anomalía de volatilidad calculada sobre la serie mensual del IPC, junto con su correspondencia aproximada.



Fuente: Elaboración propia.

Se utilizan las mismas fórmulas que se han definido para el cálculo de la reversión a la media (expresiones 3.5 y 3.6) dado que éstas permiten hacer comparaciones iterativas entre la sucesión de precios y la señal de activación se obtiene mediante la expresión 3.11. En el caso de la anomalía de volatilidad, se considera que una cresta es seguida por un valle y, por tanto, que la tendencia normal es en misma dirección (figura 3.5). Para este caso, se utilizan los valores de  $n=1$  y  $p=1$ , misma longitud de la serie de variaciones positivas y negativas.

$$signal_{VOL} = \begin{cases} 1 & \text{Si } f_{Up} = 1 \text{ y } f_{Down} = 1 \\ 0 & \text{cualquier otro caso} \end{cases} \quad (3.11)$$

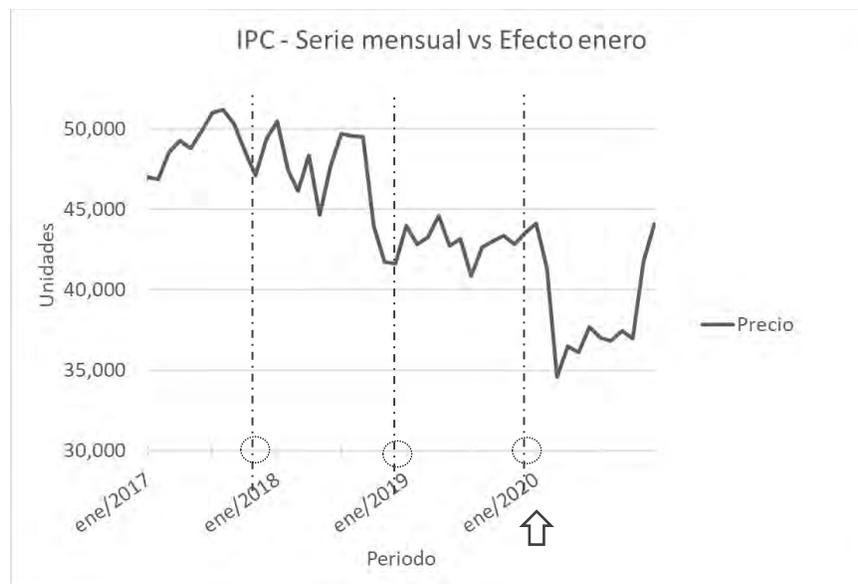
En la figura 3.5 se muestra la gráfica de la desviación estándar de los precios de la serie mensual del IPC, donde los puntos *a, b, c, d, e, f* y *g* señalan la presencia de la anomalía. Además, se puede observar su correspondencia aproximada en la serie de precios del IPC.

d) Las anomalías de calendario son calculadas de la siguiente forma: el **efecto enero**, *EEN*, es evaluado en las series de datos mensuales y su cálculo se realiza como sigue:

$$signal_{EEN} = \begin{cases} 1 & \text{Si mes es diciembre} \\ 0 & \text{Si mes no es diciembre} \end{cases} \quad (3.12)$$

La idea detrás de esta anomalía es que los precios tienden a aumentar en enero, por lo que la señal se activa con un mes de anticipación. La figura 3.6 muestra la señal trazada en cada diciembre y su correspondencia aproximada en la serie de precios mensuales del IPC. Como esta anomalía es periódica, su activación es fija de según la ocurrencia del mes correspondiente.

Figura 3.6. Representación de la señal del efecto enero en la serie de precios mensuales del IPC.



Fuente: Elaboración propia.

Dado que la señal de la anomalía no cambia con la tendencia, esta expresión es utilizada para evaluar tanto tendencia positiva como negativa. Esto mismo aplica para el efecto de fin de mes y fin de semana.

El **efecto fin de mes**,  $EFM$ , es evaluado en las series de datos semanales y la expresión para obtener la señal está dada por:

$$signal_{EFM} = \begin{cases} 1 & \text{Si semana es la cuarta del mes} \\ 0 & \text{Si semana no es la cuarta del mes} \end{cases} \quad (3.13)$$

La anomalía de fin de mes se refiere a la observación de que los precios tienden a disminuir hacia el final de cada mes, y luego incrementan nuevamente al inicio del mes siguiente. Por esta razón, la señal correspondiente se activa al término de la última semana del mes.

Por último, el **efecto fin de semana**,  $EFS$ , es evaluado en las series de datos diarios como:

$$signal_{EFS} = \begin{cases} 1 & \text{Si día es viernes} \\ 0 & \text{Si día no es viernes} \end{cases} \quad (3.14)$$

La anomalía de fin de semana se refiere a la tendencia de los precios a disminuir al final de la semana y a aumentar al comienzo de la siguiente semana. Por lo tanto, la señal se activa al final del último día de la semana.

### 3.4 Procedimiento general

El estudio se representa como un sistema de reconocimiento de patrones basado en un modelo de clasificación utilizando aprendizaje supervisado. En términos econométricos, se trata de un análisis multivariado o multifactorial con la variable dependiente de tipo categórica. Las variables independientes o componentes del modelo incluyen una ventana de rendimientos pasados ajustados al activo de referencia y una variable booleana que actúa como proxy y caracteriza la señal de la anomalía. Esto se puede expresar de la siguiente manera:

**Definición 3.1.** Sea  $t$  el índice del evento en el tiempo, se define  $t=0$  como la fecha del evento en observación o análisis. La ventana de tiempo del evento  $N$  se representa desde  $t-1$  a  $t-p$ , mientras que la subventana de estimación tiene una longitud de  $t-1$  hasta  $t=0$ , entonces:

$$Y_{t+n} = \delta_t + \alpha_1 (Y_{t-1} - x_{t-1}) + \dots + \alpha_p (Y_N - x_N) + \varepsilon_t \quad (3.15)$$

$$R_{t+n} = Y_{t+n} - X_{t+n} \quad (3.16)$$

$$C_i = \begin{cases} 1 & \text{Si } \delta_t = 1 \text{ y } R_{t+n} > 0 \\ 0 & \text{cualquier otro caso} \end{cases} \quad (3.17)$$

Donde:

- $p$  periodos de rendimientos del activo.
- $i$  clasificación del rendimiento.
- $C_i$  clase a la que pertenece la función.
- $a_i$  factor de impacto del rendimiento en la ventana.
- $Y_t$  serie de rendimientos observados del activo.
- $\delta_t$  variable que indica la presencia de la señal.
- $n$  número de periodos que abarca el rendimiento objetivo a partir de la señal que simula la tenencia de la posición.
- $t$  periodo, donde  $t_0$  es el periodo actual o de estudio.
- $\varepsilon_t$  error sistémico.
- $N$  ventana de rendimientos rezagados de  $t-1$  a  $t-p$ .
- $X_t$  rendimiento del mercado, el IPC.
- $x_t$  serie de rendimientos observados del mercado, el IPC.
- $R_t$  rendimiento ajustado al riesgo esperado.

Las clases, representadas por  $C_i$ , son dos particiones del conjunto datos: los componentes cuyos rendimientos  $Y_t$  superan al activo de referencia del mercado  $X_t$ , y aquellos que no cumplen con esta condición.

### ***3.5 Evaluación de las anomalías***

Se utilizan diferentes métodos para medir el desempeño de los clasificadores, incluyendo el uso del conjunto de datos completo, la validación cruzada y el método de retención. El primer método se emplea para evaluar el desempeño dentro de la muestra, mientras que el segundo y el tercero se utilizan para evaluar el desempeño fuera de la muestra. Se utilizan ambas alternativas para contar con una referencia de contraste de los resultados.

El criterio para asignar el grado de presencia de las anomalías  $pdeg$  está basado en el porcentaje de exactitud resultante de los modelos, representado por  $\tau$ , que se ha mencionado a lo largo del documento. Se asigna de acuerdo al siguiente criterio:

$$pdeg = \begin{cases} alta & 90 < \tau < 100 \\ media & 80 < \tau < 89 \\ baja & 60 < \tau < 79 \\ nula & \tau < 59 \end{cases} \quad (3.18)$$

### 3.6 *Análisis de las anomalías*

El análisis de las anomalías se lleva a cabo mediante el procedimiento siguiente:

- a) Antes de iniciar el análisis es importante considerar algunas inicializaciones y precondiciones:
  - a. Descargar las cotizaciones diarias, semanales y mensuales del IPC y de los activos de análisis del año 2005 a 2020.
  - b. Redondear los decimales de los precios de las series al cierre a dos posiciones, lo mismo que los valores de los cálculos sobre las anomalías que se realicen.
  - c. Dado que la intención de este trabajo no es localizar las configuraciones optimas de los clasificadores se utiliza la configuración por omisión de Weka.
  - d. En caso que los registros presenten precios de las series al cierre en nulos o no numéricos estos se remueven.
- b) El siguiente procedimiento es usado para analizar una anomalía y una serie. Para analizar el resto se deberá repetir para cada caso: serie, empresa, anomalía, rezago.
  - a. Escoger una de las empresas y emparar cada uno de los registros de la serie con la del activo de referencia, el IPC, por fecha.
  - b. Calcular los rendimientos de ambas series, la empresa seleccionada y el IPC, sobre el precio para cada registro usando la expresión 3.1.
  - c. Obtener el rendimiento ajustado al riesgo, es decir, respecto al IPC, usando la expresión 3.16.
  - d. Asignar la clase a cada registro según el criterio de etiquetado de clase de la expresión 3.17.

- e. Seleccionar una anomalía y aplicar el algoritmo correspondiente sobre el precio de cada uno de los registros de la empresa. Crear la variable *dummy* para cada registro y asignarle el resultado de la evaluación de la señal de presencia de la anomalía.
- f. Dado que el objetivo es probar la anomalía, se remueven los componentes cuyo rendimiento ajustado es superior al mercado y que no presentaron señal de anomalía:  $signal_i=1$  y  $C_i=1$ . Y se extrae la matriz de variables dependientes e independiente resultante.
- g. Evaluar en Weka para un clasificador dado y obtener los estadísticos de desempeño. Finalmente, registrar los resultados en una tabla comparativa.

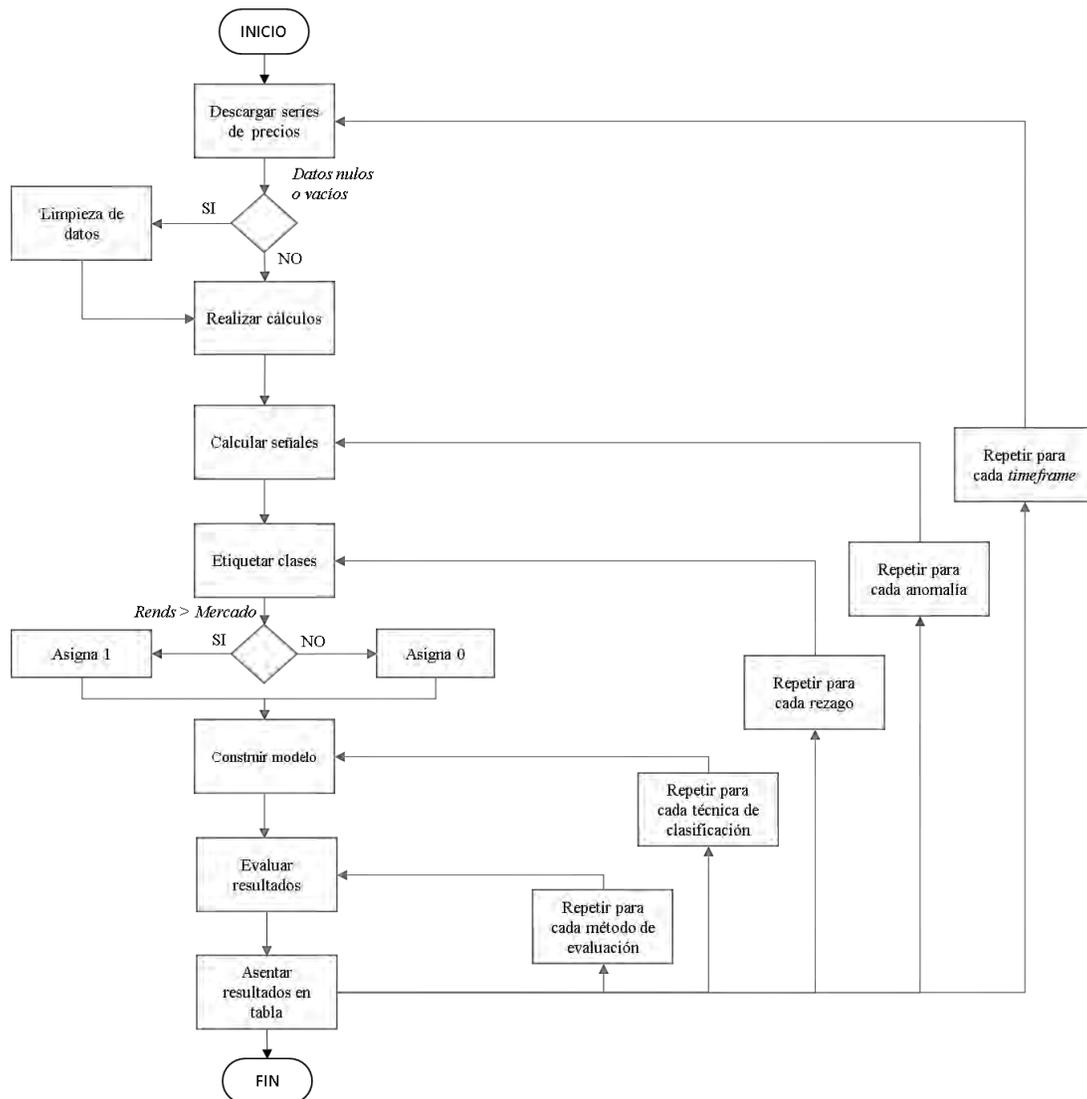
El diagrama de flujo del procedimiento descrito se observa en la figura 3.7 y el pseudo código que indica la ejecución de cada parte del procedimiento es el siguiente:

```

// Ejecuta toda la batería de pruebas
Function obtenerAnalisis()
Begin
  For cada timeframe //diario, semanal y mensual
  Begin
    descargarSeries()
    calcularRendimientos()
    For cada anomalía //MOM, REV, VOL, EEN, EFM y EFS
    Begin
      calcularSenial()
      obtenerRendimientoMercadoPromedio()
      For cada rendimientoConRezago // hasta n=10
      Begin
        etiquetarClases()
        For cada técnicaDeClasificación //logictic, mlp, svm y rf
        Begin
          construirModelo()
          For cada métodoEvaluación // crossVal, fullSet y holdout
          Begin
            evaluarResultado()
            asentarDatos()
          End For
        End For
      End For
    End For
  End For
End Function

```

Figura 3.7. Diagrama de flujo del procedimiento de análisis.



Fuente: Elaboración propia.

El procedimiento de análisis detallado, que incluye paso a paso desde la adquisición de los datos hasta su limpieza y ejecución en la plataforma Weka, se muestra en el anexo A.

### Resumen

En este capítulo se ha presentado la metodología de análisis y el núcleo principal del trabajo. Se han explicado los criterios para calcular las anomalías de estudio, cuyas expresiones se

personalizaron a partir de la definición intuitiva de trabajos en el estado del arte, en particular, el cálculo de la reversión a la media en precios y volatilidad. El resto, como el *momentum* y las anomalías de calendario, mantienen el criterio general. También se ha detallado el procedimiento de construcción del modelo y de las fases que lo componen, expresado en términos de un proceso típico de clasificación en el reconocimiento de patrones.

En el próximo capítulo se expondrán los casos de prueba utilizados para aplicar la metodología revisada y se describirán los resultados obtenidos. Asimismo, se presentarán las conclusiones y se discutirá brevemente sobre los hallazgos obtenidos en comparación con algunos de las investigaciones del estado del arte.

## Capítulo 4

### RESULTADOS Y DISCUSIONES

En el capítulo anterior se describió la metodología utilizada para el estudio de las anomalías en el mercado mexicano. Se propuso el uso de un sistema de clasificación con aprendizaje supervisado para detectar la presencia de las anomalías y se definieron los criterios para su obtención.

En esta sección, se presentan los resultados de las pruebas realizadas siguiendo los criterios mencionados en el capítulo anterior. Estas pruebas respaldan empíricamente la hipótesis presentada que originó el estudio de las anomalías en el mercado mexicano.

Como se mencionó anteriormente, las metodologías de aprendizaje automático se utilizaron para obtener el grado de exactitud del modelo, que se interpreta como el grado de presencia y de confianza de las anomalías. Cuanto más cercano esté el modelo al 100%, mayor será su ajuste y, por lo tanto, la anomalía que lo compone se considera un factor diferenciador en el estimador. En consecuencia, se considera que la anomalía es fuerte y útil. Por otro lado, si la anomalía presenta porcentajes más bajos, se traduce en una disminución de la confianza y la certeza en términos de la obtención de rendimientos superiores a los del mercado.

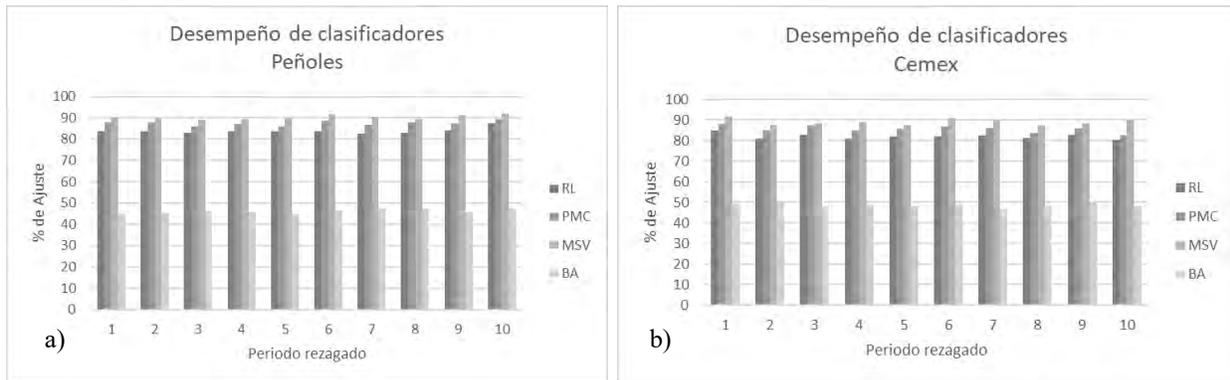
#### ***4.1 Ejecución de pruebas***

##### ***4.1.1 Desempeño del modelo***

En cuanto a la evaluación de los modelos de clasificación, se encontró que las Máquinas de Soporte Vectorial (MSV) ofrecieron los mejores porcentajes de recuperación en las pruebas de las anomalías realizadas. En las figuras 4.1 y 4.2 se muestran los resultados promedio de

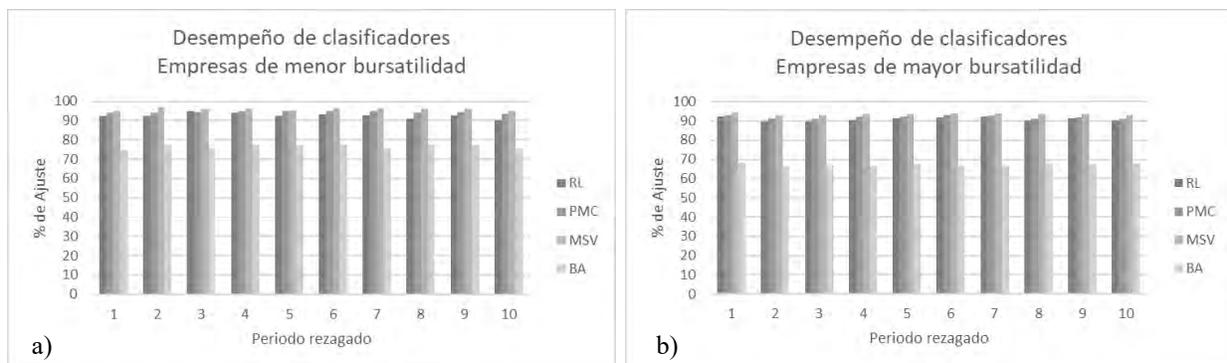
los porcentajes de ajuste de los clasificadores obtenidos mediante validación cruzada para Peñoles y Cemex, así como para los portafolios de empresas de la menor y mayor bursatilidad, respectivamente. Se puede observar que el desempeño de los modelos no varía drásticamente entre sí y, de hecho, sus resultados son similares. Las MSV superaron por muy poco a la Regresión logística (RL) y al Perceptrón Multi Capa (PMC), con una diferencia promedio del 4% en Peñoles y Cemex y del 2% para los portafolios de empresas. En contraste, los Bosques Aleatorios (BA) obtuvieron los porcentajes más bajos, alrededor del 50% para Peñoles y Cemex y alrededor del 75% para los portafolios de empresas.

Figura 4.1. Porcentajes de ajuste promedio de los clasificadores evaluados con validación cruzada para Peñoles y Cemex.



Fuente: Elaboración propia.

Figura 4.2. Porcentajes de ajuste promedio de los clasificadores evaluados con validación cruzada para los portafolios de las empresas de mayor y menor bursatilidad.

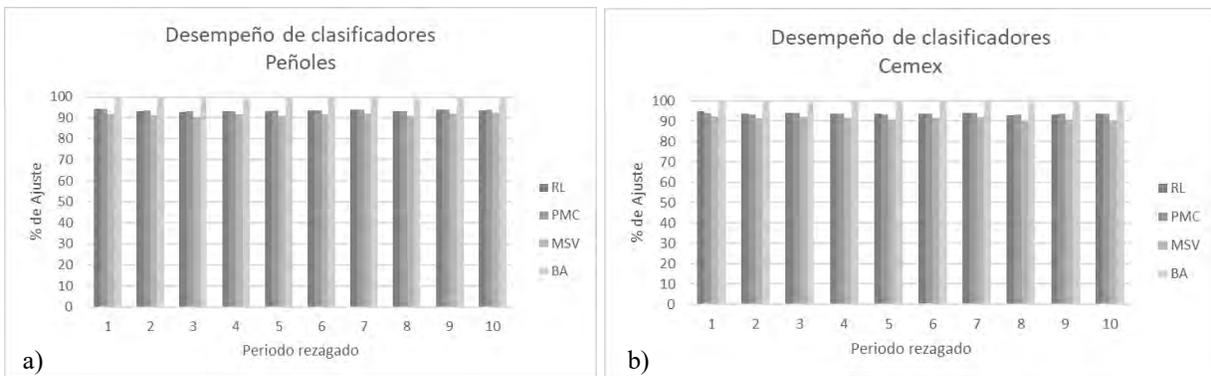


Fuente: Elaboración propia.

En el caso de la evaluación de los modelos con el conjunto de datos completo, se observó que los BA tuvo un mejor ajuste que la RL y el PMC, con una diferencia de alrededor del 6% en Peñoles y Cemex, y del 3% para los portafolios de empresas, como se puede apreciar en las gráficas de las figuras 4.3 y 4.4. Por otro lado, las MSV tuvieron el porcentaje de ajuste más bajo, ligeramente por debajo del resto de los clasificadores.

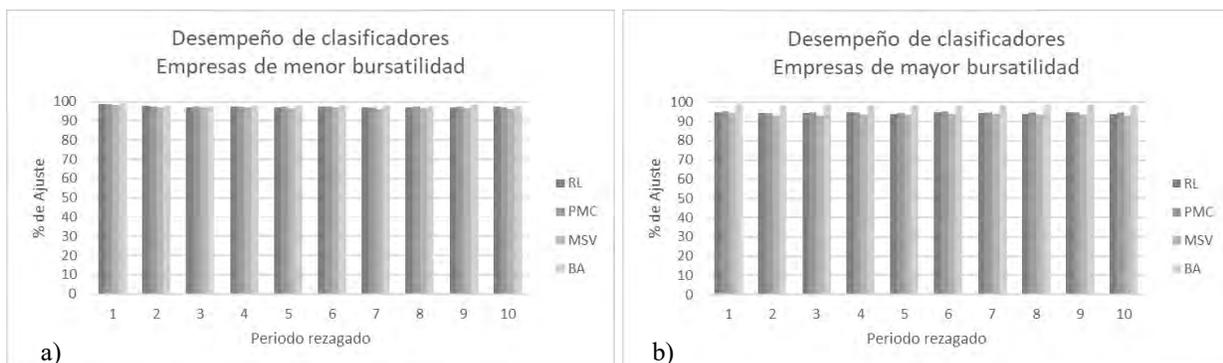
En los próximos apartados, se presentan las tablas y gráficas que hacen referencia a los resultados obtenidos con las MSV, ya que estos modelos obtuvieron los mejores porcentajes de ajuste. Todas las gráficas mostradas en esta sección, y en el resto del documento, hacen referencia a los datos y figuras que se encuentran en el anexo B.

Figura 4.3. Porcentajes de ajuste promedio de los clasificadores evaluados con el conjunto de datos completo para Peñoles y Cemex.



Fuente: Elaboración propia.

Figura 4.4. Porcentajes de ajuste promedio de los clasificadores evaluados con el conjunto de datos completo para los portafolios de empresas.



Fuente: Elaboración propia.

### ***4.1.2 Análisis de las anomalías***

En este apartado se muestran los resultados obtenidos en las pruebas realizadas para detectar la presencia de las anomalías sobre los dos casos de prueba considerados: para las dos empresas de menor y mayor bursatilidad del IPC, Peñoles y Cemex respectivamente, así como en los dos portafolios conformados por las empresas mexicanas dentro de la BMV con la menor y mayor bursatilidad, las ubicadas en primer decil ( $D_1$ ) y las ubicadas en el último ( $D_9$ ), respectivamente.

#### **Caso de empresa de menor y mayor bursatilidad del IPC**

Para este primer caso de prueba, se analizan las series de precios de Peñoles y Cemex que corresponden a las empresas de mayor y menor operación dentro de las que componen al IPC, cuyas operaciones se encuentran dentro del periodo de 2005 a 2020.

El resumen de porcentajes obtenidos para Peñoles se presenta en las gráficas de las figuras 4.5 a 4.8. En ellas, se muestran los porcentajes de ajuste obtenidos para la evaluación del *momentum* mediante validación cruzada. Se puede apreciar que el porcentaje de ajuste es homogéneo en datos diarios, mientras que para los datos semanales y mensuales presenta una mayor variabilidad. En particular, los porcentajes obtenidos a partir de la serie diaria son bastante estables y se sitúan por encima del 90%, con caídas en la serie mensual que llegan hasta el 80%. Por otro lado, la reversión a la media presenta porcentajes ligeramente inferiores a los obtenidos con el *momentum*. Respecto a la anomalía de volatilidad, los porcentajes obtenidos en las diferentes series son uniformes y se encuentran en el rango de 80% y 90%. Finalmente, para las anomalías de calendario, los porcentajes obtenidos son uniformes en el rango de 87% y 94%.

Figura 4.5 Presencia de la anomalía de *momentum* en Peñoles mediante validación cruzada.



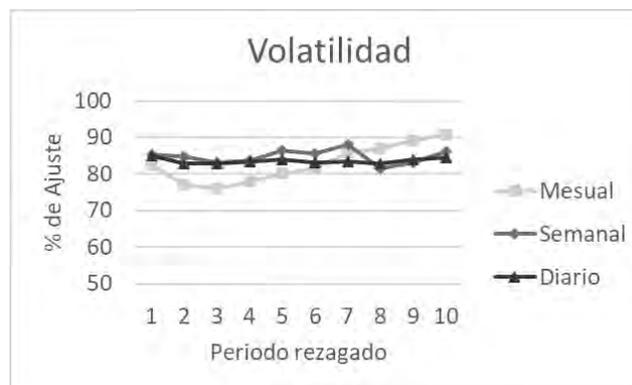
Fuente: Elaboración propia.

Figura 4.6 Presencia de la anomalía de reversión a la media en Peñoles mediante validación cruzada.



Fuente: Elaboración propia.

Figura 4.7 Presencia de la anomalía de volatilidad en Peñoles mediante validación cruzada.



Fuente: Elaboración propia.

Figura 4.8 Presencia de anomalías de calendario en Peñoles mediante validación cruzada.



Fuente: Elaboración propia.

El análisis de los modelos obtenidos para las anomalías en Cemex se muestra en las gráficas de las figuras 4.9 a 4.12. Al igual que en el caso de Peñoles, se observa que el modelo de *momentum* presenta los mayores porcentajes de ajuste, superando al 93% para las series diarias y semanales. Sin embargo, en la serie mensual se presentan algunos valles que caen hasta el 30%. Por su parte, la reversión a la media muestra porcentajes por encima del 90%, con series más uniformes. Las anomalías de calendario también muestran porcentajes por encima del 90%, aunque por debajo para algunos rezagos. Finalmente, la anomalía de volatilidad presenta un rango de porcentajes entre 75% y 90%.

Figura 4.9 Presencia de la anomalía de *momentum* en Cemex mediante validación cruzada.



Fuente: Elaboración propia.

Figura 4.10 Presencia de la anomalía de reversión a la media en Cemex mediante validación cruzada.



Fuente: Elaboración propia.

Figura 4.11 Presencia de anomalía de volatilidad en Cemex mediante validación cruzada.



Fuente: Elaboración propia.

Figura 4.12 Gráfica de la presencia de calendario en Cemex evaluando con validación cruzada.



Fuente: Elaboración propia.

### Caso de portafolios de las empresas con la mayor y menor bursatilidad de la BMV

Los portafolios de este caso de prueba se construyeron seleccionando a las empresas mexicanas de la BMV con la mayor y menor bursatilidad y cuya participación en la bolsa cubre el periodo de estudio conformado de 2005 a 2020. Las cotizaciones de las empresas se agruparon para crear una base de datos más amplia y así aumentar el número de señales de las anomalías.

Los resultados de los porcentajes de ajuste de las anomalías probadas en el portafolio de empresas de menor bursatilidad se observan en las gráficas de las figuras 4.13 a la 4.16. Se puede observar que el *momentum* presenta porcentajes uniformes en las tres series que se mantienen por encima del 97%. Le sigue la reversión a la media con porcentajes por encima del 97%, aunque presenta picos amplios en la serie semanal que alcanzan el 70%. A continuación, la anomalía de volatilidad tiene un rango entre el 90% y 100%, con las tres series uniformes. Finalmente, las anomalías de calendario muestran porcentajes superiores al 90% para las tres series.

Figura 4.13 Presencia de la anomalía de *momentum* en el portafolio de empresas de menor bursatilidad mediante validación cruzada.



Fuente: Elaboración propia.

Figura 4.14 Presencia de la anomalía de reversión a la media en el portafolio de empresas de menor bursatilidad mediante validación cruzada.



Fuente: Elaboración propia.

Figura 4.15 Presencia de la anomalía de volatilidad en el portafolio de empresas de menor bursatilidad mediante validación cruzada.



Fuente: Elaboración propia.

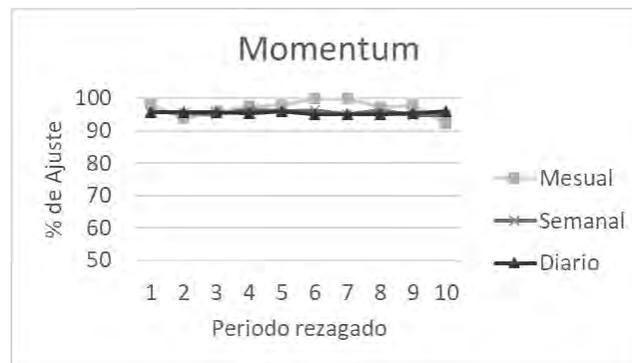
Figura 4.16 Presencia de las anomalías de calendario en el portafolio de empresas de menor bursatilidad mediante validación cruzada.



Fuente: Elaboración propia.

Los resultados de los modelos para el portafolio de empresas de mayor bursatilidad se muestran en las figuras 4.17 a la 4.20. En cuanto a la anomalía de *momentum*, se observa un porcentaje de ajuste uniforme en las tres series, alrededor del 96%. Le sigue la reversión a la media con porcentajes alrededor del 93%, también uniforme en las tres series. Ligeramente por debajo se encuentran las anomalías de calendario con un porcentaje de ajuste de alrededor del 92%, y finalmente la anomalía de volatilidad con porcentajes en el rango del 90%, todos ellos con las tres series estables en los mismos porcentajes y rangos.

Figura 4.17 Presencia de la anomalía *momentum* en el portafolio de empresas de mayor bursatilidad mediante validación cruzada.



Fuente: Elaboración propia.

Figura 4.18 Presencia de la anomalía de reversión a la media en el portafolio de empresas de mayor bursatilidad mediante validación cruzada.



Fuente: Elaboración propia.

Figura 4.19 Presencia de la anomalía de volatilidad en el portafolio de empresas de mayor bursatilidad mediante validación cruzada.



Fuente: Elaboración propia.

Figura 4.20 Presencia de las anomalías de calendario en el portafolio de empresas de mayor bursatilidad mediante validación cruzada.



Fuente: Elaboración propia.

Como se ha observado, la anomalía más fuerte en los casos estudiados ha sido el *momentum*, ya que ha presentado los porcentajes más altos, alrededor del 94% en Peñoles y Cemex si se considera la serie diaria, y alrededor del 99% en los portafolios. La reversión a la media le sigue con porcentajes del 8% y 3%, para ambos casos, respectivamente. Por otro lado, la anomalía de volatilidad y las anomalías de calendario presentaron resultados opuestos para los casos estudiados: en Peñoles y Cemex, la anomalía de volatilidad fue la que mostró los porcentajes más bajos, ligeramente por encima del 80%, mientras que las anomalías de calendario oscilaron alrededor del 90% en ambas empresas. En cambio, en el caso de los portafolios, las anomalías de calendario mostraron los porcentajes más bajos, ligeramente

por debajo del 90%, mientras que la anomalía de volatilidad se ubicó justo por encima del 92% y 94%, respectivamente.

### ***4.1.3 Rendimientos ajustados***

En la sección anterior se analizaron y discutieron los resultados de los modelos para determinar el grado de presencia de las anomalías. Sin embargo, estos resultados por sí solos no son suficiente si no se indican los beneficios que las anomalías detectadas pueden ofrecer a los interesados en términos de rendimientos ajustados. En este apartado, se describen los rendimientos lineales promedio obtenidos a partir de la detección de las anomalías, con el fin de evaluar el riesgo de cada periodo analizado.

El objeto de comparar un activo con otro menos riesgoso es explorar la posibilidad de obtener beneficios mayores a partir de las anomalías detectadas. Por lo tanto, los rendimientos presentados se calculan como la diferencia entre el activo en cuestión y el activo menos riesgoso utilizado como referencia. La figura 4.21 muestra un ejemplo de la comparación de los rendimientos promedio una vez que se ha identificado la anomalía en el activo en cuestión y se ha comparado con el rendimiento del índice. Se puede observar que el rendimiento del índice se mantiene en la misma dirección durante más tiempo, mientras que el rendimiento del activo es más variable. A partir de esto, se puede inferir que mantener una posición por periodos cortos reduce el riesgo.

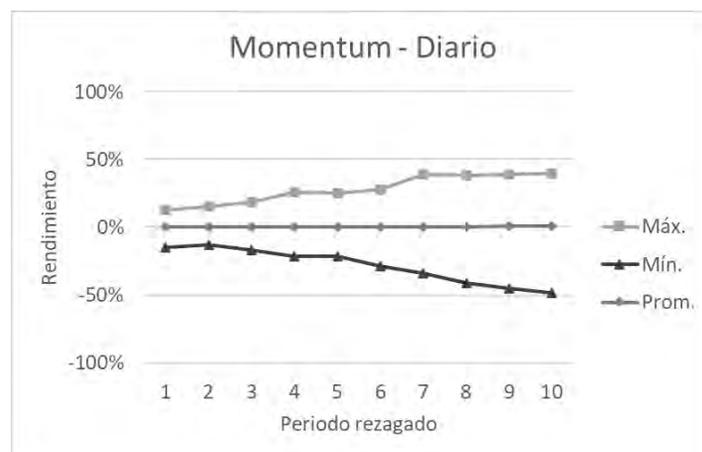
Figura 4.21 Rendimiento promedio de la serie mensual de Cemex e IPC para periodos posteriores a la señal de reversión a la media.



Fuente: Elaboración propia.

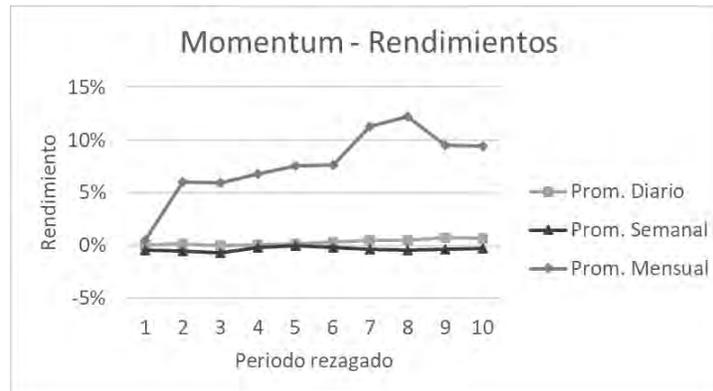
Los porcentajes de ajustes mencionados en secciones anteriores indican que para todos los periodos rezagados, los ajustes son altos sin destacar alguno en particular. No obstante, al analizar los rendimientos, se puede notar que la brecha entre el rendimiento máximo positivo y negativo se amplía a medida que se alejan de los primeros periodos. La figura 4.22 muestra un ejemplo de esto para la anomalía de *momentum* de la serie diaria de Peñoles.

Figura 4.22 Rendimientos promedio, máximos y mínimos para la anomalía de *momentum* en la serie de precios diarios de Peñoles.



Fuente: Elaboración propia.

Figura 4.23 Rendimientos promedio de las series mensual, semanal y diaria de Peñoles para los diferentes periodos rezagados de la anomalía *momentum*.

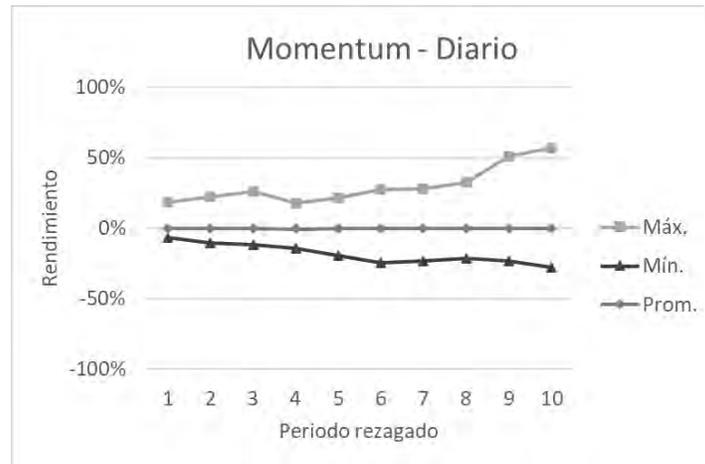


Fuente: Elaboración propia.

En la figura 4.23 se muestran los rendimientos promedio de las series diaria, semanal y mensual de Peñoles para la anomalía de *momentum*. Se observa que los rendimientos de la serie mensual son consistentemente positivos para todos los periodos rezagados, mientras que los rendimientos de la serie semanal son en su mayoría negativos. Por otro lado, los rendimientos de la serie diaria son ligeramente positivos.

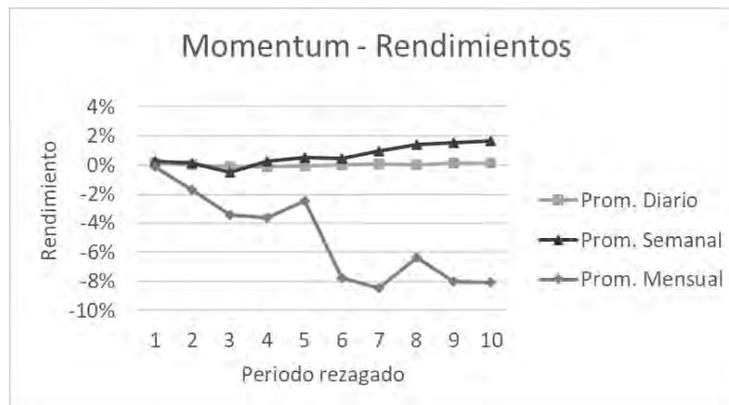
Se puede observar en las figuras 4.24 y 4.25 un ejemplo del caso de Cemex, en el que se aprecia que las diferencias entre el rendimiento máximo y mínimo aumentan a medida que el periodo rezagado se aleja del origen de la anomalía. En cuanto a los promedios comparados, se puede inferir que el rendimiento mensual es negativo, mientras que los rendimientos semanales y diarios son ligeramente positivos.

Figura 4.24 Rendimientos promedio, máximos y mínimos de la serie diaria de Cemex posteriores a la presencia de la anomalía *momentum*.



Fuente: Elaboración propia.

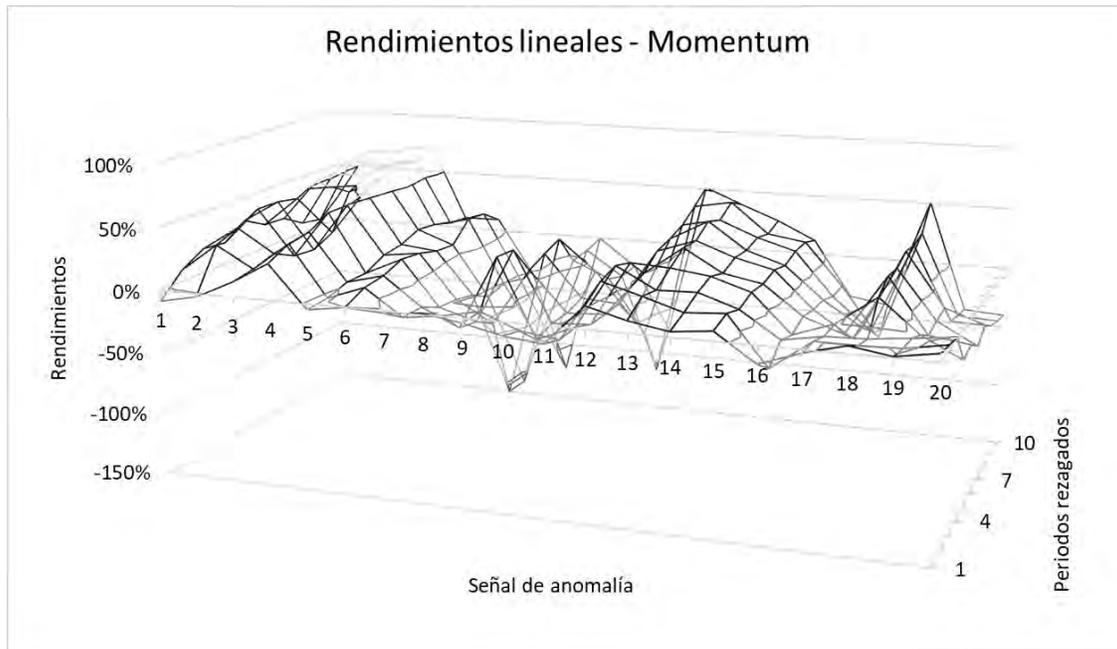
Figura 4.25 Rendimientos promedio de las series mensual, semanal y diaria de Cemex para los diferentes periodos rezagados de la anomalía *momentum*.



Fuente: Elaboración propia.

En la figura 4.26 se ilustra el comportamiento de los rendimientos una vez que se identifica la anomalía de *momentum* en la serie mensual de Cemex. La gráfica presenta los rendimientos posteriores a la detección de la anomalía, suponiendo un comportamiento lineal, con el fin de ejemplificar la posible relación entre la anomalía y el rendimiento. Por otro lado, en la figura 4.27 se observan los rendimientos promedio para cada rezago, donde se puede notar que estos caen a partir del periodo 4 aunque posteriormente parece volver a recuperarse.

Figura 4.26 Proyección de los rendimientos para periodos rezagados largos en la serie mensual de Cemex.



Fuente: Elaboración propia.

Figura 4.27 Rendimientos promedio de la anomalía *momentum* en la serie mensual de Cemex.



Fuente: Elaboración propia.

#### 4.1.4 Discusión

Como se mencionó en capítulos anteriores, el IPC ha exhibido un comportamiento creciente con una tendencia positiva durante gran parte de su historia. Los mercados emergentes, dadas

las características que se les atribuyen en la literatura, son candidatos a ser evaluados con anomalías que aprovechen mejor estos movimientos. Por lo tanto, las anomalías que mejor se desempeñan son aquellas que aprovechan la tendencia de los precios, como el caso del *momentum* y de la reversión a la media. El *momentum* aprovecha la fuerza de un movimiento en una misma dirección, mientras que la reversión a la media presenta dos casos: el primer caso, el precio se desvía de su tendencia y luego la retoma, apoyando al *momentum*; el segundo caso, el precio se mueve en una dirección durante un tiempo y repentinamente cambia de tendencia, es decir, comienza una divergencia en los precios. En este caso, se podría considerar que el movimiento regresa a una media más lenta.

La propuesta de cálculo de la anomalía de volatilidad, en términos de una regresión a la media sobre la desviación estándar, supone que, una vez que se presenta una desviación pronunciada en los movimientos de precios, ésta regresa a un periodo de estabilidad o de movimientos con dirección determinada. Por lo tanto, se infiere que esta característica corresponde a un apoyo de tendencias. Esto se puede confirmar con los resultados de ajuste, cuyos porcentajes le siguieron al *momentum* y a la regresión de la media en los precios.

En general, se observó que todas las anomalías presentaron porcentajes de ajuste muy similares tanto en la evaluación con todo el conjunto de datos, como con la validación cruzada. Estos porcentajes no variaron demasiado entre los diferentes modelos y se encontraron por encima del 75%, lo cual indica un grado de presencia de medio a alto de acuerdo al criterio establecido en el capítulo anterior para catalogar la anomalía (ecuación 3.18).

Los porcentajes de ajuste para las anomalías de calendario y volatilidad son muy similares entre sí. En las acciones de Peñoles y Cemex, las anomalías de calendario presentaron un mejor ajuste que la volatilidad, mientras que en el portafolio de mayor bursatilidad, las anomalías de calendario tuvieron el mejor ajuste. En el portafolio de menor bursatilidad, la volatilidad tuvo mejor ajuste que las anomalías de calendario. A pesar de esto, los porcentajes de ajuste en general están cercanos al 90%.

Los resultados de la anomalía de *momentum* muestran porcentajes de ajuste superiores al 90% en la mayoría de los casos, con excepción de Cemex en la serie mensual con tendencia positiva, donde solo alcanza un 33% con 3 periodos de rezago. Esto podría deberse a la falta de datos históricos disponibles, ya que, para las series mensuales, y en particular para el *momentum*, el número de señales detectadas fue muy reducido en los casos de Peñoles y Cemex: alrededor de 12 señales. A pesar de esto, en general, los casos explorados confirman los resultados de Días Ruiz, herrerías y Vásquez (2020), donde se sugiere la presencia de esta anomalía en el mercado mexicano.

Los resultados de la anomalía de reversión a la media se encuentran justamente debajo de la anomalía de *momentum*, con porcentajes de ajuste superiores al 84% en la mayoría de los casos, excepto en el portafolio de las empresas con menor bursatilidad en tendencia positiva en la serie semanal, donde alcanzó un 66%. Al igual que con la anomalía de *momentum*, la falta de datos suficientes para ejecutar adecuadamente la prueba podría ser la causa del bajo desempeño del clasificador, ya que solo se contaron con alrededor de 10 señales. De acuerdo al criterio de presencia, esta anomalía se puede clasificar como de “presencia alta”. Aunque este tipo de anomalías se podría pensar que aprovechan más los movimientos cíclicos. Lo que se observó a través del método de cálculo propuesto es que estas anomalías apoyan los movimientos con tendencia, confirmación o inicio de ésta, lo que se refleja en los altos porcentajes de ajuste obtenidos.

Los porcentajes de ajuste de la anomalía de volatilidad son superiores al 72% en Peñoles y Cemex, y superiores al 87% en los portafolios de empresas de mayor y menor bursatilidad. La serie mensual de Peñoles alcanza los porcentajes mínimos. Es importante destacar que esta anomalía se calcula en función de la desviación estándar y no infiere un movimiento de precios en alguna dirección o tendencia. Por tanto, puede ser de poca utilidad si se usa de manera aislada. Sin embargo, los resultados de este estudio sugieren que puede ser útil al combinarse con otra anomalía que apoye la dirección de los precios o confirmación de tendencia, como el *momentum*.

Con relación a las anomalías de calendario, se observaron porcentajes por encima del 83% en su mayoría, a excepción del efecto enero que presentó un mínimo del 61% en Cemex para tendencia negativa y del 66% en tendencia positiva. Los efectos de fin de mes y de fin de semana se mantuvieron sobre el 86% en todos los casos, incluyendo Peñoles, Cemex y los portafolios de empresas. El efecto enero mostró una gran variabilidad, llegando incluso al 100% de ajuste, pero también alcanzando los porcentajes más bajos de casi el 60%, lo que, según el criterio de categorización lo ubica en “presencia media”. Estos resultados confirman los hallazgos de estudios previos de Cabello y Ortiz (2003), Rodríguez y Morales (2009), y contradicen a Garay Alvarado y Demmler (2019) quienes indican no contar con elementos suficientes para llegar a una conclusión. Respecto al efecto de fin de mes, se observó un porcentaje de alrededor del 87% categorizándolo en “presencia media”, lo que contradice el resultado obtenido por Duarte, Sierra y Garcés (2016) y confirma los hallazgos de los estudios de Rodríguez y Morales (2009), García (2011) y Kristjanpoller y Arenas (2015). Por otro lado, el efecto de fin de semana oscila entre el 87% y el 95% ubicándolo en la categoría de “presencia media”. Estos resultados se oponen a lo concluido en el estudio de Duarte, Sierra y Garcés (2016), quienes encontraron poca o nula presencia.

Además, al analizar las gráficas del comportamiento de los rendimientos lineales de la figura 4.26, se puede observar que la variabilidad de los rendimientos aumenta de manera significativa a medida que los rezagos van creciendo, especialmente a partir del cuarto periodo rezagado. Los primeros tres periodos rezagados parecen ser los de menor riesgo en términos de la caída máxima que podría incurrirse. Sin embargo, es importante tener en cuenta que estos ejercicios son ilustrativos, ya que dentro de las variables independientes que componen los modelos se incluyen los rendimientos ajustados, además de la anomalía.

Se evaluaron algunos de los clasificadores más destacados del estado del arte y se encontró que las MSV tuvieron el mejor desempeño y consistencia en general. Con la excepción de los BA, cuyo desempeño fue muy variable (desde 100% de ajuste en modelos con el conjunto de datos completo hasta el 30% mediante validación cruzada). El resto de los clasificadores presentaron porcentajes de desempeño similares tanto para los datos dentro de la muestra como para la validación cruzada. Es posible que esto se deba a que las metodologías

utilizadas encontraron relaciones suficientemente separables linealmente, lo que permitió que las MSV obtuvieran un mejor desempeño que los otros clasificadores y, por lo tanto, los porcentajes de desempeño de los modelos fueron más consistentes y ligeramente superiores.

### ***Resumen***

En este capítulo se han presentado los resultados obtenidos utilizando los criterios y procedimientos de análisis enunciados en el capítulo anterior. Los ejemplos que se exhibieron son muestras representativas de todos los conjuntos de prueba obtenidos, y a partir de ellos se ha determinado el grado de presencia de las anomalías mediante el umbral propuesto para la categorización del porcentaje de exactitud del modelo. Se han considerado dos casos: dos empresas del IPC y dos portafolios de empresas mexicanas de la BMV. El objetivo de esta comparación es probar si las anomalías ofrecen poder de discriminación que se suma a los modelos y puedan actuar como indicador de rendimientos futuros.

En el siguiente capítulo se presentan las conclusiones a las que se ha llegado a partir de los datos obtenidos y resultados de los casos probados. Se exponen las ideas finales que se relacionan con el objetivo de este trabajo y, sobre todo, que responden a las preguntas de investigación planteados al inicio del estudio.

## CONCLUSIONES

La presente investigación se centra en probar HME en el contexto del mercado mexicano para las seis anomalías específicas: *momentum*, reversión a la media, volatilidad, efecto enero, efecto de fin de mes y efecto de fin de semana. Se lleva a cabo un análisis detallado de estas anomalías mediante un procedimiento que integra métodos cuantitativos no lineales de aprendizaje automático. Se establece un criterio de categorización del grado de presencia de las anomalías. También se determinan los periodos de rendimientos promedio ajustados al riesgo considerando al IPC como activo de referencia para el periodo 2005 a 2020. Los resultados preliminares del presente trabajo han sido presentados en foros del área económico-financieras, tanto nacionales como internacionales, y han despertado un gran interés debido a la metodología propuesta y resultados obtenidos (Chávez Sánchez et al., 2021a y 2021b).

En respuesta a los objetivos de investigación, se concluye lo siguiente:

Obj. 1: *Analizar las evidencias de la presencia de las anomalías de mercado en el mercado mexicano para conocer su influencia y cómo éstas han sido evaluadas.*

- a) *Las evidencias de los estudios sobre la influencia de las anomalías en el mercado mexicano presentan resultados no concluyentes sobre su impacto en los rendimientos.*

Durante la investigación, se constata que el mercado mexicano cuenta con información histórica limitada. Esto se debe, en parte, a la cantidad reducida de empresas listadas. Hasta 2022, tan solo había 126 empresas mexicanas en comparación con las 143 que figuran en la Bolsa Mexicana de Valores (BMV). Esto contrasta significativamente con otras bolsas, como la de Brasil, que cuenta con 550 compañías, o la de Nueva York, que alberga alrededor de 3,200 empresas. Esta

disparidad en la cantidad de empresas, combinada con el constante aumento y rotación de compañías en la BMV, dificulta la realización de análisis en los mismos horizontes de tiempo. Además, la escasez y la calidad variable de los datos presentes en el mercado mexicano imponen limitaciones adicionales, ya que no es posible analizar una gran cantidad de anomalías reportadas, especialmente aquellas que requieren información cruzada entre empresas.

Sin embargo, se ha constatado que las anomalías estudiadas desafían la HME al mostrar su presencia en el mercado mexicano. En particular, se ha observado una fuerte presencia de *momentum*, una presencia moderada de reversión a la media y volatilidad, según el criterio de categorización propuesto. Además, se ha identificado la presencia de las anomalías de calendario, como el efecto enero, fin de mes y fin de semana.

*Obj. 2: Emplear las técnicas de aprendizaje automático regresión logística, perceptrón multicapa, máquinas de soporte vectorial y bosques aleatorios como mecanismos de evaluación de ajuste de los modelos planteados para establecer el grado de presencia de las anomalías.*

b) *El aprendizaje automático ayuda a identificar relaciones no lineales entre las anomalías de mercado y los rendimientos futuros de los activos del mercado mexicano para confirmar su grado de influencia.*

Entre las metodologías seleccionadas, se encontró que las MSV lograron los mejores ajustes de los modelos, y son los resultados obtenidos con esta metodología los que se presentan en la investigación. Las MSV se utilizan como clasificador para estimar la categoría de presencia de las anomalías. La RL y el PMC obtuvieron porcentajes de desempeño muy similares y solo ligeramente inferiores a los de la MSV. Por otro lado, los BA, que suelen tener un desempeño alto en modelos de análisis de datos financieros, exhibieron resultados con alta variabilidad y menos consistentes en comparación con el resto de las metodologías.

Obj. 3: *Mostrar los periodos futuros de mayor influencia de las anomalías en el mercado mexicano.*

c) *Las anomalías impactan positivamente en los rendimientos futuros de los activos del mercado mexicano en el corto plazo,*

Además de analizar la presencia de las anomalías, un aspecto consiste en determinar los rendimientos promedio ajustados para validar su utilidad. En este análisis, se evalúan los rendimientos promedio de diferentes periodos, considerando una estrategia de “comprar y mantener”. Los resultados señalan que los primeros cuatro periodos son los que ofrecen los mejores rendimientos promedio ajustados al riesgo en general.

Se ha demostrado que, a pesar de ser un mercado emergente con limitados registros históricos, es posible obtener resultados relevantes en el mercado mexicano para la toma de decisiones considerando las anomalías existentes. Este trabajo confirma y extiende análisis anteriores relacionados a las anomalías *momentum*, reversión a la media, volatilidad y calendario revisadas, mostrando su presencia en otros términos más allá de los estadísticos y cálculos típicos. Los resultados obtenidos indican que las anomalías estudiadas son relevantes para explicar los movimientos de los precios en las empresas mexicanas que cotizan en la BMV. Sin embargo, es importante destacar que estos porcentajes de ajuste no aseguran la efectividad de las estrategias basadas en estas anomalías, ya que se requiere llevar a cabo un análisis de rendimiento y riesgo de las estrategias implementadas para evaluar su efectividad.

### ***Trabajo futuro***

Con motivo de la investigación realizada se identifican algunas extensiones del presente trabajo en adición al estudio y análisis de nuevas variantes:

- *Poder predictivo de las anomalías.*

La construcción del modelo puede ser revisado para identificar la capacidad predictiva de las anomalías. Esto genera la posibilidad de que las anomalías pueden predecir algún tipo de característica o movimiento de los precios o sus rendimientos.

- *Modelos multifactoriales de las anomalías presentadas*

Las anomalías analizadas pueden estudiarse de manera conjunta para evaluar si la combinación de estas podría dar lugar a un modelo de estimación de rendimientos con un ajuste más sólido que los modelos individualmente explorados.

- *Estudio en otros mercados y activos*

El modelo de trabajo propuesto pudiera ser replicado para analizar otros mercados, tipos de activos e incluso realizar comparaciones entre diferentes índices en términos de estimación de rendimientos.

- *Otras anomalías con el mismo procedimiento*

Aplicar el modelo de trabajo propuesto para analizar otras anomalías y evaluar el grado de influencia en los rendimientos futuros de los activos.

- *Estrategias de inversión con las anomalías*

Se pueden realizar estudios sobre la implementación integral de las anomalías en estrategias de inversión con el fin de demostrar la utilidad práctica del modelo explorados.

- *Optimización de las anomalías*

Es posible encontrar las versiones optimizadas de las anomalías estudiadas para determinar el mejor periodo y ventana de datos. Esto con el fin de revisar si el ajuste de los modelos pudiera mejorar o incluso que el ajuste pudiera resultar en rendimientos mayores.

## REFERENCIAS

- Akwolaga, T. (2020). Assessing the Random Walk Hypothesis in Emerging Markets: Evidence from the Stock Exchange of Mauritius. *International Journal of Advances in Engineering and Management*, 2(6), 209-217.
- Alexeev, V., y Tapon, F. (2011). Testing weak form efficiency on the Toronto Stock Exchange. *Journal of Empirical Finance*, 18(4), 661-691.
- Ansah, K., Denwar, I. W., y Appati, J. K. (2022). Intelligent Models for Stock Price Prediction: A Comprehensive Review. *Journal of Information Technology Research*, 15(1), 17.
- Argade, S., Chothe, P., Gawande, A., y Joshi, S. (2022). Machine Learning in Stock Market Prediction: A Review. *Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2022*.
- Ariel, R. (1987). A Montly Effect in Stock Returns. *Journal of Financial Economics*, 18(1), 161-174.
- Ashadi, S., Hadavandia, E., Mehmanpazirb, F., y Nakhostin, M. M. (2012). Hybridization od Evolutionary Levenberg-Marquardt Neural Networks and Data Preprocessing for Stock Market Prediction. *Knowledge-based Systems*, 35, 245-258.
- Baker, M., Bradley, B., y Wurgler, J. (2011). Benchmarks as Limits to Arbitrage: Understanding the Low-Volatility Anomaly. *Financial Analysts Journal*, 67(1), 40-54.
- Balvers, R., y Wu, Y. (2006). Momentum and Mean Reversion Across National Equity Markets. *Journal of Empirical Finance*, 13(1), 24-48.
- Barberis, N., y Huang, M. (2009). Preferences with frames: A new utility specification that allows for the framing of risks. *Journal of Economic Dynamics and Control*, 33(8), 1555-1576.

- Barberis, N., y Thaler, R. (2003). A Survey of Behavioral Finance. En G. M. Constantinides, M. Harris, & R. M. Stulz (Eds.), *Handbook of the Economics of Finance: Financial Markets and Asset Pricing* (Vol. 1B, pp. 1053-1128). Elsevier.
- Basu, S. (1977). Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis. *Journal of Finance*, 32(3), 663-682.
- Bhattacharya, U., Daouk, H., Jorgenson, B., y Kehr, C. H. (2000). When an Event is not an Event: The Curious Case of an Emerging Market. *Journal of Financial Economics*, 55(1), 69-101.
- Black, F. (1972). Capital Market Equilibrium with Restricted Borrowing. *The Journal of Business*, 45(3), 444-455.
- Black, F., Jensen, M. C., y Scholes, M. (1972). The Capital Asset Pricing Model: Some Empirical Tests. En M. C. Jensen (Ed.), *Studies in the Theory of Capital Markets* (pp. 79-121). New York, NY: Praeger Publishers.
- BMV, G. (2019). *El Índice de Precios y Cotizaciones y su importancia para el mercado*. Recuperado el 02 de Julio de 2022, de Blog BMV: <https://blog.bmv.com.mx/2019/03/el-indice-de-precios-y-cotizaciones/>
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics*, 31(3), 307-327.
- Breiman, L. (2001). Random forests. *Machine Learning*(45), 5-32.
- Brock, W., Lakonishok, J., y LeBaron, B. (1992). Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. *The Journal of Finance*, 47(5), 1731-1764.
- Brun, X., y Moreno, M. (2008). *Análisis y selección de inversiones en mercados financieros*. Barcelona, España: Profit.
- Cabello, A., y Ortíz, E. (2003). Day of the week and month of the year anomalies in the mexican stock market. *Revista Mexicana de Economía y Finanzas*, 2(3), 217-241.
- Campbell, J., Lo, A., y MacKinlay, A. (2012). *The Econometrics of Financial Markets*. New Jersey: Princeton University Press.
- Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance*, 52(1), 57-82.

- Cervantes Zepeda, M. (1999). What Explains the Returns in the Mexican Stock Market? *Tesis Doctoral*. Monterrey, Nuevo León, México: ITESM. Recuperado el 20 de Octubre de 2021, de <https://repositorio.tec.mx/handle/11285/572615?show=full&locale-attribute=es>
- Chan, L., Jegadeesh, N., y Lakonishok, J. (1996). Momentum Strategies. *The Journal of Finance*, 51(5), 1681-1713.
- Chávez Sánchez, C. O. (2012). Discriminación robusta de imágenes basada en momentos geométricos y el clasificador Gamma. *Tesis de maestría*. Cd. de México, México: Centro de Investigación en Computación, Instituto Politécnico Nacional. Recuperado el 10 de Marzo de 2020, de Tesis IPN: <https://tesis.ipn.mx/handle/123456789/16012>
- Chavez Sanchez, C. O., Morales Castro, A., Rodríguez Castellanos, A., y García Salgado, O. (2021a). Bolsa Mexicana de Valores: Presencia de Anomalías de Mercado y su Relevancia. *XXXV Congreso Anual de AEDEM Digitalización de las organizaciones: ventaja competitiva o adaptación forzada*, 133-150.
- Chávez Sánchez, C. O., Morales Castro, A., Rodríguez Castellanos, A., y García Salgado, O. (2021b). Bolsa Mexicana de Valores: la EMH y sus anomalías. *Ponencia presentada al 1er. Congreso Internacional de Grupos de Investigación - CIGI 2021 y 3er. Encuentro de Cuerpos Académicos*. Guadalajara, Jalisco: Universidad de Guadalajara y Verano TIC, 18-20 octubre.
- Chen, C., Huang, C., y Lai, H. (2009). The impact of data snooping on the testing of technical analysis: an empirical study of Asian markets. *Journal of Asian Economics*, 20(5), 580-591.
- Cheong, C. W., y Isa, Z. (2007). Test of random walk hypothesis under drift and structural break-a nonparametric approach. *World applied sciences journal*, 2(6), 674-681.
- Chopra, R., y Sharma, G. D. (2021). Application of Artificial Ontelligence in Stock Market Forecastiong: A Critique, Review and Research Agenda. *Journal of Risk and Finantial Management*, 14(256).
- Chordia, T., Subrahmanyam, A., y Tong, Q. (2014). Have Capital Markets Anomalies Attenuated in the Recent Era of High Liquidity and Trading Activity? *Journal of Accounting and Economics*, 58(1), 41-58.

- Cross, F. (1973). The Behavior of Stock Prices on Fridays and Mondays. *Financial Analysts Journal*, 29(6), 67-69.
- Cruz, I. (2007). Pronósticos en el mercado de derivados utilizando RNA y modelos ARIMA: una aplicación al Cete de 91 días en el MexDer. *MexDer*. México, D.F., México. Recuperado el 24 de Enero de 2020, de <http://www.mexder.com.mx/>
- Cutler, D., Poterba, J., y Summers, L. (1991). Speculative Dynamics. *The Review of Economics Studies*, 58(3), 529-549.
- Damodaran, A. (2012). *Investment Valuation: Tools and Techniques for Determining the Value of Any Asset*. Hoboken, New Jersey: Wiley Finance.
- De Bondt, W., y Thaler, R. (1985). Does the Stock Market Overreact? *The Journal of Finance*, 40(3), 793-805.
- Delcey, T. (2019). Samuelson vs Fama on the Efficient Market Hypothesis: The Point of View of Expertise. *Economía-Historie/Epistémologie/Philosophie*, 9(1), 37-58.
- Díaz Ruiz, P., Herrerías, R., y Vásquez, A. (2020). Anomalies in Emerging Markets: The Case of Mexico. *The North American Journal of Economics and Finance*, 53, 59. <https://doi.org/10.1016/j.najef.2020.101188>
- Dimson, E., y Marsh, P. (1984). An analysis of brokers' and analysts' unpublished forecasts of UK stock returns. *The Journal of Finance*, 39(5), 1257-1292.
- Dorantes Dosamantes, C. A. (2013). The Relevance of Using Accounting Fundamentals in the Mexican Stock Market. *Journal of Economics, Finance and Administrative Science*, 18, 2-10.
- Duarte Duarte, J. B., Sierra Suárez, K. J., y Garcés Carreño, L. D. (2016). Anomalías de Mercado en la Bolsa de Valores de México. *XVIII Congreso Internacional de Contaduría, Administración y Finanzas*, 15.
- Engle, R. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4), 987-1007.
- Fama, E. (1965). The behavior of stock-market prices. *The Journal of Business*, 38(1), 34-105.
- Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 353-417.

- Fama, E. F., y French, K. R. (1988). Permanent and Temporary Components of Stock Prices. *The Journal of Political Economy*, 96(2), 246-273.
- Fama, E. F., y French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, XLVII(2).
- Fama, E. F., y French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33(1), 3-56.
- Fama, E. F., y MacBeth, J. D. (1973). Risk, Return and Equilibrium: Empirical Test. *Journal of Political Economy*, 81(3), 607-636.
- Fama, E., y French, K. (1996). Multifactor Explanations of Asset Pricing Anomalies. *The Journal of Finance*, 51(1), 55-84.
- Fernández, R., Palma, B., y Rho, C. (2021). A sentiment-based risk indicator for the mexican financial sector. *Latin american journal of central banking*, 2(3), 100036.
- Ferreira, F., Gandomi, A., y Cardoso, R. (2021). Artificial Intelligence Applied to Stock Market Trading: A Review. *IEEE Access*, 30898-30917.
- Fisher, L. (1966). Some new stock-market indexes. *Journal of businesses*, 39, 191-225.
- Frankfurter, G. M., y McGoun, E. G. (2001). Anomalies in Finance: What are They and What are They Good for? *International Review of Financial Analysis*, 10(4), 407-429.
- French, K. (1980). Stock Returns and the Weekend Effect. *Journal of Financial Economics*, 8(1), 55-69.
- Friedman, M. (1953). The Case For Flexible Exchanges Rates. *Essays in Positive Economies*.
- Gajapathy, V., y Sathyanarayana, K. (2021). Testing Weak Form of Market Efficiency During Pandemic. *Palarch's Journal of Archaeology of Egypt/Egyptology*, 18(9), 248-258.
- Garay Alvarado, M., y Demmler, M. (2019). Analysis of the January Effect in Time Series of Mexican Stock Market indexes. *Mercados y Negocios*, 40(20), 43-62.
- García Blandón, J. (2011). Return Seasonality in Emerging Markets: Evidence From Latin America. Em J. Batten, & P. G. Szilagyi, *The Impact of the Global Financial Crisis on Emerging Financial Markets* (pp. 405-423). Leeds, U.K.: Emerald Group.
- García, M. C., Jalal, A. M., Garzón, L. A., y López, J. M. (2013). Métodos para Predecir Índices Bursátiles. *Ecos de Economía*, 17(37), 51-82.

- Gibbons, M., y Hess, P. (1981). Day of the Week Effects and Asset Returns. *The Journal of Business*, 54(4), 579-596.
- Gimba, V. (2012). Testing the weak-form efficiency market hypothesis: Evidence from Nigerian stock market. *CBN Journal of Applied Statistics*, 2(1), 117-136.
- Gómez Martínez, R., Prado Román, C., y Saz Peñas, L. J. (2020). Inteligencia artificial como herramienta de análisis en gestión value. *Recta@*, 21, 89-104. <https://doi.org/10.24309/recta.2020.20.2.01>
- Green, J., Hand, J. R., y Zhang, F. (2012). The Supraview of Return Predictive Signals. *Review of Accounting Studies*, 18(3), 692–730.
- Grupo BMV. (2018). *La historia del IPC y su importancia para el mercado*. Recuperado el 20 de Febrero de 2022, de BMV: <https://blog.bmv.com.mx/>
- Grupo BMV. (2020). *Informe Anual Integrado 2020*. Recuperado el 10 de Septiembre de 2022, de BMV: <https://www.bmv.com.mx/>
- Gultekin, M., y Gultekin, N. (1983). Stock market seasonality: international evidence. *Journal of finance and economics*, 12, 469-481.
- Gutiérrez, M. (2016). *Diccionario económico*. Recuperado el 16 de Junio de 2022, de Expansión: <http://www.expansion.com/diccionario-economico/teoria-del-mercado-eficiente.html>
- Hall, M., Frank, E., Holmes, G., Pfahringer, P. R., y Witten, I. H. (2009). The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 11, 10-18.
- Harvey, C. (1995). Predictable Risk and Returns in Emerging Markets. *Review of Financial Studies*, 8(3), 773-816.
- Harvey, C. R., Liu, Y., y Zhu, H. (2016). ... and the Cross-Section of Expected Returns. *The Review of Financial Studies*, 29(1), 5-68.
- Hawawini, G., y Keim, D. (1995). On the Predictability of Common Stocks Returns: Worldwide Evidence. *Handbooks in operations research and management science*, 9, 487-544.
- Herrera, M. J., y Lockwood, L. J. (1994). The Size Effect in the Mexican Stock Market. *Journal of Banking & Finance*, 18(4), 621-632.

- Hinojosa, R., Trejo, L., Hervert, L., Hernández, N., y González, E. (2021). Mexican Stock Return Prediction with Differential Evolution for Hyperparameter Tuning. *Advances in computational intelligence, MICAI, 184*, 355-368.
- Hirsch, R. (2021). Risk and Trouble: Adam Smith on Profit and the Protagonists of Capitalism. *American Journal of Political Science, 65*(1), 166-179.
- Ho, K., Zheng, L., y Zhang, Z. (2012). Volume, volatility and information linkages in the stock and option markets. *Review of Financial Economics, 21*(4), 168-174.
- Holmes, G., Witten, I. H., y Donkin, A. (1994). WEKA: A Machine Learning Wokbench. *Proceedings of ANZIIS '94 - Australian New Zealand Intelligent Information Systems Conference*.
- Hong, K. (2014). Asymmetric Impact of Financial Integration to International Nonsynchronous Trading Effects in Developed and Emerging Equity Markets. *Theoretical Economics Letters, 4*(7), 517-525.
- Hua, Z., Wang, Y., Xu, X., Zhang, B., y Liang, L. (2007). Predicting Corporate Financial Distress Based on Integration of Support Vector Machine and Logistic Regression. *Expert Systems with Applications, 33*(2), 434-440.
- Hull, M., y McGroarty, F. (2014). Do emerging markets become more efficient as they develop? Long memory persistence in equity indices. *Emerging Markets Review, 18*, 45-61.
- Infosel Financiero. (2020). Recuperado el 09 de Febrero de 2021, de <https://www.infoselfinanciero.com.mx/>
- Jegadeh, N. (1990). Evidence of Predictable Behavior of Security Returns. *The Journal of Finance, 45*(3), 881-898.
- Jegadeh, N., y Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficient. *The Journal of Finance, 48*(1), 65-91.
- Jensen, M. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics*(12), 33-56.
- Kahneman, D., y Tversky, A. (1984). Choices, Values and Frames. *American Psychologist, 39*(4), 341-350.
- Kendall, M. (1953). The analysis of economic time series, part 1: series. *Journal of the Royal Statistical Society. Series A (General), 116*(1), 11-34.

- Khursheed, A., Naeem, M., Ahmed, A., y Mustafa, F. (2020). Adaptive market hypothesis: An empirical analysis of time-varying market efficiency of cryptocurrencies. *Cogent Economics & Finance*, 8(1), 1-15.
- Kristjanpoller, W., y Arenas Yanez, T. (2015). El Efecto Fin de Mes en los Principales Mercados Accionarios Latinoamericanos. *Contaduría y Administración*, 60(1), 53-86.
- Kuhe, D. A., y Audu, S. D. (2019). Modelling Volatility Mean Reversion in Stock Market Prices: Implications for Long-Term investment. *Nigerian Journal of Scientific Research*, 15(1), 131-139.
- Kumar, M., y Thenmozhi, M. (2006). Forecasting Stock Index Movement: A Comparison of Support Vector Machines and Random Forest. *Indian Institute of Capital Markets 9th*, (p. 16).
- Lakonishok, J., y Levi, M. (1982). Weekend Effects on Stock Returns: A Note. *The Journal of Finance*, 37(3), 883-889.
- Latif, M., Arshad, S., Fatima, M., y Farooq, S. (2011). Market Efficiency, Market Anomalies, Causes, Evidences and Some Behavioral Aspects of Market Anomal. *Research Journal of Finance and Accounting*, 2(9/10), 1-13.
- Lehmann, B. (1990). Fads, Martingales and Market Efficiency. *The Quarterly Journal of Economics*, 105(1), 1-28.
- Lekovic, M. (2018). Evidence for and against the validity of efficient market hypothesis. *Economics themes*, 56(3), 369-387.
- Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, 47(1), 13-37.
- Liu, Q., Wang, C., Zhang, P., y Zheng, K. (2021). Detecting stock market manipulation via machine learning: Evidence from China Securities Regulatory Commission punishment cases. *International Review of Financial Analysis*, 78, 101887.
- Lo, A. W. (2005). Reconciling efficient markets with behavioral finance: the adaptive market hypothesis. *Journal of Investment Consulting*, 7(2), 21-44.

- Loke, K. S. (2017). Impact of Financial Ratios and Technical Analysis on Stock Price Prediction Using Random Forests. *2017 International Conference on Computer and Drone Applications (IConDA)*, (pp. 38-42). Kuching, Malaysia.
- López, F., y Rodríguez, D. (2010). El efecto enero en las principales bolsas latinoamericanas de valores. *Contaduría y Administración*, *64*(1), 25-46.
- Malkiel, B. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, *17*(1), 59-82.
- Marshall, B., Young, M., y Rose, L. (2006). Candlestick technical trading strategies: can they create value for investors? *Journal of Banking & Finance*, *30*(8), 2303-2323.
- McLean, R., Pontiff, D., y Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, *71*(1), 5-32.
- Montalván, D., Barrón, R., y Godoy, S. (2018). Generating trading strategies in the mexican stock market: a pattern recognition approach. *Research in computer science*, *147*(12), 107-114.
- Moore, A. (1964). Some characteristics of changes in common stock pices. En P. H. Cootner (Ed.), *The random character of stock market prices*. Boston, MA: The MIT Press.
- Moore, D. (2021). Did You Really Beat the Market? A Practical and Parsimonious Approach to Evaluating Risk-Adjusted Performance. *Journal of Mathematical Finance*, *11*(3), 554-577.
- Morgan Stanley. (2021). *MSCI Emerging Markets Index (USD)*. Recuperado el 1 de Mayo de 2021, de Morgan Stanley: <https://www.msci.com/documents/10199/c0db0a48-01f2-4ba9-ad01-226fd5678111>
- Mossin, J. (1966). Equilibrium in a Capital Asset Market. *Econometrica*, *34*(4), 768-783.
- Munir, A. F., Sukor, M. E., y Shaharuddin, S. S. (2022). Adaptive Market Hypothesis and Time-varying Contrarian Effect: Evidence from Emerging Stock Markets of South Asia. *SAGE Open*, *12*(1), 1-16.
- Nti, I., Adekoya, A., y Weyori, B. (2020). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, *53*(4), 3007-3057.
- Picasso, A., Merello, S., Yukun, M., Oneto, L., y Cambria, E. (2019). Technical analysis and sentiment embeddings for market trend prediction. *Expert Systems with Applications*, *135*(30), 60-70.

- Poterba, J. M., y Summers, L. H. (1988). Mean Reversion in Stock Prices: Evidence and Implications. *The Journal of Financial Economics*, 22(1), 27-59.
- Potvin, J., Soriano, P., y Valleé, M. (2004). Generating trading rules on the stock markets with genetic programming. *Computers & Operations Research*, 31, 1033-1047.
- Ramírez Morales, J. J., Ruíz Yepes, E. A., y Gutiérrez Castañeda, B. E. (2015). *Hipótesis de Mercados Eficientes en el Mercado Accionario Colombiano a través del Estudio de Eventos*. XX Congreso Internacional de Contaduría, Administración e Informática.
- Ritter, J. (1988). The Buying and Selling Behavior of Individual Investors at the Turn of the Year. *The Journal of Finance*, 43(3), 701-717.
- Roberts, H. V. (1967). *Statistical versus clinical prediction of the stock market*. Chicago: Universidad de Chicago.
- Rodríguez, D., y Morales, A. (2009). Los efectos "día de la semana" y "mes del año" para los rendimientos de una muestra de acciones que cotizan en la bolsa mexicana de valores. *XV Congreso Internacional de Contaduría, Administración e Informática*, 1-20.
- Rodríguez, M. D., Cortéz, K. A., Méndez, A. B., y Garza, H. H. (2015). Análisis de portafolio por sectores mediante el uso de algoritmos genéticos: caso aplicado a la bolsa mexicana de valores. *Revista Contaduría y Administración*, 60(1), 87-112.
- Roldán, J., y García, M. (2022). A procedure for testing the hypothesis of weak efficiency in financial markets: a Monte Carlo simulation. *Statistical Methods & Applications*, 31, 1289-1327.
- Roseff, M., y Kinney, W. (1976). Capital market seasonality: the case of stock market returns. *Journal of finance and economics*, 3, 379-402.
- Ross, S. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory*, 13(3), 341-360.
- Rossi, M., y Gunardi, A. (2018). Efficient Market Hypothesis And Stock Market Anomalies: Empirical Evidence In Four European Countries. *The Journal of Applied Business Reseach*, 34(1), 183-192.
- Rubistein, M. (2001). Rational Markets: Yes or No? The Affirmative Case. *Financial Analysts Journal*, 57(3), 15-29.
- Samuelson, P. (1965). Proof that Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review*, 6, 41-49.

- Sarin, A. B., y Sharma, S. (2023). Artificial Intelligence in Behavioral Finance for Investment Decision-making. En B. Bhardwaj, D. Sharma, & M. C. Dhiman (Eds.), *AI and Emotional Intelligence for Modern Business Management* (pp. 194-212). Hershey: IGI Global.
- Scholes, M., y Williams, J. (1977). Estimating Betas from Nonsynchronous Data. *Journal of Financial Economics*, 5(3), 309-327.
- Semaan, D., Harb, A., y Kassem, A. (2014). Forecasting exchange rates: Artificial neural networks vs regression. *Third International Conference on e-Technologies and Networks for Development (ICeND)*, 156 - 161.
- Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *The Journal of Finance*, 19(3), 425-442.
- Shreve, S. E. (1992). Martingales and the Theory of Capital Asset Pricing. En L. D. Davisson, G. j. MacFarlane, A. H. Kwakernaak, J. T. Massey, Y. Z. Tsyppkin, A. J. Viterbi, & P. Kall (Eds.), *System Modelling and Optimization, Lecture Notes in Control and Information Sciences* (Vol. 180, pp. 809-823). Berlín, Heidenberg: Springer.
- Shynkevich, A. (2012). Short-term predictability of equity returns along two style dimensions. *Journal of Empirical Finance*, 19(5), 675-685.
- Sonkavde, G., Dharrao, D., Bongale, A., Deokate, S., Doreswamy, D., y Bhat, S. (2023). Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications. *International Journal of Financial Studies*, 11(3), 94.
- Sood, S., Gupta, Y. K., y Bhushan, P. (2023). Exploring the Efficacy of Artificial Intelligence Techniques in Predicting Stock Market Trends. En B. Bhardwaj, D. Sharma, & M. C. Dhiman (Eds.), *AI and Emotional Intelligence for Modern Business Management* (pp. 241-253). Hershey, PA: IGI Global.
- Theil, H., y Leenders, C. (1965). Tomorrow on the Amsterdam Stock Exchange. *The Journal of Business*, 38(3), 277-284.
- Timmermann, A., y Granger, C. W. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting*, 20(1), 15-27.
- Tversky, A., y Thaler, R. H. (1990). Anomalies: Preference Reversals. *Journal of Economics Perspectives*, 4(2), 201-211.

- Uribe Gil, J., y Ulloa Villegas, I. (2011). Revisando la Hipótesis de los Mercados Eficientes: Nuevos Datos, Nuevas Crisis y Nuevas Estimaciones. *Cuadernos de Economía*, XXX(55), 127-154.
- Villareal-Samaniego, D., y Santillán, R. J. (2023). Day of the Week Effect and Adaptive Markets in Latin America Stock Indices. *Estudios Gerenciales*, 39(168), 286-296.
- Weiwei, J. (2021). Applications of Deep Learning in Stock Market Prediction: Recent Progress. *Expert Systems with Applications*, 184, 115537.
- Wold, H. (1938). A Study in the Analysis of Stationary Time Series. Uppsala, Sweden: Almqvist and Wiksell.
- Wong, W. (2020). Editorial Statement and Research Ideas for Efficiency and Anomalies in Stock Markets. *Economies, MDPI Journal*, 8(20), 1-4.
- Wu, Y., y Balvers, R. (2000). Mean Reversion across National Stock Markets and Parametric Contrarian Investment Strategies. *The Journal of Finance*, 55(2), 745-772.

## **Anexo A**

### **PROCEDIMIENTO DE ANÁLISIS.**

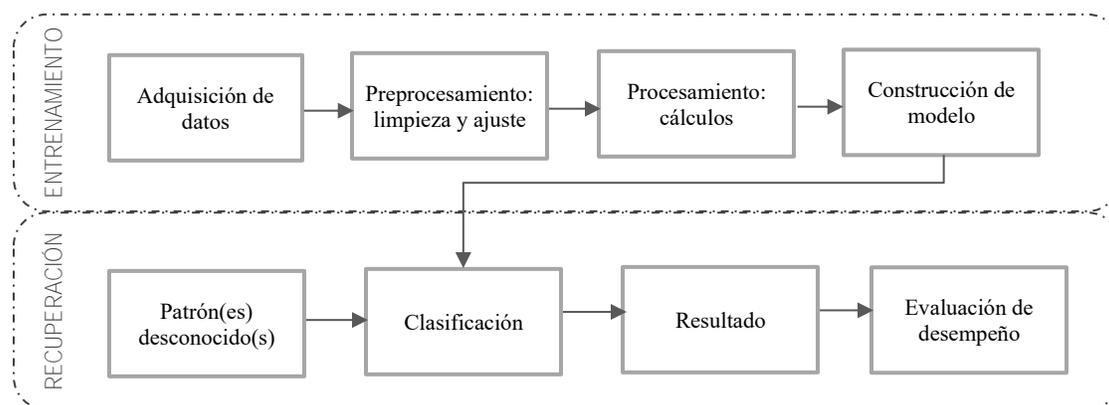
Este apartado describe el proceso manual que se realiza para llevar a cabo el análisis de las anomalías, desde la preparación de los datos, hasta la parametrización y organización de los cálculos. Además, se explica el uso de la plataforma Weka en su versión 3.9.4 (Hall et al., 2009) para llevar a cabo las baterías de pruebas de los diferentes métodos de evaluación del modelo propuesto.

En términos generales, este anexo consiste en la preparación de datos, la generación de los modelos y la utilización del ambiente. Además, se muestra el extracto de código en lenguaje de programación java que automatiza el proceso anterior y ejecuta toda una batería de pruebas. Java se utiliza para integrar las librerías ofrecidas por la plataforma, lo que permite usar los clasificadores, filtros de datos y métodos de evaluación de manera automática e iterativa.

#### ***Modelo de análisis***

El procedimiento comienza con un esquema de análisis de variables dependientes categóricas, las cuales pueden clasificarse en dos categorías: la primera abarca rendimientos superiores al mercado, mientras que la segunda engloba rendimientos inferiores al mercado. Se emplea el aprendizaje supervisado como el método para construir las reglas de conocimiento. Esto implica una primera fase de entrenamiento o aprendizaje, en la que se desarrolla el modelo utilizando datos históricos. Luego, en una segunda fase de recuperación o prueba, se utiliza el modelo creado para evaluar su capacidad de clasificar elementos desconocidos (véase la figura A.1). Finalmente, se evalúa el modelo resultante de cada técnica de análisis para obtener sus métricas de desempeño.

Figura A.1. Proceso de ejecución del estudio.



Fuente: Elaboración propia.

A continuación, se describen los elementos de modelo de análisis para la ejecución manual de un caso de prueba:

### 1. Obtención de datos

Los datos de las acciones se descargan de la plataforma de preferencia. En este trabajo, se utilizó Infosel Financiero (2020) porque ésta es una plataforma gratuita<sup>17</sup>. Se deben seguir los siguientes pasos:

1. Localizar el IPC mexicano:
  - a. Obtener los datos históricos desde el 1ro de enero de 2005 hasta el 31 de diciembre de 2020.
  - b. Descargar los archivos de las series diaria, semanal y mensual para el periodo indicado.
2. Localizar Cemex, Peñoles y las empresas que componen los portafolios de mayor y menor bursatilidad, y seguir los pasos 1.a y 1.b para cada empresa.

### 2. Preprocesamiento

En esta fase se realiza la limpieza de los datos y se establecen las correspondencias por fecha entre las series de cada empresa y índice de referencia del mercado (IPC). Las series pueden contener datos vacíos y sus fechas pueden no coincidir entre el IPC y las empresas, lo que puede generar registros con empate de fechas. En este trabajo se ha

<sup>17</sup> Un inconveniente de este tipo de información es que los datos no son del todo precisos o pueden contener elementos vacíos o fuera de rango. Sin embargo, el objetivo de este trabajo es analizar las generalidades de los movimientos para mostrar la aplicabilidad de las anomalías, mas no la precisión de estas, cuyo objeto de estudio pudiera ser un trabajo futuro.

decidido seguir el criterio de eliminación de registros con datos vacíos. El proceso consta de los siguientes pasos:

1. Emparejar la serie de la empresa con la del IPC.
2. Ordenar las series de forma ascendente.
3. Filtrar los registros vacíos o nulos y eliminarlos.
4. Ejecutar del paso 1 al 3 para ambas empresas y para cada serie: diaria, semanal y mensual.

Cabe destacar que se utiliza Microsoft Excel 2016 como herramienta para el manejo de los datos.

### 3. *Procesamiento*

Se realizan los cálculos y operaciones intermedias de los datos para su comparación y filtrado:

1. Se obtienen los rendimientos de la serie de la empresa y del IPC y se colocan en una columna nueva en la fila de cada fecha. Se recomienda nombrar adecuadamente las columnas para facilitar la comprensión de los datos.
2. Se calculan las anomalías y se agregan en una nueva columna de nombre *dummy* colocando un *1* en la fila donde se detecta anomalía y *0* en caso contrario.
3. Se transponen los siguientes 10 rendimientos<sup>18</sup> para cada fila y se etiquetan asignando *1* si superan al IPC o *0* si no. A las columnas de estos rendimientos se les nombra como *Lag1*, *Lag2*, *Lag3*, ..., hasta *Lag10*.
4. A la matriz resultante compuesta de los 10 periodos de la serie de rendimientos ( $t-1$  hasta  $t-10$ ), la variable de la señal *dummy* y los 10 periodos etiquetados como *Lag* se guarda en un archivo tipo *csv* (*Comma Separated Values*). Es importante asegurarse de que los datos estén correctamente separados por comas para que Weka pueda identificarlos.
5. Se ejecutan los pasos del 1 al 6 para cada serie y cada empresa.

### 4. *Creación de archivo arff*

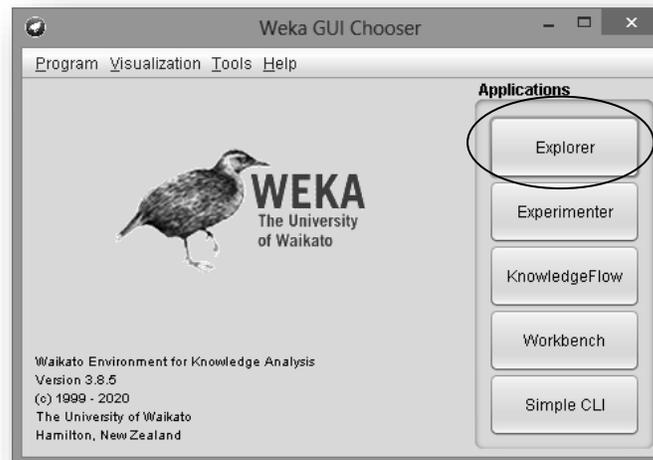
---

<sup>18</sup> Note que el efecto de fin de mes, donde la ventana de tiempo es semanal, y el efecto fin de semana, donde la ventana de tiempo es diaria, el periodo de análisis queda recortado a 4 semanas y 5 días, respectivamente. Esto por la naturaleza de los periodos (el cambio de mes es cada 4 semanas y 5 días para el cambio de semana).

El archivo *csv* debe ser convertido en formato estructurado enriquecido para ser procesado en Weka. Este archivo es de tipo *arff*. Esto se realiza desde la herramienta de la siguiente forma:

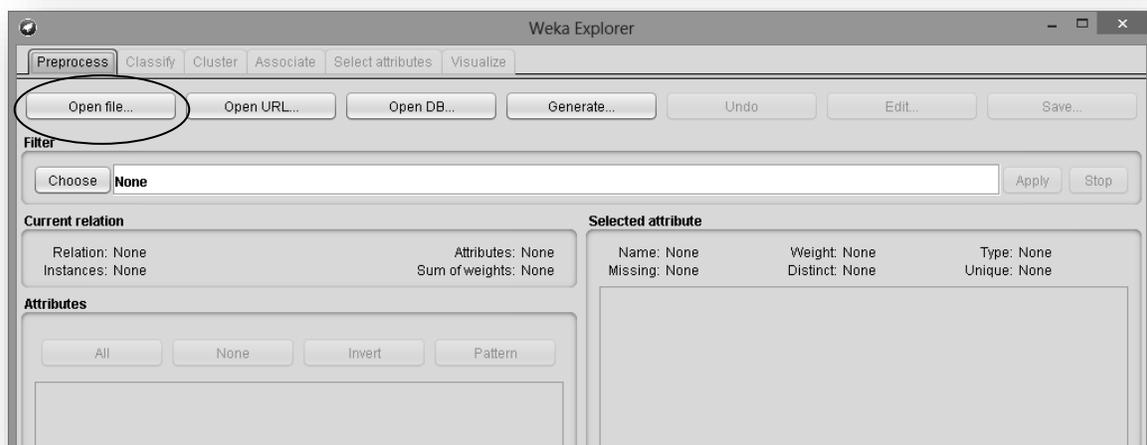
1. Al abrir Weka se muestra la pantalla de la figura A.2. Se selecciona la opción “Explorer” para mostrar la vista de trabajo para la preparación de los modelos de la figura A.3.

Figura A.2. Pantalla principal de Weka en ejecución de aplicación de escritorio.



Fuente: Imagen obtenida de la ejecución de Weka (Hall et al., 2009).

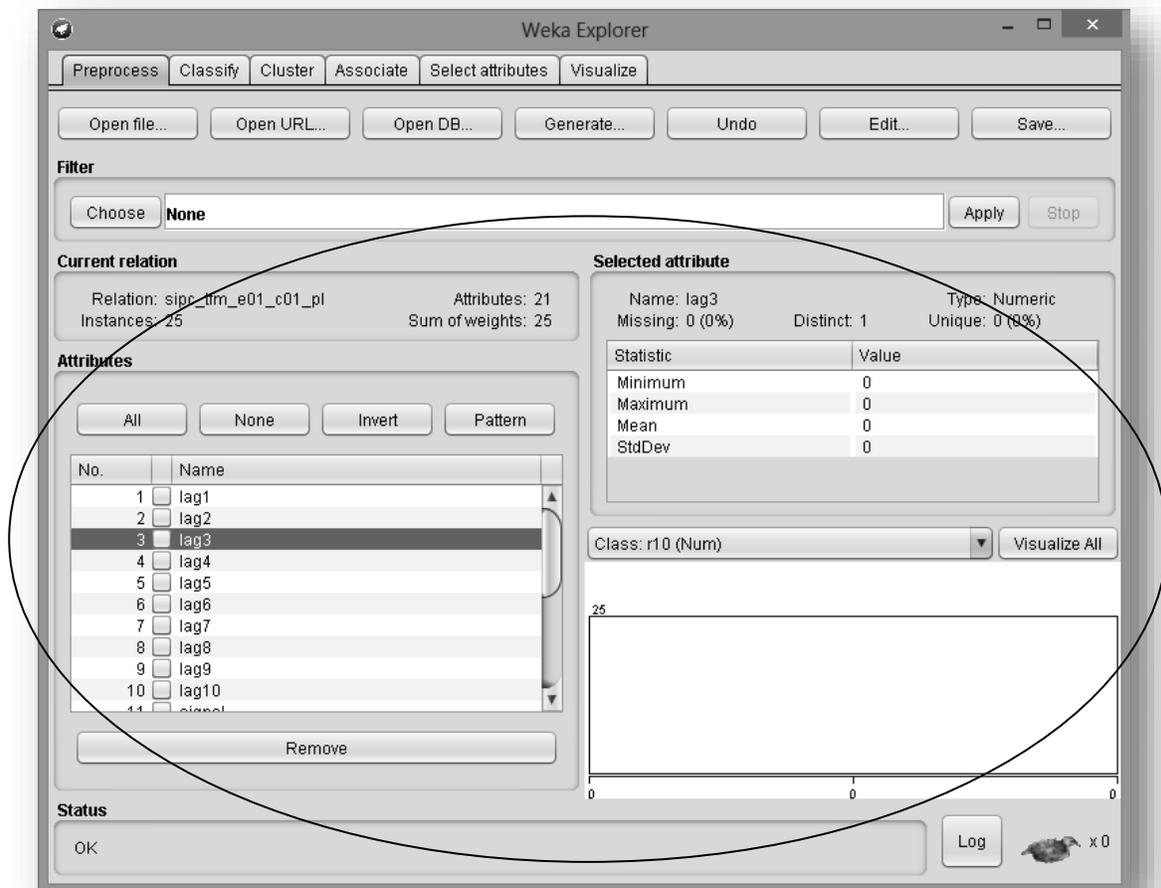
Figura A.3. Preparación de prueba.



Fuente: Imagen obtenida de la ejecución de Weka (Hall et al., 2009).

2. Dentro de la pestaña “Preprocess”, seleccionar la opción “Open file”. Esto mostrará una ventana emergente para buscar el archivo que al seleccionarlo regresará a la ventana llenando los apartados con los factores independientes y los dependientes configurados. Como se muestra en la figura A.4.

Figura A.4. Creación de archivo de configuración de datos arff.



Fuente: Imagen obtenida de la ejecución de Weka (Hall et al., 2009).

3. Se debe tener presente que el archivo abierto es de formato general por lo que se deberá seleccionar la opción etiquetada “Save” para crear un archivo con formato más preciso en términos de lenguaje de etiquetado de Weka con extensión *arff*. Al crear el archivo se tiene como resultado un archivo de texto con encabezados que indican el tipo de variables analizadas junto con metadatos necesarios para

que la plataforma tipifique la información. Este archivo puede ser editado a mano para ajustarlos a conveniencia: nombre, tipo, categoría, etc (figura A.5).

Figura A.5. Descripción de los datos y su especificación.

```

@relation sips_tfm_e01_c01_pl

@attribute up01 {0,1}
@attribute up02 {0,1}
@attribute up03 {0,1}
@attribute up04 {0,1}
@attribute up05 {0,1}
@attribute up06 {0,1}
@attribute up07 {0,1}
@attribute up08 {0,1}
@attribute up09 {0,1}
@attribute up10 {0,1}
@attribute signal {1,0}
@attribute m01 numeric
@attribute m02 numeric
@attribute m03 numeric
@attribute m04 numeric
@attribute m05 numeric
@attribute m06 numeric
@attribute m07 numeric
@attribute m08 numeric
@attribute m09 numeric
@attribute m10 numeric

@data
0,0,0,0,1,1,1,1,1,1,1.5,2,0.7,0,-0.8,-1.6,0.3,-0.3,0.3,-0.8
1,1,1,1,1,1,1,1,0,1,-2.7,1.5,1.3,1,1.5,2,0.7,0,-0.8,-1.6
1,1,0,0,0,0,0,0,0,1,1,-2.1,0.8,-1.3,0.5,-1,-2.7,1.5,1.3,1,1.5
1,1,1,1,1,1,1,1,1,1,2.4,-0.4,6,0,-2.6,-2.1,0.8,-1.3,0.5,-1
1,1,1,1,1,1,1,1,1,1,0.3,0.4,-2.9,-1.5,2.4,-0.4,6,0,-2.6,-2.1
0,0,0,1,1,1,1,1,1,1,0.8,-1,1.4,2.8,-4.3,-1.2,-1.1,-1.6,-1.5,0.
3
1,1,1,1,1,1,1,1,1,1,1,-4,-1.1,0,0.9,0.8,-1,1.4,2.8,-4.3,-1.2

```

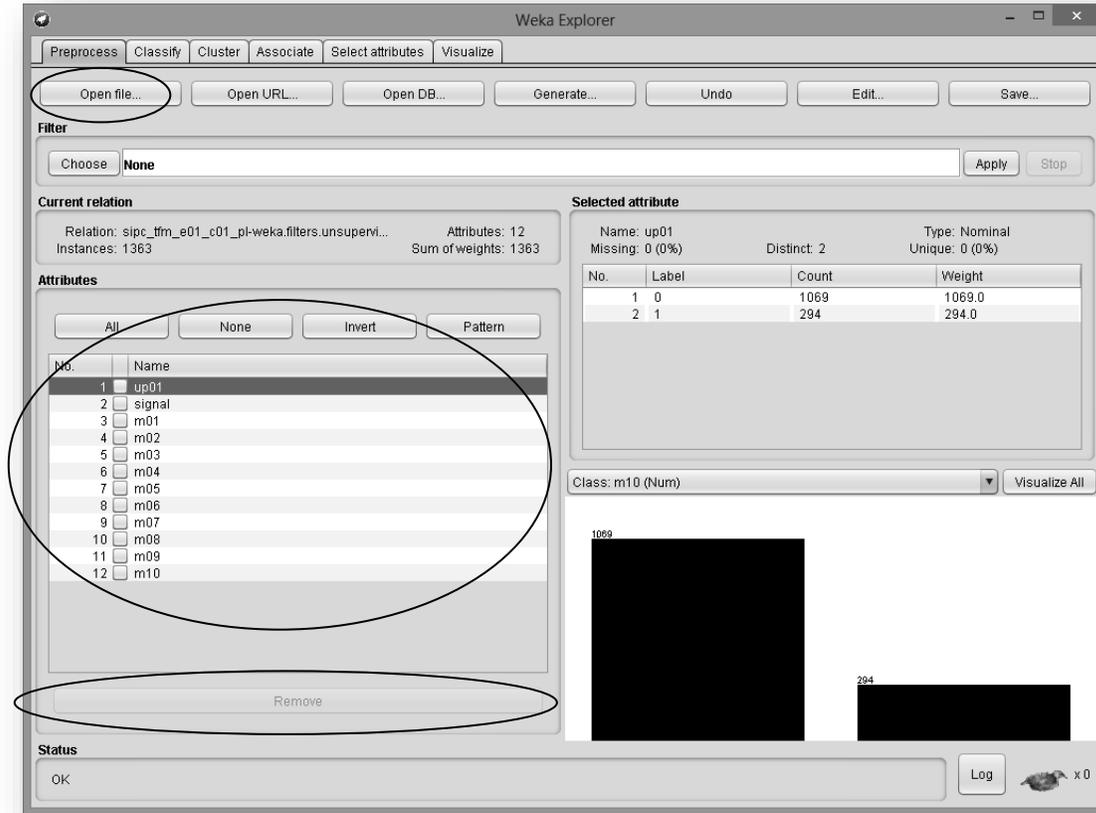
Fuente: Elaboración propia.

## 5. Ejecución de análisis

Para ejecutar un análisis es necesario seleccionar las variables en el apartado “Attributes” que fungirán como variables independientes, así como la dependiente de la siguiente forma:

1. En la pantalla “Explorer”, y dentro la pestaña actual de “Preprocess” se selecciona la opción etiquetada “Open” para localizar y abrir el archivo *arff* a analizar (el creado en el apartado anterior). Al abrir el archivo se llenan los apartados de la pestaña y se procede con la selección de las variables que no se deseen usar en el análisis. Se selecciona la opción “Remove” para quitarlas de la prueba (figura A.6).

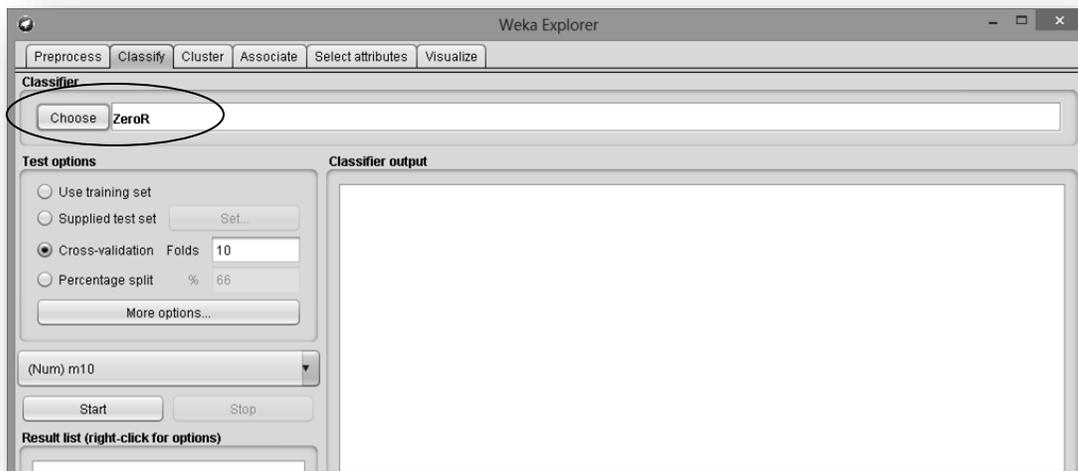
Figura A.6. Selección y filtrado de variables del modelo.



Fuente: Imagen obtenida de la ejecución de Weka (Hall et al., 2009).

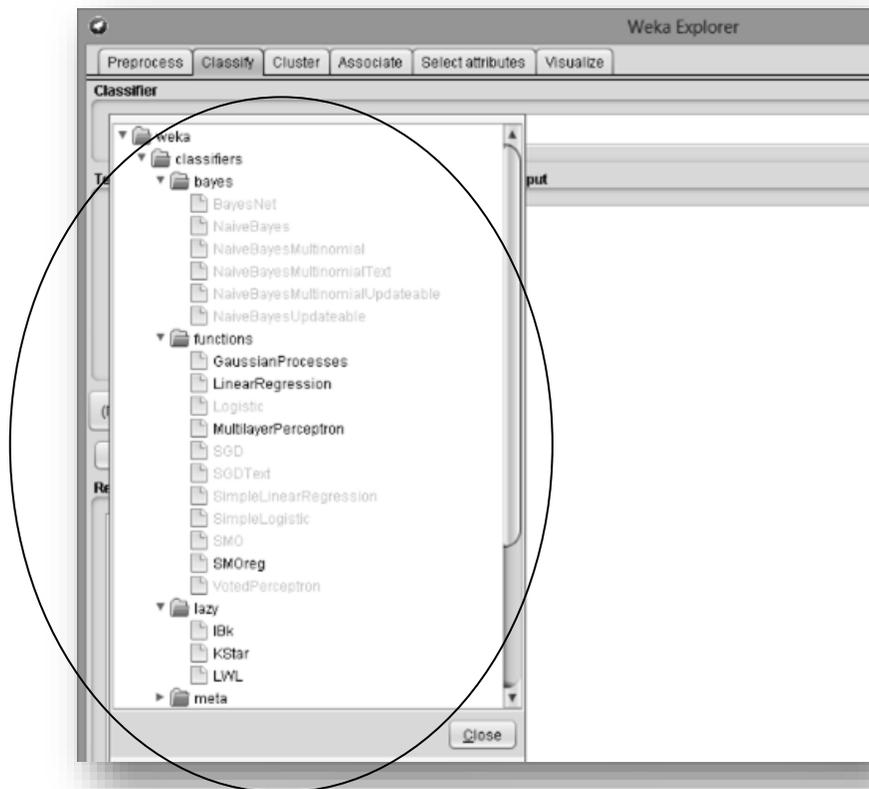
- Una vez que se tienen los atributos para la prueba se cambia de vista seleccionando la pestaña "Classify". Esta acción muestra la vista del configurador del modelo de clasificación, como se muestra en la figura A.7. En esta misma vista, con la opción "Choose" se selecciona el clasificador deseado de la lista emergente de algoritmos, como se observa en la figura A.8.

Figura A.7. Configuración de modelo de análisis.



Fuente: Imagen obtenida de la ejecución de Weka (Hall et al., 2009).

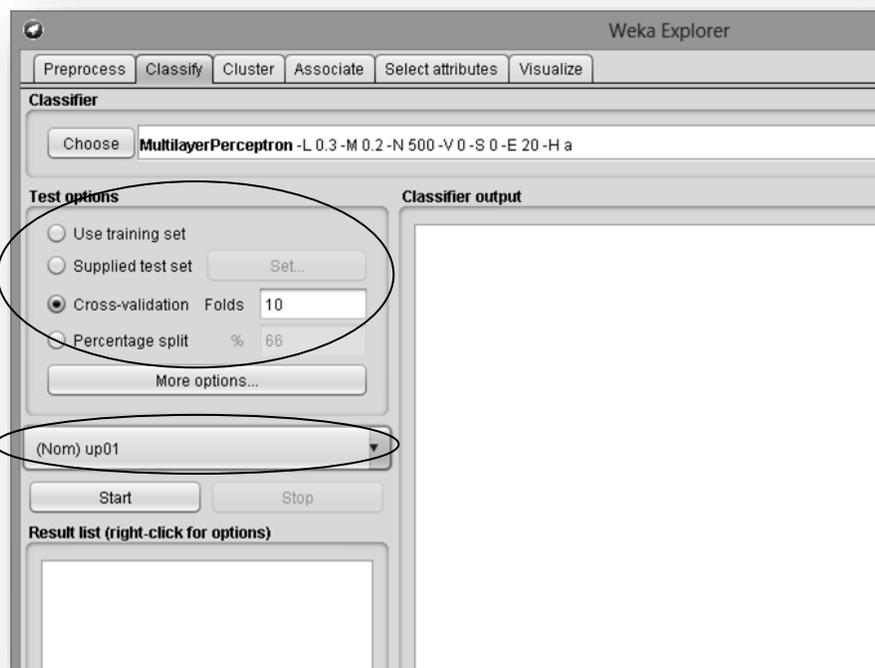
Figura A.8. Selección de clasificador.



Fuente: Imagen obtenida de la ejecución de Weka (Hall et al., 2009).

3. El tipo de evaluación del modelo se indica con las opciones del apartado “Test options”. Se selecciona la evaluación deseada y, posteriormente, en la lista desplegable, debajo de este grupo de opciones, se escoge la variable del listado que fungirá como la clase o variable dependiente a pronosticar. Como se observa en la figura A.9. Para el ejemplo de la figura A.9, la configuración muestra la selección del clasificador “MultilayerPerceptron” con su configuración por omisión, “Cross-validation” como método de evaluación del modelo y a la variable “up10” como la variable dependiente que hace referencia a la clase a pronosticar.

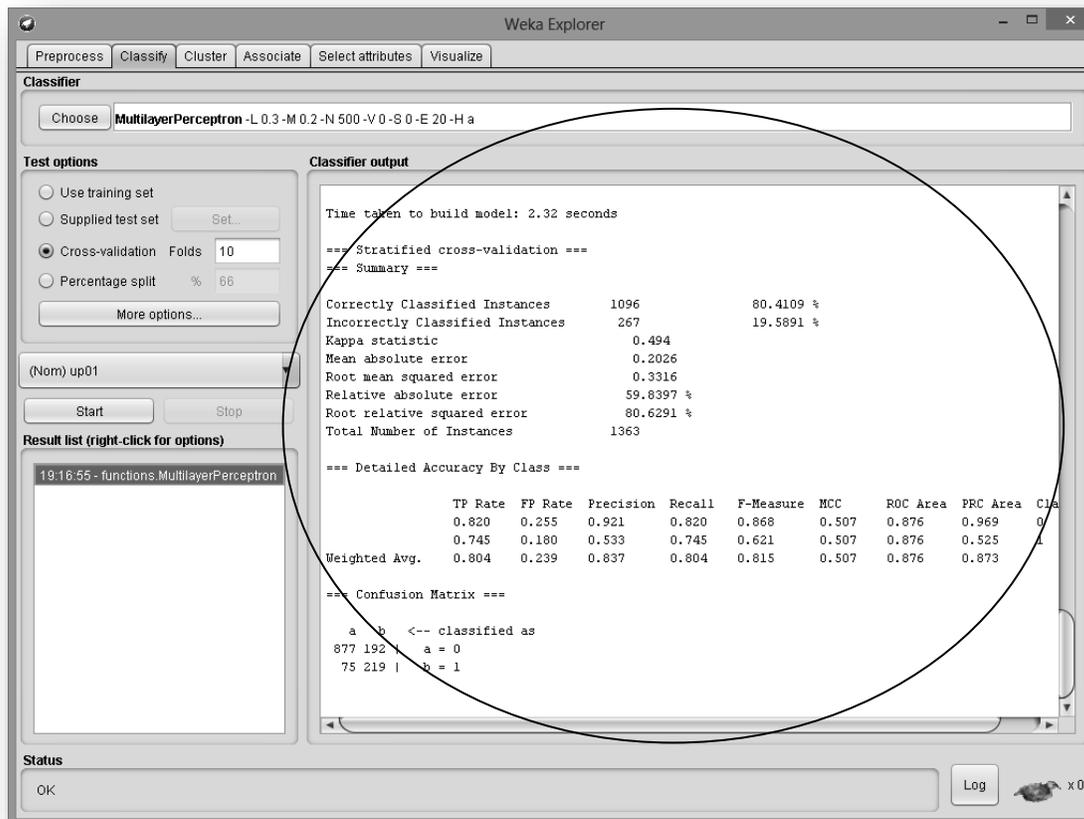
Figura A.9. Selección de método de evaluación de modelo.



Fuente: Imagen obtenida de la ejecución de Weka (Hall et al., 2009).

4. Para iniciar el análisis se selecciona la opción “Start” y después de unos momentos, que dependerá de la cantidad de variables y datos históricos del archivo cargado, se muestra el resultado como se observa en la figura A.10. Entre los datos relevantes están los porcentajes de exactitud de la clasificación, la matriz de confusión, algunos datos estadísticos y el cálculo de errores.

Figura A.10. Ejecución de análisis de un modelo.



Fuente: Imagen obtenida de la ejecución de Weka (Hall et al., 2009).

5. El procedimiento se repite para cada método de evaluación, para cada clasificador y para cada variable dependiente que se desee probar.

### ***Automatización del análisis***

El apartado anterior describe el proceso paso a paso o manual que se utiliza para verificar y ejecutar un único análisis. Sin embargo, una vez que se tienen los archivos *arff* de cada conjunto de prueba, es posible realizar todos los análisis de manera automatizada mediante la ejecución iterativa de los modelos con las librerías de Weka y el lenguaje de programación java. El siguiente extracto de código en lenguaje java pertenece al programa desarrollado para automatizar el proceso de filtrado de atributos de los conjuntos de pruebas y la evaluación de los clasificadores mencionados anteriormente:

## 1. Importar las librerías de Weka y se inicializan las variables de clase:

```

import weka.classifiers.Classifier;
import weka.classifiers.evaluation.Evaluation;
import weka.classifiers.functions.Logistic;
import weka.classifiers.functions.MultilayerPerceptron;
import weka.classifiers.functions.SMO;
import weka.classifiers.meta.LogitBoost;
import weka.core.Debug;
import weka.core.Instances;
import weka.core.converters.ConverterUtils.DataSource;
import weka.filters.Filter;
import weka.filters.unsupervised.attribute.Normalize;
import weka.filters.unsupervised.attribute.Remove;

/**
 * Inicialización de variables.
 */
static int[] attrs = {0,1,2,3,4,5,6,7,8,9};
static int[] schemeMethodsIndices = {1,2,3,6};
static int[] evalMethodsIndices = {1,2,3};
static String[] schemeMethodsIndicesLabels = {"Logistic", "ML", "SMO", "Logit", "SVM", "RF"};
static String[] evalMethodsIndicesLabels = {"FS", "CV", "SP", "OS"};
static int _NSCHEMES = 4;
static int _NEVALUATION_METHODS = 3;
static int _NLAGS = 10;
static boolean _FILTROENABLED = true;

```

## 2. Se colocan los archivos *arff* en un folder de preferencia y se declaran en el código como se muestra:

```

/**
 * Rutina principal para la ejecución.
 */
public static void main(String[] args) {
    String path = "C:\\Users\\usuario\\Desktop\\ejercicios\\csvs\\ipc\\";
    String file;
    ExecuteTest exec = new ExecuteTest();
    Map<String, String> results;
    for(int i=0; i<files.length; i++) {
        file = srcStock + files[i];
        System.out.print("\nAnalizando file=" + file);
        exec = new ExecuteTest();
        results = exec.fullTest(path + file);
        exec.printResults(file, results);
    }
}

```

## 3. Los métodos de código con la lógica de ejecución se detallan a continuación:

```

/**
 * Ejecuta la batería de pruebas completa.
 */
public Map<String, String> fullTest(String file) {
    DataSource source;
    Instances dataset, newDataset, baseDataset;
    Classifier scheme = null;
    Map<String, String> results = new HashMap<String, String>();
    Remove rm;
    SubsetByExpression sbe;
    String[] options;
    int[] valuesInClasses;
    SpreadSubsample sss;
    long minimumClassSize;
    RemoveDuplicates rd;
}

```

```

    try {
        source = new DataSource(file);
        dataset = source.getDataSet();
        for(int lag=0; lag<attrs.length; lag++) {
            rm = new Remove();
            rm.setAttributeIndicesArray(removeElementsFromArray(attrs, lag));
            rm.setInputFormat(dataset);
            baseDataset = Filter.useFilter(dataset, rm);
            baseDataset.setClassIndex(0); //el primer atributo: lag restante
            rd = new RemoveDuplicates();
            rd.setInputFormat(baseDataset);
            newDataset = Filter.useFilter(baseDataset, rd);
            sbe = new SubsetByExpression();
            options = new String[] {"-E", "(ATT2 is '1') or (ATT1 is '0' and ATT2 is '0')"};
            sbe.setOptions(options);
            sbe.setInputFormat(baseDataset);
            newDataset = Filter.useFilter(baseDataset, sbe);
            // selecciona el numero de elementos de la clase con menos elementos
            valuesInClasses = newDataset.attributeStats(0).nominalCounts;
            if (valuesInClasses[0]>valuesInClasses[1])
                minimumClassSize = valuesInClasses[1];
            else
                minimumClassSize = valuesInClasses[0];
            newDataset.randomize(new Random(10));
            sss = new SpreadSubsample();
            sss.setMaxCount(minimumClassSize);
            sss.setInputFormat(newDataset);
            newDataset = Filter.useFilter(newDataset, sss);
            if(!_FILTROENABLED)
                newDataset = baseDataset;
            System.out.print("\n\nclass=" + newDataset.classAttribute().name());
            String res, key;
            for (int i=0; i<_NSCHEMES; i++) {
                //scheme = getScheme(i+1);
                scheme = getScheme(schemeMethodsIndices[i]);
                for (int j=0; j<_NEVALUATION_METHODS; j++) {
                    res = "-";
                    try {
                        res = getEvaluation(evalMethodsIndices[j],
                            scheme, newDataset, baseDataset);
                    } catch (Exception e) {
                        e.printStackTrace();
                    }
                    key = buildKey(lag, i+1, j+1);
                    results.put(key, res);
                }
            }
        } catch (Exception e) {
            e.printStackTrace();
        }
        return results;
    }
}

/**
 * Construye el identificador con el que se ubica cada prueba en la estructura de datos.
 */
public String buildKey(int lag, int scheme, int method) {
    String k = "l" + lag + "_s" + scheme + "_m" + method;
    return k;
}

/**
 * Ejecuta una prueba/clasificación
 * Se le pasan los parámetros del tipo de prueba que se desea ejecutar de las 3 focalizadas.
 */
public String getEvaluation(int method, Classifier scheme, Instances newDataset, Instances
baseDataset) throws Exception {

```

```

String pje = null;
Evaluation eval = null;
Filter filter = null;
Instances tmp = null;
SubsetByExpression sbe;
String[] options;
int[] valuesInClasses;
SpreadSubsample sss;
long minimumClassSize;
RemoveDuplicates rd;
switch(method) {
    case 1:
        // Full set
        eval = new Evaluation(newDataset);
        scheme.buildClassifier(newDataset);
        eval.evaluateModel(scheme, newDataset);
        pje = String.format("%.2f", eval.pctCorrect());
    break;
    case 2:
        // Cross validation
        eval = new Evaluation(newDataset);
        scheme.buildClassifier(newDataset);
        eval.crossValidateModel(scheme, newDataset, 10, new Random(1));
        pje = String.format("%.2f", eval.pctCorrect());
    break;
    case 3:
        // Split 66%
        filter = new Normalize();
        tmp = new Instances(newDataset);
        // divide dataset to train dataset 80% and test dataset 20%
        int trainSize = (int) Math.round(tmp.numInstances() * 0.66);
        int testSize = tmp.numInstances() - trainSize;
        tmp.randomize(new Debug.Random(1));
        //Normalize dataset
        filter.setInputFormat(tmp);
        Instances datasetnor = Filter.useFilter(tmp, filter);
        Instances traindataset = new Instances(datasetnor, 0, trainSize);
        Instances testdataset = new Instances(datasetnor, trainSize,
        testSize);
        eval = new Evaluation(traindataset);
        scheme.buildClassifier(traindataset);
        eval.evaluateModel(scheme, testdataset);
        pje = String.format("%.2f", eval.pctCorrect());
    break;
    case 4:
        // Testing set
        filter = new Normalize();
        tmp = new Instances(baseDataset);
        tmp.randomize(new Debug.Random(1));
        //Normalize dataset
        filter.setInputFormat(tmp);
        tmp = Filter.useFilter(tmp, filter);
        sbe = new SubsetByExpression();
        options = new String[] {"-E", "(ATT2 is '1') or (ATT1 is '0' and
        ATT2 is '0')"};
        sbe.setOptions(options);
        sbe.setInputFormat(tmp);
        tmp = Filter.useFilter(tmp, sbe);
        rd = new RemoveDuplicates();
        rd.setInputFormat(tmp);
        tmp = Filter.useFilter(tmp, rd);
        // selecciona el numero de elementos de la clase con menos elementos
        valuesInClasses = tmp.attributeStats(0).nominalCounts;
        if (valuesInClasses[0]>valuesInClasses[1])
            minimumClassSize = valuesInClasses[1];
        else
            minimumClassSize = valuesInClasses[0];
        tmp.randomize(new Random(10));
        sss = new SpreadSubsample();

```

```

        sss.setMaxCount(minimumClassSize * 2);
        sss.setInputFormat(tmp);
        newDataset = Filter.useFilter(tmp, sss);
        eval = new Evaluation(newDataset);
        scheme.buildClassifier(newDataset);
        eval.evaluateModel(scheme, tmp);
        pje = String.format("%.2f", eval.pctCorrect());
    }
    return pje;
}

/**
 * Devuelve el esquema de clasificación a emplear en la prueba.
 * Configura y asigna las propiedades iniciales al clasificador deseado.
 */
public Classifier getScheme(int index) throws Exception {
    Classifier scheme = null;
    switch (index) {
        case 1:
            scheme = new Logistic();
            ((Logistic) scheme).setOptions(weka.core.Utils.splitOptions("-R
            1.0E-8 -M -1 -num-decimal-places 4"));
            break;
        case 2:
            scheme = new MultilayerPerceptron();
            ((MultilayerPerceptron) scheme)
            .setOptions(weka.core.Utils.splitOptions("-L 0.3 -M 0.2 -N 500 -V 0
            -S 0 -E 20 -H a"));
            break;
        case 3:
            scheme = new SMO();
            ((SMO) scheme).setOptions(weka.core.Utils.splitOptions(
            "-C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
            \\weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C
            250007\" -calibrator \\weka.classifiers.functions.Logistic -R 1.0E-8
            -M -1 -num-decimal-places 4"));
            break;
        case 4:
            scheme = new LogitBoost();
            ((LogitBoost) scheme).setOptions(weka.core.Utils.splitOptions(
            "-P 100 -L -1.7976931348623157E308 -H 1.0 -Z 3.0 -O 1 -E 1 -S 1 -I
            10 -W weka.classifiers.trees.DecisionStump"));
            break;
        case 5:
            scheme = new RandomForest();
            ((RandomForest) scheme).setOptions(weka.core.Utils.splitOptions(
            "-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1"));
            break;
        case 6:
            scheme = new LibSVM();
            ((LibSVM) scheme).setOptions(weka.core.Utils.splitOptions(
            "-S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1
            -model \\\"C:\\\\Program Files\\Weka-3-8-6\\ -seed 1"));
            break;
    }
    return scheme;
}

/**
 * Elimina del archivo arff los atributos dependientes dejando únicamente
 * el que se desea evaluar.
 */
public static int[] removeElementsFromArray(int[] attrs, int index) {
    int[] newArray = new int[attrs.length - 1];
    for (int i = 0, j = 0; i < attrs.length; i++) {
        if (i == index)
            continue;
    }
}

```

```

        newArray[j++] = attrs[i];
    }
    return newArray;
}

```

4. Al ejecutar el programa, los resultados son mostrados en pantalla como texto separado por comas para poderlos importar en Excel como tablas de datos. Esto es realizado por el siguiente método:

```

/**
 * Los resultados recolectados correspondientes a una tabla de pruebas son
 * depositados en la salida estándar con formato predefinido y en csv.
 */
public StringBuilder printResults(String file, Map<String, String> results) {
    String evalSt1 = evalMethodsIndicesLabels[evalMethodsIndices[0]-1];
    String evalSt2 = evalMethodsIndicesLabels[evalMethodsIndices[1]-1];
    String evalSt3 = evalMethodsIndicesLabels[evalMethodsIndices[2]-1];
    StringBuilder sb = new StringBuilder();
    sb.append(file + ",,,,,,,,,,\n");
    sb.append(", "+ schemeMethodsIndicesLabels[schemeMethodsIndices[0]-1] + ",,"
    + ", "+ schemeMethodsIndicesLabels[schemeMethodsIndices[1]-1] + ",,"
    + ", "+ schemeMethodsIndicesLabels[schemeMethodsIndices[2]-1] + ",,"
    + ", "+ schemeMethodsIndicesLabels[schemeMethodsIndices[3]-1] + "\n");
    sb.append("LAG," + evalSt1 + "," + evalSt2 + "," + evalSt3 + ","
    + evalSt1 + "," + evalSt2 + "," + evalSt3 + ","
    + evalSt1 + "," + evalSt2 + "," + evalSt3 + ","
    + evalSt1 + "," + evalSt2 + "," + evalSt3 + "\n");
    for(int i=0; i<_NLAGS; i++) {
        sb.append((i+1));
        for(int j=1; j<=_NSCHEMES; j++) {
            for(int k=1; k<=_NEVALUATION_METHODS; k++) {
                sb.append(", " + results.get("1"+ i + "_s" + j + "_m" + k));
            }
            sb.append("");
        }
        sb.append("\n");
    }
    System.out.print("\n");
    System.out.print(sb);
    return sb;
}

```

5. La salida en la consola de ejecución da como resultado una matriz en texto con formato csv con todos los análisis de una misma serie, como se observa en la figura A.11. En la figura A.12 se muestra un ejemplo del resultado de la anomalía de momentum en el IPC de la serie mensual con tendencia positiva del archivo csv visualizado en Excel.

Figura A.11. Salida de la ejecución de una batería de pruebas automatizadas en lenguaje java.

```

Analizando file=cemexcpo_tfMens_sMomentum_avg10_m05_ma00_pL_vSeries.arff
class=up01
class=up02
class=up03
class=up04
class=up05
class=up06
class=up07
class=up08
class=up09
class=up10
cemexcpo_tfMens_sMomentum_avg10_m05_ma00_pL_vSeries.arff,,,,,,,,,
,LOGISTIC,,,ML,,,SMO,,,LOGIT
LAG,FS,CV,SP,FS,CV,SP,FS,CV,SP,FS,CV,SP
1,100.00,80.95,80.95,100.00,84.13,71.43,84.13,84.13,76.19,98.41,73.02,66.67
2,96.83,82.54,85.71,98.41,80.95,76.19,82.54,82.54,85.71,100.00,71.43,85.71
3,95.24,76.19,85.71,100.00,77.78,85.71,85.71,80.95,85.71,100.00,77.78,85.71
4,96.83,80.95,80.95,96.83,82.54,80.95,85.71,76.19,80.95,98.41,76.19,80.95
5,90.48,77.78,76.19,96.83,74.60,85.71,84.13,74.60,76.19,100.00,76.19,66.67
6,90.48,77.78,76.19,96.83,74.60,85.71,84.13,74.60,76.19,100.00,76.19,66.67
7,90.48,77.78,76.19,96.83,74.60,85.71,84.13,74.60,76.19,100.00,76.19,66.67
8,88.89,79.37,61.90,96.83,79.37,71.43,85.71,77.78,80.95,100.00,79.37,76.19
9,88.89,79.37,61.90,96.83,79.37,71.43,85.71,77.78,80.95,100.00,79.37,76.19
10,87.30,76.19,76.19,95.24,74.60,80.95,87.30,87.30,85.71,98.41,77.78,85.71
    
```

Fuente: Elaboración propia.

Figura A.12. Resultado de tabla en Excel de importar una batería de prueba en formato csv.

cveIPC_tfMens_sMomentum_avg10_m05_ma00_pL_vSeries.arff												
	LOGISTIC			ML			SMO			LOGIT		
LAG	FS_LT	CV_LT	SP_LT	FS_ML	CV_ML	SP_ML	FS_SO	CV_SO	SP_SO	FS_LG	CV_LG	SP_LG
1	86.00	72.00	64.71	96.00	78.00	58.82	82.00	82.00	76.47	100.00	68.00	76.47
2	84.00	70.00	52.94	98.00	74.00	47.06	84.00	84.00	88.24	100.00	68.00	64.71
3	92.00	74.00	70.59	96.00	78.00	64.71	84.00	84.00	82.35	100.00	74.00	64.71
4	90.00	76.00	70.59	100.00	70.00	82.35	84.00	84.00	88.24	98.00	76.00	64.71
5	90.00	78.00	76.47	100.00	62.00	70.59	84.00	84.00	88.24	100.00	68.00	70.59
6	100.00	82.00	88.24	98.00	70.00	88.24	84.00	84.00	82.35	98.00	68.00	76.47
7	96.00	76.00	70.59	98.00	66.00	70.59	80.00	80.00	76.47	98.00	74.00	70.59
8	90.00	80.00	52.94	98.00	74.00	52.94	80.00	62.00	64.71	100.00	72.00	64.71
9	90.00	68.00	58.82	98.00	68.00	64.71	80.00	70.00	64.71	96.00	68.00	64.71
10	92.00	74.00	70.59	98.00	68.00	70.59	82.00	78.00	82.35	98.00	68.00	82.35

Fuente: Elaboración propia.

Donde:

- *LAG* es el número de periodo del rendimiento que se desea estimar una vez aparecida la señal.
- *LOGISTIC*, *ML*, *SMO* y *LOGIT* son los clasificadores utilizados: Regresión logística, Perceptrón multicapa, Máquina de soporte vectorial y Método Logit, respectivamente.

- *FS*, *CV* y *SP* son las claves para indicar el método de evaluación del modelo: *FS* de *full set* del inglés o conjunto completo, *CV* para *cross-validation* o validación cruzada y *SP* para *split* o método de retención que consiste en partir en tercios el conjunto fundamental, de los cuales dos tercios son usados para entrenamiento y el restante para prueba.

6. La ejecución complementaria de la lectura de archivos de texto con las matrices de registros para procesarlas en Weka, es realizada con los extractos de las rutinas siguientes:

```

/**
 * Rutina principal de ejecución del programa.
 */
public static void main(String[] args) {
    executeFromScratchTest();
}

/**
 * Ejecuta un análisis completo a partir de archivos de texto con las matrices construidas en
 * archivos de texto
 */
public static void executeFromScratchTest() {
    List<IntegrationFileBean> preparedStructures, integratedStructures, arffStructures;
    preparedStructures = prepareStructures("HIGHS", _positionType, _timeFrame,
    _anomalyType, _stocksHigh);
    integratedStructures = integrateFiles(basePath, sourcePath, integrationFolder,
    preparedStructures);
    arffStructures = createArffFiles(basePath, integrationFolder, arffFolder,
    templateFolder, templateFile, integratedStructures);
    analyzeConWeka(basePath, resultsFolder, arffStructures);
    preparedStructures = prepareStructures("LOWS", _positionType, _timeFrame,
    _anomalyType, _stocksLow);
    integratedStructures = integrateFiles(basePath, sourcePath, integrationFolder,
    preparedStructures);
    arffStructures = createArffFiles(basePath, integrationFolder, arffFolder,
    templateFolder, templateFile, integratedStructures);
    analyzeConWeka(basePath, resultsFolder, arffStructures);
}

/**
 * Ejecuta el análisis con Weka
 */
public static void analyzeConWeka(String basePath,
    String resultFolder,
    List<IntegrationFileBean> arffFilesToAnalyze) {
    String file;
    ExecuteTest exec = new ExecuteTest();
    Map<String, String> results;
    StringBuilder sb;
    for(IntegrationFileBean bean: arffFilesToAnalyze) {
        file = bean.getFullFile();
        if(!file.endsWith(".arff"))
            continue;
        System.out.print("\nAnalizando file=" + file);
        exec = new ExecuteTest();
        results = exec.fullTest(file);
        sb = exec.printResults(file, results);
        createFileFromText(basePath+resultFolder+bean.getRawFileName()+".csv", sb);
    }
}

```

```

    }
}

/**
 * Obtiene las estructuras de archivos y sus conjuntos para integrarlos.
 * Clasifica los nombres de archivos por tock y los integra según anomaly, frame y position.
 * Solo agrega nombres sin extension y sin path
 */
public static List<IntegrationFileBean> prepareStructures(
    String prefix,
    String[] positionType,
    String[] timeFrame,
    String[] anomalyType,
    String[] stocks) {
    String strFileReading, strFileWriting;
    List<String> filesToIntegrate;
    List<IntegrationFileBean> filesToProcess = new ArrayList<IntegrationFileBean>();
    IntegrationFileBean bean;
    for(String pT: positionType) {
        for(String tF: timeFrame) {
            for(String aT: anomalyType) {
                strFileWriting = prefix + "-" + tF + "_" + aT + "-" + pT;
                System.out.println("prepareIntegration: Creando: " +
                    strFileWriting);
                filesToIntegrate = new ArrayList<String>();
                for(String sT: stocks) {
                    strFileReading = sT + "-" + tF + "_" + aT + "-" +
                        pT;
                    filesToIntegrate.add(strFileReading);
                    System.out.println("prepareIntegration: Agregando: "
                        + strFileReading);
                }
                bean = new IntegrationFileBean(strFileWriting, null,
                    filesToIntegrate);
                filesToProcess.add(bean);
                System.out.println("prepareIntegration: Archivos integrados: "
                    + filesToIntegrate.size());
            }
        }
    }
    return filesToProcess;
}

/**
 * Prepara las estructuras de archivos y sus conjuntos para integrarlos.
 * Clasifica los nombres de archivos por tock y los integra según anomaly, frame y position.
 */
public static IntegrationFileBean prepareIntegration(
    String basePath,
    String sourcePath,
    String resultFolder,
    IntegrationFileBean rawEstructure) {
    String strFileReading, strFileWriting;
    List<String> filesFormattedToIntegrate = new ArrayList<String>();
    IntegrationFileBean preparedEstructure;
    strFileWriting = basePath + resultFolder + rawEstructure.getRawFileName() + ".csv";
    System.out.println("prepareIntegration: Creando: " + strFileWriting);
    for(String sourceFile: rawEstructure.getFiles()) {
        strFileReading = basePath + sourcePath + sourceFile + ".csv";
        filesFormattedToIntegrate.add(strFileReading);
        System.out.println("prepareIntegration: Agregando: " + strFileReading);
    }
    preparedEstructure = new IntegrationFileBean(rawEstructure.getRawFileName(),
        strFileWriting, filesFormattedToIntegrate);
    System.out.println("prepareIntegration: Archivos integrados: " +
        filesFormattedToIntegrate.size());
    return preparedEstructure;
}

```

```

/**
 * Lee cada archivo de stock y los integra según anomaly, frame y position.
 */
public static List<IntegrationFileBean> integrateFiles(String basePath,
String sourcePath,
String resultFolder,
List<IntegrationFileBean> rawMainFilesToIntegrate) {
    int nFilesCreated = 0;
    List<String> result;
    IntegrationFileBean preparedStructure;
    List<IntegrationFileBean> integratedStructures = new
    ArrayList<IntegrationFileBean>();
    for(IntegrationFileBean rawEstructure: rawMainFilesToIntegrate) {
        preparedStructure = prepareIntegration(basePath, sourcePath, resultFolder,
        rawEstructure);
        System.out.println("integrateFiles: Agregando: " +
        rawEstructure.getRawFileName());
        result = createFileFromSetOfFilees(preparedStructure.getFullFile(),
        preparedStructure.getFiles());
        System.out.println("integrateFiles: Archivos integrados: " + result.size());
        integratedStructures.add(new
        IntegrationFileBean(preparedStructure.getRawFileName(),
        preparedStructure.getFullFile(), null));
    }
    return integratedStructures;
}

/**
 * Integra el template de weka.
 * @param integratedFiles nombre de archivos integrados sin extension ni ruta
 */
public static List<IntegrationFileBean> createArfffFiles(String basePath,
String integratedFolder,
String arfffFolder,
String templateFolder,
String templateFile,
List<IntegrationFileBean> integratedEstructures) {
    String strFileReading, strFileWriting;
    List<IntegrationFileBean> result = new ArrayList<IntegrationFileBean>();
    List<String> arfffAndCsvFilesToIntegrate;
    for(IntegrationFileBean preparedStructure: integratedEstructures) {
        strFileWriting = basePath + arfffFolder + preparedStructure.getRawFileName()
        + ".arfff";
        System.out.println("Procesando: " + strFileWriting);
        arfffAndCsvFilesToIntegrate = new ArrayList<String>();
        strFileReading = basePath + templateFolder + templateFile;
        arfffAndCsvFilesToIntegrate.add(strFileReading);
        strFileReading = preparedStructure.getFullFile();
        arfffAndCsvFilesToIntegrate.add(strFileReading);
        createFileFromSetOfFilees(strFileWriting, arfffAndCsvFilesToIntegrate);
        result.add(new IntegrationFileBean(preparedStructure.getRawFileName(),
        strFileWriting, arfffAndCsvFilesToIntegrate));
    }
    System.out.println("Archivos arfff creados: " + result.size());
    return result;
}

/**
 * Crea el archivo dado y le integra la lista de archivos del argumento.
 */
public static List<String> createFileFromText(String newFile,
StringBuilder text) {
    String strFileWriting, err;
    File fWriting;
    FileWriter fw = null;
    BufferedWriter bw = null;
    List<String> result = new ArrayList<String>();
    strFileWriting = newFile;
    System.out.println("Creando: " + strFileWriting);

```

```
fWritting = new File(strFileWritting);
if(fWritting.exists()) {
    System.out.println("El archivo ya existe! Favor de borrarlo: " +
        strFileWritting);
    return result;
}
try {
    fWritting.createNewFile();
} catch (IOException e1) {
    e1.printStackTrace();
    System.out.println("Problema al crear el archivo destino.");
    return result;
}
try {
    fw = new FileWriter(fWritting);
    bw = new BufferedWriter(fw);
} catch (IOException e1) {
    e1.printStackTrace();
    System.out.println("Problema al abrir el archivo destino.");
    return result;
}
err = appendFile(text, bw);
System.out.println("Resultado Add integrated file: " + err);
if(!err.startsWith("R00")) {
    System.out.println("Error al agregar archivo...");
    System.out.println("Cancelando creación de archivo: " + strFileWritting);
    try {
        bw.close();
        fw.close();
    } catch (IOException e) {
        e.printStackTrace();
        System.out.println("Problema al cerrar el buffer destino.");
    }
    return result;
}
try {
    bw.close();
    fw.close();
} catch (IOException e) {
    e.printStackTrace();
    System.out.println("Problema al cerrar el buffer destino.");
}
System.out.println("Cración exitosa.");
return result;
}
```

## **Anexo B**

### **RESULTADOS EXTENDIDOS**

En este apartado se presentan los resultados completos de los análisis revisados y mencionados en las secciones anteriores. Se muestran las tablas de los porcentajes de los modelos de clasificación de cada uno de los métodos a contrastar, gráficas comparativas de los periodos rezagados, las gráficas de rendimientos lineales de las anomalías y gráficas descriptivas de las señales. Esto para ayudar a ejemplificar la procedencia de los datos y su discusión.

Los porcentajes obtenidos representan el grado de ajuste de las variables independientes, incluyendo la de la anomalía, respecto a los rendimientos ajustados al riesgo. Los números consecutivos en las gráficas de los porcentajes de ajuste representan el periodo rezagado que se está probando, es decir, el número de periodos que pudiera anticiparse el pronóstico con la señal de la anomalía.

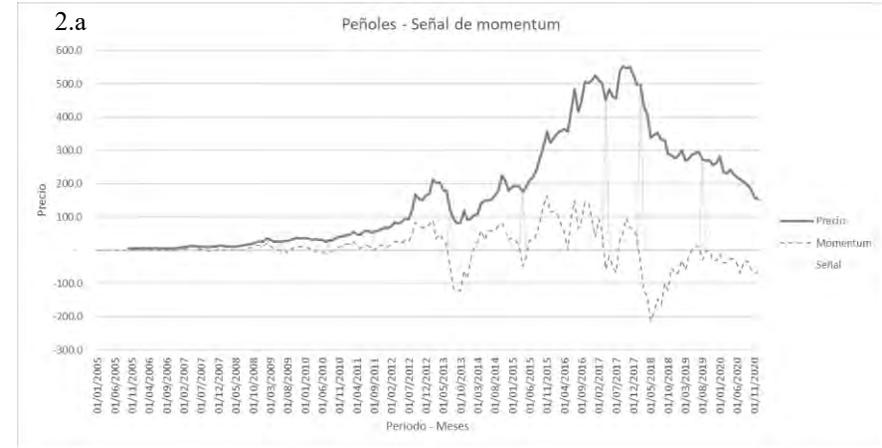
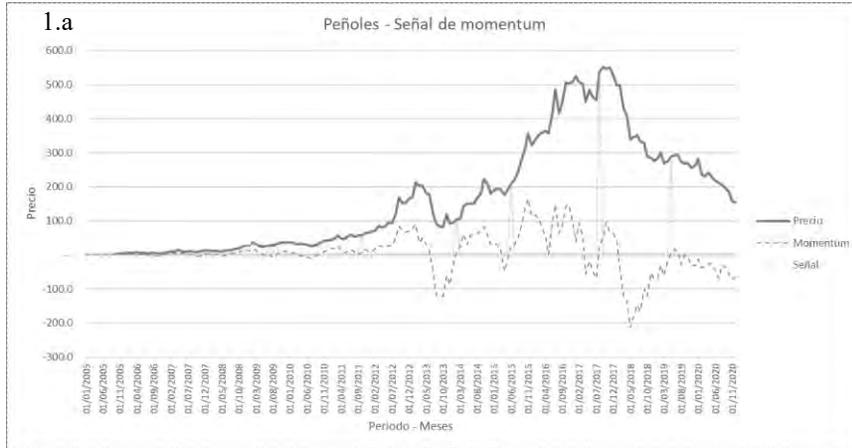
Algunos elementos importantes, que se manejan con siglas en las gráficas, y que no se indica su significado en extenso son:

- CC – Conjunto completo. Indica que la evaluación del modelo se realizó con el conjunto entero de datos.
- VC – Validación cruzada. Hace referencia al método de validación cruzada de 10 particiones.
- CP – Conjunto partido. Significa que se usaron dos tercios del conjunto para entrenar y el tercio restante para probar. Se le conoce como método de retención.

Figura B.1. Ejemplo de las señales en la serie mensual de Peñoles: *momentum* y reversión a la media.

1) Señales en tendencia positiva

2) Señales en tendencia negativa

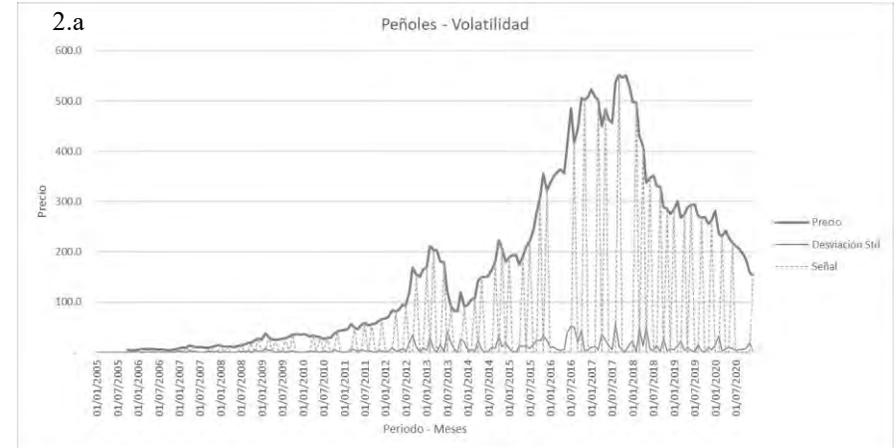
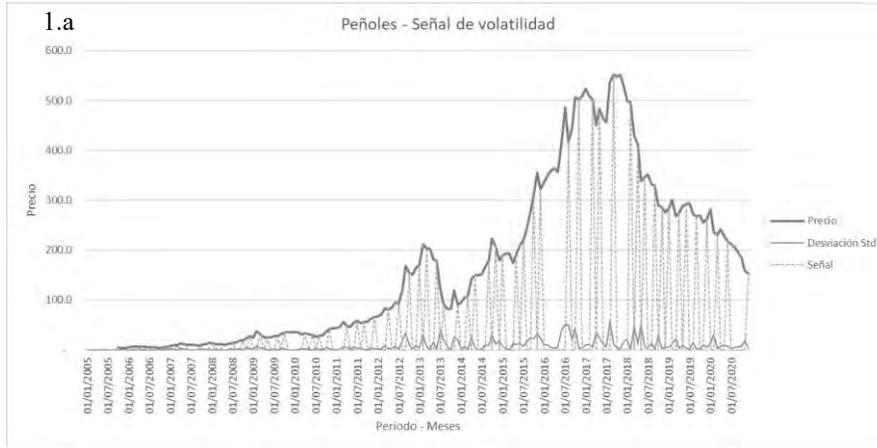


Fuente: Elaboración propia.

Figura B.2. Ejemplo de las señales en la serie mensual de Peñoles: volatilidad y efecto de calendario.

1) Señales en tendencia positiva

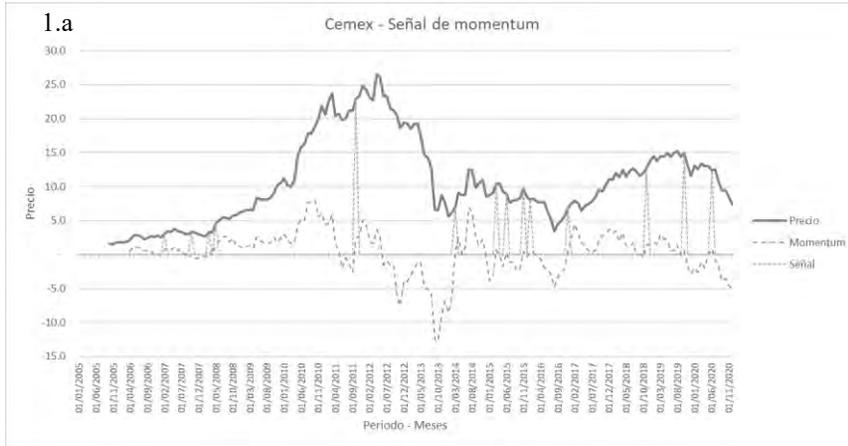
2) Señales en tendencia negativa



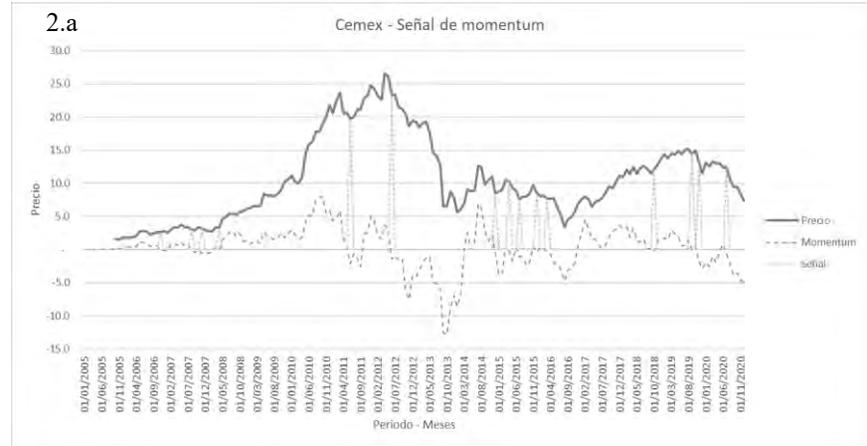
Fuente: Elaboración propia.

Figura B.3. Ejemplo de las señales en la serie mensual de Cemex: *momentum* y reversión a la media.

1) Señales en tendencia positiva



2) Señales en tendencia negativa

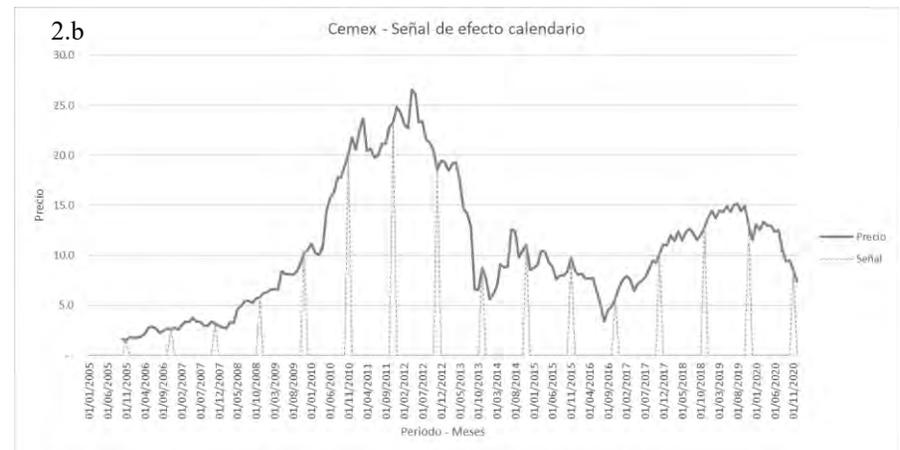
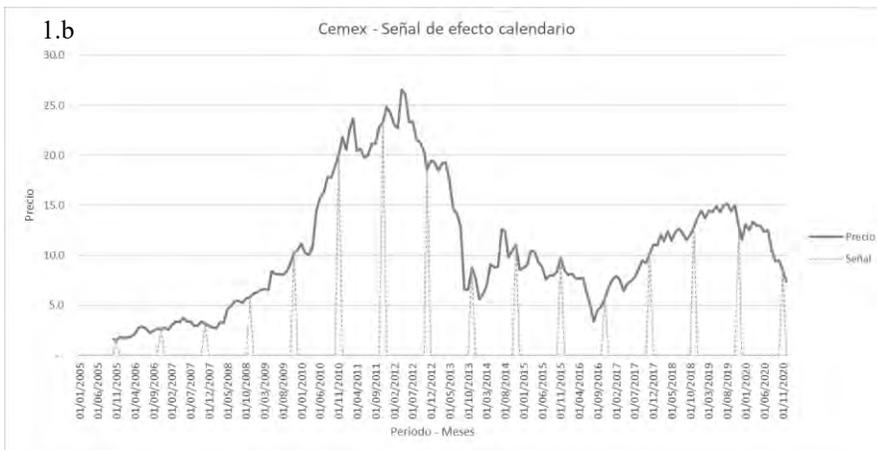
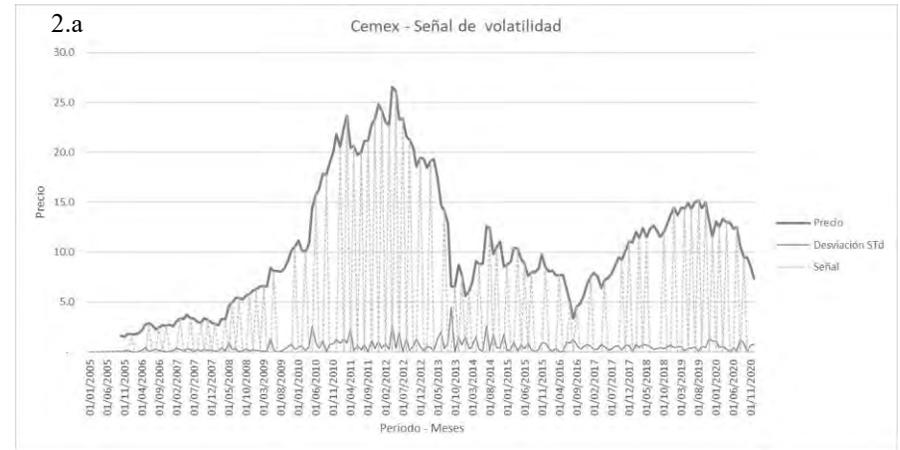
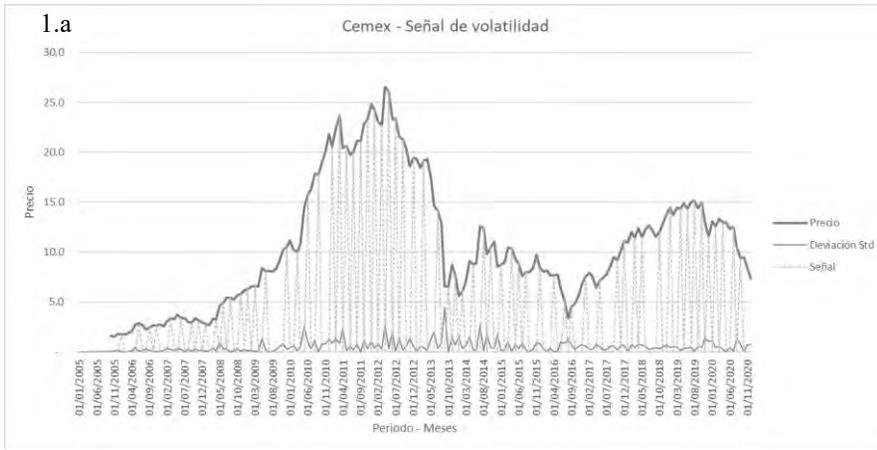


Fuente: Elaboración propia.

Figura B.4. Ejemplo de las señales en la serie mensual de Cemex: volatilidad y efecto de calendario.

1) Señales en tendencia positiva

2) Señales en tendencia negativa



Fuente: Elaboración propia.

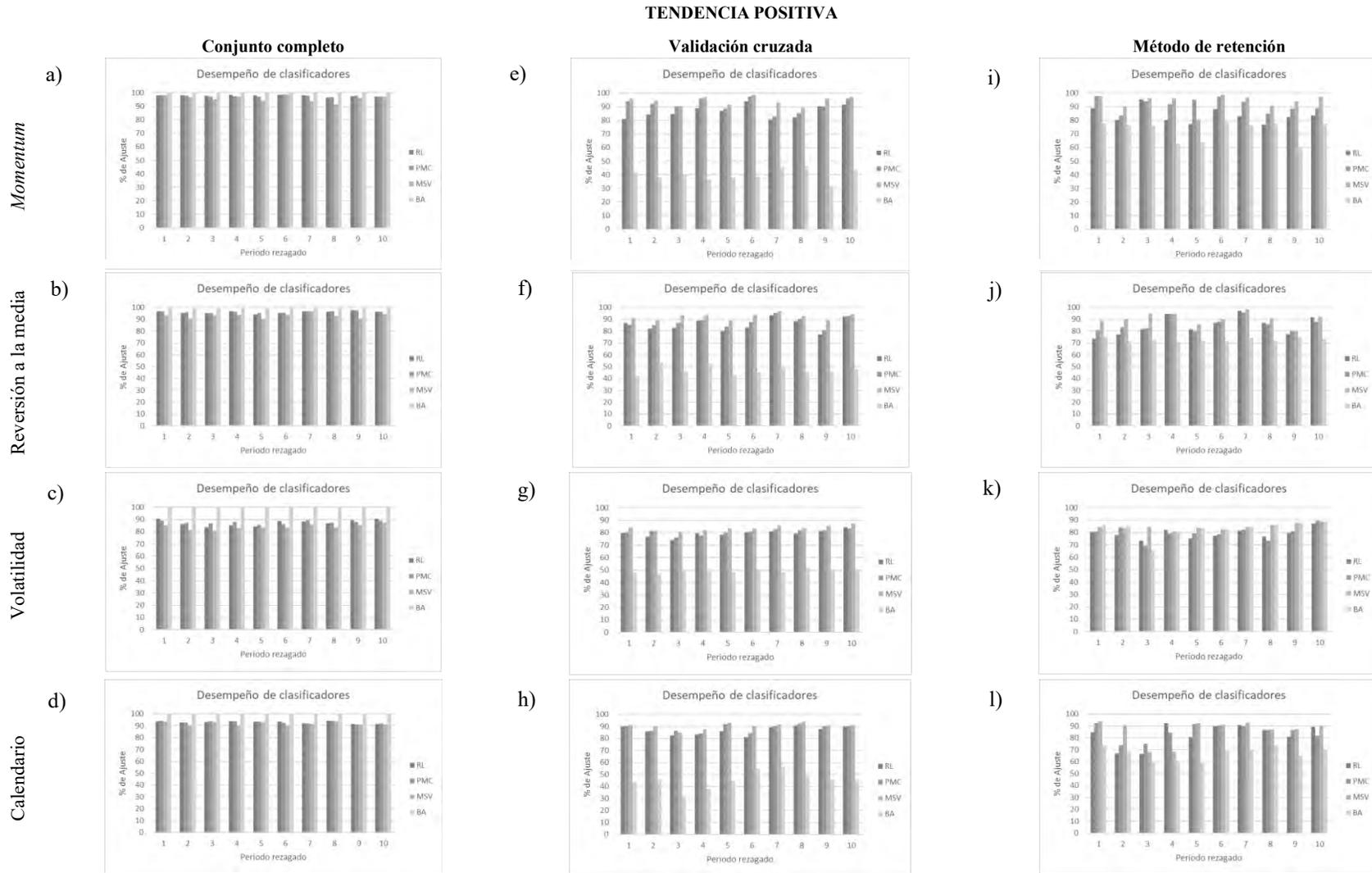
Tabla B.1. Porcentajes de ajuste de los modelos para Peñoles en tendencia positiva: resultados de la clasificación por anomalía.

	Series mensuales												Series semanales												Series diarias																																																																																																																																																																																																																																																																																																																																																																																
	Regresión Logística			Perceptrón multicapa			M. de soporte vectorial			Bosques aleatorios			Regresión Logística			Perceptrón multicapa			M. de soporte vectorial			Bosques aleatorios			Regresión Logística			Perceptrón multicapa			M. de soporte vectorial			Bosques aleatorios																																																																																																																																																																																																																																																																																																																																																																							
	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP																																																																																																																																																																																																																																																																																																																																																												
Momentum	a)	1	100.0	64.3	80.0	100.0	92.9	100.0	100.0	92.9	100.0	100.0	28.6	40.0	1	100.0	85.7	94.7	100.0	96.4	100.0	100.0	100.0	100.0	100.0	42.9	100.0	1	94.2	92.7	92.0	94.2	93.3	93.8	94.2	94.2	93.8	100.0	53.3	93.8	2	100.0	78.6	60.0	100.0	92.9	60.0	100.0	92.9	80.0	100.0	28.6	40.0	2	100.0	81.0	85.0	98.3	91.4	95.0	96.6	96.6	95.0	100.0	37.9	95.0	2	94.3	92.7	95.3	94.9	92.0	95.3	94.3	94.3	95.3	100.0	48.1	95.3	3	100.0	85.7	100.0	100.0	85.7	100.0	100.0	85.7	100.0	100.0	28.6	40.0	3	100.0	75.9	94.4	98.2	92.6	88.9	92.6	92.6	94.4	100.0	40.7	94.4	3	93.6	92.0	92.5	93.6	92.0	92.5	93.3	93.3	94.4	100.0	51.3	94.4	4	100.0	91.7	75.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	16.7	0.0	4	100.0	80.4	73.7	96.4	92.9	84.2	96.4	92.6	94.7	100.0	42.9	94.7	4	95.3	94.7	91.7	95.7	94.7	91.7	95.3	95.3	93.6	100.0	50.0	93.6	5	100.0	83.3	50.0	100.0	83.3	100.0	91.7	83.3	50.0	100.0	16.7	0.0	5	100.0	84.9	90.9	97.0	89.4	95.5	97.0	97.0	100.0	100.0	48.5	100.0	5	94.2	92.9	90.5	94.2	93.2	90.5	93.9	93.9	91.4	100.0	48.4	91.4	6	100.0	92.9	80.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	21.4	40.0	6	100.0	95.2	90.5	100.0	98.4	100.0	100.0	100.0	100.0	100.0	45.2	100.0	6	95.8	93.3	93.8	96.1	94.6	92.9	96.1	96.1	96.4	100.0	49.1	96.4	7	100.0	71.4	80.0	100.0	71.4	100.0	92.9	92.9	100.0	100.0	42.9	40.0	7	100.0	76.6	77.3	98.4	82.8	86.4	93.8	92.2	95.5	100.0	46.9	95.5	7	94.8	93.3	91.0	94.8	94.8	94.6	94.8	94.8	94.6	100.0	47.2	94.6	8	100.0	71.4	80.0	100.0	71.4	80.0	85.7	78.6	80.0	100.0	35.7	40.0	8	95.0	81.7	60.0	96.7	91.7	85.0	95.0	95.0	100.0	100.0	56.7	100.0	8	94.8	93.3	91.0	94.2	92.6	90.1	94.2	94.2	91.9	100.0	46.6	91.9	9	100.0	91.7	100.0	100.0	91.7	100.0	100.0	100.0	100.0	100.0	0.0	0.0	9	94.6	82.1	52.6	96.4	82.1	68.4	91.1	91.1	84.2	100.0	53.6	84.2	9	97.4	96.3	94.2	97.4	97.2	97.5	97.4	97.4	97.5	100.0	42.9	97.5	10	100.0	92.9	80.0	100.0	100.0	80.0	100.0	100.0	100.0	100.0	28.6	40.0	10	100.0	93.1	85.0	100.0	98.3	100.0	100.0	100.0	100.0	100.0	55.2	100.0	10	91.3	88.4	85.7	91.9	89.0	86.7	91.3	91.3	91.4	100.0	48.4	91.4		
	Reversión a la media	b)	1	100.0	81.3	40.0	100.0	75.0	60.0	93.8	87.5	80.0	100.0	37.5	40.0	1	96.7	87.8	90.3	96.7	88.9	90.3	93.3	93.3	93.6	100.0	40.0	93.6	1	93.2	91.4	91.1	93.5	91.8	91.8	93.0	93.0	92.5	100.0	49.3	92.5	2	100.0	70.0	57.1	100.0	75.0	71.4	90.0	85.0	85.7	100.0	60.0	28.6	2	92.5	85.0	81.5	95.0	88.8	85.2	90.0	90.0	92.6	100.0	51.3	92.6	2	92.9	91.4	92.1	92.9	91.2	92.8	92.7	92.7	92.8	100.0	49.3	92.8	3	100.0	69.2	88.9	100.0	80.8	88.9	96.2	96.2	100.0	100.0	38.5	33.3	3	94.1	89.3	69.0	94.1	89.3	69.0	92.9	92.9	93.1	100.0	51.2	93.1	3	91.3	89.8	86.4	91.3	90.5	89.3	91.0	91.0	91.4	100.0	48.3	91.4	4	100.0	86.4	100.0	100.0	86.4	100.0	95.5	95.5	100.0	100.0	50.0	28.6	4	95.0	87.0	91.2	94.0	88.0	91.2	91.0	91.0	91.2	100.0	57.0	91.2	4	95.1	93.3	92.4	95.1	93.6	92.4	94.6	94.6	93.2	100.0	51.3	93.2	5	100.0	68.2	71.4	100.0	77.3	57.1	90.9	86.4	71.4	100.0	31.8	28.6	5	87.8	78.9	77.4	91.1	80.0	87.1	85.6	85.6	90.3	100.0	47.8	90.3	5	95.1	93.4	96.4	95.1	94.4	96.4	94.4	94.4	95.7	100.0	49.3	95.7	6	100.0	72.7	85.7	100.0	86.4	85.7	95.5	95.5	85.7	100.0	36.4	28.6	6	92.4	85.9	80.7	93.5	84.8	83.9	92.4	92.4	90.3	100.0	48.9	90.3	6	93.2	90.7	94.1	94.0	92.5	94.1	94.2	93.2	94.8	100.0	50.5	94.8	7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	50.0	28.6	7	94.3	85.2	100.0	94.3	90.9	93.3	94.3	94.3	100.0	100.0	46.6	100.0	7	96.6	94.3	92.0	96.1	94.6	93.5	95.6	95.6	94.9	100.0	52.0	94.9	8	100.0	86.4	85.7	100.0	86.4	71.4	90.9	90.9	85.7	100.0	36.4	28.6	8	96.6	87.5	83.3	96.6	92.1	93.3	94.3	94.3	93.3	100.0	48.9	93.3	8	93.1	91.0	91.6	93.4	92.2	92.3	93.1	93.1	93.7	100.0	51.9	93.7	9	100.0	50.0	42.9	100.0	55.0	42.9	80.0	75.0	42.9	100.0	40.0	28.6	9	96.5	86.1	93.1	96.5	91.9	100.0	96.5	96.5	100.0	100.0	45.4	100.0	9	95.8	94.9	96.6	95.8	95.6	97.3	95.6	95.6	97.3	100.0	52.8	97.3	10	100.0	100.0	100.0	100.0	95.0	85.7	95.0	95.0	85.7	100.0	45.0	28.6	10	94.6	83.7	83.9	94.6	89.1	83.9	93.5	93.5	96.8	100.0	48.9	96.8	10	94.9	93.0	92.1	94.9	93.7	94.3	94.7	94.7	94.3	100.0	50.0	94.3	
		Volatilidad	c)	1	100.0	71.2	77.8	94.2	73.1	77.8	84.6	82.7	83.3	100.0	40.4	88.9	1	87.1	83.8	78.1	87.9	83.8	80.5	85.4	85.4	84.2	100.0	50.4	84.2	1	85.1	84.4	84.9	85.1	84.6	84.4	85.0	85.0	85.7	100.0	52.5	85.7	2	87.5	64.6	68.8	91.7	77.1	81.3	77.1	77.1	81.3	100.0	41.7	87.5	2	87.4	82.9	80.0	86.9	84.7	86.7	84.7	84.7	84.0	100.0	48.2	84.0	2	83.3	82.6	84.8	83.4	82.7	84.3	82.9	82.9	84.5	100.0	50.9	84.5	3	82.6	58.7	56.3	91.3	65.2	43.8	76.1	76.1	87.5	100.0	47.8	31.3	3	84.9	81.1	84.0	86.1	80.7	84.0	83.2	83.2	85.2	100.0	50.4	85.2	3	83.3	82.6	80.4	83.5	82.8	80.2	82.9	82.9	80.4	100.0	52.5	80.4	4	88.0	74.0	82.4	96.0	70.0	76.5	80.0	78.0	76.5	100.0	42.0	76.5	4	83.8	81.3	79.3	84.2	80.4	78.1	83.8	83.8	81.7	100.0	52.1	81.7	4	84.0	83.5	84.5	84.4	83.3	83.4	83.5	83.5	84.5	100.0	52.4	84.5	5	82.0	68.0	58.8	86.0	72.0	70.6	80.0	80.0	82.4	100.0	44.0	82.4	5	86.8	83.3	80.0	87.2	84.6	81.3	86.3	85.9	82.5	100.0	48.3	82.5	5	84.2	83.5	87.0	84.5	84.0	85.4	84.0	84.0	86.5	100.0	52.5	86.5	6	96.3	74.1	72.2	88.9	75.9	77.8	81.5	81.5	83.3	100.0	48.2	83.3	6	86.4	84.3	76.8	86.8	84.7	75.6	85.5	85.5	81.7	100.0	50.8	81.7	6	83.6	83.0	83.3	83.3	82.7	82.5	83.0	83.0	82.7	100.0	53.0	82.7	7	94.0	74.0	76.5	94.0	80.0	76.5	86.0	86.0	82.4	100.0	40.0	82.4	7	88.4	86.6	86.8	90.6	85.7	88.2	88.0	88.0	88.2	100.0	51.8	88.2	7	83.4	82.9	81.4	83.5	83.0	81.9	83.3	83.3	82.1	100.0	53.4	82.1	8	95.7	76.1	68.8	95.7	84.8	62.5	87.0	87.0	93.8	100.0	50.0	93.8	8	81.9	78.6	77.8	82.4	77.7	74.1	81.5	81.5	81.5	100.0	50.8	81.5	8	83.4	82.9	83.9	83.5	82.6	83.9	82.9	82.9	83.9	99.9	52.1	83.9	9	100.0	80.4	75.0	95.7	80.4	81.3	89.1	89.1	100.0	100.0	50.0	100.0	9	83.6	79.7	78.5	83.6	81.5	77.2	83.2	83.2	78.5	100.0	48.7	78.5	9	84.6	84.2	85.0	84.8	83.7	84.8	84.1	84.2	85.0	100.0	53.3	85.0	10	100.0	86.4	93.3	95.5	81.8	100.0	90.9	90.9	93.3	100.0	45.5	93.3	10	87.0	82.9	82.1	87.4	84.2	83.3	86.2	86.2	86.9	100.0	50.0	86.9	10	84.8	84.3	86.0	84.8	84.2	85.8	84.5	84.5	85.8	99.9	53.4	85.8
			Calendario	d)	1	100.0	92.9	80.0	100.0	92.9	100.0	100.0	92.9	100.0	100.0	35.7	40.0	1	92.4	88.0	85.7	93.5	89.7	88.9	91.3	91.3	90.5	100.0	44.0	90.5	1	89.7	88.8	88.9	89.7	89.1	88.0	89.6	89.6	91.6	100.0	50.3	91.6	2	100.0	83.3	33.3	100.0	83.3	50.0	94.4	94.4	100.0	100.0	38.9	33.3	2	89.4	87.1	82.8	90.0	87.7	86.2	88.2	87.7	87.9	100.0	46.5	87.9	2	88.5	88.1	85.5	89.0	88.4	85.5	88.5	88.5	85.1	100.0	50.2	85.1	3	100.0	75.0	25.0	100.0																																																																																																																																																																																																																																																																																																																		

Tabla B.2. Porcentajes de ajuste de los modelos para Peñoles en tendencia negativa: resultados de la clasificación por anomalía.

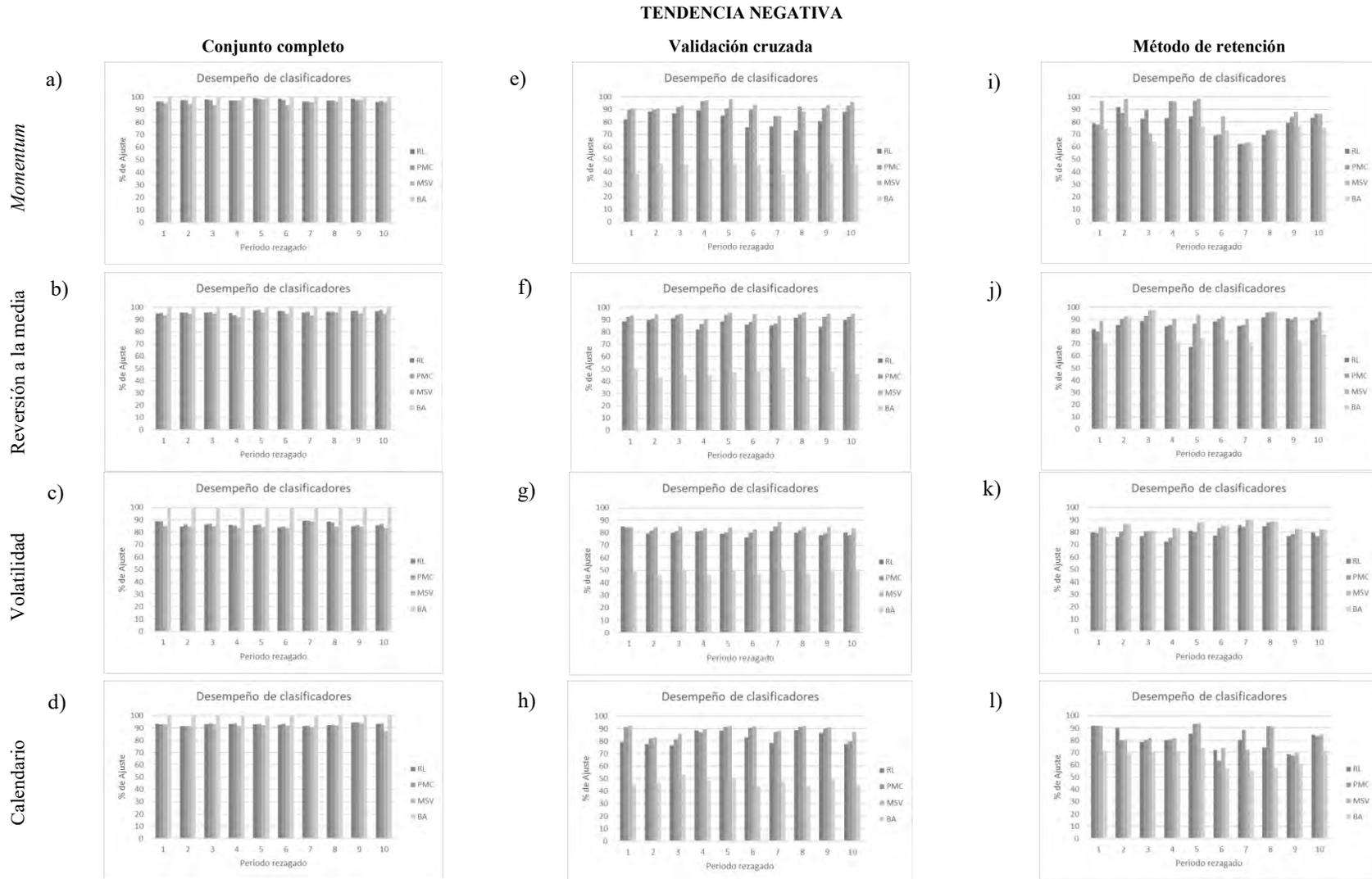
	Series mensuales												Series semanales												Series diarias																					
	Regresión Logística			Perceptrón multicapa			M. de soporte vectorial			Bosques aleatorios			Regresión Logística			Perceptrón multicapa			M. de soporte vectorial			Bosques aleatorios			Regresión Logística			Perceptrón multicapa			M. de soporte vectorial			Bosques aleatorios												
	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP				
a)	1	100.0	75.0	66.7	100.0	87.5	66.7	100.0	87.5	100.0	100.0	25.0	33.3	100.0	87.5	66.7	100.0	87.5	100.0	100.0	25.0	33.3	100.0	87.5	66.7	100.0	87.5	66.7	100.0	87.5	66.7	100.0	87.5	66.7	100.0	87.5	66.7	100.0	87.5	66.7	100.0	87.5	66.7	100.0	87.5	66.7
	2	100.0	80.0	100.0	100.0	80.0	66.7	90.0	80.0	100.0	100.0	40.0	33.3	100.0	80.0	66.7	100.0	80.0	66.7	100.0	40.0	33.3	100.0	80.0	66.7	100.0	80.0	66.7	100.0	80.0	66.7	100.0	80.0	66.7	100.0	80.0	66.7	100.0	80.0	66.7	100.0	80.0	66.7	100.0	80.0	66.7
	3	100.0	83.3	75.0	100.0	91.7	100.0	91.7	91.7	25.0	100.0	41.7	0.0	100.0	83.3	66.7	100.0	91.7	100.0	100.0	41.7	0.0	100.0	83.3	66.7	100.0	91.7	100.0	100.0	83.3	66.7	100.0	91.7	100.0	100.0	83.3	66.7	100.0	91.7	100.0	100.0	83.3	66.7			
	4	100.0	90.9	85.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	50.0	33.3	100.0	90.9	85.7	100.0	100.0	100.0	100.0	50.0	33.3	100.0	90.9	85.7	100.0	100.0	100.0	100.0	90.9	85.7	100.0	100.0	100.0	100.0	90.9	85.7	100.0	100.0	100.0	100.0	90.9	85.7			
	5	100.0	75.0	66.7	100.0	87.5	100.0	100.0	100.0	100.0	100.0	50.0	33.3	100.0	75.0	66.7	100.0	87.5	100.0	100.0	50.0	33.3	100.0	75.0	66.7	100.0	87.5	100.0	100.0	75.0	66.7	100.0	87.5	100.0	100.0	75.0	66.7	100.0	87.5	100.0	100.0	75.0	66.7			
	6	100.0	50.0	33.3	100.0	90.0	33.3	90.0	90.0	66.7	100.0	30.0	33.3	100.0	50.0	33.3	100.0	90.0	33.3	100.0	30.0	33.3	100.0	50.0	33.3	100.0	90.0	33.3	100.0	90.0	33.3	100.0	90.0	33.3	100.0	90.0	33.3	100.0	90.0	33.3	100.0	90.0	33.3			
	7	100.0	50.0	0.0	100.0	66.7	0.0	100.0	66.7	0.0	100.0	16.7	0.0	100.0	50.0	0.0	100.0	66.7	0.0	100.0	16.7	0.0	100.0	50.0	0.0	100.0	66.7	0.0	100.0	66.7	0.0	100.0	66.7	0.0	100.0	66.7	0.0	100.0	66.7	0.0	100.0	66.7	0.0			
	8	100.0	37.5	33.3	100.0	87.5	33.3	100.0	75.0	33.3	100.0	25.0	33.3	100.0	37.5	33.3	100.0	87.5	33.3	100.0	25.0	33.3	100.0	37.5	33.3	100.0	87.5	33.3	100.0	87.5	33.3	100.0	87.5	33.3	100.0	87.5	33.3	100.0	87.5	33.3	100.0	87.5	33.3			
	9	100.0	62.5	66.7	100.0	87.5	66.7	100.0	87.5	66.7	100.0	50.0	33.3	100.0	62.5	66.7	100.0	87.5	66.7	100.0	50.0	33.3	100.0	62.5	66.7	100.0	87.5	66.7	100.0	87.5	66.7	100.0	87.5	66.7	100.0	87.5	66.7	100.0	87.5	66.7	100.0	87.5	66.7			
	10	100.0	90.0	66.7	100.0	100.0	66.7	100.0	100.0	66.7	100.0	40.0	33.3	100.0	90.0	66.7	100.0	100.0	66.7	100.0	40.0	33.3	100.0	90.0	66.7	100.0	100.0	66.7	100.0	100.0	66.7	100.0	100.0	66.7	100.0	100.0	66.7	100.0	100.0	66.7						
b)	1	100.0	90.9	85.7	100.0	95.5	85.7	95.5	95.5	85.7	100.0	54.6	28.6	100.0	90.9	85.7	100.0	95.5	85.7	100.0	54.6	28.6	100.0	90.9	85.7	100.0	95.5	85.7	100.0	95.5	85.7	100.0	95.5	85.7	100.0	95.5	85.7	100.0	95.5	85.7	100.0	95.5	85.7			
	2	100.0	92.9	80.0	100.0	92.9	90.0	96.4	96.4	90.0	100.0	35.7	90.0	100.0	92.9	80.0	100.0	92.9	90.0	100.0	35.7	90.0	100.0	92.9	80.0	100.0	92.9	80.0	100.0	92.9	80.0	100.0	92.9	80.0	100.0	92.9	80.0	100.0	92.9	80.0						
	3	100.0	91.7	87.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	41.7	100.0	100.0	91.7	87.5	100.0	100.0	100.0	100.0	41.7	100.0	100.0	91.7	87.5	100.0	100.0	100.0	100.0	91.7	87.5	100.0	100.0	100.0	100.0	91.7	87.5									
	4	100.0	76.9	88.9	96.2	84.6	88.9	92.3	88.5	88.9	100.0	38.5	33.3	100.0	76.9	88.9	96.2	84.6	88.9	92.3	38.5	33.3	100.0	76.9	88.9	96.2	84.6	88.9	92.3	88.5	88.9	100.0	76.9	88.9	96.2	84.6	88.9	92.3	88.5	88.9						
	5	100.0	85.0	42.9	100.0	95.0	71.4	95.0	95.0	85.7	100.0	50.0	28.6	100.0	85.0	42.9	100.0	95.0	71.4	95.0	50.0	28.6	100.0	85.0	42.9	100.0	95.0	71.4	95.0	95.0	85.7	100.0	85.0	42.9	100.0	95.0	71.4	95.0	95.0	85.7						
	6	100.0	77.3	85.7	100.0	81.8	85.7	95.5	95.5	85.7	100.0	40.9	28.6	100.0	77.3	85.7	100.0	81.8	85.7	95.5	40.9	28.6	100.0	77.3	85.7	100.0	81.8	85.7	95.5	95.5	85.7	100.0	77.3	85.7	100.0	81.8	85.7	95.5	95.5	85.7						
	7	100.0	77.3	85.7	100.0	77.3	85.7	90.9	90.9	85.7	100.0	54.6	28.6	100.0	77.3	85.7	100.0	77.3	85.7	90.9	54.6	28.6	100.0	77.3	85.7	100.0	77.3	85.7	100.0	77.3	85.7	100.0	77.3	85.7	100.0	77.3	85.7	100.0	77.3	85.7						
	8	100.0	95.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	37.5	100.0	100.0	95.8	100.0	100.0	100.0	100.0	100.0	37.5	100.0	100.0	95.8	100.0	100.0	100.0	100.0	100.0	95.8	100.0	100.0	100.0	100.0	100.0	95.8	100.0									
	9	100.0	72.7	85.7	100.0	90.9	85.7	95.5	95.5	85.7	100.0	40.9	28.6	100.0	72.7	85.7	100.0	90.9	85.7	95.5	40.9	28.6	100.0	72.7	85.7	100.0	90.9	85.7	95.5	95.5	85.7	100.0	72.7	85.7	100.0	90.9	85.7	95.5	95.5	85.7						
	10	100.0	88.5	88.9	100.0	92.3	88.9	96.2	96.2	100.0	100.0	38.5	44.4	100.0	88.5	88.9	100.0	92.3	88.9	96.2	38.5	44.4	100.0	88.5	88.9	100.0	92.3	88.9	96.2	96.2	100.0	100.0	88.5	88.9	96.2	96.2	100.0									
c)	1	100.0	90.9	73.3	100.0	90.9	73.3	88.6	88.6	86.7	100.0	43.2	86.7	100.0	90.9	73.3	100.0	90.9	73.3	88.6	43.2	86.7	100.0	90.9	73.3	100.0	90.9	73.3	100.0	90.9	73.3	100.0	90.9	73.3	100.0	90.9	73.3	100.0	90.9	73.3						
	2	85.7	73.2	63.2	91.1	78.6	73.7	85.7	85.7	89.5	100.0	39.3	89.5	100.0	85.7	73.2	91.1	78.6	73.7	85.7	39.3	89.5	100.0	85.7	73.2	91.1	78.6	73.7	85.7	85.7	89.5	100.0	85.7	73.2	91.1	78.6	73.7	85.7	85.7	89.5						
	3	91.1	76.8	73.7	91.1	80.4	84.2	87.5	87.5	84.2	100.0	44.6	84.2	100.0	91.1	76.8	91.1	80.4	84.2	87.5	44.6	84.2	100.0	91.1	76.8	91.1	80.4	84.2	87.5	87.5	84.2	100.0	91.1	76.8	91.1	80.4	84.2	87.5	87.5	84.2						
	4	91.1	78.6	52.6	87.5	80.4	57.9	83.9	83.9	79.0	100.0	39.3	79.0	100.0	91.1	78.6	87.5	80.4	57.9	83.9	39.3	79.0	100.0	91.1	78.6	87.5	80.4	57.9	83.9	83.9	79.0	100.0	91.1	78.6	87.5	80.4	57.9	83.9	83.9	79.0						
	5	92.6	75.9	83.3	94.4	79.6	83.3	88.9	88.9	100.0	100.0	46.3	100.0	100.0	92.6	75.9	94.4	79.6	83.3	88.9	46.3	100.0	100.0	92.6	75.9	94.4	79.6	83.3	88.9	88.9	100.0	100.0	92.6	75.9	94.4	79.6	83.3	88.9	88.9	100.0						
	6	80.0	62.0	58.8	84.0	72.0	76.5	80.0	80.0	82.4	100.0	38.0	82.4	100.0	80.0	62.0	84.0	72.0	76.5	80.0	38.0	82.4	100.0	80.0	62.0	84.0	72.0	76.5	80.0	80.0	82.4	100.0	80.0	62.0	84.0	72.0	76.5	80.0	80.0	82.4						
	7	98.2	79.6	83.3	98.2	88.9	77.8	96.3	96.3	94.4	100.0	48.2	94.4	100.0	98.2	79.6	98.2	88.9	77.8	96.3	48.2	94.4	100.0	98.2	79.6	98.2	88.9	77.8	96.3	96.3	94.4	100.0	98.2	79.6	98.2	88.9	77.8	96.3	96.3	94.4						
	8	96.0	76.0	94.1	94.0	78.0	94.1	84.0	84.0	94.1	100.0	40.0	94.1	100.0	96.0	76.0	94.0	78.0	94.1	84.0	40.0	94.1	100.0	96.0	76.0	94.0	78.0	94.1	84.0	84.0	94.1	100.0	96.0	76.0	94.0	78.0	94.1	84.0	84.0	94.1						
	9	86.5	71.2	72.2	88.5	73.1	72.2	86.5	86.5	83.3	100.0	46.2	83.3	100.0	86.5	71.2	88.5	73.1	72.2	86.5	46.2	83.3	100.0	86.5	71.2	88.5	73.1	72.2	86.5	86.5	83.3	100.0	86.5	71.2	88.5	73.1	72.2	86.5	86.5	83.3						
	10	85.7	75.0	73.7	91.1	67.9	63.2	82.1	82.1	79.0	100.0	48.2	79.0	100.0	85.7	75.0	91.1	6																												

Figura B.5. Gráficas comparativas del desempeño promedio de los clasificadores con las series de Peños por anomalía en tendencia positiva.



Fuente: Elaboración propia.

Figura B.6. Gráficas comparativas del desempeño promedio de los clasificadores con las series de Peñoles por anomalía en tendencia negativa.



Fuente: Elaboración propia.

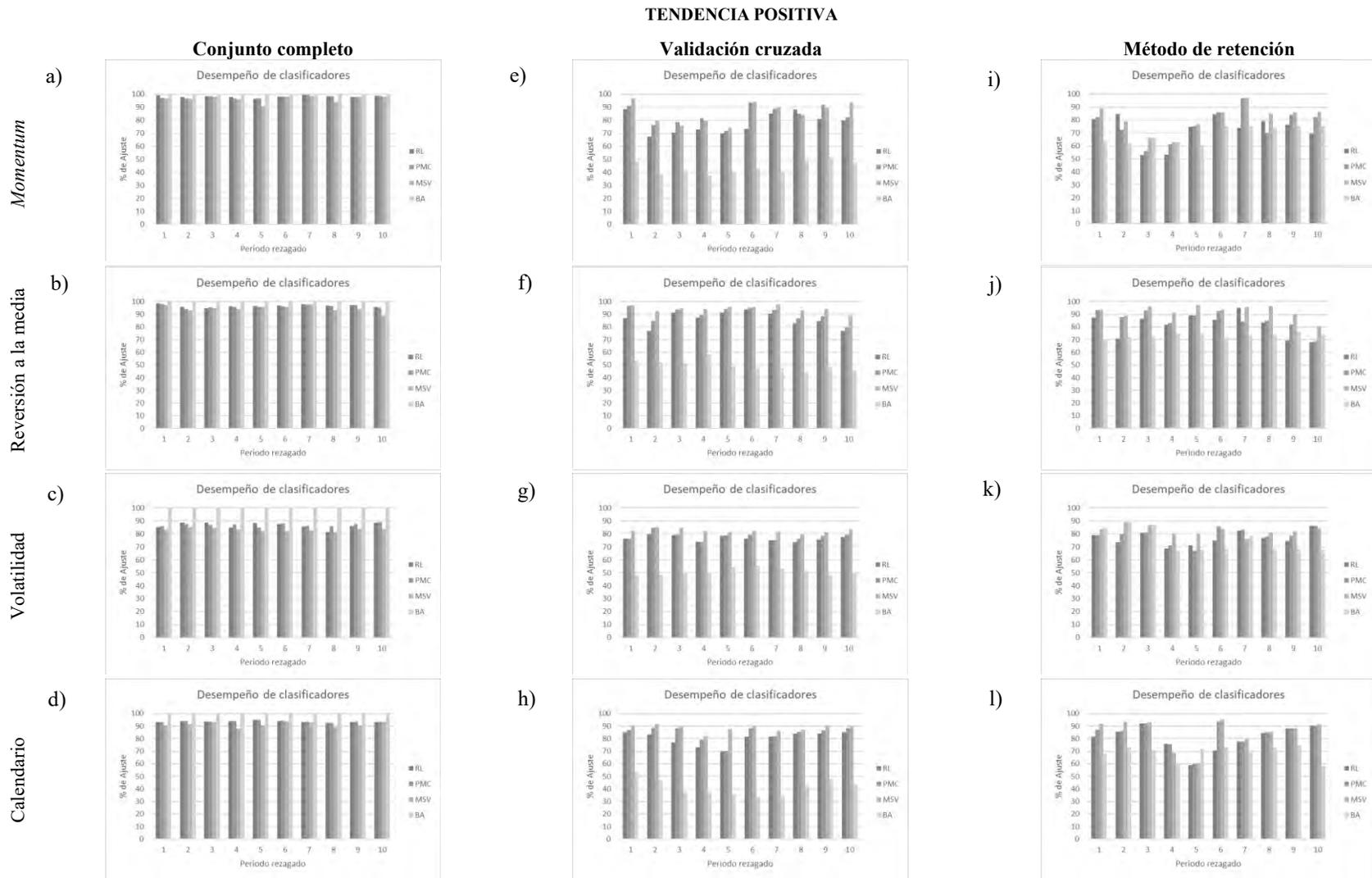


Tabla B.4. Porcentajes de ajuste de los modelos para Cemex en tendencia negativa: resultados de la clasificación por anomalía.

	Series mensuales												Series semanales												Series diarias																																																																																																																																																																																																																																																																																																																																																																
	Regresión Logística			Perceptrón multicapa			M. de soporte vectorial			Bosques aleatorios			Regresión Logística			Perceptrón multicapa			M. de soporte vectorial			Bosques aleatorios			Regresión Logística			Perceptrón multicapa			M. de soporte vectorial			Bosques aleatorios																																																																																																																																																																																																																																																																																																																																																							
	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP																																																																																																																																																																																																																																																																																																																																									
Momentum	1	100.0	85.7	80.0	100.0	78.6	80.0	92.9	85.7	80.0	100.0	35.7	40.0	1	100.0	80.7	66.7	96.8	88.7	81.0	95.2	95.2	95.2	100.0	50.0	95.2	1	95.6	93.2	88.9	96.2	93.8	89.9	95.2	95.2	92.9	100.0	52.7	92.9	2	100.0	94.4	83.3	100.0	94.4	83.3	100.0	100.0	83.3	100.0	55.6	33.3	2	94.8	89.6	89.0	94.0	92.5	90.1	93.7	93.7	91.2	100.0	48.9	91.2	3	100.0	95.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	65.0	28.6	3	100.0	86.4	90.9	98.5	95.5	95.5	98.5	98.5	95.5	100.0	50.0	95.5	3	96.3	94.8	94.5	96.6	95.5	95.6	96.3	96.3	96.7	100.0	47.8	96.7	4	100.0	88.9	66.7	100.0	100.0	66.7	100.0	100.0	66.7	100.0	72.2	33.3	4	100.0	84.7	83.3	97.2	91.7	83.3	94.4	94.4	83.3	100.0	54.2	83.3	4	97.6	94.9	95.0	97.6	95.2	92.9	95.9	95.9	96.5	100.0	45.6	95.0	5	100.0	85.0	85.7	100.0	90.0	100.0	100.0	100.0	100.0	100.0	65.0	28.6	5	95.8	80.6	83.3	95.8	90.3	87.5	94.4	94.4	100.0	100.0	48.6	100.0	5	97.7	94.2	95.5	97.7	96.5	97.7	97.3	97.3	97.7	99.6	52.3	97.7	6	100.0	85.0	42.9	100.0	100.0	71.4	100.0	100.0	71.4	100.0	70.0	28.6	6	97.1	78.6	79.2	97.1	92.9	91.7	95.7	95.7	91.7	100.0	41.4	91.7	6	96.4	94.3	96.8	96.4	94.6	99.0	95.7	95.7	99.0	100.0	52.1	99.0	7	100.0	95.5	85.7	100.0	95.5	85.7	100.0	100.0	100.0	100.0	45.5	28.6	7	97.4	89.5	73.1	97.4	92.1	92.3	93.4	93.4	92.3	100.0	48.7	92.3	7	97.9	94.8	91.8	97.6	96.9	92.8	97.2	97.2	93.8	99.7	54.6	93.8	8	100.0	59.1	42.9	95.5	72.7	28.6	90.9	90.9	85.7	100.0	63.6	28.6	8	94.4	79.2	79.2	95.8	83.3	87.5	90.3	90.3	87.5	100.0	55.6	87.5	8	94.9	92.8	93.9	94.9	92.8	95.0	94.5	94.5	97.0	100.0	50.0	97.0	9	100.0	86.4	85.7	100.0	95.5	100.0	95.5	95.5	100.0	100.0	68.2	28.6	9	98.7	92.1	88.5	98.7	97.4	92.3	97.4	97.4	92.3	100.0	52.6	92.3	9	98.7	96.4	93.3	98.7	96.1	94.3	97.4	97.1	96.2	100.0	46.8	97.1	10	100.0	40.9	71.4	95.5	50.0	100.0	95.5	95.5	100.0	100.0	68.2	28.6	10	100.0	97.1	91.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	44.1	100.0	10	98.0	96.6	98.0	98.0	97.0	98.0	97.6	97.3	98.0	100.0	45.6	98.0
	Reversión a la media	1	100.0	94.4	83.3	100.0	94.4	100.0	100.0	94.4	83.3	100.0	57.6	33.3	1	100.0	85.3	73.9	98.5	97.1	69.6	98.5	98.5	100.0	100.0	45.6	100.0	1	95.8	94.3	94.4	95.8	93.9	95.8	95.5	95.5	96.5	99.8	47.4	96.5	2	100.0	91.7	87.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	70.8	100.0	2	100.0	84.7	87.5	98.6	88.9	87.5	94.4	94.4	87.5	100.0	48.6	87.5	3	100.0	67.9	70.0	100.0	78.6	80.0	92.9	85.7	80.0	100.0	42.9	40.0	3	98.7	89.2	88.0	98.7	97.3	96.0	97.3	97.3	96.0	100.0	43.2	96.0	4	100.0	80.8	77.8	100.0	88.5	88.9	96.2	92.3	88.9	100.0	50.0	88.9	4	98.8	89.3	82.8	97.6	92.9	93.1	96.4	96.4	93.1	100.0	48.8	93.1	5	100.0	83.3	87.5	100.0	91.7	100.0	95.8	95.8	100.0	100.0	41.7	50.0	5	92.2	78.9	87.1	92.2	85.6	93.6	91.1	91.1	96.8	100.0	42.2	96.8	6	100.0	81.8	85.7	100.0	86.4	100.0	95.5	95.5	100.0	100.0	45.5	28.6	6	92.5	85.0	88.9	91.3	80.0	81.5	92.5	92.5	88.9	100.0	61.3	88.9	7	100.0	65.4	77.8	100.0	76.9	77.8	88.5	88.5	88.9	100.0	34.6	33.3	7	93.5	84.8	77.9	93.5	89.1	90.3	91.3	90.2	93.6	100.0	47.8	93.6	8	100.0	73.3	60.0	100.0	80.0	70.0	93.3	93.3	90.0	100.0	50.0	90.0	8	95.7	85.1	84.4	93.6	87.2	87.5	92.6	92.6	93.8	100.0	44.7	93.8	9	100.0	66.7	90.0	96.7	70.0	90.0	90.0	90.0	100.0	100.0	53.3	100.0	9	95.8	89.6	72.7	95.8	95.8	93.9	94.8	94.8	97.0	100.0	42.7	97.0	10	100.0	80.0	80.0	100.0	83.3	90.0	93.3	93.3	100.0	100.0	40.0	100.0	10	94.8	89.6	72.7	93.8	91.7	84.9	93.8	93.8	90.9	100.0	45.8	90.9																																																																																																							
		Volatilidad	1	98.3	78.3	60.0	93.3	85.0	70.0	88.3	88.3	85.0	100.0	46.7	85.0	1	82.9	79.6	84.2	84.2	80.4	84.2	81.7	81.7	81.7	100.0	55.0	81.7	1	84.0	83.2	85.1	84.0	83.5	85.9	83.7	83.7	86.2	99.8	53.5	86.2	2	86.4	75.8	68.2	87.9	78.8	59.1	77.3	77.3	72.7	100.0	40.9	72.7	2	82.0	79.3	80.0	84.7	78.4	80.0	82.0	82.0	85.3	100.0	54.5	85.3	3	88.9	77.8	70.8	91.7	76.4	75.0	80.6	80.6	70.8	100.0	56.9	70.8	3	86.1	83.3	84.1	85.7	82.6	83.0	83.7	83.7	83.0	100.0	52.3	83.0	4	87.5	78.1	45.5	95.3	78.1	68.2	82.8	82.8	81.8	100.0	57.8	81.8	4	82.7	80.0	80.7	83.5	78.9	84.1	82.3	82.3	86.4	100.0	56.2	86.4	5	88.7	74.2	57.1	90.3	82.3	66.7	83.9	83.9	85.7	100.0	59.7	85.7	5	85.7	82.1	84.9	85.7	84.1	83.7	84.1	84.1	84.9	100.0	56.0	84.9	6	90.9	80.3	86.4	90.9	83.3	77.3	84.9	84.9	86.4	100.0	57.6	86.4	6	85.8	82.7	83.7	85.8	82.7	83.7	83.9	83.9	88.4	100.0	46.5	88.4	7	93.1	80.6	83.3	93.1	86.1	83.3	88.9	88.9	87.5	100.0	59.7	87.5	7	81.5	78.4	83.7	82.7	79.9	83.7	81.9	81.9	88.4	100.0	54.7	88.4	8	88.2	80.3	65.4	93.4	82.9	76.9	86.8	86.8	88.5	100.0	57.9	88.5	8	87.5	83.8	77.2	87.1	84.9	79.4	85.7	85.7	85.9	100.0	49.6	85.9	9	84.7	72.2	66.7	90.3	76.4	58.3	83.3	83.3	79.2	100.0	58.3	79.2	9	83.0	79.6	80.8	83.3	81.1	81.1	81.1	80.7	85.6	100.0	52.7	85.6	10	87.1	74.3	62.5	88.6	71.4	66.7	81.4	81.4	79.2	98.6	54.3	79.2	10	83.1	78.9	81.8	83.1	80.8	81.8	81.5	81.5	85.2	100.0	47.3	86.4																																																																																																						
			Calendario	1	100.0	62.5	33.3	100.0	62.5	33.3	100.0	62.5	0.0	100.0	62.5	33.3	1	90.6	85.5	76.6	89.9	85.5	83.0	88.4	88.4	85.1	100.0	52.2	85.1	1	89.2	88.5	84.6	89.5	88.5	85.5	89.1	89.1	86.3	100.0	50.3	86.3	2	100.0	75.0	75.0	100.0	100.0	100.0	100.0	91.7	100.0	100.0	41.7	0.0	2	91.6	86.4	82.7	90.9	86.4	82.7	89.6	89.6	92.3	100.0	49.4	92.3	3	100.0	55.6	33.3	94.4	55.6	33.3	88.9	88.9	66.7	100.0	38.9	33.3	3	93.5	88.6	90.5	94.0	89.7	87.3	92.4	92.4	92.1	100.0	57.1	92.1	4	100.0	92.9	80.0	100.0	100.0	80.0	100.0	100.0	80.0	100.0	35.7	40.0	4	95.4	93.1	88.1	95.4	92.5	88.1	92.5	92.5	89.8	100.0	55.2	89.8	5	100.0	92.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	42.9	40.0	5	92.0	87.7	89.1	92.6	85.8	87.3	90.1	90.1	89.1	100.0	56.2	89.1	6	100.0	41.7	25.0	100.0	58.3	25.0	91.7	83.3	75.0	100.0	50.0	0.0	6	93.6	89.5	89.7	93.6	90.7	89.7	92.4	92.4	94.8	100.0	50.0	94.8	7	100.0	68.8	60.0	100.0	81.3	80.0	87.5	81.3	100.0	100.0	50.0	40.0	7	91.3	88.6	88.9	90.8	89.1	90.5	90.8	90.8	90.5	100.0	55.4	90.5	8	100.0	75.0	50.0	100.0	75.0	75.0	91.7	75.0	25.0	100.0	41.7	0.0	8	87.1	85.4	83.6	86.5	83.2	90.2	86.0	86.0	91.8	100.0	50.0	91.8	9	100.0	71.4	40.0	100.0	71.4	60.0	92.9	92.9	100.0	100.0	35.7	40.0	9	92.5	85.5	82.5	90.3	88.2	85.7	90.3	90.3	88.9	100.0	50.5	88.9	10	100.0	68.8	40.0	100.0	68.8	40.0	93.8	93.8	100.0	100.0	31.3	40.0	10	87.6	83.7	80.3	87.6	83.7	68.9	87.6	87.6	86.9	100.0	53.4	86.9																																																																																																					

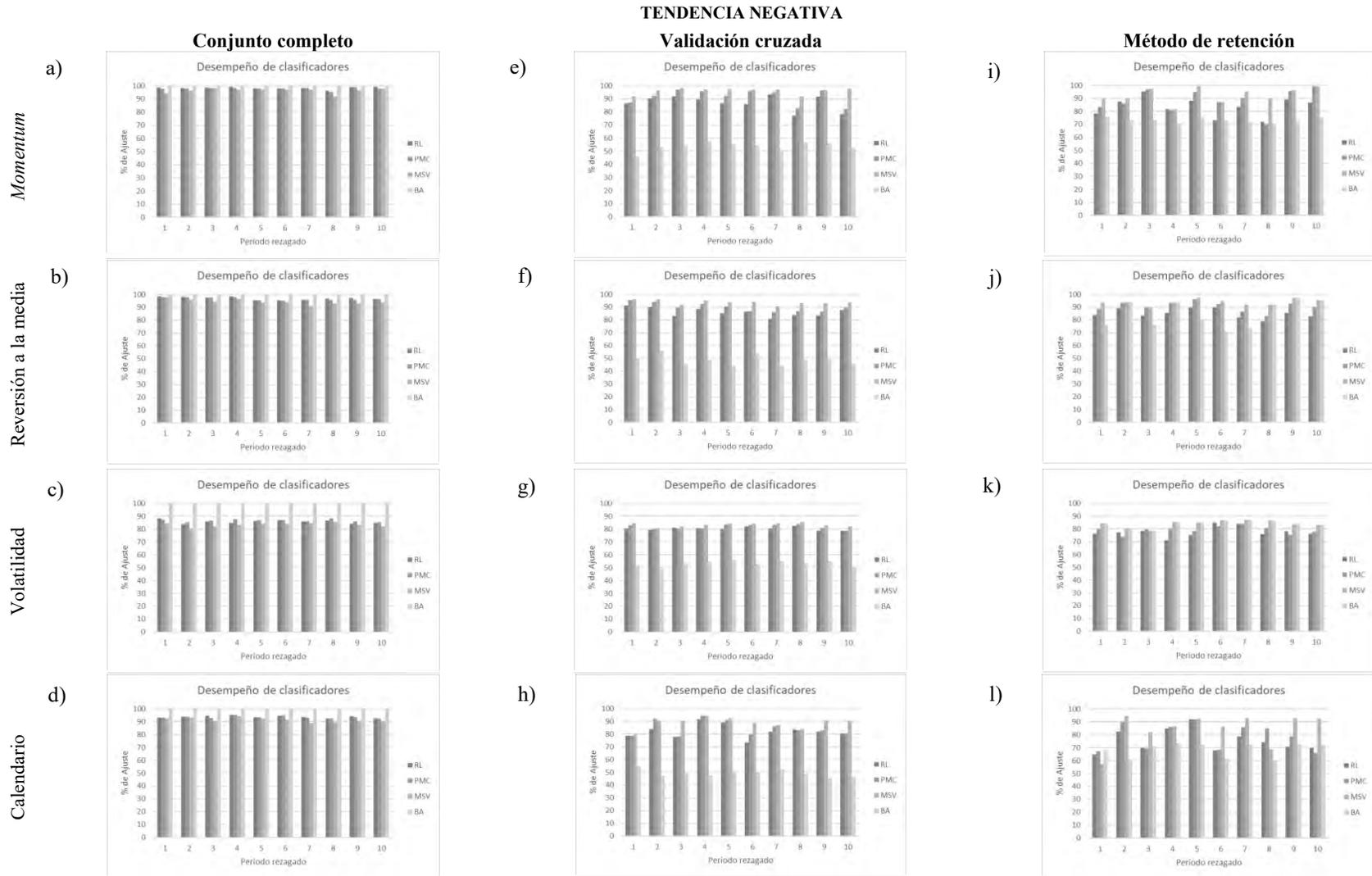
Fuente: Elaboración propia.

Figura B.7. Gráficas comparativas del desempeño promedio de los clasificadores con las series de Cemex por anomalía en tendencia positiva.



Fuente: Elaboración propia.

Figura B.8. Gráficas comparativas del desempeño promedio de los clasificadores con las series de Cemex por anomalía en tendencia negativa.

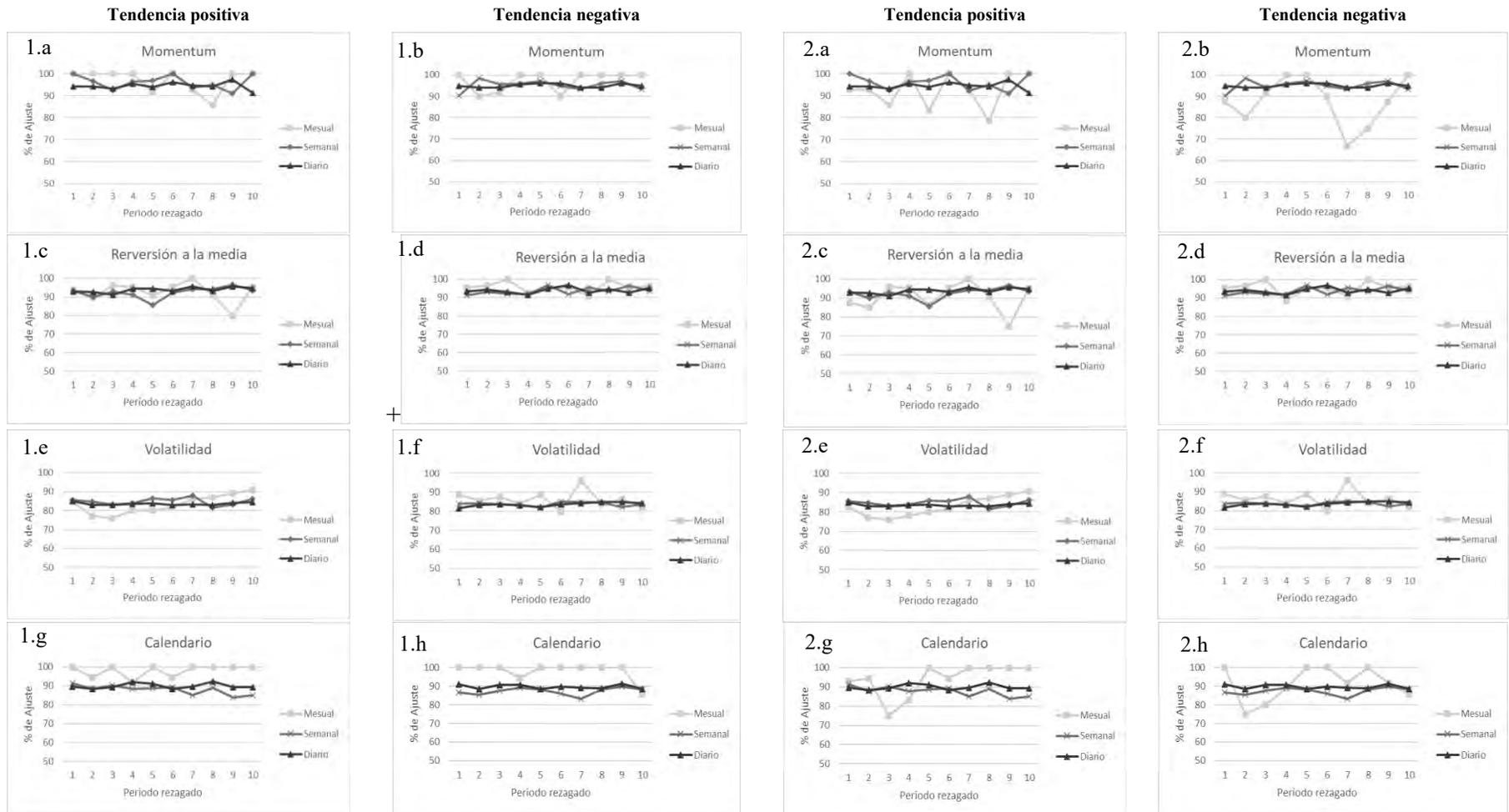


Fuente: Elaboración propia.

Figura B.9. Porcentajes de exactitud obtenido con el modelo de Máquina de soporte vectorial en las series de Peñoles.

1) Conjunto completo

2) Validación cruzada

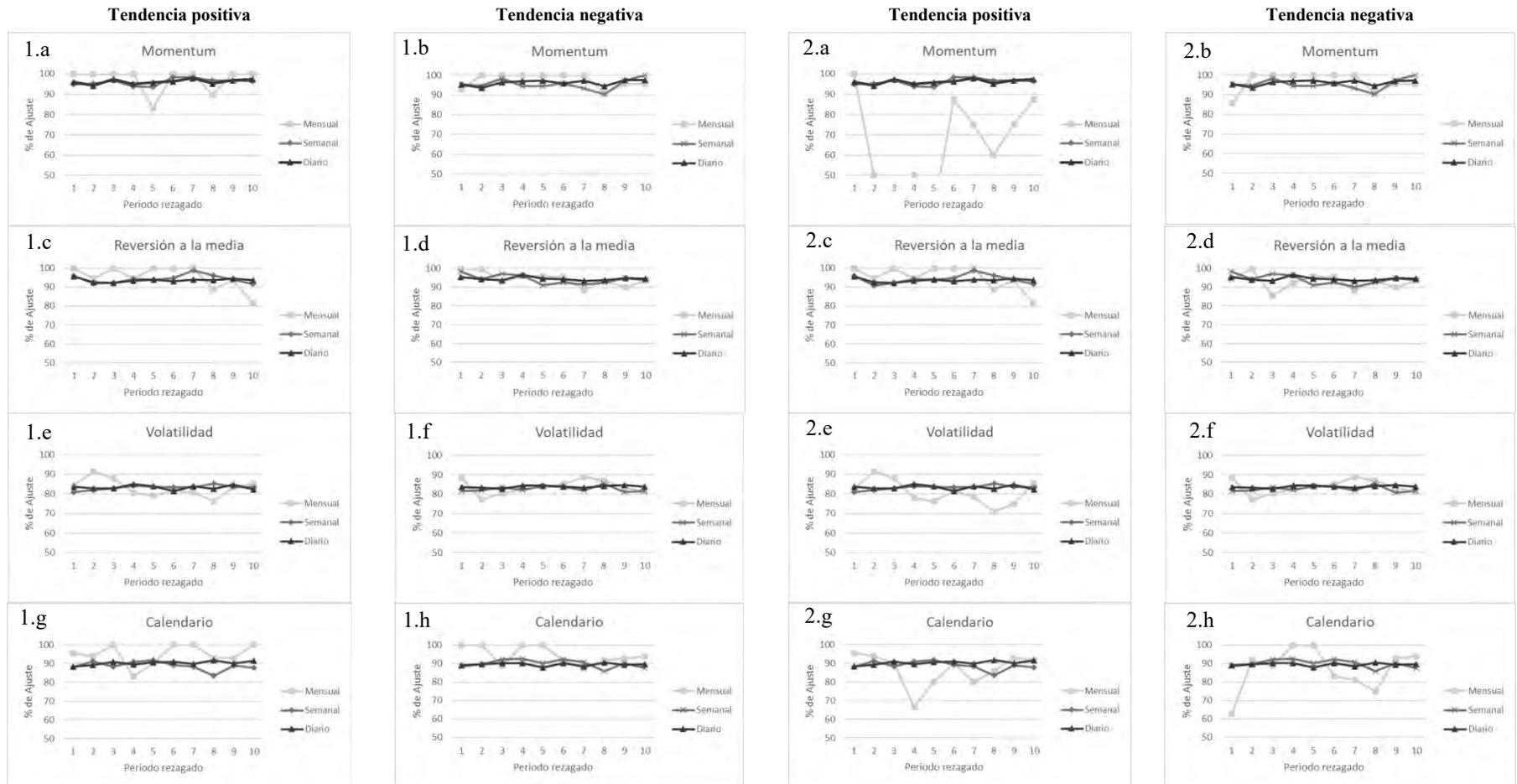


Fuente: Elaboración propia.

Figura B.10. Porcentajes de exactitud obtenido con el modelo de Máquina de soporte vectorial en las series de Cemex.

1) Conjunto completo

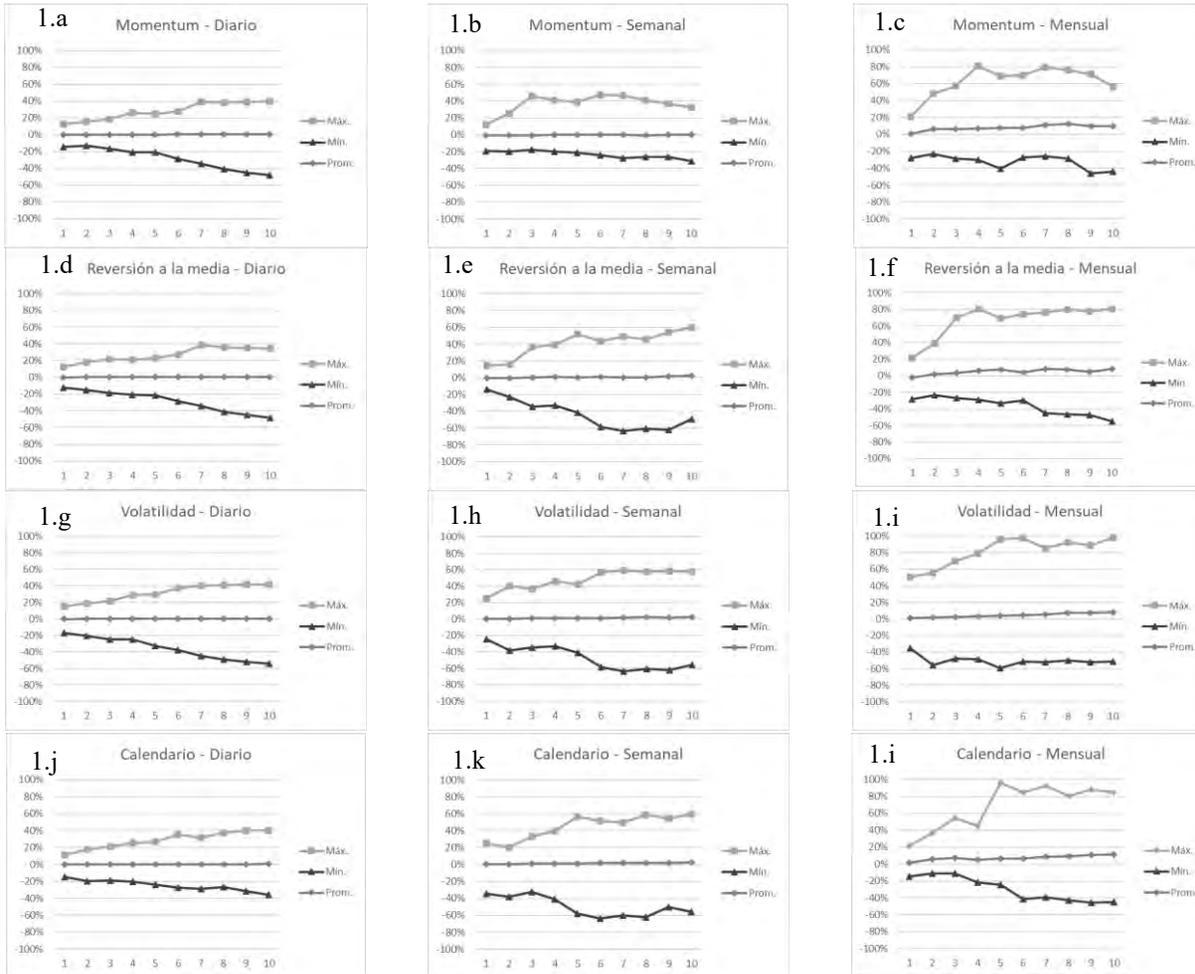
2) Validación cruzada



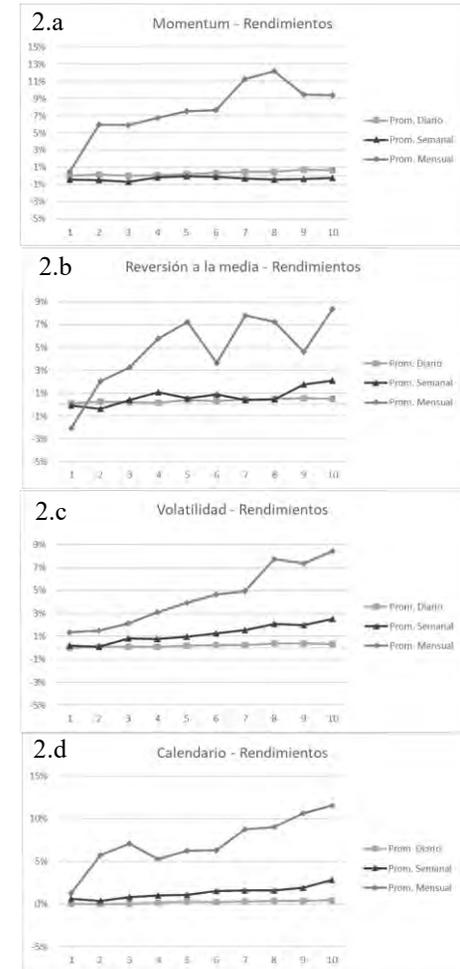
Fuente: Elaboración propia.

Figura B.11. Rendimientos lineales ajustados al riesgo de Peñoles en tendencia positiva.

1) Rendimientos máximos, mínimos y promedios por periodo rezagado



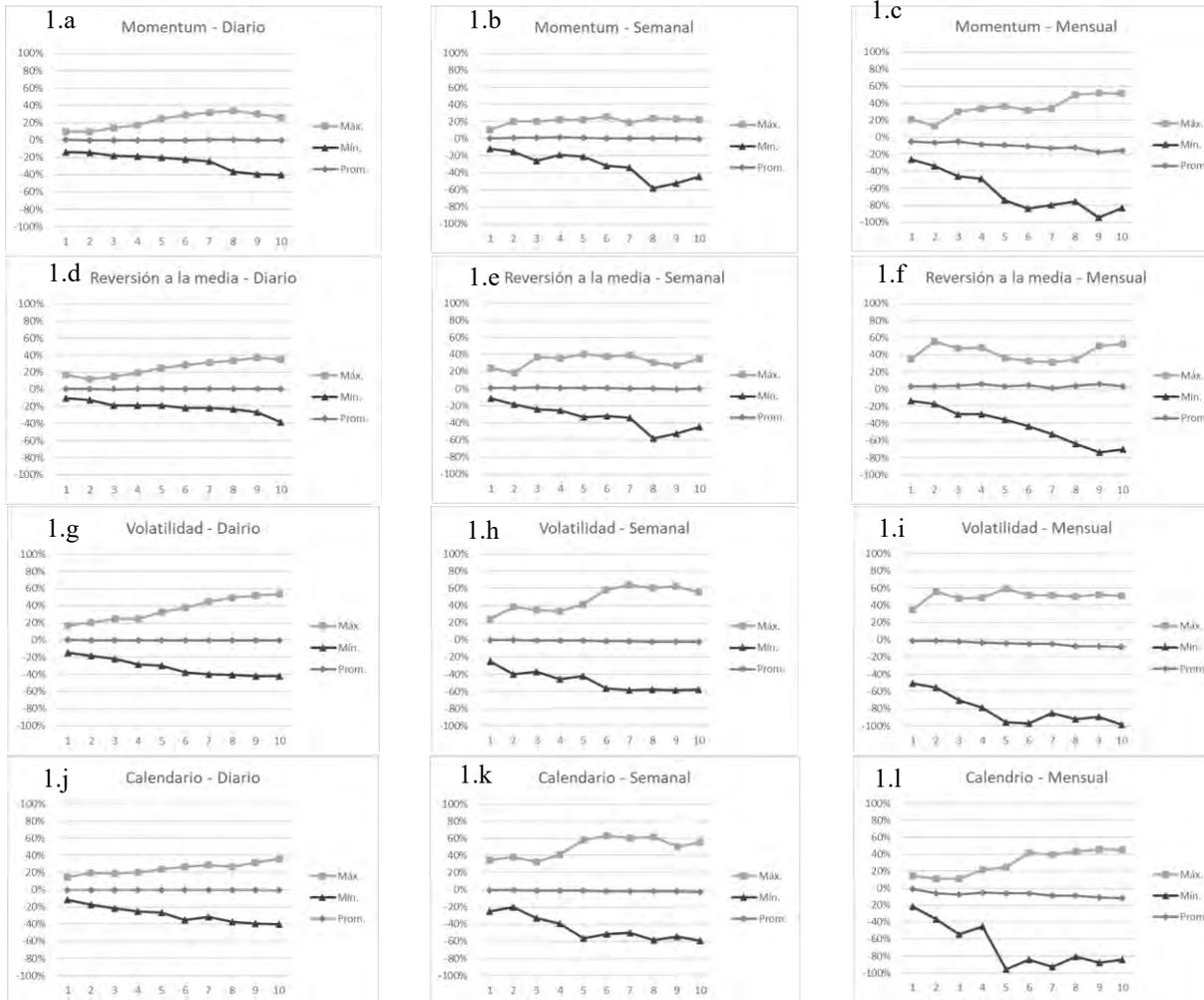
2) Rendimientos promedio ampliados



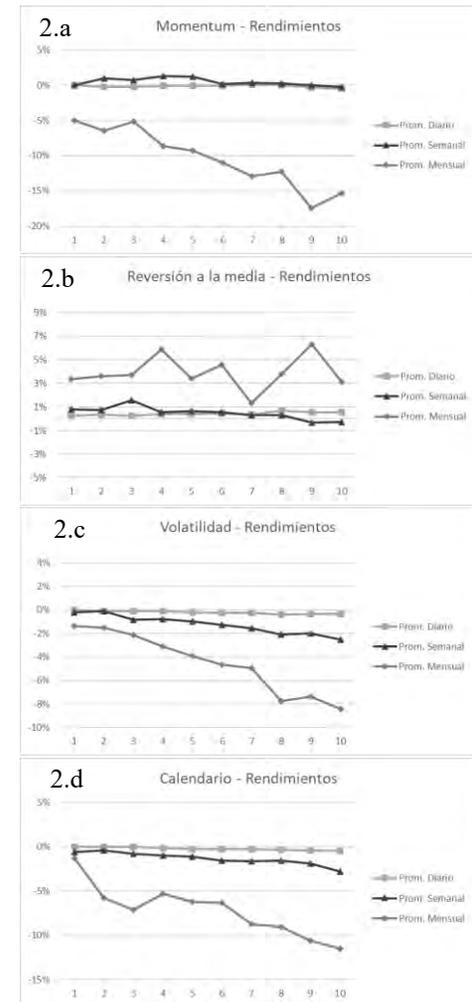
Fuente: Elaboración propia.

Figura B.12. Rendimientos lineales ajustados al riesgo de Peñoles en tendencia negativa.

1) Rendimientos máximos, mínimos y promedios por periodo rezagado



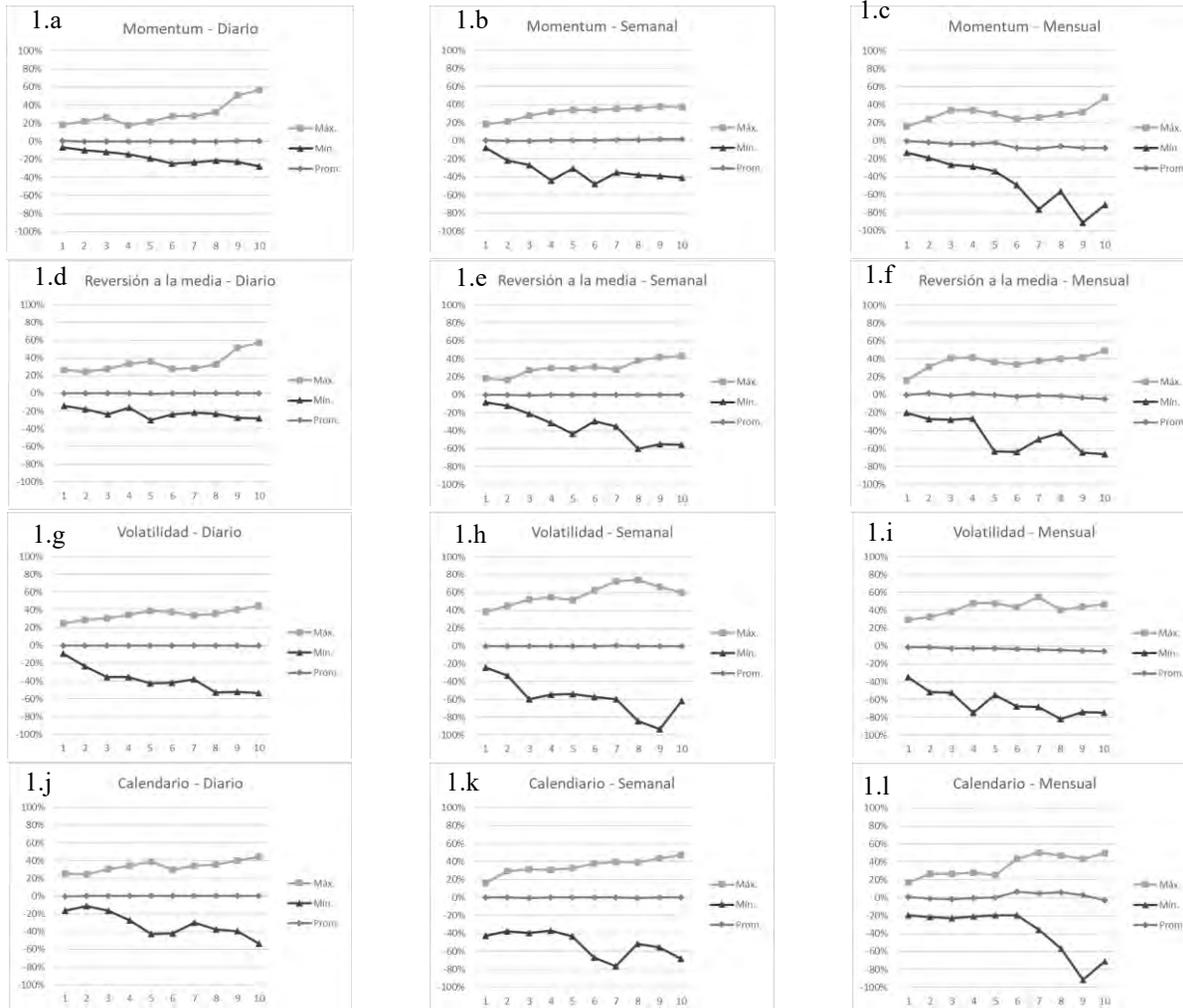
2) Rendimientos promedio ampliados



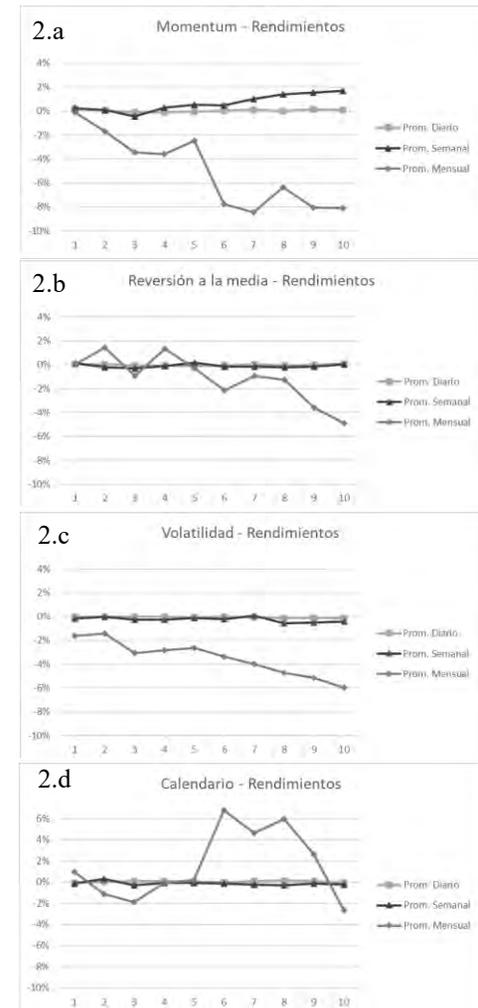
Fuente: Elaboración propia.

Figura B.13. Rendimientos lineales ajustados al riesgo de Cemex en tendencia positiva.

1) Rendimientos máximos, mínimos y promedios por periodo rezagado



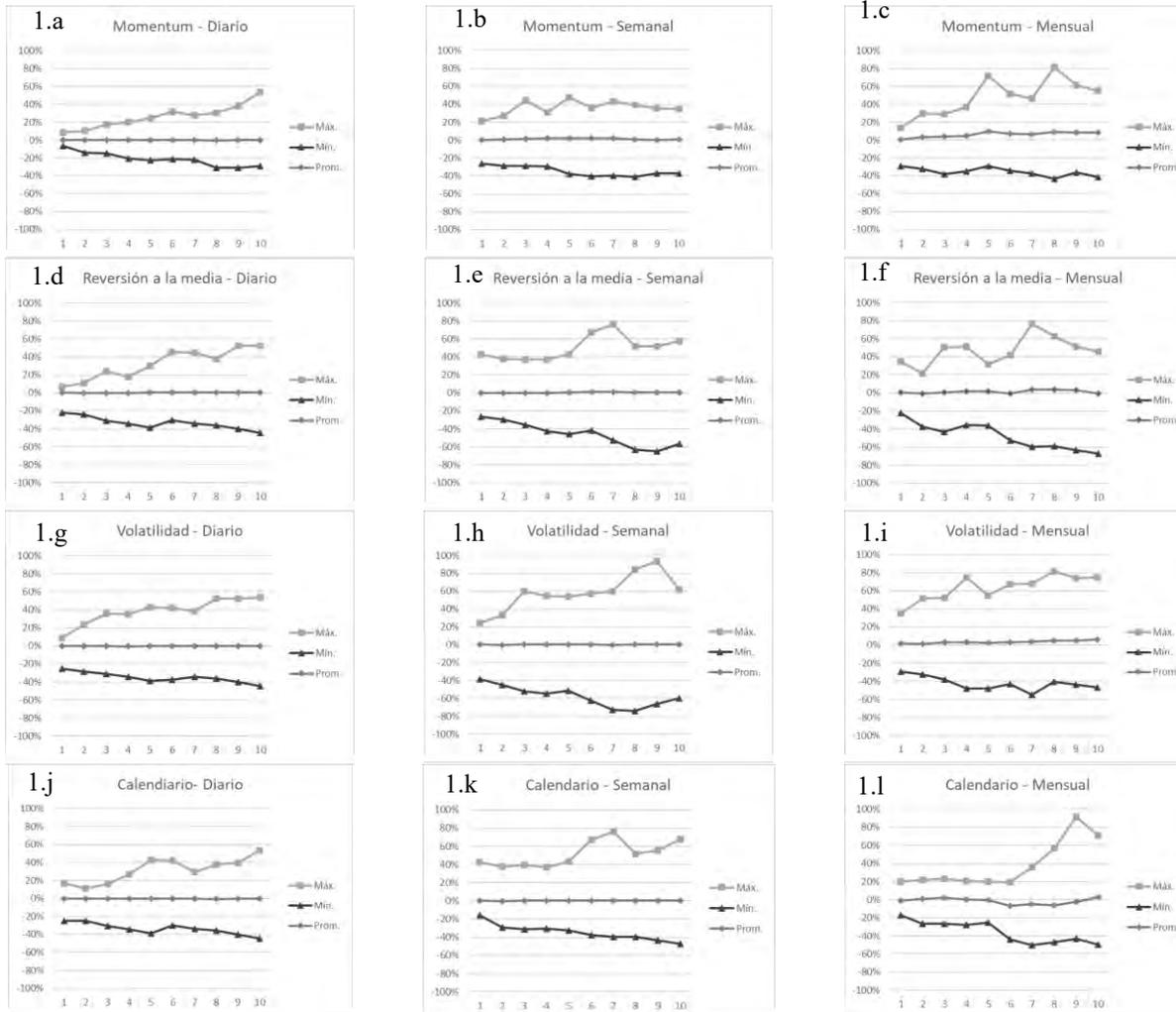
2) Rendimientos promedio ampliados



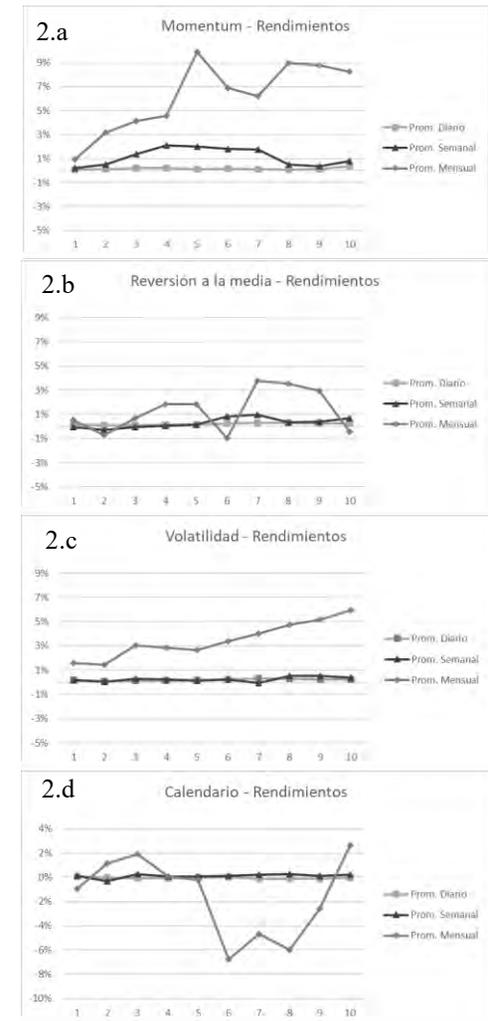
Fuente: Elaboración propia.

Figura B.14. Rendimientos lineales ajustados al riesgo de Cemex en tendencia negativa.

1) Rendimientos máximos, mínimos y promedios por periodo rezagado



2) Rendimientos promedio ampliados



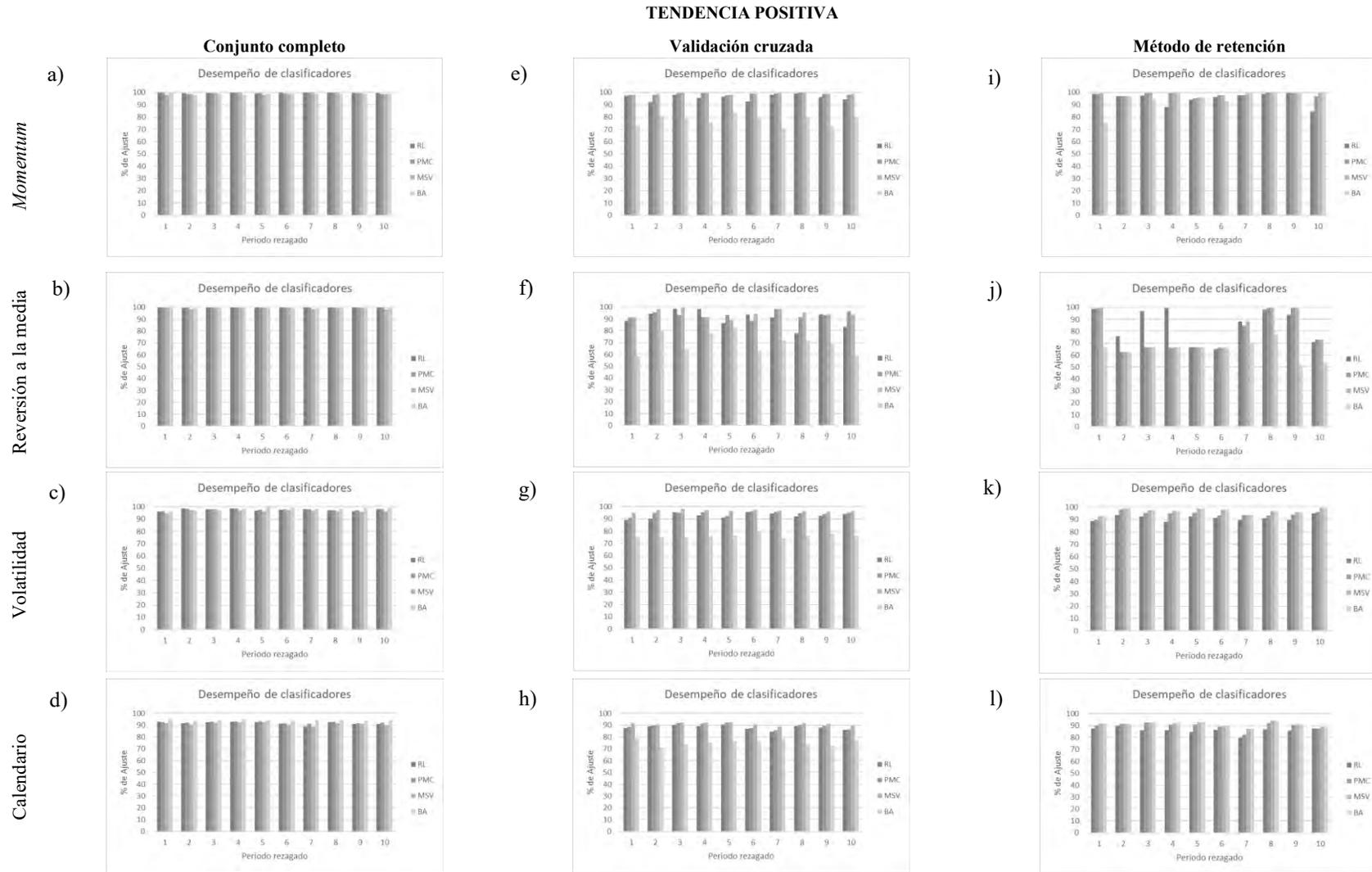
Fuente: Elaboración propia.

Tabla B.5. Porcentajes de ajuste de los modelos para el portafolio de empresas con menor bursatilidad en tendencia positiva.

	Series mensuales												Series semanales												Series diarias																		
	Regresión Logística			Perceptrón multicapa			M. de soporte vectorial			Bosques aleatorios			Regresión Logística			Perceptrón multicapa			M. de soporte vectorial			Bosques aleatorios			Regresión Logística			Perceptrón multicapa			M. de soporte vectorial			Bosques aleatorios									
	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	
a) <i>Momentum</i>	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	2	100.0	78.3	93.8	97.8	95.7	93.8	97.8	97.8	93.8	95.7	65.2	93.8	95.7	65.2	93.8	95.7	65.2	93.8	95.7	65.2	93.8	95.7	65.2	93.8	95.7	65.2	93.8	95.7	65.2	93.8	95.7	65.2	93.8	95.7	65.2	93.8	95.7	65.2	93.8	95.7	65.2	93.8
	3	100.0	97.6	92.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.6	64.3	100.0	97.6	64.3	100.0	97.6	64.3	100.0	97.6	64.3	100.0	97.6	64.3	100.0	97.6	64.3	100.0	97.6	64.3	100.0	97.6	64.3	100.0	97.6	64.3	100.0	97.6	64.3	100.0
	4	100.0	97.8	81.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	95.7	63.0	100.0	95.7	63.0	100.0	95.7	63.0	100.0	95.7	63.0	100.0	95.7	63.0	100.0	95.7	63.0	100.0	95.7	63.0	100.0	95.7	63.0	100.0	95.7	63.0	100.0	95.7	63.0	100.0
	5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.2	72.2	100.0	97.2	72.2	100.0	97.2	72.2	100.0	97.2	72.2	100.0	97.2	72.2	100.0	97.2	72.2	100.0	97.2	72.2	100.0	97.2	72.2	100.0	97.2	72.2	100.0	97.2	72.2	100.0
	6	100.0	80.0	92.9	97.5	97.5	92.9	97.5	97.5	92.9	97.5	97.5	92.9	97.5	70.0	92.9	97.5	70.0	92.9	97.5	70.0	92.9	97.5	70.0	92.9	97.5	70.0	92.9	97.5	70.0	92.9	97.5	70.0	92.9	97.5	70.0	92.9	97.5	70.0	92.9	97.5	70.0	92.9
	7	100.0	95.7	93.8	100.0	97.8	93.8	100.0	100.0	100.0	100.0	100.0	100.0	97.8	50.0	100.0	97.8	50.0	100.0	97.8	50.0	100.0	97.8	50.0	100.0	97.8	50.0	100.0	97.8	50.0	100.0	97.8	50.0	100.0	97.8	50.0	100.0	97.8	50.0	100.0	97.8	50.0	100.0
	8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.8	73.9	100.0	97.8	73.9	100.0	97.8	73.9	100.0	97.8	73.9	100.0	97.8	73.9	100.0	97.8	73.9	100.0	97.8	73.9	100.0	97.8	73.9	100.0	97.8	73.9	100.0	97.8	73.9	100.0
	9	100.0	97.8	100.0	100.0	97.8	100.0	100.0	97.8	100.0	100.0	97.8	100.0	97.8	63.0	100.0	97.8	63.0	100.0	97.8	63.0	100.0	97.8	63.0	100.0	97.8	63.0	100.0	97.8	63.0	100.0	97.8	63.0	100.0	97.8	63.0	100.0	97.8	63.0	100.0	97.8	63.0	100.0
	10	100.0	85.4	68.8	97.9	95.8	93.8	97.9	97.9	100.0	97.9	62.5	100.0	97.9	62.5	100.0	97.9	62.5	100.0	97.9	62.5	100.0	97.9	62.5	100.0	97.9	62.5	100.0	97.9	62.5	100.0	97.9	62.5	100.0	97.9	62.5	100.0	97.9	62.5	100.0	97.9	62.5	100.0
b) <i>Reversión a la media</i>	1	100.0	93.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	63.3	100.0	100.0	63.3	100.0	100.0	63.3	100.0	100.0	63.3	100.0	100.0	63.3	100.0	100.0	63.3	100.0	100.0	63.3	100.0	100.0	63.3	100.0	100.0	63.3	100.0	100.0	63.3	100.0			
	2	100.0	85.7	80.0	100.0	89.3	90.0	96.4	96.4	90.0	100.0	53.6	90.0	100.0	53.6	90.0	100.0	53.6	90.0	100.0	53.6	90.0	100.0	53.6	90.0	100.0	53.6	90.0	100.0	53.6	90.0	100.0	53.6	90.0	100.0	53.6	90.0	100.0	53.6	90.0			
	3	100.0	96.7	90.0	100.0	96.7	100.0	100.0	100.0	100.0	100.0	73.3	100.0	100.0	73.3	100.0	100.0	73.3	100.0	100.0	73.3	100.0	100.0	73.3	100.0	100.0	73.3	100.0	100.0	73.3	100.0	100.0	73.3	100.0	100.0	73.3	100.0	100.0	73.3	100.0			
	4	100.0	96.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	71.4	100.0	100.0	71.4	100.0	100.0	71.4	100.0	100.0	71.4	100.0	100.0	71.4	100.0	100.0	71.4	100.0	100.0	71.4	100.0	100.0	71.4	100.0	100.0	71.4	100.0	100.0	71.4	100.0			
	5	100.0	93.3	100.0	100.0	96.7	100.0	100.0	100.0	100.0	100.0	76.7	100.0	100.0	76.7	100.0	100.0	76.7	100.0	100.0	76.7	100.0	100.0	76.7	100.0	100.0	76.7	100.0	100.0	76.7	100.0	100.0	76.7	100.0	100.0	76.7	100.0	100.0	76.7	100.0			
	6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	38.5	100.0	100.0	38.5	100.0	100.0	38.5	100.0	100.0	38.5	100.0	100.0	38.5	100.0	100.0	38.5	100.0	100.0	38.5	100.0	100.0	38.5	100.0	100.0	38.5	100.0	100.0	38.5	100.0			
	7	100.0	88.5	100.0	100.0	96.2	88.9	96.2	96.2	100.0	100.0	76.9	77.8	100.0	76.9	77.8	100.0	76.9	77.8	100.0	76.9	77.8	100.0	76.9	77.8	100.0	76.9	77.8	100.0	76.9	77.8	100.0	76.9	77.8	100.0	76.9	77.8	100.0	76.9	77.8			
	8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	62.5	100.0	100.0	62.5	100.0	100.0	62.5	100.0	100.0	62.5	100.0	100.0	62.5	100.0	100.0	62.5	100.0	100.0	62.5	100.0	100.0	62.5	100.0	100.0	62.5	100.0	100.0	62.5	100.0			
	9	100.0	100.0	85.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	68.2	57.1	100.0	68.2	57.1	100.0	68.2	57.1	100.0	68.2	57.1	100.0	68.2	57.1	100.0	68.2	57.1	100.0	68.2	57.1	100.0	68.2	57.1	100.0	68.2	57.1	100.0	68.2	57.1			
	10	100.0	77.3	85.7	100.0	90.9	85.7	95.5	95.5	85.7	100.0	50.0	28.6	100.0	50.0	28.6	100.0	50.0	28.6	100.0	50.0	28.6	100.0	50.0	28.6	100.0	50.0	28.6	100.0	50.0	28.6	100.0	50.0	28.6	100.0	50.0	28.6	100.0	50.0	28.6			
c) <i>Volatilidad</i>	1	92.9	89.3	82.8	94.1	88.1	86.2	92.9	92.9	89.7	98.8	67.9	89.7	98.8	67.9	89.7	98.8	67.9	89.7	98.8	67.9	89.7	98.8	67.9	89.7	98.8	67.9	89.7	98.8	67.9	89.7	98.8	67.9	89.7	98.8	67.9	89.7	98.8	67.9	89.7			
	2	99.0	91.0	88.2	99.0	91.0	97.1	95.0	95.0	100.0	99.0	61.0	100.0	99.0	61.0	100.0	99.0	61.0	100.0	99.0	61.0	100.0	99.0	61.0	100.0	99.0	61.0	100.0	99.0	61.0	100.0	99.0	61.0	100.0	99.0	61.0	100.0	99.0	61.0	100.0			
	3	96.0	88.0	79.4	96.0	87.0	88.2	96.0	96.0	94.1	99.0	68.0	94.1	99.0	68.0	94.1	99.0	68.0	94.1	99.0	68.0	94.1	99.0	68.0	94.1	99.0	68.0	94.1	99.0	68.0	94.1	99.0	68.0	94.1	99.0	68.0	94.1	99.0	68.0	94.1			
	4	99.1	92.7	78.4	99.1	95.5	94.6	97.3	97.3	100.0	99.1	70.0	100.0	99.1	70.0	100.0	99.1	70.0	100.0	99.1	70.0	100.0	99.1	70.0	100.0	99.1	70.0	100.0	99.1	70.0	100.0	99.1	70.0	100.0	99.1	70.0	100.0	99.1	70.0	100.0			
	5	94.9	88.1	92.5	97.5	89.8	95.0	95.8	95.8	97.5	100.0	67.0	97.5	100.0	67.0	97.5	100.0	67.0	97.5	100.0	67.0	97.5	100.0	67.0	97.5	100.0	67.0	97.5	100.0	67.0	97.5	100.0	67.0	97.5	100.0	67.0	97.5	100.0	67.0	97.5			
	6	95.5	91.8	75.7	96.4	90.9	81.1	95.5	95.5	94.6	99.1	70.0	94.6	99.1	70.0	94.6	99.1	70.0	94.6	99.1	70.0	94.6	99.1	70.0	94.6	99.1	70.0	94.6	99.1	70.0	94.6	99.1	70.0	94.6	99.1	70.0	94.6	99.1	70.0	94.6			
	7	98.2	93.8	81.6	98.2	95.5	94.7	97.3	97.3	94.7	98.2	60.7	94.7	98.2	60.7	94.7	98.2	60.7	94.7	98.2	60.7	94.7	98.2	60.7	94.7	98.2	60.7	94.7	98.2	60.7	94.7	98.2	60.7	94.7	98.2	60.7	94.7	98.2	60.7	94.7			
	8	97.4	92.2	87.2	97.4	92.2	87.2	96.6	95.7	97.4	100.0	66.4	97.4	100.0	66.4	97.4	100.0	66.4	97.4	100.0	66.4	97.4	100.0	66.4	97.4	100.0	66.4	97.4	100.0	66.4	97.4	100.0	66.4	97.4	100.0	66.4	97.4						
	9	92.9	87.5	76.3	94.6	86.6	86.8	92.9	92.9	92.1	100.0	63.4	92.1	100.0	63.4	92.1	100.0	63.4	92.1	100.0	63.4	92.1	100.0	63.4	92.1	100.0	63.4	92.1	100.0	63.4	92.1	100.0	63.4	92.1	100.0	63.4	92.1						
	10	99.1	92.7	100.0	98.2	94																																					

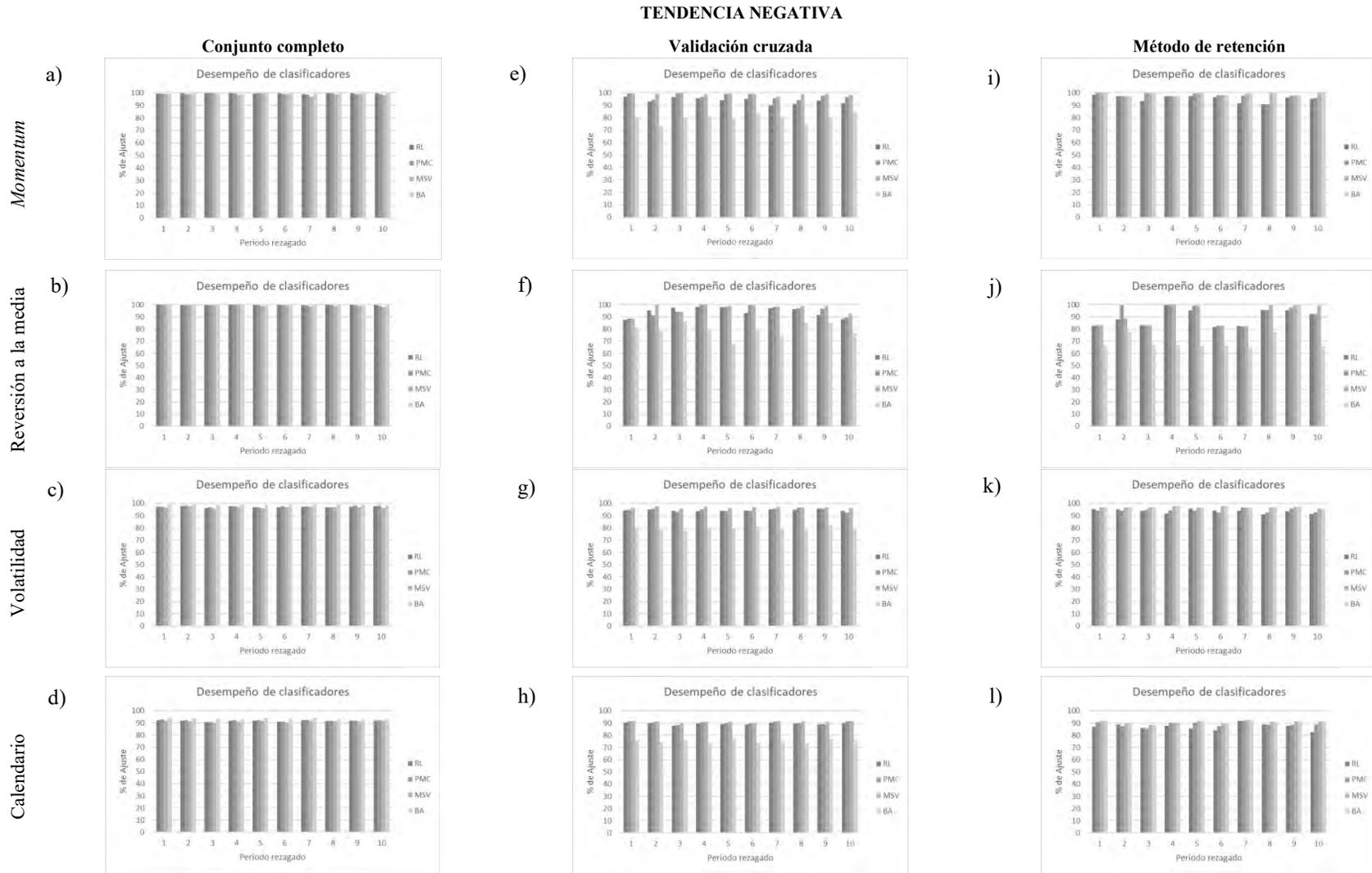


Figura B.15. Desempeño promedio de los clasificadores para el portafolio de empresas con menor bursatilidad en tendencia positiva.



Fuente: Elaboración propia.

Figura B.16. Desempeño promedio de los clasificadores para el portafolio empresas con menor bursatilidad en tendencia negativa.



Fuente: Elaboración propia.

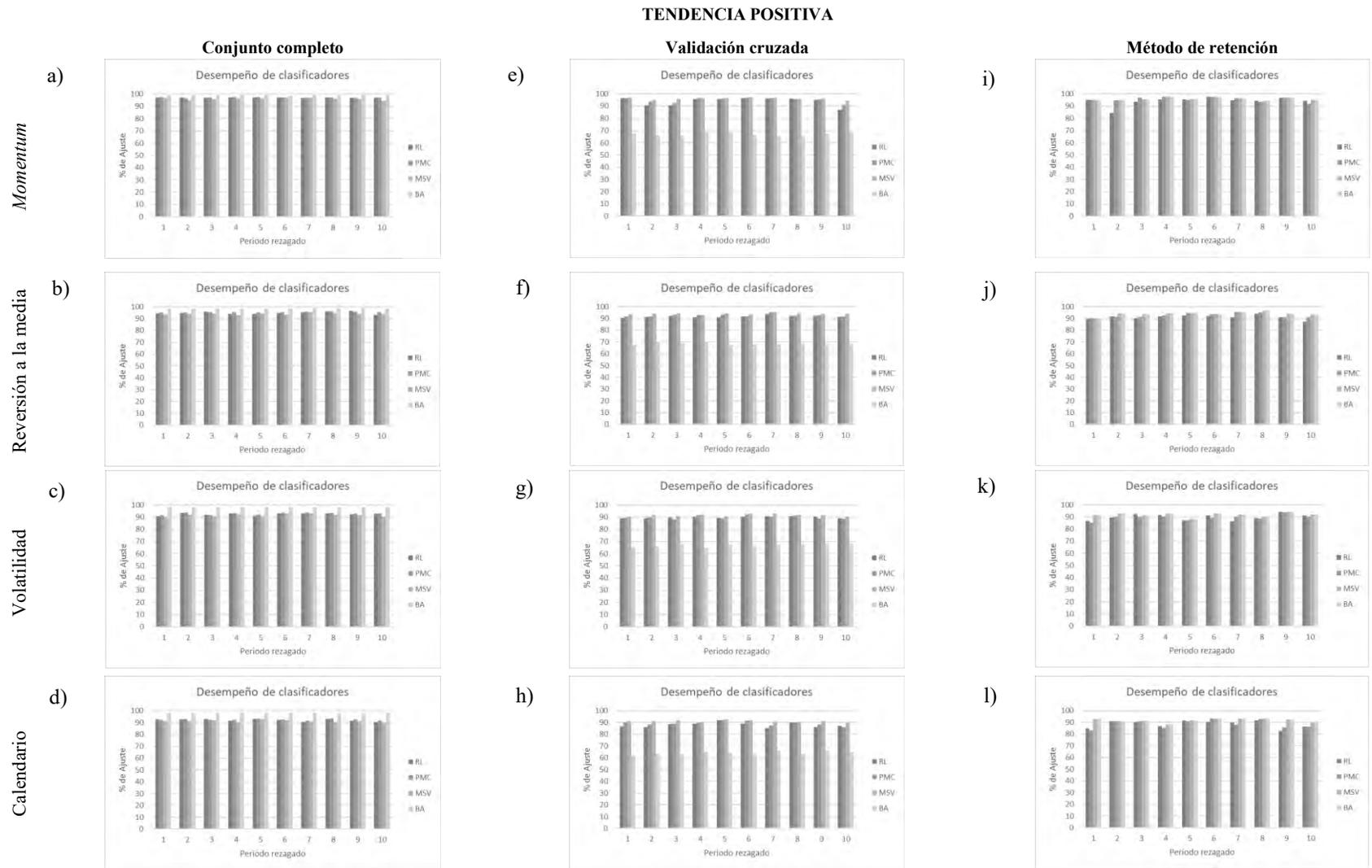
Tabla B.7. Porcentajes de ajuste de los modelos para el portafolio de empresas de mayor bursatilidad en tendencia positiva.

	Series mensuales												Series semanales												Series diarias																																																																																																																																																																																																																																																											
	Regresión Logística			Perceptrón multicapa			M. de soporte vectorial			Bosques aleatorios			Regresión Logística			Perceptrón multicapa			M. de soporte vectorial			Bosques aleatorios			Regresión Logística			Perceptrón multicapa			M. de soporte vectorial			Bosques aleatorios																																																																																																																																																																																																																																																		
Rezagos	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP	CC	VC	CP																																																																																																																																																																																																																																	
Momentum	1	100.0	98.3	95.0	100.0	98.3	95.0	98.3	98.3	95.0	100.0	48.3	95.0	1	96.1	95.9	94.2	96.5	95.6	94.2	96.1	96.1	94.9	100.0	66.4	94.9	1	94.4	94.4	94.2	94.8	94.4	93.6	94.4	94.4	94.2	99.8	68.1	94.2	2	95.5	95.5	94.3	95.5	95.0	94.3	95.5	95.5	95.6	100.0	69.2	95.6	3	95.2	94.8	96.2	96.1	94.6	94.3	95.2	95.2	96.2	100.0	68.6	96.2	4	96.6	96.6	98.2	97.0	96.4	98.2	96.6	96.6	98.2	100.0	65.6	98.2	5	95.9	95.3	96.0	96.5	95.1	96.0	95.9	95.9	97.1	100.0	66.6	97.1	6	96.2	96.2	97.2	96.6	95.9	97.2	96.2	96.2	97.2	100.0	67.7	97.2	7	95.2	95.0	93.8	95.6	94.8	93.8	95.2	95.2	93.8	100.0	67.0	93.8	8	96.4	96.0	95.5	96.4	95.3	94.4	96.2	96.2	95.5	100.0	65.4	95.5	9	94.9	94.8	94.9	95.3	94.9	94.9	94.9	94.9	95.4	99.8	67.3	95.4	10	94.8	94.6	94.9	95.0	94.4	94.3	94.8	94.8	96.0	99.8	66.8	96.0																																																																																																																								
	Reversión a la media	10	100.0	81.0	83.9	100.0	81.0	83.9	100.0	81.0	83.9	100.0	47.1	83.9	10	94.4	94.4	94.2	94.8	94.4	93.6	94.4	94.4	94.2	99.8	68.1	94.2	1	94.4	94.4	94.2	94.8	94.4	93.6	94.4	94.4	94.2	99.8	68.1	94.2	2	94.0	93.8	93.6	94.1	93.2	92.3	94.0	94.0	95.0	100.0	71.6	95.0	3	93.8	93.6	93.9	94.1	93.2	93.4	93.8	93.8	94.3	100.0	68.8	94.3	4	95.7	95.5	95.6	96.0	94.9	94.7	95.7	95.7	95.6	100.0	68.5	95.6	5	94.1	94.1	92.6	95.0	94.0	93.1	94.1	94.1	92.6	100.0	69.2	92.6	6	94.7	94.6	93.3	95.4	94.4	94.1	94.7	94.7	94.1	100.0	69.1	94.1	7	94.3	94.0	94.1	94.6	93.0	94.1	94.2	94.2	94.1	99.9	67.4	94.1	8	95.1	94.9	95.0	95.4	94.5	95.0	95.1	95.1	95.4	100.0	68.2	95.4	9	94.7	94.4	93.1	94.7	94.3	93.1	94.3	94.3	93.1	100.0	68.0	93.1	10	93.9	93.9	93.0	94.2	93.0	93.0	93.9	93.9	93.0	100.0	68.0	93.0																																																																																																																							
		Volatilidad	1	89.5	83.7	75.9	90.7	86.1	72.4	87.2	87.2	89.7	100.0	45.4	89.7	1	91.4	91.2	93.2	91.7	90.9	92.0	91.3	91.3	93.2	100.0	65.5	93.2	1	91.8	91.8	91.6	91.8	91.6	91.5	91.8	91.8	91.6	95.1	84.8	91.6	2	97.8	84.4	83.9	97.8	87.8	87.1	93.3	93.3	93.6	100.0	44.4	93.6	2	91.7	91.7	92.0	91.8	91.4	91.6	91.7	91.7	92.2	95.0	85.3	92.2	3	93.1	86.3	94.3	92.2	82.4	88.6	90.2	90.2	91.4	100.0	51.0	91.4	3	91.6	91.6	90.9	91.6	91.5	90.7	91.6	91.6	90.9	94.8	84.9	90.9	4	96.4	87.3	91.9	96.4	91.8	89.2	93.6	93.6	94.6	100.0	45.5	94.6	4	92.0	92.0	91.7	92.0	91.8	91.5	92.0	92.0	91.7	95.0	85.4	91.7	5	92.0	86.0	82.4	94.0	86.0	82.4	90.0	90.0	85.3	100.0	51.0	85.3	5	91.4	91.4	91.0	91.6	91.3	91.0	91.4	91.4	91.0	94.4	85.4	91.0	6	96.2	87.5	88.6	98.1	94.2	82.9	95.2	95.2	91.4	100.0	50.0	91.4	6	91.7	91.7	92.4	91.8	91.6	92.1	91.7	91.7	92.4	94.9	84.9	92.4	7	97.1	90.4	77.1	98.1	89.4	88.6	96.2	96.2	94.3	100.0	51.0	94.3	7	91.2	91.2	91.3	91.8	90.9	90.8	91.3	91.3	91.3	99.9	65.8	91.3	8	91.3	91.2	91.3	91.4	90.9	90.5	91.3	91.3	91.6	100.0	64.4	91.6	8	91.8	91.8	91.1	91.9	91.6	91.0	91.8	91.8	91.1	94.6	84.7	91.1	9	93.1	88.2	97.1	95.1	84.3	97.1	91.2	91.2	97.1	100.0	52.9	97.1	9	91.7	91.7	92.3	91.7	91.6	92.3	91.7	91.7	92.3	94.4	85.6	92.3	10	95.0	85.0	88.2	96.0	84.0	85.3	89.0	89.0	91.2	100.0	55.0	91.2	10	92.4	92.4	92.1	92.5	92.2	92.0	92.4	92.4	92.1	94.8	85.7	92.1	
			Calendario	1	100.0	82.0	76.5	98.0	92.0	70.6	96.0	96.0	100.0	100.0	40.0	100.0	1	88.8	88.5	88.1	89.2	88.4	87.8	88.6	88.6	88.3	99.8	64.2	88.3	1	89.4	89.4	89.8	89.6	89.3	89.7	89.4	89.4	89.9	94.0	81.0	89.9	2	100.0	80.4	93.8	100.0	87.0	93.8	95.7	95.7	93.8	100.0	43.5	93.8	2	87.5	87.5	89.7	88.5	87.0	89.2	87.7	87.7	89.7	99.8	65.0	89.7	3	100.0	86.8	91.3	97.1	88.2	95.7	97.1	97.1	95.7	100.0	44.1	95.7	3	88.8	88.5	88.2	89.4	88.2	86.8	88.6	88.6	88.2	99.9	63.3	88.2	4	95.3	87.5	81.8	96.9	90.6	77.3	92.2	92.2	86.4	100.0	48.4	86.4	4	88.4	88.3	87.2	88.7	87.8	87.2	88.4	88.4	87.2	99.8	64.6	87.2	5	100.0	96.9	95.5	100.0	96.9	95.5	98.4	98.4	95.5	100.0	46.9	95.5	5	89.2	89.1	90.2	90.1	89.0	87.9	89.2	89.2	90.2	99.8	64.3	90.2	6	98.6	88.6	91.7	98.6	97.1	100.0	97.1	97.1	100.0	100.0	42.9	100.0	6	88.3	88.2	89.0	89.1	88.1	88.6	88.3	88.3	89.2	99.9	63.9	89.2	7	91.7	76.7	90.0	95.0	83.3	85.0	93.3	93.3	100.0	100.0	53.3	100.0	7	89.2	89.0	89.0	89.9	89.3	88.8	89.1	89.1	89.3	99.4	62.2	89.3	8	100.0	91.4	95.0	100.0	91.4	100.0	93.1	93.1	100.0	100.0	48.3	100.0	8	89.2	88.9	89.3	90.3	88.5	88.4	89.1	89.1	89.3	99.4	61.2	89.3	9	96.6	79.3	70.0	98.3	86.2	80.0	94.8	94.8	100.0	100.0	53.5	100.0	9	88.5	88.4	86.7	89.3	88.0	86.7	88.5	88.5	86.7	99.4	62.8	86.7	10	93.1	82.8	80.0	96.6	79.3	80.0	91.4	91.4	90.0	100.0	51.7	90.0	10	88.1	87.7	88.5	89.0	88.0	88.3	87.8	87.8	88.7	99.5	61.2	88.7

Fuente: Elaboración propia.

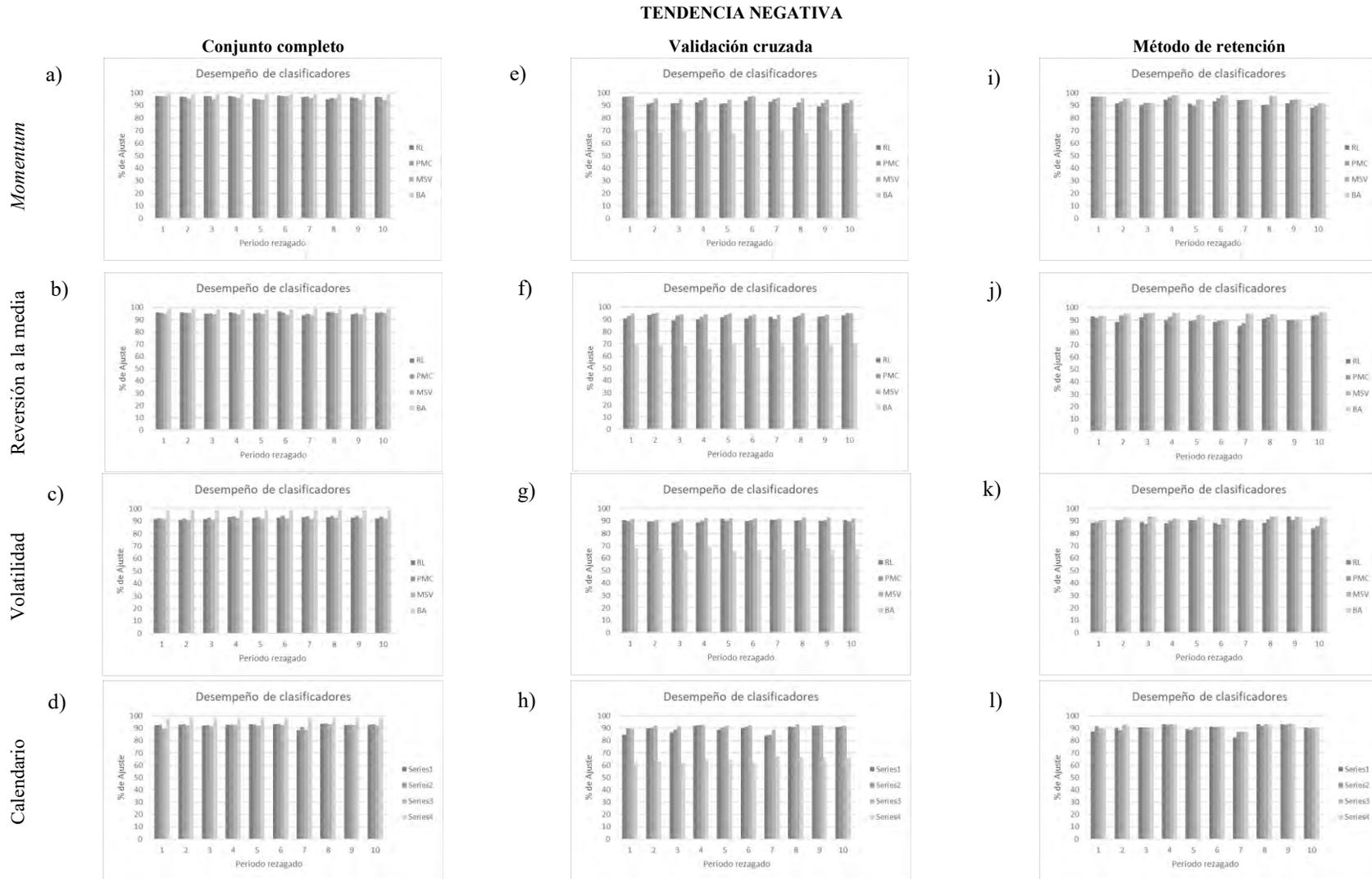


Figura B.17. Desempeño promedio de los clasificadores para el portafolio de empresas con mayor bursatilidad en tendencia positiva.



Fuente: Elaboración propia.

Figura B.18. Desempeño promedio de los clasificadores para el portafolio de empresas con mayor bursatilidad en tendencia negativa.

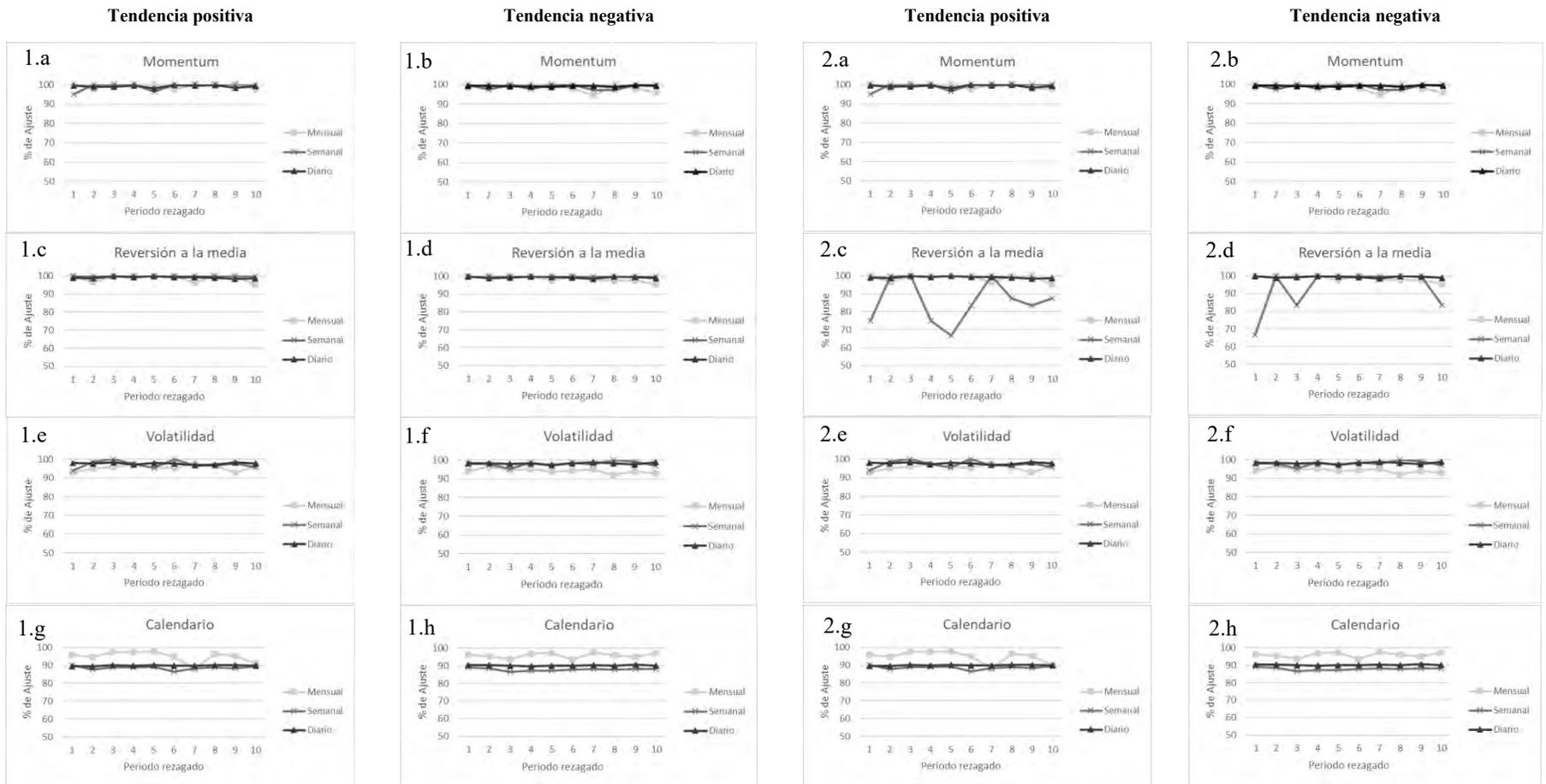


Fuente: Elaboración propia.

Figura B.19. Porcentajes de exactitud obtenido con el modelo de Máquina de soporte vectorial para el portafolio de empresas con menor bursatilidad.

1) Conjunto completo

2) Validación cruzada

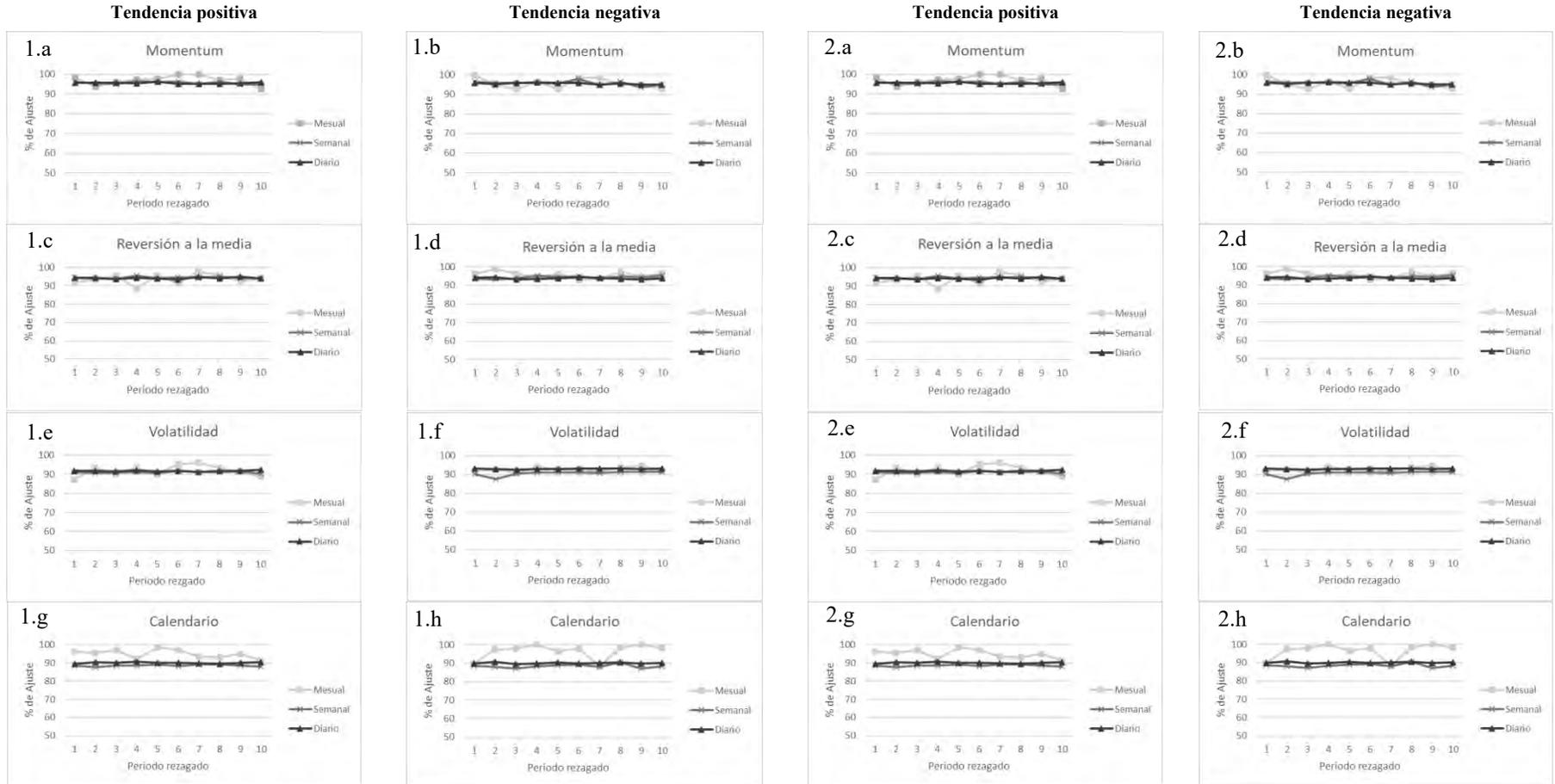


Fuente: Elaboración propia.

Figura B.20. Porcentajes de exactitud obtenido con el modelo de Máquina de soporte vectorial para el portafolio de empresas con mayor bursatilidad.

1) Conjunto completo

2) Validación cruzada



Fuente: Elaboración propia.