

Benemérita Universidad Autónoma de Puebla
Facultad de Ciencias Físico Matemáticas
Licenciatura en Actuaría



Revisión del modelo de valuación de activos CAPM y
Fama-French de tres factores: estudios de casos
en el mercado financiero mexicano y análisis comparativo.

TESIS

Que para obtener el título de:

Licenciada en Actuaría

PRESENTA:

C. Sara Josefina Ramírez Sánchez

Directores de Tesis: **Dra. Gladys Denisse Salgado Suárez (B5M - BUAP)**
Dr. José Rubén Conde Sánchez (FCFM - BUAP)

Junio 2024. Puebla, Pue. Méx.

Agradecimientos

A Dios:

Que me permitió lograrlo, me otorgó la fuerza, sabiduría, paciencia y salud para culminar con éxito este proyecto. Su bondad me ha acompañado en cada paso del camino, iluminando mi mente y fortaleciendo mi espíritu.

A mí misma:

Agradezco profundamente a mi propia persona por la perseverancia, la disciplina y el compromiso que he demostrado a lo largo de este proceso. Reconozco los retos que he enfrentado y las satisfacciones que he obtenido, todo ello como parte de mi crecimiento personal y profesional.

A mi distinguido jurado:

Expreso mi más sincero agradecimiento al Dr. Cruz Suárez Hugo Adán, al Dr. Tajonar Sanabria Francisco Solano, y al Dr. Conde Mones José Julio, por haber aceptado ser parte de mi jurado calificador. Agradezco su valioso tiempo, sus comentarios constructivos y su atenta evaluación de mi trabajo. Sus aportes han sido fundamentales para mejorar la calidad de esta tesis.

A mis asesores:

A la Dra. Gladys Denisse Salgado Suárez y al Dr. José Rubén Conde Sánchez, les extiendo mi más profundo agradecimiento por su invaluable asesoría y guía durante todo el proceso de investigación y elaboración de esta tesis. Sus conocimientos, experiencia y paciencia han sido pilares fundamentales para el desarrollo de este trabajo, gracias por no rendirse conmigo y hacer un equipo impecable.

A mi madre, Dra. Guillermina Sánchez López:

Mi güera, la razón de ser mejor cada día, gracias por tu amor, apoyo, conocimiento, educación y compañía inquebrantable, tus enseñanzas han sido la luz que ha iluminado mi camino. Gracias por creer en mí, por motivarme a perseguir mis sueños y por enseñarme que con esfuerzo y dedicación todo es posible. Este logro es tan tuyo como mío.

A mi hermana, Ana Guillermina Ramírez Sánchez:

Hanita, gracias por tu apoyo incondicional, tu alegría contagiosa y tu constante creencia en mis capacidades. Tus palabras de aliento y tu presencia en los momentos difíciles han sido un motor fundamental para seguir adelante. Te amo, hermanita.

A mi María José Rojas Peralta:

Gracias por tu compañía, amor, ayuda y cuidado incondicional. Sin ti no lo hubiera logrado, gracias por ser un pilar indispensable en mi vida y por siempre creer en mí, muchas veces más que yo misma. Te agradezco por existir.

A las instituciones, a mis maestros y a la vida:

A todas las instituciones que han contribuido a mi formación académica y profesional, a mis maestros que han compartido sus conocimientos y experiencias, y a la vida en sí misma por otorgarme un nuevo día.

Gracias a todos por ser parte de este camino.

Dedicatoria

A mi madre, la Dra. Guillermina Sánchez, mi mayor orgullo e inspiración.

Índice general

Resumen	I
1. Introducción	1
1.1. Planteamiento del problema	2
1.2. Motivación	3
1.3. Justificación	3
1.4. Objetivos	4
1.5. Metodología propuesta	5
1.6. Organización de la tesis	7
1.7. Estado del arte	7
2. Marco teórico	9
2.1. Conceptos actuariales	9
2.1.1. Valores	9
2.1.2. Portafolio de inversión	10
2.1.3. Riesgo	12
2.1.4. Rentabilidad	14
2.1.5. Retorno	15
2.1.6. Tasa de interés	16
2.1.7. Diversificación	16
2.2. Conceptos matemáticos	17
2.2.1. Variable aleatoria	17
2.2.2. Esperanza matemática	18
2.2.3. Varianza	19
2.2.4. Covarianza	20
2.2.5. Modelos de regresión	21
2.2.6. P-Value	33
2.2.7. Variable Dummy	33
2.2.8. Prueba de razón de verosimilitudes	34
2.2.9. Prueba de correlación	35
2.2.10. Modelo Lasso	36
2.2.11. Parámetro error estándar	37
2.2.12. Modelo autorregresivo integrado de medias móviles de orden p,d,q ARI- MA (p,d,q)	37

3. Modelos de valoración de activos financieros	39
3.1. Capital Asset Pricing Model (CAPM)	39
3.2. Una derivación sencilla del modelo CAPM: problema de optimización	45
3.3. El modelo de 3 factores de Fama y French	49
4. Estudios de caso	55
4.1. Caso 1: Aplicación del modelo CAPM en el mercado financiero mexicano	55
4.2. Caso 2: Aplicación del modelo CAPM con variaciones en el mercado financiero mexicano	58
4.3. Caso 3: Modelo Fama and French de tres factores	68
5. Conclusiones	79
Bibliografía	81
Apéndice A	87

Índice de figuras

1.1. Diagrama de la metodología propuesta.	6
2.1. Nube de puntos. En las abscisas, los valores de la variable independiente X y, en las ordenadas, los de la variable dependiente Y	22
3.1. Frontera eficiente sin prestar ni pedir prestado.	43
3.2. Frontera eficiente permitiendo prestar o pedir prestado.	43
3.3. Línea de mercado.	44
3.4. Maximización de la utilidad del inversionista con la existencia de un activo libre de riesgo	46

Índice de cuadros

3.1. Forma de agrupación de los portafolios respecto a la relación Book/Market [39].	52
4.1. Recopilación de los parámetros estimados: α , β_1 y β_2 de cada acción.	64
4.2. Se muestran los intervalos de confianza calculados.	67
4.3. Orden decreciente de las acciones por su relación Book/Market y capitalización bursátil.	70
4.4. Conformación de portafolios con base en la relación Book/Market y capitalización bursátil.	72
4.5. Resultados del cálculo de los rendimientos esperados de las acciones por cada modelo.	75

Resumen

La presente tesis aborda dos modelos prominentes en la valoración de activos en diversos mercados financieros: el Modelo de Precios de Activos de Capital (CAPM) y el Modelo de Fama-French de tres factores. Se desarrollan tres casos de estudio: el primero con el modelo CAPM estricto, el segundo con el modelo CAPM con modificaciones y, por último, con el modelo Fama-French de tres factores. Estos casos se despliegan a lo largo de cinco fases: la primera, recolección y limpieza de datos, a continuación, segmentación y pruebas, estas son realizadas con ayuda del lenguaje de programación estadística RStudio, como tercera fase, el reprocesamiento con los debidos ajustes para la correcta implementación e interpretación de resultados y finalmente, las dos últimas fases, análisis, valoración y presentación de los resultados.

El fundamento de la importancia de realizar estos tres casos de estudio radica en la determinación del rendimiento esperado futuro de 28 acciones pertenecientes a la Bolsa Mexicana de Valores, la observación de su contraste con los rendimientos reales futuros y la utilización de estos casos para tomar decisiones de inversión, debido a que la elección de una cartera de inversión óptima requiere tener certeza de parámetros como el riesgo y la rentabilidad de los activos. Se consideran variables como la tasa libre de riesgo (Certificados de la Tesorería de la Federación CETES), la prima del mercado, considerada por el Índice de precios y cotizaciones (IPC) de la Bolsa Mexicana de Valores, el Producto Interno Bruto (PIB) y la tasa Tasa de Interés Interbancaria de Equilibrio (TIIE).

Los modelos CAPM y Fama-French, son herramientas útiles y populares para la valoración de activos. El CAPM establece que la rentabilidad esperada de un activo está en función de su parámetro β , la prima de riesgo de mercado y la tasa libre de riesgo. Por otro lado, el modelo Fama-French de tres factores, se basa en que la rentabilidad de un activo está explicada por tres parámetros: riesgo de mercado, tamaño y valor.

Palabras Clave: Modelo CAPM, Modelo Fama-French de tres factores, activos, riesgos, IPC, PIB, TIIE.

Capítulo 1

Introducción

En el ámbito financiero, la estimación del rendimiento esperado de los activos y la toma de decisiones de inversión informadas son aspectos cruciales para los inversores y analistas. Desde los estudios pioneros de Harry Markowitz en 1952, el Modelo de Valoración de Activos de Capital (CAPM) y el modelo de Fama y French de tres factores han sido fundamentales al proporcionar una relación entre el riesgo y el rendimiento esperado de un activo [48].

Particularmente, en el contexto del mercado financiero mexicano, la aplicación del Modelo CAPM ha suscitado interrogantes sobre su precisión y relevancia, lo que ha llevado a propuestas de adaptación para mejorar su exactitud. Una de estas propuestas incluye la incorporación de variables adicionales y la sustitución de otras, como el popular modelo Fama-French de tres factores, que considera variables como el riesgo de mercado, el tamaño y el valor de la empresa correspondiente a la acción [28].

Esta tesis tiene como objetivo estudiar y obtener el rendimiento esperado de un grupo de acciones a través de la aplicabilidad del modelo de Valoración de Activos CAPM y el modelo Fama-French de tres factores en el mercado financiero mexicano por medio de la presentación de tres casos de estudio.

El primero, el modelo CAPM original, el segundo, considerando modificaciones del estudio original como la sustitución de la tasa de mercado por el índice de precios y cotizaciones (IPC), la creación de una variable instrumental y la incorporación de variables influyentes en el comportamiento del mercado financiero mexicano, esto como consecuencia de las particularidades, regularidades y características únicas que posee dicho mercado.

Por último, el tercer caso toma como referencia el modelo de Fama-French de 3 factores, el cual ha recibido percepciones favorecedoras en el sentido de explicación, robustez y aplicabilidad [54] debido a que ofrece una perspectiva complementaria al incorporar factores adicionales como el tamaño de la compañía y el valor contable-precio, lo cual puede mejorar la capacidad de explicación de las anomalías del mercado.

Para realizar satisfactoriamente el objetivo mencionado, se utiliza una muestra de 28 empresas que pertenecen a la Bolsa Mexicana de Valores durante el periodo del 01 de enero de 2014 al 01 de enero de 2023, con el fin de recopilar datos históricos necesarios de los precios de las acciones, se aprovechan recursos electrónicos como BANXICO, Yahoo Finance, Investiing y la página oficial de CETES directo.

En adición, para la conformación del caso CAPM modificado se destaca la integración del periodo de contingencia ambiental COVID 19, el cual resultó impactante para la sociedad en general y en particular para los factores financieros, este es mostrado a través de una variable Dummy.

Además, se espera que este estudio contribuya al conocimiento y comprensión teórica de estos modelos y de forma práctica por medio de la creación de los casos de estudio en el mercado financiero mexicano, además de proporcionar herramientas útiles para la toma de decisiones de inversión por parte de los integrantes del entorno financiero.

1.1. Planteamiento del problema

El análisis y la predicción del rendimiento de los activos son aspectos cruciales en el ámbito financiero, no solo para los inversionistas y analistas, sino también para la academia. El CAPM ha sido ampliamente utilizado para estimar el rendimiento esperado de un activo en relación con el riesgo sistemático asociado. Sin embargo, a pesar de su aceptación teórica, se ha cuestionado su capacidad para explicar adecuadamente el comportamiento de los mercados financieros reales [37].

Por otro lado, el modelo financiero de Fama-French de tres factores, desarrollado por Eugene Fama y Kenneth French, ha ganado popularidad al proporcionar un análisis más completo de las anomalías del mercado que el CAPM no puede explicar. Esto se logra al incorporar factores adicionales como el tamaño de la empresa y el valor contable-precio y ofrece una visión completa de los determinantes de la rentabilidad de un activo [28].

En el contexto del mercado financiero mexicano, existen características únicas que pueden influir en el comportamiento de los activos. A pesar de la popularidad del modelo CAPM y Fama-French de tres factores, su aplicación en este contexto específico, especialmente con modificaciones, no ha sido ampliamente explorada. Por lo tanto, surge la necesidad de investigar y comparar la eficacia del modelo CAPM tradicional, el CAPM modificado y el modelo de Fama-French de tres factores en el mercado financiero mexicano.

Este estudio busca cerrar esta brecha de conocimiento y proporcionar una comprensión profunda de comparación en su aplicabilidad en un contexto único como el mexicano. Por lo que se tiene la siguiente pregunta de investigación:

¿En qué medida difiere el rendimiento real de las acciones seleccionadas del mercado financiero mexicano del rendimiento esperado, cuando se utiliza el modelo CAPM (en sus versiones

tradicional y modificada) frente al modelo de tres factores de Fama-French, considerando las particularidades del contexto local?

1.2. Motivación

Los modelos financieros son indispensables para la toma de decisiones en inversiones. En particular, en el mercado financiero mexicano, el CAPM y el modelo Fama-French de tres factores son relevantes para estimar el rendimiento esperado de una acción en función de su riesgo sistemático. Además, tienen un amplio uso en el ámbito empresarial para calcular el costo de capital y tomar decisiones de financiamiento [24].

De acuerdo con los autores Rodríguez y López en [67], en los últimos años, el mercado financiero mexicano ha experimentado un crecimiento significativo, lo que ha generado una creciente necesidad de emplear herramientas óptimas que permitan a los inversores tomar decisiones de inversión informadas y analizadas. Entre estas herramientas se encuentran el modelo CAPM y el modelo de Fama-French de tres factores.

Aunque existen estudios que relacionan estos modelos con el mercado financiero mexicano como en [72] y [73], aún queda un amplio campo por explorar, especialmente en cuanto a la consideración de variables características de México a lo largo de períodos de tiempo específicos, así como la posibilidad de realizar ciertas modificaciones e incluir variables adicionales.

La motivación para este estudio radica en la importancia de proporcionar herramientas óptimas para la toma de decisiones de inversión en el mercado financiero mexicano, así como en contribuir al desarrollo y comprensión de modelos financieros como el CAPM y el modelo de Fama-French de tres factores. Asimismo, se busca generar información útil para inversores y empresas financieras, promoviendo un mayor entendimiento y aplicación de estos modelos en el contexto mexicano.

1.3. Justificación

Este trabajo busca ofrecer una serie de beneficios significativos. En primer lugar, pretende facilitar para los inversores, la capacidad de tomar decisiones de inversión óptimas basadas en proyecciones sólidas, lo que les permite tener una perspectiva amplia y confiable de los activos a considerar. En segundo lugar, al emplear los modelos financieros, se proporciona un argumento adicional para calcular los rendimientos esperados, lo que brinda seguridad al inversor al tener una opción adecuada para su perfil de riesgo.

En tercer lugar, se persigue fomentar el fortalecimiento de la cultura de inversión financiera en México mediante la exposición de tres casos de estudio aplicando los modelos a acciones de la Bolsa Mexicana de Valores. Este enfoque busca facilitar el acceso a conceptos financieros

fundamentales para el público mexicano, promoviendo así una comprensión más amplia y una participación más informada en el mercado de valores. Además, esta tesis pretende contribuir a la comprensión del modelo CAPM y el modelo de Fama-French de tres factores, añadiendo tres casos prácticos de su ajuste apropiado para su aplicación en este enfoque específico. Se espera que esta contribución proporcione información valiosa para la comunidad financiera interesada.

1.4. Objetivos

Objetivo general

Calcular el rendimiento esperado mediante la aplicación del Modelo CAPM en su forma original, el Modelo CAPM con modificaciones, y el Modelo de Tres Factores de Fama-French en el contexto del mercado financiero mexicano.

Objetivos particulares

- Realizar una revisión de la literatura relacionada con los modelos financieros y su aplicación en algún otro país de Latinoamérica, con el fin de obtener un amplio conocimiento sobre los análisis e implementaciones existentes en el campo.
- Seleccionar y preparar adecuadamente los datos necesarios para llevar a cabo tres casos de estudio sobre la aplicación de los modelos CAPM y Fama-French de tres factores en el mercado financiero mexicano, asegurando la inclusión de variables relevantes para un ajuste preciso de los modelos.
- Desarrollar y validar un código en el lenguaje de programación RStudio que permita la extracción de datos, la realización de pruebas estadísticas, la presentación y comparación de resultados de manera eficiente y precisa.
- Calcular los parámetros β para cada modelo, lo que permitirá su uso en el cálculo del rendimiento esperado de las acciones y facilitará la comparación entre los modelos CAPM y Fama-French de tres factores en el contexto específico del mercado financiero mexicano.
- Analizar y documentar cuidadosamente los resultados obtenidos en los casos de estudio, presentando conclusiones claras y fundamentadas que contribuyan a una mejor comprensión de la efectividad de los modelos CAPM y Fama-French de tres factores en el mercado financiero mexicano y a la toma de decisiones de inversión informadas en este contexto.

1.5. Metodología propuesta

Para lograr el objetivo planteado, se propone una metodología de 5 fases:

Fase 1. Recolección y limpieza de datos:

- Obtención de los precios de las acciones cotizadas en la Bolsa Mexicana de Valores desde el 01 de enero de 2014 al 01 de enero de 2023.
- Selección de 28 acciones para los casos de estudio, junto con la recopilación de datos sobre variables relevantes como: Tasa de interés de los Certificados de la Tesorería de la Federación (CETES), el índice de Precios y Cotizaciones (IPC), el tipo de cambio, la circulación de la moneda metálica de 10 pesos, la tasa de interés interbancaria (TIIE), el índice Standard & Poor's 500 (S&P 500) y los valores del Producto Interno Bruto (PIB).

Fase 2. Segmentación y pruebas:

- Realización de pruebas para evaluar la idoneidad del PIB como variable instrumental en el contexto del mercado mexicano.
- Estimación de modelos de regresión para analizar la relación entre rendimientos de acciones y variables económicas importantes, considerando también, para el segundo caso, una variable *dummy* para un evento significativo como la pandemia COVID-19.
- Cálculo de parámetro β utilizando los modelos de valoración de activos y elaboración de pruebas estadísticas para evaluar la significancia de los resultados.
- Segmentación de acciones según capitalización bursátil y relación Book/Market, seguido del cálculo de las variables Small Minus Big (SMB) y High Minus Low (HML), así como la prima de mercado y la tasa libre de riesgo diaria.

Fase 3. Reprocesamiento:

- Identificación y manejo de la multicolinealidad entre variables independientes mediante el modelo Least Absolute Shrinkage and Selection Operator (LASSO).
- Estimación de tres diferentes percentiles para el tercer caso, 50 % para la capitalización bursátil, 33 % y 66 % para la relación Book/Market.

Fase 4. Análisis de resultados:

- Interpretación estadística y econométrica de los modelos estimados y las pruebas realizadas, evaluando la significancia de los coeficientes y el ajuste del modelo.
- Cálculo de los intervalos de confianza.
- Cálculo del error porcentual para comparar el rendimiento esperado calculado con el real y evaluar la precisión de los modelos.

Fase 5. Valoración y presentación de resultados:

- Evaluación de la importancia práctica de los hallazgos obtenidos, considerando su impacto financiero, la toma de decisiones y su consistencia con evidencias previas.
- Identificación de limitaciones y áreas de mejora para futuras investigaciones.

La metodología propuesta se centra en la aplicación de los modelos a una cartera de acciones de la Bolsa Mexicana de Valores. Esta metodología implica la utilización del software estadístico *RStudio*, que permite modelar, llevar a cabo pruebas estadísticas, obtener datos financieros y realizar consultas. Además, analiza operaciones econométricas y estadísticas en su rango de posibilidades de uso.

Dicha metodología inicia con la recolección y limpieza de datos, en la cual se observa la tendencia del precio de las acciones, el periodo de estadía en la bolsa y la continuidad de precios, además, se considera la interpolación de los valores faltantes para obtener el periodo completo de los datos requeridos, seguido de la aplicación de pruebas estadísticas y los modelos CAPM, CAPM con modificaciones y Fama-French, finaliza con el análisis, interpretación, valoración y presentación de los resultados. La metodología propuesta se encuentra en la Figura 1.1.

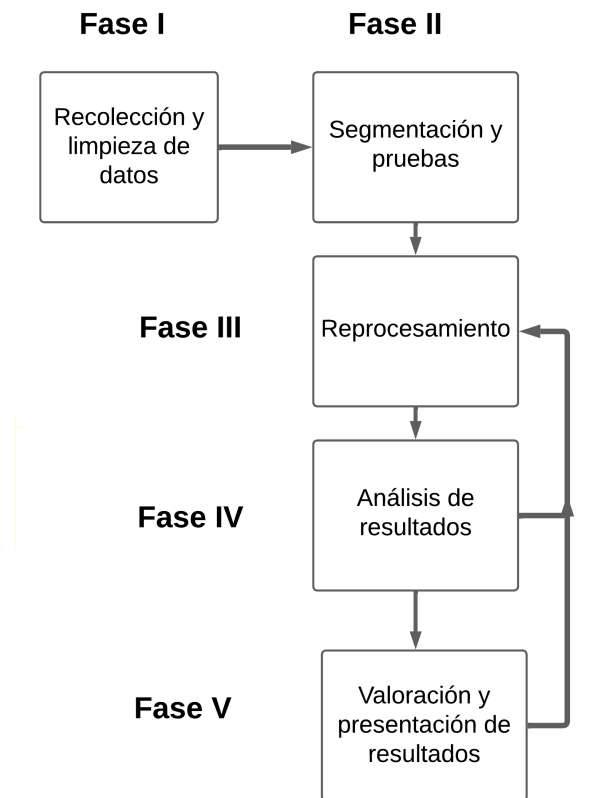


Figura 1.1: Diagrama de la metodología propuesta.

1.6. Organización de la tesis

La presente tesis se encuentra organizada en 6 capítulos, cuyo contenido se describe a continuación:

- En el Capítulo 1 se encuentra la introducción, la cual presenta la motivación, justificación, los objetivos, el diagrama a bloques propuesto para el desarrollo de la tesis y el estado del arte.
- En el capítulo 2 se explican los conceptos necesarios para la comprensión del trabajo y las herramientas requeridas para desarrollarlo.
- En el Capítulo 3 se externan los modelos financieros CAPM y Fama-French.
- En el Capítulo 4 se muestra la aplicación de los tres modelos en los casos de estudios realizados, centrándose en los rendimientos esperados de la cartera de acciones.
- En el Capítulo 5 se presentan las conclusiones pertinentes.

1.7. Estado del arte

En esta sección, se presenta una revisión crítica y actualizada de la literatura sobre los modelos financieros, en particular, los modelos CAPM y Fama-French. Nuestro objetivo es identificar las áreas de oportunidad en el conocimiento para realizar una línea de investigación útil.

Los modelos CAPM y Fama-French son ampliamente utilizados en el campo de las finanzas para estimar el rendimiento esperado de las acciones en función de su riesgo [24].

En primer lugar, la teoría del modelo CAPM se basa en el equilibrio de mercado debido a que se introducen los conceptos de rendimiento libre de riesgo y prima de riesgo, que es proporcional al parámetro β de la acción [12].

Black, Jenden y Scholes en [12], emplean la metodología de series temporal y de corte transversal para realizar la implementación del modelo CAPM, además, incluyen excesos de rentabilidad sobre la tasa libre de riesgo, como conclusión, obtienen una constante significativamente positiva y un coeficiente de β cuantiosamente inferior a la prima de riesgo de rentabilidad media del mercado.

Las autoras Común y Huaman en [21], realizan una propuesta de investigación sobre el modelo y distintas pruebas empíricas, evaluaciones y aplicaciones del modelo en diferentes contextos financieros, particularmente en modelos emergentes, además de tener un enfoque mixto con características cuantitativas y cualitativas para la investigación.

Por su parte, Lázaro en [45], elabora un análisis del Modelo de Valoración de Activos con cartera de mercado y los modelos derivados, como el modelo de 3 factores de Fama y French, el cual compara la habilidad explicativa de los factores de riesgo de cada modelo, concluyendo que el modelo CAPM tiene suma popularidad e importancia, y el modelo de 3 factores muestra una gran capacidad explicativa en comparación con otros modelos analizados, como es el modelo Carhart.

Existen diversos estudios de aplicación del modelo CAPM al mercado financiero mexicano. Por ejemplo, los autores Botello y Guerrero en [16] calculan el modelo de Valoración de Activos Financieros para Colombia antes y después de la implementación de las Normas Internacionales de Información Financiera (NIIF), estimando los coeficientes β desapalancados y apalancados para cada entidad bancaria, con el fin de evaluar el riesgo de los inversionistas basado en información contable, obtuvieron resultados significativos, los coeficientes β apalancados fueron un 15 y 20 por ciento mayores que los coeficientes sin apalancar, lo que indica que el riesgo financiero en los bancos colombianos es positivo.

Por otra parte, el autor Mateo en [50], aplicó el modelo CAPM a activos del IBEX 35 durante un periodo de cinco años, analizando rendimientos esperados y observados, estimando los parámetros β y α de las compañías estudiadas, concluyen que β es crucial en el modelo CAPM y permite clasificar los activos en ofensivos ($\beta > 1$) y defensivos ($\beta < 1$), en la actualidad, este modelo sigue siendo el más utilizado por empresas y profesionales.

Respecto a las limitaciones del modelo de valuación de activos CAPM, se consideran argumentos de autores como Hasan, Fama y Barrezueta en [11], [28] y [37] los cuales indican que existen ciertas limitaciones que tienen como consecuencia resultados que carecen de cuantiosa precisión, algunos de los hallazgos más importantes presentados por los autores son: resulta complicado (en ocasiones se menciona imposible) identificar y, por lo tanto, medir el portafolio de mercado debido a la variación y fluctuación que presentan los títulos en cuestión de periodos de tiempo, huecos de información y permanencia en la Bolsa Mexicana de Valores, como solución a este punto, usualmente se trabaja con índices que proporcionen información de los movimientos más importantes del mercado financiero.

Un argumento adicional es la prueba empírica de Black, Jensen y Scholes en [12], la cual muestra que el coeficiente β que se estima por el modelo CAPM no es ideal para representar la realidad, por ejemplo, existen activos con un β menor a 1 que presentan una mayor volatilidad que el mercado.

Asimismo, se asevera que utilizar la prima de mercado como única variable exógena es insuficiente debido a que el riesgo de un título es afectado por diversos factores además de sus precios históricos y el mercado. También no permite considerar factores importantes en materia de valoración de empresas, por ejemplo el precio del activo, el tamaño y crecimiento esperado de las compañías, y variables como el PIB y la tasa TIEE [76].

Capítulo 2

Marco teórico

El siguiente capítulo de la presente tesis, tiene como objetivo principal establecer las bases teóricas que sustentan la investigación. Se presenta una revisión exhaustiva de la literatura existente sobre los conceptos básicos que construyen el modelo y la aplicación de este, con el fin de contextualizar la investigación y proporcionar un marco de referencia sólido para el tratamiento de los datos.

2.1. Conceptos actuariales

En esta sección, se explican los conceptos actuariales necesarios para establecer un contexto del entorno actuarial y financiero a utilizar.

2.1.1. Valores

Según el autor Gup en [36], el término “inversión” se refiere a la asignación de fondos con el propósito de adquirir valores o bienes que generen utilidad o interés. Por otro lado, de acuerdo con el autor Amezcua en [8], existe una distinción notable entre los bienes reales y los valores. El primero abarca objetos tangibles como bienes raíces, joyas y vehículos, mientras que el segundo se refiere a derechos vendidos, como acciones y participaciones en sociedades. En este sentido, las acciones se dividen en dos categorías principales:

- **Valores negociables:**

Se pueden realizar transacciones con ellos en la bolsa de valores o en mercados informales, estos están conformados por casas de bolsa y corredores, por otra parte, es posible realizarlas a través de mercados conocidos u organizados, los cuales pueden sustituirse como mercados continuos debido a que realizan transacciones con sus acciones continuamente.

Algunas características importantes del mercado continuo son: frecuencia de ventas (desplazamiento de acciones), dispersión estrecha (diferencia entre el precio que sugiere el vendedor y el ofrecido por el comprador), cambios mínimos en precios (a menor dispersión, menor variación del precio), ejecución rápida (en cuanto se tome la decisión de realizar alguna transacción con la acción, ponerla en el mercado) y liquidez (capacidad de transformar los valores en dinero en un plazo de tiempo corto).

Como es posible notar, invertir en valores negociables tiene ventajas como la facilidad de encontrar información actual sobre el precio de las acciones, invertir con distintos montos, los mínimos costos de comisión con respecto a otro tipo de bienes y regulación de seguridad.

- Valores no negociables:

Se encuentran en el mercado como valores intangibles, por ejemplo, valores tipo depósito o cuenta de ahorros en un banco, las ventajas que contienen son: cuentas respaldadas por el gobierno en su mayoría, cuentas muy líquidas o relativamente líquidas dependiendo del plazo establecido y la nula restricción de las tasas de interés ofrecidas, no siendo así con las tasas cobradas.

2.1.2. Portafolio de inversión

Para comenzar, la teoría del portafolio fue desarrollada en 1952 por Markowitz [48], en esta se detallaba la razón de elección de portafolios con la observación de maximizar el rendimiento esperado a un cierto nivel de riesgo, o equivalentemente, minimizar el riesgo a un nivel esperado de rendimiento.

De otro modo, los inversionistas elegirían los activos de un portafolio por su popularidad o favoritismo, sin embargo, un portafolio de inversión óptimo requiere de activos eficientes que ofrezcan el máximo rendimiento esperado a un nivel de riesgo mínimo con un cierto grado de rentabilidad, no siempre populares o en tendencia para el inversionista o asesor financiero.

En términos matemáticos, un portafolio de inversión se puede definir como una colección de activos financieros, donde cada activo tiene una proporción específica del valor total del portafolio. Supongamos que un portafolio contiene n activos [49].

Definimos:

x_i : El peso del activo i en el portafolio, donde $i = \{1, 2, \dots, n\}$,

r_i : Retorno o rentabilidad esperada del activo i , donde $i = \{1, 2, \dots, n\}$,

x : Vector de pesos del portafolio $x = [x_1, x_2, \dots, x_n]$,

r : Vector de retornos esperados $r = [r_1, r_2, \dots, r_n]$.

El retorno esperado de un portafolio R_p se calcula como la suma ponderada de los retornos

esperados de los activos individuales:

$$R_p = x^T r = \sum_{i=1}^n x_i r_i,$$

En donde:

- R_p es el retorno promedio que se espera obtener de todo el portafolio de inversión.
- x^T es el vector transpuesto de los pesos de cada activo en el portafolio. La transposición se refiere a que estamos cambiando el vector de filas a columnas para que pueda multiplicarse adecuadamente con el vector de retornos esperados.

Como bien afirman Gysel y Samano en [8], los portafolios de inversión, ocasionalmente llamados carteras de inversión, son conjuntos de activos o valores en los cuales una persona o asesor financiero decide emplear su capital. Esto se logra a través de una elección óptima, es decir, es necesario conocer los instrumentos existentes en el mercado, el riesgo máximo que está dispuesto a afrontar y el rendimiento que aspire a lograr.

Se clasificarán los activos en cuatro tipos:

- Instrumentos de deuda o renta fija: Se conoce y fija el rendimiento (por ejemplo, los CETES), se establece un plazo, puede ser corto, mediano o largo.
- Instrumentos de renta variable: El rendimiento depende de distintos factores endógenos y exógenos, por lo cual no es fijo (por ejemplo, las acciones), usualmente se establece a mediano y largo plazo, sin embargo, no es indispensable que tengan un plazo definido.
- Productos derivados: Son activos complejos que protegen al inversionista de riesgos como el alza de precios de productos y las tasas de interés.
- Metales: Los más conocidos son oro y plata, generalmente se adquieren a largo plazo.

Para elegir el portafolio óptimo, además de las características propias del inversionista, se analiza: el plazo al que se quiere invertir, la liquidez deseada, el objetivo de la inversión, preferencia sobre los activos (deuda, renta variable, futuros, derivados, metales, bienes raíces, fibras, etc.) y el riesgo que sea capaz de afrontar.

Al mismo tiempo, es necesario que el inversionista busque alternativas que le ofrezcan mínimamente una tasa de rendimiento superior a la inflación para conservar el valor de su inversión en el tiempo, además de estar consciente de la relación que existe entre riesgo y rendimiento. De acuerdo con el autor Rodrik en [69], la diversificación económica se refiere a la estrategia de acrecentar la gama de productos, servicios o actividades en las que una empresa, industria, persona, o país participa.

Tipos de portafolios de inversión

Los autores Durango y Delgado en [25], enfatizan sobre la agrupación de los portafolios con base en el grado de riesgo al que los inversionistas se arriesguen a experimentar (Perfil del inversionista):

- Perfil conservador: Su principal objetivo es obtener rendimientos estables sin perder el valor de su inversión en el tiempo, por lo cual, simpatizan con un bajo grado de rentabilidad y una minimización del riesgo, de acuerdo con Gonzalez en [34], un ejemplo de este perfil es un bono corporativo, por ejemplo los CETES, debido a que su rendimiento es ligeramente superior a la inflación y mantiene el valor del dinero en el tiempo.
- Perfil moderado: Para la autora Alas en [4], la clase de inversionistas con este perfil busca un equilibrio riesgo-rentabilidad, tienden a ser cautelosos, sin embargo arriesgan un poco más que los inversionistas con perfil conservador y un poco menos que los inversionistas arriesgados.

Con el fin de incrementar utilidades, no se inquietan por perder un pequeño porcentaje de capital invertido, un ejemplo en el cual existen activos de renta fija y variable en porcentajes diferentes (productos de Certificados de Depósito a Término (CDTs) y acciones, locales o internacionales) es un portafolio diversificado.

- Perfil arriesgado: Jaramillo en [44] asevera que los inversionistas con este perfil están dispuestos a sacrificar una disminución del riesgo, a cambio de mayor rentabilidad. Tienen un nivel de cobertura ante el riesgo mínimo, por lo que es posible perder todo o un segmento importante de la inversión. En su mayoría, tienen un nivel bajo de endeudamiento a largo plazo y se inclinan por portafolios que incluyan renta variable, derivados y moneda extranjera en un periodo de tiempo corto.

2.1.3. Riesgo

De acuerdo con el autor Pascal, el riesgo se define como una amenaza, peligro o incertidumbre a la que se ve enfrentado un inversionista por un evento u acción (puede ser del mercado o ajeno a este), en la cual existe una afectación financiera a su portafolio de manera negativa o positiva [64].

En el 2011, autores como Zvie y Kané en [13], reflexionan que la palabra riesgo tiene la misma antigüedad que la existencia humana, este concepto permite definir la probabilidad de perder algo (incluido un individuo) o de tener un resultado negativo, no deseado o dañino.

Ahora bien, al conocer la definición, surge la inquietud de saber cómo puede cuantificarse. Según Gysel y Samano en [8], el concepto de riesgo se puede representar con la varianza de los rendimientos de dicho instrumento. Estos dos conceptos tienen una equivalencia proporcional, por lo que a menor desviación, menor riesgo y viceversa, a mayor desviación, mayor riesgo.

Matemáticamente, de acuerdo con Hull en [42], la varianza mide la dispersión de los retornos de una inversión en torno a su media, se calcula como el promedio de las diferencias al cuadrado entre cada retorno y la media de los retornos.

Para una serie de retornos $r_1, r_2, \dots, r_n$, la varianza σ^2 se calcula como:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (r_i - \bar{r})^2,$$

donde:

- r_i es el retorno en el periodo i .
- $\bar{r}$ es el retorno promedio.
- n es el número de periodos.

La desviación estándar es la raíz cuadrada de la varianza y proporciona una medida de la dispersión de los retornos en las mismas unidades que los retornos, lo que la hace más intuitiva, esto es,

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (r_i - \bar{r})^2}.$$

Sin embargo, el riesgo tiene otros factores de los cuales depende además del riesgo individual de los valores que forman la cartera, uno de ellos es la relación que existe entre todos los elementos, para cuantificar esta relación es posible utilizar la covarianza de los posibles rendimientos de valores implicados.

Un portafolio de inversión posee cuantiosos riesgos, a la suma de todos ellos se le conoce como *riesgo total*, particularmente existen dos sumandos: Riesgo sistemático y Riesgo no sistemático.

El riesgo sistemático es un generalizador ya que afecta a todos los activos financieros, por ejemplo, disturbios graves causados por el hombre (atentados terroristas) como el sucedido el 11 de septiembre de 2001, desastres naturales, etc. Como es de esperarse, no se puede mitigar este riesgo, sin embargo, es posible cuantificarlo a través de la letra griega β .

Dentro de los riesgos sistemáticos, existen diversas fuentes, como es de acuerdo con Amezcua en [8], el riesgo de mercado, debido a la volatilidad del activo, pueden existir variaciones en el precio de mercado de los activos. Cabe recalcar que, aun cuando el IPC tiene variaciones, no todos los activos las tendrán proporcionalmente. Algunos de los factores que tienen la capacidad de modificar los valores de mercado son: actividad de los negocios (un incremento de las utilidades representará un incremento en el precio de mercado de sus acciones), inflación (a corto plazo, una alza de inflación representará menos liquidez y dividendos) y psicología del inversionista (tendencia de acciones favoritas cambiante).

El riesgo no sistemático, también llamado *riesgo residual*, afecta únicamente a un activo, por ejemplo, al tener un problema interno en la empresa, las acciones pertenecientes a la compañía pueden sufrir un decremento en su valor.

Una fuente común de este tipo de riesgo incluye al riesgo comercial, el cual determina la probabilidad de quiebra de negocio (pérdida total de acciones). La primera parte es insolvencia, provocada por problemas financieros internos, la segunda son medidas gubernamentales que se refieren a la regulación política de precios, cantidades, etc.

2.1.4. Rentabilidad

La rentabilidad puede considerarse, a nivel general, como la capacidad o aptitud de la empresa de generar un excedente a partir de un conjunto de inversiones efectuadas [7].

Forero y Bohórquez en [32], establecen que el principal indicador de rentabilidad es el ROA (Rentabilidad Operativa del Activo), definido por la ecuación (2.1):

$$ROA = \frac{\text{Utilidad neta : } UN}{\text{Activos totales : } AT} \quad (2.1)$$

El cual refleja la capacidad de generar ingresos en el sector económico específico de la empresa. En este sentido, la empresa busca aumentar su actividad económica utilizando activos con el fin de impulsar las ventas y, por ende, aumentar las utilidades.

De acuerdo con Álvarez en [7], debido a esta razón, se puede aseverar que la rentabilidad es una puntualización del resultado conseguido a partir de una actividad económica de transformación, de producción, y/o de intercambio. El excedente aparece en el periodo final del intercambio. Es por ello que el cálculo del resultado obtiene una significación concreta en tanto se coteja con los factores implicados para su obtención: los recursos económicos y financieros. El autor asegura que la distinción entre recursos económicos y financieros conforma la estructura fundamental sobre la rentabilidad financiera y económica, otra forma de definir este concepto es la relación que existe entre el resultado obtenido y los medios que se utilizan para ejecutar dicho resultado (capitales-económicos o financieros-empleados). Una manera lógica de calcularla es: (Resultado o excedente)/ Inversión.

En el numerador se pueden considerar conceptos como a) valor añadido; b) resultado de la explotación; c) cash-flow bruto; d) beneficio procedente de actividades ordinarias; e) beneficio neto (después de impuestos); f) dividendos. Mientras que en el denominador existen diferentes conceptos, no siempre relacionados, que pueden ser utilizados: a) Activo total; b) Activo fijo bruto; c) Activo fijo neto; d) Fondos propios; e) Pasivo fijo (fondos propios + deudas a largo plazo); f) Capitalización bursátil (valor de mercado).

Debido a la conformación de los componentes, es posible mencionar que el concepto de la rentabilidad es complejo debido en primer lugar, a que contiene elementos técnicos, económicos

y financieros y en segundo, a la falta de identificación de factores que puedan repercutir en la variación de los niveles de rentabilidad. Existen dos tipos de rentabilidad:

- Rentabilidad económica: para calcularla, se relaciona el resultado alcanzado por la empresa (independientemente de la raíz de los recursos financieros implicados) y los activos que se utilizan para lograr el resultado, es decir, $\text{Rentabilidad económica} = \text{resultado del periodo} / \text{activo actual}$.

Equivalentemente, utilizando esta razón y sustituyendo el término 'resultado del periodo' con el término 'explotación' (actividad económica propia de la empresa), existen diversas formas en las que los componentes de la razón pueden ser sustituidos para evaluar la procedencia de la rentabilidad, por ejemplo, con el margen de beneficio en función de la explotación y ventas netas o intercambiar el concepto del denominador por la cifra de producción, entre otras.

- Rentabilidad financiera: siguiendo con la idea de Álvarez en [7], las inversiones realizadas en la empresa por los propietarios se calculan a través de la suma total de los fondos propios existentes al final del ejercicio, en otras palabras, $\text{Rentabilidad financiera} = \text{resultado neto} / \text{fondos propios}$.

De manera similar, estos elementos pueden descomponerse en diferentes razones. Es importante destacar que la rentabilidad financiera es el resultado del margen de beneficio, que mide el rendimiento global de la empresa a través del resultado final obtenido, y la rotación de los fondos propios, que evalúa la eficiencia de la utilización del capital propio en relación con la cifra de negocio alcanzada.

2.1.5. Retorno

Para comenzar a definir el concepto de retorno, podemos establecer que el autor Damodaran en [23] expresa que el retorno supuesto sobre cualquier activo que no sea libre de riesgo se puede expresar como la suma del retorno que ofrece un activo libre de riesgo más una cantidad adicional que contrarreste los riesgos asumidos. Los retornos financieros se refieren a las ganancias o pérdidas generadas por una inversión durante un período de tiempo específico. Estos retornos son una medida clave para evaluar el rendimiento de activos financieros como acciones, bonos, fondos mutuos u otros instrumentos de inversión.

Zvie y Kané en [14] mencionan que la manera más sencilla de expresar el retorno financiero es la fórmula básica para calcular el retorno de un activo en particular:

$$R = \frac{V_f}{V_i} \times 100\%, \quad (2.2)$$

en la cual:

R es el retorno porcentual.

V_f es el valor final del activo en un periodo dado.

V_i es el valor inicial del activo en un periodo dado.

El retorno porcentual puede indicarnos ganancias (al ser positivo), pérdidas (al ser negativo) o sin cambios (0), de esta manera se evalúa el desempeño de las inversiones y es posible tomar decisiones sobre la contemplación de activos en el portafolio. Adicionalmente, existen otras maneras de calcular los retornos, algunas de ellas son rendimiento total anualizado y ajustado al riesgo.

Una de las principales diferencias entre rentabilidad y retorno, es que la primera se utiliza para medir la eficiencia y capacidad de una empresa para generar beneficios a partir de sus recursos (activos y fondos propios) y el segundo, se emplea para evaluar el rendimiento de una inversión específica, incluyendo tanto ganancias como pérdidas en el valor del activo durante un periodo [13]. Sin embargo, en ocasiones se utilizan análogamente.

2.1.6. Tasa de interés

Uno de los conceptos más importantes en los modelos financieros es la tasa de interés, como introduce Carrizo en [20], es la cantidad que se aporta en una unidad de tiempo por cada unidad de capital invertido, también es considerada como el interés de una unidad de moneda en un periodo de tiempo dado, y por último, es el rendimiento de la unidad capital en una unidad de periodo de tiempo.

Es posible notar que los dos conceptos recurrentes en dichas definiciones son la unidad de capital y la unidad de tiempo, una de las razones es porque el propósito de las tasas es medir las variaciones en una unidad.

Al mismo tiempo, una tasa de interés es un dato (precio), en el que se observa el costo por usar los recursos ajenos, en particular el dinero, es decir es el costo del dinero (prestado) [66].

2.1.7. Diversificación

Siguiendo al autor Aguilar en [3], en diversas ocasiones para establecer un modelo de inversión, los individuos se basan en tendencias de mercado sustentadas por la popularidad y el seguimiento a esquemas productivos, lo cual, si bien puede tener algunos casos de éxito, la probabilidad de que el riesgo aumente, es cuantiosa.

De otro modo, los autores Castillo y Requena en [33], infieren que la relación no es lineal, debido a que en el comienzo, las compañías en su mayoría optan por diversificar con el fin de disminuir riesgos.

La diversificación tiene amplias ventajas, por ejemplo, Ramanujam, Vasudevan y Varadarajan en [63] mencionan que la diversificación es una introducción para las empresas o compañías para un nuevo giro de mercado, mediante el desarrollo de gestiones internas o externas, lo cual podría conducir a mejoras en la estructura administrativa, sistemas y otros procesos directivos.

De acuerdo con la aseveración de Palich y Cardinal en [61], la diversificación y el desempeño se pueden co-relacionar proporcionalmente, es decir, si incrementa la diversificación, los resultados también aumentan.

Acerca de las características del desempeño en relación a la diversificación, existe una elasticidad menor de la rentabilidad, ya que se involucran una serie de factores adicionales como la calidad y la adaptabilidad a las necesidades directas de la empresa. No obstante, el precio de los productos ofertados (cartera de productos) no determina la situación de mercado, esto de acuerdo al autor Fierro en [18].

Por otra parte, para Huerta en [41], los distintos modelos reconocen que los acrecentamientos en la diversificación no pueden tener relación con el incremento en el desempeño, al menos no de forma continua.

El autor Fierro en [18], analiza las características del desempeño en función de la diversificación, se puede hablar de una menor elasticidad de la rentabilidad, ya que se valoran múltiples elementos adicionales como la calidad y la adaptación a las necesidades de la empresa. Sin embargo, esto no significa que el precio de los productos ofertados (cartera de productos) determine la situación de mercado.

2.2. Conceptos matemáticos

En esta sección se incluyen los conceptos matemáticos necesarios para construir el algoritmo, evaluar los resultados de las pruebas e interpretarlos. Es crucial comprender estos conceptos, ya que representan la base teórica sobre la cual se desarrollará la aplicación. La comprensión de estos fundamentos no solo facilitará el desarrollo del algoritmo, sino que también permitirá una evaluación e interpretación de los resultados en el contexto de su aplicación práctica.

2.2.1. Variable aleatoria

Sea (Ω, F, P) un espacio de probabilidad donde:

- Ω es el conjunto de todos los posibles resultados del experimento.
- F es una σ -álgebra sobre Ω (colección de subconjuntos de Ω que incluye Ω mismo, es cerrada bajo la complementación y bajo uniones numerables).
- P es una medida de probabilidad asignada a los eventos en F .

Una función

$$X : \Omega \mapsto \mathbb{R}$$

es una variable aleatoria si para cada conjunto Borel B en $\mathbb{R}$, la preimagen $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ pertenece a $\mathcal{F}$, es decir, X es $\mathcal{F}$ -medible [26]. En otras palabras, es una función de Ω en $\mathbb{R}$ que tiene la propiedad de satisfacer una condición de medibilidad. Al conjunto de posibles valores que puede tomar X se le llama Rango o Recorrido de X y se denota con R_X , al ser un conjunto, este puede ser finito, infinito numerable o infinito no numerable [27]. Existen dos tipos de variables aleatorias:

- Discretas: Su recorrido es finito o infinito numerable, es decir, el recorrido de X es $R_X = \{x_1, x_2, \dots, x_n, \dots\}$ y además, para todo $x_i \in R_X$, se le asocia un número real $p(x_i) = P(X = x_i)$ llamado probabilidad de x_i , tal que:
 $0 \leq p(x_i) \leq 1$ para todo $x_i \in R_X$, y
 $\sum_{i=1}^{\infty} p(x_i) = 1$.

A $p(x_i)$ se le conoce como función de probabilidad puntual (denotada por fdpp) de X

- Continuas: Su recorrido R_X es infinito y no numerable, es decir, R_X puede ser: (a, b) , $[a, b)$, $(a, b]$, $[a, b]$, $(-\infty, \infty)$, con $a, b \in \mathbb{R}$, $a < b$ y además existe una función $f(x)$ llamada función de densidad tal que cumple:
 $0 \leq f(x) \leq 1$, para todo $x \in R_X$, y
 $\int_{-\infty}^{\infty} f(x)dx = 1$.

2.2.2. Esperanza matemática

Conforme con Contreras y Cañada en [26], sea $X : \Omega \mapsto \mathbb{R}$ una variable aleatoria discreta con función de probabilidad puntual $p(x_j)$ y recorrido $R_X = \{x_1, x_2, \dots\}$. La esperanza matemática de X se denota como $E[X]$ y se define como:

$$E[X] = \sum_{j=1}^{\infty} x_j p(x_j), \quad (2.3)$$

siempre que la serie infinita (2.3) sea absolutamente convergente, es decir, $\sum |x_j|p(x_j) < \infty$.

Algunos puntos a mencionar son: esta suma comprende a todos los posibles valores que puede tomar X y el número de sumandos puede ser finito o infinito (depende del conjunto de valores de la variable aleatoria) [65].

Por otra parte, cuando X es una v.a. continua con función de densidad de probabilidad $f(x)$, la esperanza matemática de X se define por:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx \quad (2.4)$$

siempre que la integral impropia (2.4) sea absolutamente convergente, es decir, $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$ [53].

Cuando la condición de absoluta convergencia no se cumple, se le denomina no existente a la esperanza, en otras palabras, la esperanza de una variable aleatoria es un número que indica el promedio ponderado de todos los valores posibles que puede obtener una variable X . Comúnmente es conocida como media, valor esperado o valor promedio y es nombrada por la letra griega μ .

En el lado opuesto, si la integral o suma de las ecuaciones no es convergente, se dice que la variable no tiene esperanza finita [65].

2.2.3. Varianza

A manera de definición, Rincón en [65], asevera que es una medida del grado de dispersión de los distintos valores posibles para la variable. Comúnmente se denota por σ^2 , en particular, a la raíz de la varianza, σ , se le conoce como desviación estándar. Se denota por $Var(X)$ y se puede definir como sigue. En caso de que X sea una variable aleatoria discreta:

$$Var(X) = \sum_x (x - E(X))^2 p(x). \quad (2.5)$$

Cuando X es una variable aleatoria continua:

$$Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx. \quad (2.6)$$

En conjunto, la varianza puede definirse como $Var(X) = E[(X - E(X))^2]$, como observación, si la suma o integral no existe, la variable aleatoria no tiene varianza finita.

Supóngase la proposición siguiente: Sean X y Y dos variables aleatorias, y sea c una constante. Entonces

- $Var(X) \geq 0$,
- $Var(c) = 0$,
- $Var(cX) = c^2 Var(X)$,
- $Var(X + c) = Var(X)$,
- $Var(X) = E(X^2) - E^2(X)$,
- En general, $Var(X + Y) \neq Var(X) + Var(Y)$.

2.2.4. Covarianza

Debido a la necesidad de analizar al menos dos variables al mismo tiempo, Hernández y Espinosa en [38], mencionan que surge una cuestión innata que conduce a conocer el grado de relación y coherencia que poseen. Por lo cual, se pretende conocer si los valores de una variable se modifican conforme a los de la otra, o si no lo se modifica. Este suceso puede ser cuantificado a través de la covarianza.

En resumen, se define como sigue:

$$Cov(X, Y) \equiv \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]. \quad (2.7)$$

en la cual:

E representa el operador de la esperanza matemática.

μ_X y μ_Y son las medias poblacionales de X y Y .

Asimismo, las expresiones que siguen son equivalentes: $Cov(X, Y)$ y σ_{XY} .

Ahora, considerando la covarianza de una variable con ella misma, tendremos:

$$Cov(X, X) \equiv \sigma_{XX} = \sigma_X^2 = E[(X - \mu_X)(X - \mu_X)] = E(X - \mu_X)^2, \quad (2.8)$$

lo cual se traduce en que la covarianza de una variable con ella misma es la varianza. Dependiendo del tipo de variable a considerar (continua o discreta), la ecuación se actualizaría como sigue:

$$Cov(X, Y) = \sum_X \sum_Y (X - \mu_X)(Y - \mu_Y) f_{XY}. \quad (2.9)$$

ó como

$$Cov(X, Y) = \iint_{-\infty}^{\infty} (X - \mu_X)(Y - \mu_Y) f_{XY} dx dy. \quad (2.10)$$

En esta última ecuación, f_{XY} es la función de probabilidad conjunta.

Si consideramos el producto, $(X - \mu_X)(Y - \mu_Y)$ y “*si los valores de X y Y tienden a ser mayores o menores que los de las medias poblacionales, la covarianza tendrá signo positivo; por el contrario, si los valores de X tienden a ser mayores o menores que los de su media, mientras que los valores de Y varían de forma inversa, la covarianza tendrá signo negativo.*” [38]

La magnitud del resultado tiene una estrecha relación con las diferencias en $(X - \mu_X)(Y - \mu_Y)$.

2.2.5. Modelos de regresión

Son una técnica utilizada para modelar la relación entre variables. En el momento en que existe la motivación de evaluar la relación que contiene una variable de interés (comúnmente denominada variable dependiente Y), con un conjunto de variables (variables independientes $X_1, X_2, \dots, X_n$), resulta indispensable la aplicación de los modelos de regresión, los cuales permiten la obtención de información sobre cómo una variable de interés Y , variable dependiente, varía cuando una de las independientes lo hace, en otras palabras, se describen como sigue.

$X = (X_1, X_2, \dots, X_n)$: variable controlada, predictora independiente, explicativa.

Y : variable respuesta, predicción, dependiente [17].

$$Y = f(X_1, X_2, \dots, X_n) + \varepsilon. \quad (2.11)$$

Estos modelos pueden tener fines explicativos, es decir, se busca evaluar la afectación del cambio en unas características determinadas (variables independientes) sobre otra característica en particular (variable dependiente) o fines predictivos, estimar o aproximar el valor de una característica (variable dependiente) en relación con los valores que pueden tomar en conjunto otra serie de características (variables independientes) [56].

Existen diversas opciones para estimar un modelo de regresión, los más comunes por facilidad de aplicación e interpretación son el modelo de regresión lineal y el modelo de regresión logística, la utilización de estos depende del tipo de variable dependiente, también conocida como variable respuesta, que se pretenda estimar, si es una variable continua, el modelo de regresión frecuentemente utilizado es de regresión lineal.

Por otro lado, si la variable de interés es dicotómica (es decir, toma dos valores como hombre/mujer, sí/no, día/noche), comúnmente se utiliza la regresión logística. Además, es posible mencionar otros tipos de modelos de regresión como la regresión no lineal o la regresión ordinal.

Regresión lineal

Como se mencionó, actualmente existen diversas técnicas de regresión en función del tipo de variables y de la relación que conforman, la más elemental y utilizada es la regresión lineal. Este tipo de regresión, supone que la relación entre dos variables tiene una forma lineal (o bien, es posible linealizarse mediante alguna transformación de las variables) [35]. En general, para mencionar que existe regresión en los valores de una variable con respecto a otra, debe existir alguna línea, (denominada línea de regresión) que notoriamente se ajusta en una cierta medida a los valores observados. Algunos de los usos que se le da a la regresión es la identificación de relaciones potencialmente causales o cuando no hay interrogantes sobre su relación causal,

la predicción de una variable a partir de otra. Esta relación es significativamente más fácil de observar en variables de tipo determinista, debido a que una define perfectamente el valor de la otra, ya que al realizar las mediciones, los resultados estarán muy cerca de la línea de regresión; esta distancia de los puntos analizados con la línea es denominada error de medición. Sin embargo, existen otras variables que tienen factores de afectación aleatorios (múltiples), en ocasiones desconocidos; en este caso, a esta variabilidad se le agrega aquella introducida por los factores adicionales posibles.

Para introducir el procedimiento a seguir, se divide en cuatro etapas:

1. Trazar o dibujar los puntos en un gráfico cartesiano que muestre la relación entre las variables como primera aproximación.
2. Determinar la ecuación de la línea que se adhiera de mejor manera a dichos puntos.
3. Posteriormente, calcular la variabilidad de la muestra en función de la línea de regresión previamente calculada.
4. Realizar inferencias, por ejemplo, a través de gráficos de coordenadas en los que es posible obtener “nubes de puntos”, ver Figura 2.1.

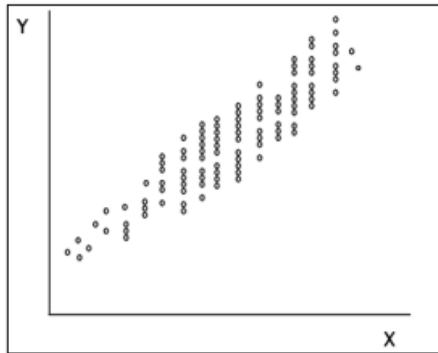


Figura 2.1: Nube de puntos. En las abscisas, los valores de la variable independiente X y, en las ordenadas, los de la variable dependiente Y .

El modelo de regresión parte de algunas suposiciones acerca de las variables y de las muestras experimentales [19]:

Supuesto 1 Sea f un modelo de regresión, se tiene:

- a) *Linealidad:* La relación existente entre X y Y es lineal.
- b) *Homogeneidad:* El valor promedio del error es cero, $E[\varepsilon] = 0$.
- c) *Homocedasticidad:* La varianza de los errores es constante, $Var(\varepsilon) = \sigma^2$.

d) *Independencia:* El término del error ε_i no está correlacionado con las variables independientes observadas $x_i, x_{i+1}, \dots, x_k$, es decir, $E[\varepsilon_i | x_i, x_{i+1}, \dots, x_k] = 0$ donde $x_i, x_{i+1}, \dots, x_k$ son las variables independientes del modelo.

e) *Normalidad:* Los errores siguen una distribución normal, $\varepsilon \sim N(0, \sigma)$.

Además, 1) La media de la población de Y dado un valor de X crece (o decrece) linealmente a medida que X aumenta. 2) Para cualquier valor de X , los posibles valores de Y se distribuyen normalmente. 3) La desviación estándar de la población de Y alrededor de su media a un valor dado de X es la misma para todos los valores de X . En otras palabras, el modelo supone aleatoriedad de la muestra, normalidad, linealidad y homogeneidad de las varianzas [22].

Regresión lineal simple

Se conoce como regresión simple al cálculo de la ecuación correspondiente a la línea que mejor describe la relación entre la respuesta y la variable que la explica [19]. Principalmente, se pretende explicar o predecir la variable dependiente (Y) a partir de una única variable independiente, (x_1), es decir, tiene el objetivo de predecir y/o estimar los valores de la variable Y (dependiente), a partir de la obtención de la función lineal de la variable independiente. Como mencionan los autores Rodríguez y Mora en [68], matemáticamente es Y expresada como sigue:

$$Y = \beta_0 + \beta_1 x_1 + \varepsilon, \quad (2.12)$$

en donde:

Y es la variable a predecir;
 β_0 y β_1 , son parámetros desconocidos a estimar;
 ε es el error que cometemos en la predicción de los parámetros.

Nota: Este modelo cumple el Supuesto 1.

Para comprobar, si la línea de regresión es significativa, se debe comprobar la siguiente hipótesis:

Supuesto 2 *Teniendo un modelo lineal $Y = \beta_0 + \beta_1 x_1 + \varepsilon$, $\beta_1 \neq 0$.*

Para comprobar la condición 2 debe realizarse una prueba de hipótesis con:

- $H_0 : \beta_1 = 0$,
- $H_1 : \beta_1 \neq 0$.

De acuerdo con los autores Montgomery, Peck y Vining en [55], es conveniente considerar al regresor x como una variable controlada por el analista de datos y medida con un error despreciable, mientras que la respuesta y es una variable aleatoria. Es decir, hay una distribución de probabilidad para y en cada valor posible de x . La media de esta distribución es

$$E[y|x] = \beta_0 + \beta_1 x,$$

y la varianza es

$$Var(y|x) = Var(\beta_0 + \beta_1 x + \varepsilon) = \sigma^2.$$

Así, la media de y es una función lineal de x aunque la varianza de y no depende del valor de x , además, debido a que los errores no están correlacionados, las respuestas tampoco lo están. Los parámetros β_0 y β_1 , suelen llamarse coeficientes de regresión y tienen como interpretación que la pendiente β_1 es el cambio en la media de la distribución de y producido por un cambio unitario en x . Si el rango de datos en x incluye $x = 0$, entonces la intersección β_0 es la media de la distribución de la respuesta y cuando $x = 0$. Si el rango de x no incluye al cero, entonces β_0 no tiene una interpretación práctica.

Estimación de mínimos cuadrados de los parámetros

Los parámetros β_0 y β_1 son desconocidos y deben ser estimados utilizando datos de muestra. Supongamos que tenemos n pares de datos, $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$, resultantes de un experimento controlado para recopilar datos, un estudio observacional o registros históricos existentes (estudio retrospectivo). El método de mínimos cuadrados se utiliza para estimar β_0 y β_1 , es decir, se estiman de manera en que la suma de los cuadrados de las diferencias entre las observaciones y_i y la recta sea mínima. A partir de la ecuación (2.12), podemos escribir para $i = 1, 2, \dots, n$:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i. \quad (2.13)$$

La ecuación (2.12) puede ser vista como un modelo de regresión de la población, mientras que la ecuación (2.13) es un modelo de regresión de muestra, escrita en términos de los n pares de datos (y_i, x_i) , para $i = 1, 2, \dots, n$, de esta manera, el criterio de mínimos cuadrados es:

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \quad (2.14)$$

Los estimadores de mínimos cuadrados de β_0 y β_1 , digamos $\hat{\beta}_0$ y $\hat{\beta}_1$, deben satisfacer:

$$\left. \frac{\partial S}{\partial \beta_0} \right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

y

$$\left. \frac{\partial S}{\partial \beta_1} \right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0$$

Simplificar estas dos ecuaciones, produce:

$$\begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i, \\ \hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n y_i x_i. \end{aligned} \quad (2.15)$$

Las ecuaciones (2.15), se denominan ecuaciones normales de mínimos cuadrados, su solución es:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (2.16)$$

y

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i - \frac{(\sum_{i=1}^n y_i)(\sum_{i=1}^n x_i)}{n}}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}, \quad (2.17)$$

donde

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

y

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

son los promedios de y_i y x_i , respectivamente. Por lo tanto, $\hat{\beta}_0$ y $\hat{\beta}_1$ en las ecuaciones (2.16) y (2.17) son los estimadores de mínimos cuadrados de la intersección y la pendiente, respectivamente. El modelo de regresión lineal simple ajustado es entonces

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i. \quad (2.18)$$

La ecuación (2.18) proporciona una estimación puntual de la media de y para un x particular. Dado que el denominador de la ecuación (2.17), es la suma de cuadrados corregida de x_i y el numerador es la suma corregida del producto cruzado de x_i y y_i , podemos escribir estas cantidades en otra notación como:

$$S_{xx} = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} = \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.19)$$

y

$$S_{xy} = \sum_{i=1}^n y_i x_i - \frac{(\sum_{i=1}^n y_i)(\sum_{i=1}^n x_i)}{n} = \sum_{i=1}^n y_i (x_i - \bar{x}). \quad (2.20)$$

Y una manera más compacta de escribir la ecuación (2.17) es:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}. \quad (2.21)$$

La diferencia entre el valor observado y_i y el valor ajustado correspondiente $\hat{y}_i$ es un residuo. Matemáticamente, el residuo i -ésimo es para $i = 1, 2, \dots, n$

$$\varepsilon_i = y_i - \bar{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i). \quad (2.22)$$

Para el modelo descrito antes, es necesario realizar algunos pasos que el autor [68], menciona:

- Exposición de los estadísticos para valorar la *bondad* de ajuste en el modelo de regresión lineal simple, es decir, analizar el grado de asociación lineal entre las variables, así como la proporción de variabilidad de la variable dependiente ejemplificada por la independiente. Para conocer la *bondad* de ajuste, se utilizan estadísticos como:
 - El *Coefficiente de Correlación Lineal Simple* (r) que mide el grado de asociación lineal entre dos variables, este estadístico oscila entre el intervalo $(-1, 1)$ en el cual 1 es una fuerte asociación lineal positiva, si aumentan los valores de una variable, los de la otra también y -1 es una fuerte asociación lineal negativa, es decir, si aumentan los valores de una variable, disminuyen los de la otra, por último, si es 0 o próximo a 0, se considera que no existe asociación lineal.
 - El *Coefficiente de Correlación Múltiple al Cuadrado o Coeficiente de Determinación* (R Square R^2) que se define a partir del coeficiente de correlación múltiple (R) y mide la proporción de variabilidad de la variable dependiente explicada por la variable independiente introducida o por la recta de regresión. Si el valor que resulta lo multiplicamos por 100, obtendremos el porcentaje de variabilidad explicada [68].
 - El *Coefficiente de Determinación Ajustado* (*Adjusted R Square*) que aún cuando el mencionado R^2 se utiliza como medida de ajuste al modelo, existen algunas limitaciones al incrementar el número de variables que participan en el modelo (por

ejemplo en el análisis multivariable), su valor también aumenta, por lo que puede sobrestimar el verdadero R de la población, para contribuir a este inconveniente, se utiliza el Coeficiente de Determinación Ajustado, que no tiene este incremento al añadir variables a la ecuación.

Asimismo, queda ajustado por el número de observaciones y variables independientes incluidas en la relación.

- El *Error Típico de Predicción (ETB)* que se define como la fracción de la variable dependiente que se evita explicar por falta de introducción o deficiente elección de una variable. Para calcularlo, se utiliza la desviación típica de la variable dependiente y el coeficiente de determinación ajustado.
 - El *Análisis de Varianza* que permite valorar los límites superiores de considerar adecuado el modelo de regresión lineal para estimar los valores de la variable dependiente. La tabla ANOVA proporciona el estadístico F a partir del cual es posible contrastar la H_0 de que $R^2 = 0$, o en otras palabras, que las dos variables están incorrelacionadas. Si el p-valor asociado al estadístico F es menor que el nivel de significancia (normalmente 0.05), rechazaremos la hipótesis nula planteada [68].
 - Los *Análisis de Residuales*, los residuos “ ε ”, son estimaciones de los errores, deben ser independientes y su distribución debe ser normal, es decir, los residuos observados y los esperados bajo la hipótesis de distribución normal deben ser parecidos. Tiene el objetivo de profundizar en la relación de las variables y ponderar la *bondad* de ajuste de regresión obtenida.
- Estimación de los parámetros o coeficientes de regresión: Si se confirma a través de los estadísticos que los valores poseen una asociación lineal, y posterior al análisis del carácter e intensidad de la relación entre las variables, se pueden estimar los parámetros B_0 y B_1 para efectuar predicciones de la variable dependiente. El que los estadísticos se rechacen, no implica que las variables no posean la relación lineal, es decir, no es una co-implicación. Para obtener dichos coeficientes, es posible utilizar el método de mínimos cuadrados, este tiene sus bases en la minimización de la suma de los cuadrados de los residuos con el fin de que la recta de regresión que se define tenga cercanía a la nube de puntos observados, es decir, que mejor los represente.

Los estadísticos asociados a la variable independiente que, para este momento, pasa a formar parte del modelo de regresión simple son:

1. Coeficiente de regresión β : indica el número de unidades que aumentará la variable dependiente por cada unidad que aumente la variable independiente.
2. $SE(\beta)$: error típico de β , mide la precisión de β .
3. Coeficiente β : coeficiente de regresión estandarizado, expresa la pendiente de la recta de regresión cuando todas las variables estén transformadas en puntuaciones Z , es decir, han sido estandarizadas.
4. Constante (intercepto): es el valor de y cuando todas las variables independientes x son 0.
5. Tolerancia: proporción de variabilidad no explicada por el resto de variables ($1 - R^2$), a mayor incremento de la tolerancia, más independencia en la variable.

6. Estadístico T : permite comprobar si la regresión entre una variable independiente y la dependiente es significativa. Si el p-valor asociado al estadístico T es mayor al nivel de significación, rechazaremos que la regresión sea significativa para las dos variables relacionadas.
- Se expone la secuencia de pasos que permiten determinar los anteriores puntos, se realizan pruebas congruentes que permitan obtener la mayor cantidad de información posible.

Regresión lineal múltiple

El modelo de regresión múltiple es la extensión del modelo de regresión simple a k variables explicativas [19]. La estructura del modelo de regresión múltiple es posible visualizarla en la ecuación (2.23):

$$y = f(x_1, x_2, \dots, x_k) + \varepsilon, \quad (2.23)$$

donde:

y es la variable explicada, dependiente o respuesta,

$x_1, x_2, \dots, x_k$ son las variables explicativas, regresoras o variables independientes,

$y = f(x_1, x_2, \dots, x_k)$ es la parte determinista del modelo,

ε representa el error aleatorio. Contiene el efecto sobre de todas las variables distintas de $x_1, x_2, \dots, x_k$.

El modelo de regresión lineal múltiple con k regresores tiene la forma:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon, \quad (2.24)$$

donde los parámetros β_j con $j = \{0, 1, \dots, k\}$ son llamados coeficientes de regresión y representan el cambio en la variable de respuesta y por unidad de cambio en x_j cuando todas las variables regresoras restantes $x_i (i \neq j)$ se mantienen constantes. Por esta razón, los parámetros β_j con $j = \{1, 2, \dots, k\}$, se denominan comúnmente coeficientes parciales de regresión.

Los modelos de regresión lineal múltiple se utilizan a menudo como modelos empíricos o funciones de aproximación. Es decir, la verdadera relación funcional entre y y $x_1, x_2, \dots, x_k$ es desconocida, pero sobre ciertos rangos de las variables regresoras, el modelo de regresión lineal es una aproximación adecuada a la función real.

Para una mejor comprensión, describimos el vector de variables independientes $\mathbf{X}$ y el vector de coeficientes $\boldsymbol{\beta}$:

$\mathbf{X}$ es la matriz de datos que contiene las variables independientes x_{ij} para $i = 1, 2, \dots, n$ y $j = 1, 2, \dots, p$. $\boldsymbol{\beta}$ es el vector de coeficientes asociado, compuesto por $\beta_1, \beta_2, \dots, \beta_p$.

Se puede utilizar el método de mínimos cuadrados para estimar los coeficientes de la ecuación (2.24), suponiendo que se dispone de $n > k$ observaciones disponibles, que y_i denota la

i -ésima respuesta observada y x_{ij} denota la i -ésima observación o nivel del regresor x_j , que el término de error ε en el modelo tiene $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2$, y que los errores no están correlacionados.

La ecuación (2.24), para $i = 1, 2, \dots, n$, puede escribirse como:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i. \quad (2.25)$$

La función de mínimos cuadrados es:

$$S(\beta_0, \beta_1, \dots, \beta_k) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij})^2. \quad (2.26)$$

La función S , se minimiza con respecto a $\beta_0, \beta_1, \dots, \beta_k$ y los estimadores de mínimos cuadrados de $\beta_0, \beta_1, \dots, \beta_k$, deben satisfacer:

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij}) = 0, \quad (2.27)$$

y para $j = 1, 2, \dots, k$,

$$\frac{\partial S}{\partial \beta_j} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij}) x_{ij} = 0. \quad (2.28)$$

Simplificando la ecuación (2.28), se obtienen las ecuaciones normales de mínimos cuadrados.

Es posible desarrollar un estimador de σ^2 a través de la suma de cuadrados,

$$SS_{Res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2 = e'e.$$

donde:

e es un vector columna de errores $(e_1, e_2, \dots, e_n)^T$,

e' es la transpuesta de e , un vector fila $(e_1, e_2, \dots, e_n)$,

Sustituyendo $e = y - X\hat{\beta}$, se tiene:

$$\begin{aligned} SS_{Res} &= (y - X\hat{\beta})'(y - X\hat{\beta}) \\ &= y'y - \hat{\beta}'X'y - y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \\ &= y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} \end{aligned}$$

. Debido a que $X'X\hat{\beta} = X'y$, entonces:

$$SS_{Res} = y'y - \hat{\beta}'X'y. \quad (2.29)$$

en la cual:

X es la matriz de diseño (o matriz de regresores), que incluye las observaciones de las variables independientes (predictoras). Para un modelo de regresión lineal con n observaciones y k variables predictoras, X es una matriz de dimensión $n \times k$,

X' es la transpuesta de la matriz X , se obtiene una matriz de dimensión $k \times n$,

y es el vector de observaciones de la variable dependiente (respuesta). En un modelo con n observaciones, y es un vector columna de dimensión $n \times 1$,

y' es la transpuesta del vector y , lo que convierte a y en un vector fila de dimensión $1 \times n$.

Podemos demostrar que los estimadores de máxima verosimilitud para los parámetros del modelo en la regresión lineal múltiple, cuando los errores del modelo se distribuyen de forma normal e independiente, también son estimadores de mínimos cuadrados. Consideremos el modelo:

$$y = X\beta + \varepsilon. \quad (2.30)$$

Y que los errores se distribuyen normal e independientemente con una varianza constante σ^2 o que ε se distribuye como $N(0, \sigma^2 I)$. La función de densidad normal para los errores es:

$$f(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\varepsilon_i^2\right).$$

Debido a que la función de verosimilitud es la densidad conjunta de $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, o $\prod_{i=1}^n f(\varepsilon_i)$. La función de verosimilitud es:

$$L(\varepsilon, \beta, \sigma^2) = \prod_{i=1}^n f(\varepsilon_i) = \frac{1}{(2\pi)^{(n/2)}\sigma^n} \exp\left(-\frac{1}{2\sigma^2}\varepsilon'\varepsilon\right).$$

Teniendo en cuenta que se puede escribir $\varepsilon = y - X\beta$, la función se convierte en:

$$L(y, X, \beta, \sigma^2) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right). \quad (2.31)$$

Al utilizar el logaritmo de la función de verosimilitud,

$$\ln L(y, X, \beta, \sigma^2) = -\frac{n}{2} \ln(2\pi) - n \ln(\sigma) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta).$$

Para un valor fijo de σ , el logaritmo de la función de verosimilitud, se maximiza cuando el término $(y - X\beta)'(y - X\beta)$ se minimiza. Por lo tanto, el estimador de máxima verosimilitud de β , bajo errores normales es equivalente al estimador de mínimos cuadrados $\hat{\beta} = (X'X)^{-1}X'y$. El estimador de máxima verosimilitud de σ^2 es:

$$\tilde{\sigma}^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{n}. \quad (2.32)$$

Posteriormente a la estimación de los parámetros en el modelo, se busca responder dos preguntas, ¿cuál es la idoneidad general del modelo? y, ¿qué regresores específicos parecen importantes? Las pruebas formales, explicadas a continuación, requieren que los errores aleatorios sean independientes y sigan una distribución con media $E(e_i) = 0$ y varianza $Var(\varepsilon_i) = \sigma^2$.

Pruebas de hipótesis en regresión lineal múltiple

- Prueba de significancia: es una prueba para determinar si existe una relación lineal entre la respuesta y y alguna de las variables del regresor $x_1, x_2, \dots, x_k$. Usualmente se considera que este procedimiento es una prueba general o global de la adecuación del modelo. Las hipótesis apropiadas son:

$$H_0 : \beta_1 = \dots = \beta_k = 0.$$

$$H_1 : \beta_j \neq 0 \text{ para al menos una } j.$$

Es posible notar que el rechazo de la hipótesis nula implica que al menos uno de los regresores $x_1, x_2, \dots, x_k$ contribuye significativamente al modelo. El procedimiento de prueba es una generalización del análisis de varianza utilizado en la regresión lineal. La suma total de cuadrados SS_T se divide en una suma de cuadrados debido a la regresión, SS_R , y una suma residual de cuadrados, SS_{Res} . Así

$$SS_T = SS_R + SS_{Res}.$$

- Pruebas sobre coeficientes de regresión individuales y sobre subconjuntos de coeficientes:

Después de determinar que al menos uno de los regresores tiene importancia, es necesario saber cuál o cuáles lo son. Agregar una variable a un modelo de regresión siempre causa que la suma de cuadrados para la regresión aumente y la suma de cuadrados residual disminuya.

Debemos decidir si el aumento en la suma de cuadrados de la regresión justifica el uso del regresor adicional en el modelo. La adición de un regresor también aumenta la varianza del valor ajustado $\hat{y}$, por lo que debemos ser cuidadosos e incluir solo aquellos regresores que realmente aporten valor en la explicación de la respuesta. Además, agregar un regresor no importante puede incrementar la media residual de los cuadrados, lo cual podría disminuir la utilidad del modelo.

Las hipótesis para comprobar la significancia de cualquier coeficiente de regresión individual como β_j , son:

$$H_0 : \beta_j = 0, H_1 : \beta_j \neq 0. \quad (2.33)$$

Si $H_0 : \beta_j = 0$ no se rechaza, esto indica que el regresor x_j se puede eliminar del modelo. El estadístico de prueba para esta hipótesis es:

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}. \quad (2.34)$$

en donde C_{jj} es el elemento diagonal de $(X'X)^{-1}$ correspondiente a $\hat{\beta}_j$. La hipótesis nula $H_0 : \beta_j = 0$ se rechaza si $|t_0| > t_{\alpha/2, n-k-1}$. Es importante mencionar que esta es realmente una prueba parcial o marginal porque el coeficiente de regresión $\hat{\beta}_j$ depende de todas las demás variables del regresor $x_i (i \neq j)$ que están en el modelo. Por lo tanto, esta es una prueba de la contribución de x_j dados los demás regresores en el modelo.

Intervalos de confianza en la regresión múltiple

Para construir intervalos de confianza para los coeficientes de regresión β_j , se continuará asumiendo que los errores ε_i se distribuyen de forma normal e independientemente con media cero y varianza σ^2 . Por lo tanto, las observaciones y_i se distribuyen de forma normal e independiente con media $\beta_0 + \sum_{j=1}^k \beta_j x_{ij}$ y varianza σ^2 . Dado que el estimador de mínimos cuadrados $\hat{\beta}$ es una combinación lineal de las observaciones, se deduce que $\hat{\beta}$ se distribuye normalmente con vector de medias β y matriz de covarianza $\sigma^2(X'X)^{-1}$.

Esto implica que la distribución marginal de cualquier coeficiente de regresión $\hat{\beta}_j$ es normal con media β_j y varianza $\sigma^2 C_{jj}$, donde C_{jj} es el j -ésimo elemento de la matriz diagonal $(X'X)^{-1}$. En consecuencia, cada uno de los estadísticos para $j = 0, 1, \dots, k$,

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}}, \quad (2.35)$$

posee una distribución t con $n - p$ grados de libertad. Basándonos en el resultado dado en la expresión (2.35), podemos definir un intervalo de confianza de $100(1 - \alpha) \%$ para el coeficiente de regresión β_j , $j = 0, 1, \dots, k$, como:

$$\hat{\beta}_j - t_{\alpha/2, n-p} \sqrt{\hat{\omega}^2 C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2, n-p} \sqrt{\hat{\omega}^2 C_{jj}}. \quad (2.36)$$

Considerando el error estándar del coeficiente de regresión $\hat{\beta}_j$ como:

$$se(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}}$$

La regresión lineal múltiple pretende ajustar modelos lineales o bien, que sean linealizables entre una variable dependiente y más de una variable independiente [35]. El antecedente mostrado es la aseveración de que en la regresión simple se tiene una única variable predictora. Sin embargo, en ocasiones, es necesario considerar dos o más variables regresoras o predictoras, por lo que se debe recurrir al uso de este tipo de regresión [19].

Respecto a la utilización del modelo de regresión lineal múltiple, se puede mencionar: Cuando la variable y , tiene una relación de dependencia lineal con cada una de las variables explicativas $x_1, x_2, \dots, x_k$, y cuando un regresor no basta para explicar suficientemente la variabilidad de y .

En cuestión de los supuestos del modelo de regresión lineal múltiple, al ajustarlo, se supondrá que se verifican las siguientes hipótesis:

1. Fijados los valores de las variables $x_1, x_2, \dots, x_k$, se tiene que:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i. \quad (2.37)$$

2. Cualquier par de errores, e_i y e_j , son independientes.
3. Las variables explicativas son, algebraicamente, linealmente independientes.
4. El número de datos es mayor o igual que $k + 2$.

Es importante considerar que el parámetro β_i , en regresión múltiple, representa el efecto del aumento de una unidad del regresor sobre la respuesta, y , cuando el resto de los regresores permanecen constantes [19].

2.2.6. P-Value

El valor p (P-Value) se introdujo en 1925 por el estadístico Ronald A. Fisher, quien lo utilizó en pruebas de significación y diseño de experimentos. Este valor es la probabilidad que nos permite determinar si los resultados de una prueba son significativos. En otras palabras, el valor p estima si la probabilidad de obtener los resultados observados es lo suficientemente pequeña como para rechazar la hipótesis nula. Es una probabilidad calculada basada en los datos observados, asumiendo que la hipótesis nula es verdadera [71].

De acuerdo con los autores Mendivelso y Rodríguez en [52], este valor, también conocido como *Valor crítico de la prueba estadística*, es utilizado en la investigación para analizar los resultados, es una cantidad que nos indica la inusualidad de los resultados en la muestra estudiada considerando que la hipótesis nula (H_0), es verdadera, en otras palabras, es la probabilidad de rechazar la hipótesis nula con los resultados obtenidos. Además, mencionan que cuando el p -value es muy pequeño (comparado con el valor α o nivel de confianza establecido), es posible pensar que los datos probablemente no se producirán solo por azar.

2.2.7. Variable Dummy

De acuerdo con Granados en [35], es una variable dicotómica, es decir, únicamente toma los valores 0 y 1, se construye expresamente con algún propósito informativo o interpretativo,

ocasionalmente, se construyen para obtener información sobre la influencia de cada escalón de una variable factor ordinal, es decir, que se puede introducir literalmente en un modelo o factor puro, en el cual, no es posible introducirse de ninguna manera debido a que carece de interpretación, en este último caso, es posible otorgarle características de variables Dummies.

Para realizar esto, el programa en el que se realice, excluirá una dummy con el fin de que no estén perfectamente colineadas, (expresarse como combinación de las demás), también es importante, considerar que la interpretación de los β , serán el grado de cambio de la variable dependiente cuando el individuo pasa de la variable excluida a la dummy estimada [5].

Por estas razones, existe la posibilidad de tener resultados significativos o no, en ocasiones, se menciona que para otro tipo de variables, se excluyen las no significativas, sin embargo, en el aspecto de variables dummies, se deben incluir todas mientras una sea significativa.

En otras palabras, la variable dummy permite medir el efecto de una determinada característica de los individuos (o periodos) en la muestra. Típicamente la variable dummy toma el valor de uno cuando la característica de interés está presente y cero en caso contrario.

2.2.8. Prueba de razón de verosimilitudes

Se comenzará por mencionar que un procedimiento para contrastar un juego de hipótesis es la Razón de Verosimilitud Generalizada, esta razón, se sustenta en el lema de Neyman-Pearson, junto con la estimación de máxima verosimilitud. A continuación, se define matemáticamente el estadístico de Razón de Verosimilitud Generalizada (RVG):

$$\lambda(X) = \frac{\sup_{\omega \in \Omega_0} f(X; \omega)}{\sup_{\omega \in \Omega_1} f(X; \omega)} = \frac{\sup_{\omega \in \Omega_0} L(\omega)}{\sup_{\omega \in \Omega_1} L(\omega)}, \quad (2.38)$$

donde $L(\omega)$ es la función de verosimilitud. En el momento de la utilización del estadístico λ , se debe considerar su distribución cuando H_0 es verdadera, si tiene un grado elevado de dificultad es posible utilizar la distribución asintótica [46].

Se consideran ahora las aportaciones del autor Fernández en [29], se definen dos aspectos clave: la sensibilidad (fracción de verdaderos positivos (FVP)) y especificidad (fracción de verdaderos negativos (FVN)), para definir:

- Razón de verosimilitudes positiva o cociente de probabilidades positivo: se refiere a la razón entre la fracción de verdaderos positivos con la fracción de falsos positivos de esta forma:

$$RV_+ = \frac{\text{Sensibilidad}}{1 - \text{Especificidad}}.$$

- Razón de verosimilitudes negativa o cociente de probabilidades negativo: se calcula como la razón entre la fracción de falsos negativos, y la fracción de verdaderos negativos, como sigue:

$$RV_- = \frac{1 - \text{Sensibilidad}}{\text{Especificidad}}.$$

2.2.9. Prueba de correlación

El concepto de correlación comprende dos observaciones (X y Y), y pretende mostrar el valor que toma Y para determinado valor de X , algunas de los objetivos son examinar la tendencia de dos variables al ir juntas, cuantificar la magnitud de la correlación entre dos variables y predecir valores. En un caso particular positivo, si estas variables tuvieran una correlación perfecta se podría inferir el valor de la variable Y conociendo el valor de X . Existen diversas pruebas estadísticas que permiten conocer la fuerza de asociación o correlación entre dos variables (cuantitativas y cualitativas), el resultado se expresa a través del coeficiente de correlación.

Coeficiente de Pearson

Es un indicador para describir numéricamente la fuerza, relación y dirección del lazo entre dos variables cuantitativas de distribución normal, es útil para la determinación de la tendencia de dos variables a ir juntas, como fue mencionado con anterioridad, covarianza. Su cálculo se puede enunciar como sigue:

$$r = \frac{n \sum(XY) - (\sum(x))(\sum(Y))}{\sqrt{[n \sum(x)^2 - (\sum(x))^2][n \sum(Y)^2 - (\sum(Y))^2]}}, \quad (2.39)$$

en donde el numerador es la covarianza (suma de productos xy) y el denominador es la raíz del producto de las sumas de cuadrados de ambas variables.

Para denominar al coeficiente de correlación, se utiliza una “ r ” que toma valores en el intervalo de $(-1, +1)$, con las siguientes consideraciones: un valor de 0 representa que no hay correlación, es decir, el comportamiento de una variable no tiene relación con el de la otra, un valor de correlación perfecto, ocurre en los valores de $-1, 0, +1$, lo cual conlleva a deducir que el valor de una variable se puede determinar al conocer el valor de la segunda, entre más cercano a 1 sea el coeficiente de correlación, mayor será la fuerza de asociación.

Un signo positivo implica la relación positiva de las variables, es decir, si una aumenta, la otra también lo hará, por el contrario, un signo negativo implica una relación negativa de las variables, si una aumenta, la otra disminuirá [70]. La hipótesis posible a plantear será:

- $H_0 : r = 0$ no existe correlación.
- $H_1 : r \neq 0$ existe correlación y puede ser positiva o negativa.

El autor también menciona que la presencia de una correlación estadística no implica forzosamente causalidad; es importante considerar los factores y características que pueden afectar a la relación causal.

2.2.10. Modelo Lasso

Este modelo de regresión, busca realizar una selección de variables a través de una técnica de regresión lineal, por esta razón, el autor Tibshirani en [74], propone LASSO (Least Absolute Shrinkage and Selection Operator), la cual es una técnica de regresión lineal regularizada como otras ya creadas en su campo, con la diferencia de realizar la penalización de variables, lo cual tiene consecuencias en la interpretación de los resultados.

El auge en los últimos años en la investigación y aplicación de técnicas tipo LASSO, se debe principalmente al desarrollo paralelo de algoritmos eficientes. De acuerdo con el autor Pacheco en [60], para la metodología del modelo, LASSO resuelve el problema de mínimos cuadrados con restricción sobre la Norma $L1$ del vector de coeficientes, de esta forma:

$$\min_{\beta} \sum_{i=1}^n (y_i - \sum_{j=1}^p \beta_j X_{ij})^2, \text{ siendo } \sum_{j=1}^p |\beta_j| \leq s, \quad (2.40)$$

lo cual es equivalente a

$$\sum_{i=1}^n (y_i - \sum_{j=1}^p \beta_j X_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j|. \quad (2.41)$$

donde:

y_i : Es el valor observado de la variable dependiente para la observación i .

X_{ij} : Es el valor de la j -ésima variable independiente para la observación i .

β_j : Es el coeficiente de regresión asociado con la j -ésima variable independiente.

n : Es el número total de observaciones.

p : Es el número total de variables independientes.

s : Es un parámetro que define la restricción sobre la norma $L1$ del vector de coeficientes β .

$\lambda \geq 0$: Es el parámetro de penalización que controla la cantidad de regularización aplicada al modelo. A medida que λ aumenta, la penalización por complejidad aumenta, forzando algunos coeficientes β_j a acercarse a cero o a ser exactamente cero, lo que induce la selección de variables.

$\hat{\beta}_{lasso}$: Es el estimador de los coeficientes de regresión obtenido mediante el método LASSO.

T : Es el vector de respuesta que contiene los valores observados de la variable dependiente.

Es importante destacar que s es un parámetro que controla la cantidad de penalización aplicada a la suma absoluta de los coeficientes β_j y la restricción $\sum_{j=1}^p |\beta_j| \leq s$ asegura que la suma de los valores absolutos de los coeficientes no exceda s .

Si s es grande, la restricción es menos estricta, lo que permite que más coeficientes β_j tengan valores más grandes y menos coeficientes sean forzados a ser cero. Por otra parte, si s es pequeño, la restricción es más estricta, lo que fuerza a que más coeficientes β_j sean exactamente cero, esto induce la selección de variables, eliminando aquellas que no contribuyen significativamente al modelo.

A diferencia de la regresión Ridge y mínimos cuadrados ordinarios, el estimador $\hat{\beta}_{lasso}$ es no lineal en el vector de respuesta T , por lo cual no hay una expresión limitada del mismo (excepto para el caso de un diseño ortogonal $X^T X = I_p$).

Al tomar valores crecientes de λ_0 y valores decrecientes de s , los coeficientes β_j se acercan significativamente a 0, y algunos de ellos, se anulan, es preciso decir que LASSO produce la estimación y selección de variables en forma continua y simultánea.

2.2.11. Parámetro error estándar

El autor señala que el error estándar de estimación comprende a la desviación normal de los residuos, es importante conocer y considerar en el momento del análisis al error estándar de cada parámetro debido a que se relaciona con la posibilidad de estimar los intervalos de confianza y hacer inferencias probabilísticas para interpretar en el rango en el que radican [75].

Respecto a la manera de disminuir dicho error, existen dos maneras en general: incrementar el tamaño de la muestra debido a que el error estándar es inversamente proporcional a la raíz del tamaño de la muestra, es decir, a mayor tamaño de muestra, menor variabilidad de datos, la segunda forma es reducir la variabilidad de los datos, es posible realizarse a través de diseños de muestreo estratificados, (si existen subpoblaciones o estratos más homogéneos).

En conjunto, además del tamaño de muestra, es importante seleccionar de manera correcta dicha muestra, ya que influye en la precisión de los parámetros.

2.2.12. Modelo autorregresivo integrado de medias móviles de orden p,d,q ARIMA (p,d,q)

Como su nombre lo indica, este modelo es una combinación de los modelos autorregresivos y del modelo de medias móviles, incluye un proceso de restablecimiento de inestabilidad original presente en una serie de tiempo, denominada “integración d ”. La forma general de un Modelo autorregresivo integrado de medias móviles de orden p,d,q (ARIMA) es como sigue:

$$X'_t = c + \phi_1 X'_{t-1} + \dots + \phi_p X'_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}. \quad (2.42)$$

en donde X' , expresa que sobre la serie original X_t se han aplicado d diferencias. En la estimación de los modelos ARIMA el problema parte de identificar el modelo que mejor describa el fenómeno a predecir [15].

Se busca modelar y predecir una serie de tiempo X_t mediante una combinación de sus valores pasados, los errores pasados y las diferencias entre los valores pasados. De primera instancia, en cuestión de modelado, se busca identificar los parámetros (p, d, q) que mejor describan la serie temporal X_t . Estos parámetros representan el orden del componente autorregresivo (AR), el componente de integración (I) y el componente de medias móviles (MA), los cuales se definen a continuación:

- **Componente Autorregresivo (AR):** Este componente está representado por p términos autorregresivos, que capturan la relación lineal entre los valores pasados de la serie temporal y el valor actual. En otras palabras, el componente AR modela cómo los valores previos de la serie temporal afectan al valor presente. Matemáticamente, el componente autorregresivo se representa como $AR(p)$, donde p , es el orden del término autorregresivo. La ecuación típica de un componente autorregresivo es:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (2.43)$$

donde X_t es el valor actual de la serie temporal, c es una constante, $\phi_1, \phi_2, \dots, \phi_p$ son los coeficientes autorregresivos y ε_t es el término de error.

- **Componente de Integración (I):** Este componente está representado por el parámetro d , que indica el número de diferencias necesarias para hacer estacionaria la serie temporal. La diferenciación implica calcular la diferencia entre los valores de la serie en diferentes puntos de tiempo. La idea detrás del componente de integración es transformar una serie temporal no estacionaria en una serie temporal estacionaria, lo que facilita el modelado. Matemáticamente, el componente de integración se representa como $I(d)$.
- **Componente de Medias Móviles (MA):** Está representado por q términos de medias móviles, que capturan la relación lineal entre los errores pasados y el valor actual de la serie temporal. El componente de medias móviles modela cómo los errores pasados afectan al valor actual de la serie temporal. Matemáticamente, el componente de medias móviles se representa como $MA(q)$, donde q es el orden del término de medias móviles. La ecuación típica de un componente de medias móviles es:

$$X_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2.44)$$

donde $\theta_1, \theta_2, \dots, \theta_q$ son los coeficientes de medias móviles y ε_t son los términos de error.

Posteriormente, una vez identificados los valores de p, d y q , se procede a estimar los coeficientes $\phi_1, \phi_2, \dots, \phi_p$ del modelo autorregresivo, los coeficientes $\theta_1, \theta_2, \dots, \theta_q$ del modelo de medias móviles y posiblemente el término constante c . Esto se puede hacer mediante métodos como mínimos cuadrados o de máxima verosimilitud.

Por último, al obtener los parámetros del modelo ARIMA, se utilizan para predecir futuros valores de la serie temporal X_t . Esto se realiza utilizando la estructura del modelo para proyectar los valores futuros de la serie, teniendo en cuenta los valores y errores pasados.

Capítulo 3

Modelos de valoración de activos financieros

A continuación se encontrará una explicación detallada de los populares modelos de valoración de activos CAPM y Fama-French de tres factores, basada en aseveraciones de expertos en el campo, se plantea como base para realizar la aplicación en el mercado financiero mexicano, se presentan supuestos, ecuaciones y gráficas, es importante empaparse de la información teórica para crear un contexto significativo que promueva la comprensión de los casos que se estudian.

3.1. Capital Asset Pricing Model (CAPM)

De acuerdo con el autor Alonso en [6], el modelo CAPM (siglas de Capital Asset Pricing Model), procede de los trabajos del autor Markowitz en 1952 [48], los cuales pretenden seleccionar un portafolio (con condiciones favorables) a través de un problema de maximización de utilidad bajo condiciones inciertas. Para resolver este problema, se supuso que las preferencias de los inversionistas eran externadas y en particular satisfechas con la media y varianza de los retornos (rentabilidad) del portafolio de inversión durante un periodo correspondiente. En otras palabras, el tratamiento de la elección del portafolio óptimo tiene dos pilares, también conocidos como parámetros de la distribución de los retornos: la varianza (desviación estándar al cuadrado) y la media. En consecuencia de la necesidad de demostrar una relación lineal entre dicho retorno y el riesgo de un activo, se toman a consideración algunos supuestos enlistados a continuación:

- Todos los inversionistas tienen un mismo horizonte de inversión, un periodo, en el cual intentan maximizar su riqueza escogiendo portafolios basándose en la media y varianza de la rentabilidad de sus activos.
- No hay costo de transacciones (comprar y/o vender activos), si hubiera, la rentabilidad

de cualquier activo, sería una función que depende de si el inversor lo adquirió antes del periodo de decisión.

- Todos los inversionistas pueden obtener y otorgar préstamos (en otras palabras, invertir en activos libres de riesgo como bonos gubernamentales) por una cantidad ilimitada a una tasa libre de riesgo (R_f). Además, no existen restricciones para la venta en corto de activos, es decir, pueden vender cualquier cantidad de cualquier activo.
- Los inversionistas tienen expectativas homogéneas, es decir, todos los inversionistas tienen estimados (subjetivos) idénticos de las medias, varianzas y covarianzas de los retornos de los activos.
- Todos los activos son perfectamente divisibles y líquidos, es decir, pueden tomar cualquier posición en una inversión, independientemente del tamaño de su riqueza.
- No existen impuestos, en especial impuesto de los ingresos personales en el mercado.
- Las cantidades de los activos están dadas (son fijas).
- Los inversionistas toman las decisiones sólo en términos de valores esperados y desviaciones estándares de la rentabilidad en su portafolio, en otras palabras, cada inversor toma la decisión de los componentes de su portafolio [6].

De acuerdo con los autores Gysel y Samano en [8]: La aseveración en la que se basa el modelo es que la rentabilidad está linealmente relacionada con el riesgo sistemático, este riesgo es medido y denotado por β .

Con estos supuestos como base, el retorno esperado ($E[R_j]$), de cualquier activo j está dado por:

$$E[R_j] = R_f + \lambda * \frac{Cov[R_j, R_M]}{\sigma R_M}. \quad (3.1)$$

En esta ecuación, R_f es la tasa libre de riesgo, λ es la prima de riesgo de mercado por unidad de riesgo y podemos calcularla como sigue:

$$\lambda = \frac{E[R_M] - R_f}{\sigma R_M}. \quad (3.2)$$

Considerando R_M como el retorno del portafolio del mercado, hay una inversión en cada activo que existe en el mercado y esto es directamente proporcional a su valor total.

Por otra parte, $Cov[R_j, R_M]$ refleja el valor en el que varían los retornos del activo j y el portafolio de mercado. Por último, σR_M representa la desviación estándar de los retornos del portafolio de inversión o mercado.

Ahora, siguiendo la idea de los autores Gysel y Samano en [8] y considerando el parámetro β y sus características, es posible observar en términos estadísticos la tendencia de una acción en particular a covariar en el mercado. Por ejemplo, una acción con $\beta = 1$ tiende a fluctuar de la misma manera que el mercado. Si el parámetro $\beta < 1$, las acciones tienen un menor desplazamiento que las del mercado. Para un $\beta > 1$, las acciones tienen un mayor movimiento que las del mercado. Como implicación, los inversionistas que asuman más riesgo, tendrán expectativas más altas en cuestión de rentabilidad. En otras palabras, si un inversionista tiene una cartera no riesgosa ($\beta = 0$), entonces se esperaría que tuviera la tasa de rendimientos sobre activos libres de riesgo. Y si otro inversionista tuviera una cartera riesgosa, con el mismo riesgo que la cartera de mercado, ($\beta = 1$), se esperaría que la tasa de rendimientos fuera igual que la de la cartera de mercado.

En un punto intermedio, se puede considerar a un inversionista que incluye en su cartera una proporción w de activos riesgosos y el resto $1 - w$, de activos poco riesgosos. Para calcular el riesgo de la cartera compuesta se utiliza β (este parámetro es un promedio ponderado de los β de los valores componentes, donde las ponderaciones son las proporciones de la cartera).

Considerando estas deducciones de forma matemática: El parámetro β_p es un promedio ponderado del primer β (correspondiente a la cartera de mercado) y del segundo β (correspondiente a la cartera libre de riesgo), es decir,

$$\beta_p = (1 - w)\beta_A + w\beta_B, \quad (3.3)$$

en donde β_A pertenece al activo libre de riesgo y β_B pertenece al activo con riesgo. Y debido a que el β de los activos de mercado es 1 y el del activo libre de riesgo es de 0, entonces:

$$\beta_p = w. \quad (3.4)$$

Lo cual indica que β_p es igual a la fracción del dinero invertido en la cartera riesgosa.

Además, si de los fondos del inversionista 100 % o menos es invertido en la cartera riesgosa, β estará entre los valores: $0 \leq \beta \leq 1$. Sin embargo, si pide prestado a una tasa con poco riesgo invirtiendo el importe total en la cartera riesgosa, como w es más grande que 1 y $(1 - w)$ es negativo, $\beta \geq 1$.

El rendimiento esperado de la cartera que contiene a los activos (con riesgo y libre de riesgo) es un promedio ponderado de los rendimientos esperados de las dos carteras, es decir,

$$E[R_P] = (1 - w) * R_f + w * E[R_m], \quad (3.5)$$

en donde:

$E[R_p]$ y $E[R_m]$ son los rendimientos esperados de la cartera y el índice de mercado.

R_f es la tasa libre de riesgo.

Sustituyendo la ecuación (3.4) en la ecuación (3.5), se obtiene:

$$E(R_p) = (1 - \beta_p) * R_f + \beta_p * E(R_m), \quad (3.6)$$

es decir,

$$E[R_p] = R_f + \beta_p * [(E[R_M] - R_f)], \quad (3.7)$$

en esta ecuación, $\beta_p = Cov[E[R_p], E[R_M]] / Var(E[R_M])$. Esta es la ecuación identificadora del modelo de valuación de activos de capital, CAPM, utilizada comúnmente en temas de manejo de riesgos, costo de capital (cálculo), análisis de eventos, valoración de activos, etc.

Es posible observar, que de esta manera se puede expresar la ecuación del retorno esperada. Indica que el rendimiento esperado de una cartera debe exceder la tasa de retorno libre de riesgo por una cantidad que es proporcional al β de la cartera, es decir, el retorno esperado depende de la tasa libre de riesgo y un retorno extra por el riesgo asumido ($\beta(E[R_M] - R_f)$).

El retorno extra depende directamente del coeficiente β , una medida del riesgo del activo en función de la variación de sus retornos y los del mercado, para calcularlo se utiliza la covarianza y tiene una dependencia de la diferencia entre el retorno esperado del mercado y la tasa libre de riesgo

$$(E[R_M] - R_f),$$

la cual trivialmente se conoce como **prima de riesgo de mercado**. Por lo tanto, el rendimiento esperado y el riesgo poseen una relación lineal.

De acuerdo con los autores Hincapié y López en [39], una forma alternativa de construir la ecuación del modelo CAPM, es la siguiente:

$$E[R_p] - R_f = \alpha + \beta(E[R_{mt}] - R_f) + \varepsilon_t. \quad (3.8)$$

Esta ecuación es más intuitiva y utilizada para realizar estimaciones, se incluyen dos términos adicionales a los mencionados anteriormente, ε_t , el cual indica la perturbación estocástica presente en el modelo y α , que es el intercepto de la recta de regresión, obtiene el valor de la prima del activo cuando la prima de mercado toma el valor de cero, es decir, que el mercado obtuvo una rentabilidad igual a la tasa libre de riesgo, para algunos autores, este factor no es considerado [57].

Ahora, a manera de explicación, si se considera la frontera eficiente sin pedir prestado ni prestar, es posible observar que en la siguiente figura, el segmento BC es la frontera eficiente, y el segmento ABC es la serie de portafolio de mínima varianza. Debido a que todas las personas tienen distintas expectativas, la frontera eficiente puede diferir en cada inversionista.

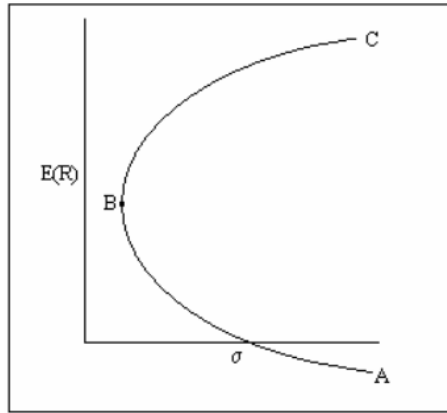


Figura 3.1: Frontera eficiente sin prestar ni pedir prestado.

En otro escenario, si se pudiera pedir prestado o prestar a una tasa con poco riesgo, el portafolio de activos con riesgo podría ser identificado sin tener en cuenta las preferencias de riesgo. Para hacerlo, este portafolio correspondería al punto tangente entre la frontera eficiente original de activos con riesgo y la línea que atraviesa el retorno con poco riesgo (Eje vertical) En la siguiente figura se puede observar este caso, siendo P_i el portafolio i -ésimo del inversor con activos de riesgo. Las preferencias de riesgo se satisfacen combinando el portafolio P_i y pidiendo prestado a una tasa baja de riesgo.

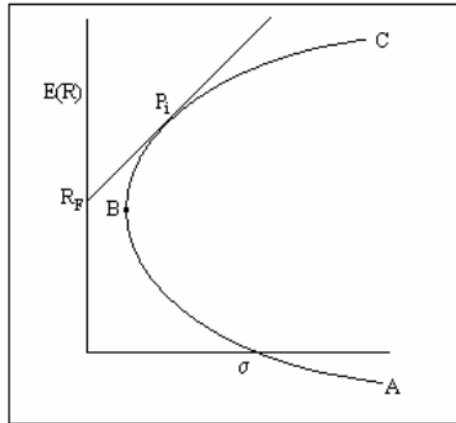


Figura 3.2: Frontera eficiente permitiendo prestar o pedir prestado.

Es importante tener en cuenta la aseveración de que si todos los inversionistas tuvieran el mismo portafolio de riesgo, el portafolio de mercado (que comprende a todos los activos del mercado) estaría en equilibrio. Por consiguiente, todos los inversionistas tendrían una combinación de dos portafolios: portafolio de mercado (M) y un activo con poco riesgo.

En la figura anterior, la línea recta es la línea de mercado de capital. Los portafolios eficientes se encuentran a lo largo de esta línea, de esto se obtiene la siguiente ecuación

$$E(R_{p*}) = R_f + \frac{[E(R_m - R_f)]}{\sigma_m} * \sigma_{p*}, \quad (3.9)$$

con p un portafolio eficiente.

Si se consideran carteras correctamente diversificadas, se utiliza el factor β , recordando los supuestos de homogeneidad y préstamos a una tasa con poco riesgo sin límite, todos conservarían el portafolio de mercado.

Como se puede observar, se ha establecido que todos los inversionistas y portafolios de inversión deben estar a lo largo de la línea de mercado, si estuvieran fuera de la línea podríamos emplear repetidamente arbitraje sin riesgo hasta que todos regresen a dicha línea.

Ahora, considerando la siguiente figura, M denota al portafolio de mercado con $\beta = 1.0$ y rendimiento R_m , si seleccionamos un segundo punto para identificar la línea de mercado, que es cuando $\beta = 0$, tendríamos la siguiente ecuación:

$$E(R_p) = a + b\beta_i. \quad (3.10)$$

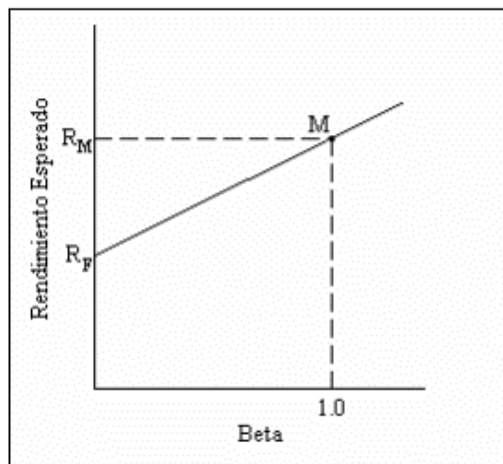


Figura 3.3: Línea de mercado.

De aquí se infiere fácilmente que: Un punto en la línea es el activo libre de riesgo $\beta = 0$, es decir,

$$R_f = a + b(0) = a.$$

Otro punto a considerar es cuando β tiene el valor de 1.0:

$$E(R_M) = a + b(1).$$

$$E(R_M) - a = b.$$

Asociando los términos previos y sustituyéndolos en la ecuación anterior, obtenemos:

$$E(R_p) = R_f + \beta_p[E(R_M) - R_f].$$

Esto es, el rendimiento esperado de cualquier activo o portafolio, en particular, entre dos activos cualesquiera, la relación puede calcularse con la diferencia de β .

Tangiblemente, el inversor tiene ahora un punto importante, puede eliminar el riesgo no sistemático a través de la diversificación pero no puede eliminar el riesgo sistemático (β). Y si tenemos β altas, se esperaría tener rendimientos mayores (aun cuando no sean en todos los intervalos de tiempo).

3.2. Una derivación sencilla del modelo CAPM: problema de optimización

Considerando ahora, la visión del autor Alonso en [6] sobre el inversionista, como es de esperarse, buscará el equilibrio óptimo entre riesgo y retorno, el cuál se ajustará a las restricciones propias del mercado y a su nivel de oposición al riesgo.

Graficando esta percepción, se presenta otra representación a dos dimensiones (el retorno en el eje vertical y el riesgo en el eje horizontal), este último representado por la desviación estándar de los retornos. *La curva FE, o frontera eficiente, representa el conjunto de todos los portafolios riesgosos eficientes, es decir, aquellos que brindan el máximo nivel esperado de retorno, para un nivel de riesgo dado* [6].

En consecuencia, con el autor Alonso en [6], las otras composiciones riesgo-retorno en el mercado se encuentran por debajo de esta.

Por otro lado, las curvas punteadas muestran las combinaciones riesgo-retorno con las que el inversionista obtiene un mismo nivel de utilidad U_i , con $i \in \{1, 2, 3\}$, mejor conocidas como curvas de indiferencia. Como se puede visualizar, el nivel de utilidad incrementa hacia arriba y a la izquierda en el plano, i.e. $U_3 > U_2 > U_1$. Por lo cual, si no hubieran activos libres de riesgo, al invertir en el portafolio B (punto de tangencia entre la frontera eficiente, línea curva FE, y la curva de indiferencia U_1), el inversionista maximizaría su utilidad, en otros términos, el portafolio B maximiza la utilidad del inversionista con el conjunto de portafolios riesgosos.

Ahora consideremos que hay un activo libre de riesgo con un retorno futuro asegurado igual a R_f y además, que el inversionista construye un conjunto de portafolios que compongan (en distintas medidas) un activo o portafolio riesgoso con el activo libre de riesgo. Estas combinaciones son denominadas línea de mercado de capital o recta CML (Capital Market Line) y por estas razones, la elección del portafolio correcto cambiará.

La elección se basa en acceder al portafolio P compuesto en un 50 % por inversión en el activo libre de riesgo y 50 % en el activo M, para alcanzar el nivel de utilidad U_2 .

Se establece que la relación entre riesgo y retorno es de carácter lineal si y sólo si los portafolios incluyen activos riesgosos y al menos uno libre de riesgo.

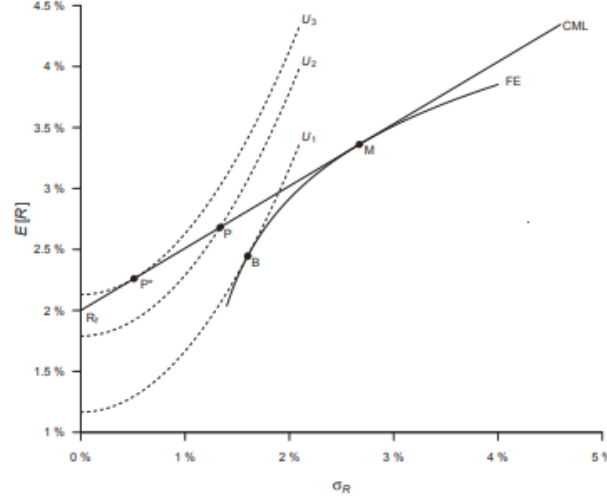


Figura 3.4: Maximización de la utilidad del inversionista con la existencia de un activo libre de riesgo

Ahora consideremos a β_p como

$$\beta_p = \frac{\sigma_{pM}}{\sigma_M^2},$$

a σ_M^2 como la varianza de los rendimientos del mercado, y a σ_{pM} la covarianza de los rendimientos de la cartera y del mercado.

Matemáticamente,

$$\sigma_{pM} = \sum_{i=1}^n [X_{pi} - E(X_p)][X_{Mi} - E(X_M)],$$

es decir,

$$\sigma_{pM} = \rho_{pM} * \sigma_p * \sigma_M,$$

siendo ρ_{pM} el coeficiente de correlación de la cartera frente al mercado.

Al sustituir β en la ecuación CAPM encontramos la siguiente ecuación:

$$E(R_p) = R_f + \left[\frac{E(R_M) - R_f}{\sigma_M} \right] \frac{\sigma_{pM}}{\sigma_M}, \quad (3.11)$$

donde el término $\frac{\sigma_{pM}}{\sigma_M}$ es el riesgo de un portafolio eficiente (cualquiera), denotado previamente por σ_p .

De acuerdo con el autor Alonso en [6], el retorno de cualquier portafolio definido en la línea del capital de mercado CML, está dado por:

$$R_p = w_{R_f} R_f + (1 - w_{R_f}) R_M, \quad (3.12)$$

donde w_{R_f} representa la proporción del portafolio invertida en el activo libre de riesgo y $(1 - w_{R_f})$ representa la proporción del portafolio invertida en M , es decir, el retorno del portafolio es una combinación lineal de los retornos de los dos activos. Entonces, la desviación estándar de la rentabilidad del portafolio estará dada por la siguiente expresión [6]:

$$\sigma_p = \sqrt{w_{R_f}^2 \sigma^2 R_f + (1 - w_{R_f})^2 \sigma^2 R_M + 2w_{R_f}(1 - w_{R_f})\rho R_f R_M \sigma R_f \sigma R_M}. \quad (3.13)$$

en la cual, $\rho R_f, R_M$, representa la correlación entre los retornos del activo libre de riesgo y del portafolio M debido a que el activo libre de riesgo no presentará ningún riesgo asociado a él, por lo tanto, $\rho R_f = 0$. Con esto en consideración, se reduce la expresión como sigue y se motiva la aseveración de que σ_p también es una combinación lineal:

$$\sigma_p = \sqrt{(1 - w_{R_f})^2 \sigma^2 R_M} = (1 - w_{R_f}) \sigma R_M. \quad (3.14)$$

Cada punto sobre la línea CML puede ser considerado una solución factible y está definida por:

$$E[R_j] = R_f + \sigma R \frac{Cov[R, R_M]}{\sigma R_M}. \quad (3.15)$$

siendo $\sigma R < \sigma R_M$,

En este sentido, el portafolio óptimo se representa por el punto P^* , el cual maximiza la utilidad bajo las condiciones de elección del inversionista llegando al nivel U_3 .

Siguiendo la idea del autor, si tomamos otro portafolio distinto a M que esté sobre la frontera eficiente e introducimos el activo libre de riesgo, este portafolio pasará a ser ineficiente.

Ahora consideremos niveles de riesgo $\sigma < \sigma M$ que estén sobre la línea de mercado de capitales; estos brindan un nivel de retorno esperado mayor, i.e., los portafolios que yacen sobre FE y por debajo de M no generan el máximo retorno esperado para un nivel de riesgo dado; por lo tanto, no son óptimos.

Para un inversionista arriesgado que puede endeudarse a tasa R_f , existe la posibilidad de acceder a portafolios sobre la CML con un riesgo mayor al de M . Esta es una de las

combinaciones que resultan factibles con un nivel mayor de retorno respecto al esperado por los que se encuentran en la frontera eficiente y con niveles de riesgo $\sigma > \sigma_M$. Por lo tanto todos los inversionistas elegirían portafolios que se encuentren sobre la línea del mercado de capitales (activos del portafolio M y en el libre de riesgo).

En otras palabras, para un mercado en equilibrio, M debería ser el portafolio de mercado, es decir un portafolio eficiente que incluya en proporción a su valor, todos los activos del mercado, ya que si no incluyera a alguno, este disminuiría mucho de precio, al punto de que como nadie invertiría en él, se volvería rentable adquirirlo. Como inferencia, el retorno esperado de cualquier portafolio eficiente P corresponderá al retorno esperado del portafolio de mercado.

Esta situación aplicaría para todo activo el cual, en equilibrio, debería estar incluido en el portafolio M . De acuerdo con lo anterior, se infiere que el retorno esperado de cualquier portafolio eficiente P estará relacionado con el retorno esperado del portafolio de mercado:

$$E[R_P] = R_f + \lambda \sigma R_P. \quad (3.16)$$

En la ecuación anterior, lambda representa la pendiente de la línea de mercado de capitales y es posible calcularla como sigue:

$$\lambda = \frac{E[R_M] - R_f}{\sigma R_M}. \quad (3.17)$$

El problema que surge es derivar la ecuación para los retornos esperados en equilibrio de un activo individual j por cuanto ya derivamos la ecuación de los retornos esperados de un portafolio.

En resumen, es común que los inversionistas tomen la varianza o desviación estándar de los retornos del portafolio para medir correctamente el riesgo. Sin embargo, los resultados muestran que la medida más apropiada para el riesgo de un activo individual es su covarianza con el portafolio de mercado $Cov[R_j, R_M]$ no siendo así su varianza $\sigma^2 R_j$. Esto lleva a considerar estrictamente la diversificación ya que es posible eliminar los efectos de la varianza de un activo en todo el portafolio pero no siendo así, con los demás activos de un portafolio.

Si el modelo se cumpliera estrictamente en la realidad, aceptar más riesgo implicaría una mayor rentabilidad, pero esto sólo ocurriría para el riesgo sistemático, que no puede eliminarse por diversificación (recordando que este está relacionado con la marcha del mercado). De lo contrario, si no se cumpliera estrictamente, los títulos podrían ubicarse alrededor de la línea propuesta debido a la falta de información precisa sobre la cartera de mercado, que solo cuenta con estimaciones. En caso de que el mercado esté desviado y no opere de manera eficiente, algunos títulos podrían encontrarse por encima de la línea (indicando que rinden más de lo esperado y están subvaluados), mientras que otros podrían situarse por debajo de la línea (indicando que están sobrevalorados). Según la lógica del modelo, esto daría lugar a la existencia de títulos subvaluados.

3.3. El modelo de 3 factores de Fama y French

El presente texto resulta de la comprensión del modelo de 3 factores de Fama y French, tiene como objetivo principal presentarlo, describirlo en detalle, explicar sus componentes y analizar su importancia en la actualidad.

Los autores Eugene F. Fama y Kenneth French en 1992, añadieron conocimiento al modelo CAPM debido a que además del retorno de mercado, introdujeron variables como el valor de mercado del patrimonio de una firma y una razón: $\frac{Book}{Market}$, mejor conocida como Book to Market para ayudar a explicar los retornos accionarios, i.e., el retorno de una acción no es descrita sólo en una dirección, sino en varias (fuentes de riesgo al invertir en el mercado accionario) [28].

Ahora, concordando con el autor Alonso en [6], el número de acciones en circulación por el precio de la acción, mejor conocido como el valor de mercado del patrimonio, trabaja como una variable explicativa importante de los retornos. Es preciso mencionar que existe una relación inversa entre el retorno y el tamaño de las firmas. La razón mencionada en el primer párrafo, $\frac{Book}{Market}$ (B/M), es una medida de dificultades financieras que puede presentar el inversionista. Un ejemplo de esto es que firmas con razones altas B/M tienen problemas financieros o tienen una mala proyección (en opinión del mercado) por lo cual, el valor de mercado de la acción de dichas firmas es pequeño, lo cual presenta un riesgo elevado, retorno mayor y costos de capital (promedio ponderado) más elevados.

Por otra parte, razones bajas B/M muestran una mejor proyección para las firmas, lo cual se relaciona con firmas sólidas y en crecimiento. Sin embargo, Fama y French advierten que el riesgo que define la variable B/M tiene otros factores determinantes como las fluctuaciones irracionales del mercado respecto al futuro y prospectos de una compañía.

En 1993, los autores estudiaron el retorno mensual mediante el análisis de 25 portafolios de acciones, creados a partir de la intersección de acciones pertenecientes al NYSE, Amex y Nasdaq, ordenadas según el tamaño y la razón B/M del patrimonio (5 quintiles de tamaño y 5 quintiles según la razón B/M del patrimonio). Los retornos de estos portafolios y el exceso de la tasa libre de riesgo constituyen la variable dependiente de la investigación.

Agregando a lo anterior, como variables independientes o explicativas, utilizaron el exceso de rendimiento calculado como la diferencia entre el rendimiento esperado de un activo y el rendimiento de la tasa libre de riesgo, de la misma forma como lo establece el modelo CAPM. La primera variable, $R_t^{(SMB)}$, muestra la diferencia entre el retorno esperado de las compañías basado en su tamaño. La segunda variable, $R_t^{(HML)}$, denota la compensación por invertir en compañías con proyecciones bajas o dificultades financieras, se calcula como la diferencia entre el retorno de firmas con alta y baja razón B/M.

Para llegar a esta conclusión de variables, se dividió por la mitad la muestra de firmas inicial según el valor de mercado del patrimonio con el fin de clasificarlas en pequeñas y grandes. Las firmas pequeñas (Small, s) se definen cuando el valor de mercado del patrimonio está por debajo de la mediana del valor del patrimonio de las firmas listadas en el NYSE. Las firmas

grandes (Big, b) se definen cuando el valor de mercado del patrimonio está por encima de dicha mediana.

Otra forma de agruparlas fue dividir las firmas en tres grupos de acuerdo con el valor de la razón B/M para cada una de ellas:

Primer grupo: 30 % de las firmas con el indicador B/M más bajo (Low, L.).

Segundo grupo: 40 % de las firmas con los indicadores B/M medios (Medium, M.).

Tercer grupo: 30 % de las firmas restantes con los indicadores más altos del B/M (High, h).

Con la intersección de estos subgrupos, crearon seis portafolios ordenados según tamaño y los tres grupos clasificados según B/M.

En particular, los seis portafolios que comprenden todas las acciones de la muestra son: S/L, S/M, S/H, B/L, B/M y B/H,

donde:

S = Small.

B = Big.

L = Low.

M = Medium.

H = High.

Por lo cual, la variable $R_t^{(SMB)}$ se calcula mensualmente como la diferencia entre el retorno promedio simple de los tres portafolios que contienen acciones pequeñas (S/L, S/M y S/H) y el retorno promedio simple de los tres portafolios que contienen las acciones más grandes (B/L, B/M y B/H).

Por otra parte, la variable $R_t^{(HML)}$ se calcula como la diferencia mensual entre el retorno promedio simple de los dos portafolios con valores mayores de la razón B/M, es decir, (S/H y B/H) y el retorno promedio simple de los dos portafolios con los valores más bajos (S/L y B/L).

El preámbulo de su modelo fue estimar el modelo CAPM para cada uno de los 25 portafolios ($i = 1, 2, 3, \dots, 25$) a través de un análisis de series de tiempo bajo la siguiente ecuación:

$$(R_{it} - R_{ft}) = \beta_1 + \beta_2(R_{Mt} - R_{ft}) + \varepsilon_t. \quad (3.18)$$

Al realizar esta estimación, observaron que el CAPM capturaba la variación de los retornos de los 25 portafolios accionarios con R^2 , y estas se encontraban entre 0.61 y 0.92, por lo cual se creó el modelo de tres factores de Fama y French, el cual tiene variables explicativas $R_t^{(SMB)}$ y $R_t^{(HML)}$ y se puede obtener a través de:

$$(R_{it} - R_{ft}) = \beta_1 + \beta_2(R_{Mt} - R_{ft}) + \beta_3 R_t^{SMB} + \beta_4 R_t^{HML} + \varepsilon_t. \quad (3.19)$$

Tras este modelo de regresión múltiple, los autores se percataron de mejoras en el R^2 , que ahora se encontraba entre 0.83 y 0.97 para los diferentes portafolios. Además, los coeficientes (β_3), que miden la sensibilidad al tamaño, disminuían en proporción al aumento del tamaño de las firmas, lo cual lleva a la conclusión de una relación inversa entre el retorno y el tamaño. De manera similar, se observa que dentro de cada grupo de tamaño quintil, los coeficientes (β_4) se incrementan a medida que se pasa del quintil más bajo al más alto (según la razón B/M), por lo tanto, las firmas con razones B/M más altas tienen mayor riesgo y representan un retorno mayor.

De acuerdo con el autor Hincapié en [39], una manera alternativa de representar este modelo es la ecuación (3.20):

$$(R_{it} - R_{ft}) = \alpha + \beta(R_{Mt} - R_{ft}) + s_i SMB_t + h_i HML_t + e_i, \quad (3.20)$$

donde:

β : En este caso, no se considera como la pendiente de la recta de regresión debido a que se presentan dos nuevas variables. Sin embargo, es posible interpretarlo como la sensibilidad de la prima del activo con respecto a cambios en la prima de mercado, considerando a lo demás constante.

SMB_t : Variable exógena calculada por medio de la consideración del tamaño de las empresas que conforman el mercado.

HML_t : Variable exógena calculada por medio de la consideración de la relación Book/Market de las empresas que conforman el mercado.

s_i, h_i : Resultan de realizar la regresión y permiten relacionar la variable endógena con las dos variables agregadas al modelo.

De acuerdo con el autor, es posible definir a estas variables como:

- **SMB**: es la diferencia en la rentabilidad de portafolios de grandes empresas con portafolios de pequeñas empresas. Es una variable que busca relacionar el factor del tamaño de las empresas y es la diferencia entre el retorno promedio de los portafolios con empresas pequeñas y el retorno promedio de los portafolios con empresas grandes.
- **HML**: es la diferencia en la rentabilidad entre portafolios de empresas con alta relación Book/Martek y portafolios de empresas con baja relación Book/Market. Tiene el objetivo de relacionar el factor de la proporción Book/Market con la diferencia entre el retorno promedio de los portafolios de empresas con alta y baja relación de este tipo.

Para calcular estas variables se sigue la siguiente metodología:

1. Se definen las acciones para la estimación del modelo.

2. Construcción de una tabla que permite construir 6 portafolios que combinan las siguientes características:

Tabla 3.1: Forma de agrupación de los portafolios respecto a la relación Book/Market [39].

		Tamaño	
		Pequeño	Grande
Relación Book to Market	Bajo	Portafolio 1	Portafolio 4
	Medio	Portafolio 2	Portafolio 5
	Alto	Portafolio 3	Portafolio 6

Para realizar la conformación de los seis portafolios, se calculan los percentiles 33 y 66 por ciento para la relación Book/Market y el percentil al 50 por ciento para el caso del tamaño o capitalización bursátil.

Capitalización bursátil = Precio de la acción * Número de acciones en circulación.

3. Se calculan los retornos de cada portafolio (todas las empresas ponderan en la misma magnitud), a través de las siguientes ecuaciones:

$$SMB = \frac{Port1 + Port2 + Port3}{3} - \frac{Port4 + Port5 + Port6}{3}. \quad (3.21)$$

$$HML = \frac{Port1 + Port4}{2} - \frac{Port3 + Port5}{2}. \quad (3.22)$$

siendo $Porti$ con $i = \{1, \dots, 6\}$, los portafolios conformados por cuatro empresas seleccionadas con base en su relación descrita en la tabla 3.1.

4. Con estas variables, se puede estimar el modelo.

Para determinar el mejor modelo empleado para describir los retornos, se basaron en aquel que presentara los interceptos más bajos, e idealmente, que su conjunto no fuera estadísticamente distinto a cero.

Siguiendo con [6], para el caso del CAPM, 11 de 25 interceptos fueron estadísticamente distintos a cero, por lo que este modelo recién elaborado, presentaba deficiencias para explicar los retornos accionarios, en mayor parte en los portafolios que incluían las acciones con valores altos de la razón B/M. Otra de las observaciones dadas fue que los interceptos de este modelo, modelo de tres factores, fueron fundamentalmente menores que los del modelo CAPM, y únicamente 3 de los 25 interceptos se encontraron estadísticamente diferentes a cero.

Posteriormente, en [1], hicieron uso del test estadístico GRS que se emplea para probar la hipótesis nula: 32 interceptos (25 portafolios de acciones considerados con anterioridad y 7 portafolios de bonos) son conjuntamente iguales a cero versus la hipótesis alterna: al menos uno de los interceptos es distinto de cero.

Por evidencias encontradas, se rechaza la hipótesis de que $R_M - R_f$ es suficiente para explicar los retornos con un nivel de confianza del 99 %, por lo tanto, a pesar de que el modelo de los tres factores presenta interceptos menores, este test rechaza la hipótesis nula de que todos los interceptos son iguales a cero [6]. Esto no demerita el modelo ya que este explica adecuadamente los retornos accionarios y en la práctica, un número muy bajo de interceptos es estadísticamente distinto a cero. También puede ser utilizado en la selección y evaluación del desempeño de portafolios, estudio de eventos y estimación del costo de capital de las compañías como el modelo CAPM.

Capítulo 4

Estudios de caso

En este capítulo, se exponen los resultados de este trabajo de tesis, a través de tres estudios de caso que se diseñaron, calcularon, desarrollaron e interpretaron con base en la teoría previamente discutida sobre el modelo CAPM y el modelo Fama-French de tres factores.

1. El primer caso se basa en el modelo CAPM estándar.
2. El segundo considera ajustes sugeridos, como la inclusión de una variable instrumental, el índice de precios y cotizaciones, y el retorno del mercado.
3. El tercer caso utiliza el modelo Fama-French de tres factores.

La implementación de los códigos para los tres casos es empleando el software RStudio por sus ventajas de ser orientado hacia la programación estadística; también se emplea a Microsoft Excel como herramienta complementaria.

Los datos se obtuvieron mediante una investigación y recopilación de fuentes electrónicas confiables como Yahoo en [77] e Investing en [43]. Posteriormente, se llevó a cabo la segmentación, procesamiento de pruebas, modelos y adaptaciones, así como el reacondicionamiento de los datos según fuera necesario. Finalmente, se procedió a la interpretación de los resultados mediante un análisis detallado y comparativo, seguido de su valoración y presentación.

4.1. Caso 1: Aplicación del modelo CAPM en el mercado financiero mexicano

En primer lugar, se realizó el caso de estudio de la aplicación estricta del modelo de valuación de activos CAPM, en el contexto financiero mexicano, utilizando tres variables principales

descritas a continuación como ejecución de la primera fase de la metodología, recolección y limpieza de datos:

Primera variable: el precio de cierre de 28 acciones pertenecientes a la Bolsa Mexicana de Valores [77]¹. Es importante destacar que la razón de utilizar precios de cierre es evitar la variación que experimentan las acciones en el transcurso del día, además de la incorporación de empresas que pertenecen a diferentes sectores, entre los que se incluyen: sector de aeropuertos, medios de comunicación, alimentos y bebidas, construcción y materiales, financiero, retail, salud, materiales básicos e infraestructura.

El pseudocódigo para la obtención de los precios se muestra en el algoritmo 1 y el código completo se visualiza en el **Apéndice A**.

Algorithm 1 Obtención de precios

Require: *cartera*

```

1: listaAcciones  $\leftarrow$  list
2: for accion in cartera do
3:   listaPreciosCierre[[accion]]  $\leftarrow$  getSymbols(accion, src = "yahoo", from = "2014-01-01", to = "2023-01-01", env = globalenv())
4: end for
5: return listaPreciosCierre

```

Segunda variable: el modelo original, considera al rendimiento del portafolio de mercado como segunda variable, sin embargo, resulta complicado identificarlo por completo, ya que existe una gran cantidad de títulos y presentan variaciones en los periodos de tiempo; por tal razón, la segunda variable es el rendimiento del IPC de la Bolsa Mexicana de Valores considerado como el rendimiento de mercado, asimismo, se pronostica este rendimiento para el periodo siguiente, a manera de realizar los cálculos necesarios para las fases siguientes que comprenden la comparación entre el rendimiento esperado y el rendimiento real.

Tercera variable: la tasa libre de riesgo, escogida como la tasa de los CETES con plazo de inversión a 28 días, debido a que esta tasa compagina en gran medida con los inversionistas conservadores que prefieren un retorno seguro con nulo riesgo. Todos los datos se recolectaron en el periodo del 01 de enero de 2014 al 01 de enero de 2023 a través de los recursos electrónicos Yahoo Finance, BANXICO y la librería quantmode de RStudio que permite extraer los datos del recurso electrónico y colocarlos en un panel bidimensional, conocido en el entorno de RStudio como DataFrame. En ese momento, se procedió a realizar la limpieza de datos para conseguir homogeneidad en la estructura de la información a introducir en los modelos, para ejecutar esta limpieza, se completaron los datos faltantes a través de interpolación en el lenguaje de programación RStudio, se planteó un ajuste en el número de acciones a considerar por motivo de huecos frecuentes y volatilidad en los datos, de manera que los resultados encontrados no presentaran discrepancias generadas por errores particulares y se alineó el periodo

¹Listado de las 28 acciones: omab, gapb, tlevisapco, asurb, grumab, cemexcpo, alfaa, alsea, AC, bimboa, bolsaa, elektra, gfnorteo, femsaubd, labb, gmexicob, gcarsoa1, gcc, chdrauib, gfinburo, kimbera, megacpo, orbia, pinfra, walmex, q, vws, volara.

de tiempo en las tres variables para realizar los cálculos en un entorno seguro, exacto y completo.

Con el fin de obtener los rendimientos de las acciones y del IPC como rendimientos del mercado en el periodo de tiempo considerado, del 01 de enero de 2014 al 01 de enero de 2023, se utilizó la herramienta de RStudio "lag" y se concentró en un DataFrame para facilitar su manejo. Además, se obtuvo la misma información para el periodo del 02 de enero al 07 de enero de 2023 y se compiló en un nuevo DataFrame para su comparación posterior. A continuación, se muestra en el algoritmo 2 el pseudocódigo para el cálculo del rendimiento de las variables mencionadas, el código completo se presenta en el **Apéndice A**.

Algorithm 2 Cálculo de rendimientos

Require: *cartera*

```

1: for accion in cartera do
2:   rendimientos_combinados  $\leftarrow$  ROC(precios_cierre_combinados, type =
     'continuous', na.pad = FALSE)
3: end for
4: return rendimientos_combinados

```

Posteriormente, se inicia la segunda fase, segmentación y pruebas, en la cual se construyó la variable endógena para cada una de las acciones restando la tasa libre de riesgo CETES de los rendimientos diarios de cada activo en el periodo considerado, obteniendo la prima de riesgo de cada acción, también conocida como exceso de rendimiento. Se conformó el modelo con los elementos extraídos, considerando los rendimientos de las acciones del mercado, la tasa libre de riesgo y el parámetro a estimar, β . Los valores de los parámetros β calculados, fueron agrupados en un vector para su posterior utilización en el análisis.

En referencia a la tercera fase del estudio (reprocesamiento), se determinó que no era necesario reajustar el modelo en este caso específico. Esto se debe a que se llevó a cabo una prueba de estacionariedad para cada serie de rendimientos, con el propósito de determinar si era requerida la aplicación de diferenciación para alcanzar dicha estacionariedad, sin embargo, los resultados de estas pruebas fueron positivos, mostrando que todas las series de rendimientos son estacionarias en su forma original. En el algoritmo 3, se presenta el pseudocódigo de la prueba de estacionariedad, en el **Apéndice A**, es posible encontrar el código completo.

Algorithm 3 Cálculo de rendimientos

Require: *exceso_rendimientos*

```

1: for col in colnames(exceso_rendimientos) do
2:   adf_test[col]  $\leftarrow$  adf.test(exceso_rendimientos[, col], alternative = 'stationary')
3: end for
4: return adf_test

```

Con el fin de implementar la cuarta fase (análisis de resultados) mencionada en la metodología, para comprobar el modelo y tener un punto de comparación con los demás casos, a través del modelo se predijo el rendimiento esperado promedio de cada una de las acciones en

el periodo del 02 al 07 de enero de 2023. La quinta y última fase, valoración y presentación de resultados, de igual manera que la comparativa obtenida entre los tres casos de estudio, se presentan al término de la explicación de casos en la Tabla 4.5.

4.2. Caso 2: Aplicación del modelo CAPM con variaciones en el mercado financiero mexicano

Como segundo caso, se consideró una modificación al modelo CAPM tomando en cuenta las limitaciones que presenta el modelo original y los agregados que realizan los autores comúnmente, como el caso de los autores Ramírez y Serna en [37], teniendo como propósito comparar los resultados y concluir si la elección de las variables modificadas es adecuada o presenta áreas de oportunidad.

Respecto a la primera fase y a efectos de realizar un caso de estudio práctico para el modelo CAPM en el mercado financiero mexicano con las variaciones a presentar, se realizó la recolección de los precios de todas las acciones que cotizan en la Bolsa Mexicana de Valores desde el 01 de enero de 2014 al 01 de enero de 2023, además de la información los CETES como un proxy de la tasa libre de riesgo en México, el IPC de la Bolsa Mexicana de Valores, tipo de cambio para México, circulación de la moneda metálica de 10 pesos, tasa TIIE, índice Standard & Poor's 500 (S&P500) y valores del PIB. A partir de esta información, se procedió a calcular las rentabilidades de cada una de las acciones utilizando, como en el caso anterior, el precio de cierre diario y de los índices mencionados, el valor diario de igual manera.

En un principio, se obtuvo la información de aproximadamente 100 acciones; sin embargo, factores como la desigualdad de fechas completas en la pertenencia en el índice, el estancamiento o los huecos de información en el periodo de tiempo, redujeron la cifra a 28 acciones. Por otra parte, el periodo de tiempo contemplado para el estudio es significativamente mayor, 9 años exactamente.

Dada la disponibilidad de la información y en consonancia con el enfoque metodológico de Black, Jensen y Scholes, se implementaron dos modificaciones importantes en relación con el modelo original. La primera consiste en sustituir el uso de portafolios de acciones por el análisis individual de las acciones, esto debido a la limitada cantidad de acciones disponibles.

La segunda modificación es para la consideración de una variable instrumental. Los autores Ramírez y Serna en [37] utilizan el β estimado del periodo pasado (cinco años) como variable instrumental del ejercicio de regresión, esto con el fin de eliminar el sesgo en la estimación, sin embargo, el presente caso de estudio, empleó como variable instrumental el PIB.

A manera de implementar la segunda fase del modelo (segmentación y pruebas), y justificar la selección del PIB como variable instrumental, se realizaron 3 pruebas estadísticas y econométricas:

4.2. CASO 2: APLICACIÓN DEL MODELO CAPM CON VARIACIONES EN EL MERCADO FINANCIERO

- Relevancia: Se procedió a realizar una evaluación de la correlación entre el (PIB) y la variable endógena, representada por el rendimiento del mercado financiero. El propósito de este análisis fue validar la pertinencia del PIB dentro del modelo, para tal fin, se empleó la prueba del coeficiente de correlación.

Los resultados obtenidos indicaron que a un nivel de confianza del 95 %, la correlación entre el PIB y el rendimiento del mercado financiero oscila entre 0.465 y 0.517. esto significa que hay una correlación positiva moderada, lo que implica que, en general, un mayor crecimiento económico, medido por el PIB, tiende a estar asociado con un rendimiento más alto en el mercado financiero. Posteriormente, esta idea se refuerza a través de la correlación puntual estimada entre ambas variables que se obtuvo por 0.491, por lo que en términos más concretos, se evidencia una asociación significativa entre estas dos variables. Se presenta en el algoritmo 4 el pseudocódigo para el desarrollo de la prueba de relevancia, el código completo se encuentra en el **Apéndice A**.

Algorithm 4 Prueba de relevancia

Require: *pib_completo*

1: *prueba_relevancia* $\leftarrow$ `cor.test(datos_mexico$PIB.Valor,datos_mexico$Rendimiento.close)`

- Prueba de exogeneidad: Se buscó asegurar que el PIB sea exógeno en el modelo, es decir, que no esté correlacionado con los términos de error, para realizar esta prueba se incluyó el PIB en una regresión lineal, con el fin de verificar si sus residuos no están correlacionados con los términos de error del modelo principal. Por esta razón, se crearon dos modelos:

El primero, modelo principal, el cual relaciona al rendimiento del mercado con el PIB, tipo de cambio, tasa TIIE y la circulación de la moneda de 10 pesos, es importante mencionar que estas variables fueron elegidas por tener relación con el rendimiento del mercado y para mejorar la capacidad explicativa del modelo, por ejemplo, el tipo de cambio es una variable económica general; las variaciones del rendimiento del mercado financiero mexicano pueden influir en el tipo de cambio a través de sus efectos en la competitividad de un país y en las expectativas de inflación [62]. En cambio, la tasa TIIE posee una relación inversa con el rendimiento del IPC en la Bolsa Mexicana de Valores. Cuando la TIIE aumenta, los costos de financiamiento para consumidores y empresas tienden al alza, lo cual conduce a menor demanda y a la motivación de establecer los precios más bajos, contribuyendo a la contención de la inflación [51]. Por último, la circulación de la moneda de 10 pesos no tiene un impacto directo sobre el IPC, sin embargo, cambios en esta circulación pueden reflejar cambios en la liquidez del sistema financiero mexicano, en particular con su rendimiento [40].

El segundo, modelo de regresión adicional, relaciona únicamente la variable del PIB con las variables establecidas en el modelo previo, es decir, el tipo de cambio, la tasa TIIE y la circulación de la moneda de 10 pesos.

La diferencia entre los modelos es la inclusión del rendimiento del mercado en el primer caso y la exclusión del mismo en el segundo, esto con el fin de considerar el contexto completo en el modelo principal y el modelo ajustado en el segundo.

Posteriormente, se procedió a la obtención de los residuos y al cálculo de la correlación entre el modelo principal y el modelo adicional. Se llevó a cabo una prueba de correlación, en la cual la hipótesis nula se estableció como la ausencia de correlación entre los residuos de ambos modelos. El resultado obtenido a partir del coeficiente de correlación de Pearson fue de $-6.250302e-17$, lo que indica una correlación extremadamente baja entre los dos conjuntos de residuos. Por otra parte, el valor p asociado a la prueba fue 1, lo que sugiere que no hay suficiente evidencia para rechazar la hipótesis nula de exogeneidad, es decir, no existe una correlación significativa entre los residuos. Por lo tanto, se concluye que la variable PIB es exógena en el modelo.

- **Prueba de Exclusión Restringida:** Se comprobó que la inclusión del PIB mejora significativamente la calidad del modelo, a través de la comparación de los estadísticos de ajuste del modelo con y sin el PIB, utilizando la prueba de razón de verosimilitudes “*LR test*” y considerando la hipótesis nula como “El modelo restringido (sin la variable PIB) es suficiente para explicar los datos”, el dato obtenido del p -value, es $Pr(> \chi^2) = 2.2e-16$, al ser un valor muy bajo, prácticamente cero, indica que hay evidencia extremadamente fuerte en contra de la hipótesis nula, es decir, sugiere que la inclusión de esta variable mejora significativamente la calidad del modelo.

Es posible conocer en el algoritmo 5 el pseudocódigo para el desarrollo de la prueba de exclusión, el código completo se encuentra en el **Apéndice A**.

Algorithm 5 Prueba de Exclusión

Require: *modelo_restringido, modelo_completo*

1: *prueba_exc* $\leftarrow$ *lrtest(modelo_restringido, modelo_completo)*

En conclusión, estas tres pruebas econométricas motivan la aseveración de que el PIB puede ser considerado una variable instrumental para el mercado financiero mexicano.

Con el propósito de proporcionar información crucial sobre la estacionariedad de las variables, en referencia al ejercicio de series de tiempo, se realiza un análisis de cada exceso de rentabilidad, empleando una prueba de raíces unitarias para obtener dicho exceso de rentabilidad. Se considera como variable de referencia al índice S&P 500 y como hipótesis nula la presencia de raíces unitarias, los datos arrojados de la prueba de raíces unitarias Dickey Fuller (ADF) Test mencionan que el p -value es muy bajo ($2.2e-16$), por lo cual existe evidencia para rechazar la hipótesis nula de la presencia de raíces unitarias, es decir, es estacionaria.

Además, después de diferenciar una vez la serie a través de la herramienta de RStudio “*lag*”, el coeficiente ‘*z.lag.1*’ obtenido es -1.017 , el cual indica la magnitud y dirección del efecto de la diferenciación en la serie de datos, al ser negativo, sugiere que la serie de datos tiende a revertirse hacia su media o nivel de equilibrio después de experimentar un choque o cambio, es decir, si la serie experimenta un aumento inusual en un periodo, es probable que en el siguiente periodo haya una disminución para regresar a su tendencia original. Esto se interpreta como un indicador de estacionariedad en la serie, lo que significa que la serie no exhibe tendencias

4.2. CASO 2: APLICACIÓN DEL MODELO CAPM CON VARIACIONES EN EL MERCADO FINANCIERO

persistentes o crecimiento exponencial a lo largo del tiempo [58].

Al mismo tiempo, considerando también el p -value mencionado, hay evidencia significativa de estacionariedad en la serie. Otro valor a considerar es el valor del estadístico, el cual es menor que el nivel de significancia 5 %, esto respalda la evidencia en contra de la presencia de raíces unitarias.

Para introducir la fase 3 de reprocesamiento, se estimaron dos modelos:

El primero se construyó utilizando el promedio de la cartera completa compuesta por las 28 acciones, permitiendo una evaluación agregada del comportamiento del mercado. El segundo modelo, en cambio, fue elaborado de manera individual para cada una de las 28 acciones que conforman la cartera, lo que permitió un análisis detallado del rendimiento específico de cada acción dentro del mercado. En otras palabras, el primero se centra en el rendimiento de la cartera completa, es decir, las 28 acciones, relacionado con el valor de la tasa libre de riesgo CETES, el rendimiento del mercado financiero mexicano, la variable instrumental PIB, la TIIE, ya que refleja las condiciones de financiamiento en el mercado interbancario.

Adicionalmente a la inclusión de una variable Dummy, como sugieren Agueldo y Uribe en [2], para establecer el cambio estructural en el mercado accionario mexicano asociado a la crisis ambiental, COVID 2019, este periodo de pandemia que representó en la sociedad, mercados financieros, tasas y factores econométricos un importante desajuste, por lo cual, se consideró esta variable con el valor de 1 en el intervalo del 16 de febrero del 2020 al 01 de enero de 2023 y 0 en el intervalo del 01 de enero de 2014 al 15 de febrero de 2020 , matemáticamente:

$$(R_t - R_{ft}) = \alpha + \beta_1(R_{mt} - R_{ft}) + \beta_2 d_t(R_{mt}^f - R_{ft}^f) + \varepsilon, \quad (4.1)$$

donde:

R_t : Rentabilidad de la cartera en el momento t .

R_{ft} : Tasa libre de riesgo en México (CETES) como un proxy, en el momento t .

α, β_1, β_2 : Parámetros a estimar.

d_t : Variable dummy que toma el valor de 0 en el periodo del 01 de enero de 2014 al 15 de febrero de 2020, y de 1 para el periodo del 16 de febrero de 2020 al 01 de enero de 2023.

R_{mt}^f : Variable instrumental PIB en el momento t .

R_{ft}^f : Tasa TIIE en el momento t .

ε : Término de perturbación estocástica.

Análogamente, se puede expresar como

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 X_3 + \varepsilon, \quad (4.2)$$

donde:

Y : Diferencia entre el rendimiento de la cartera en el tiempo t y la tasa libre de riesgo CETES.

β_0 : Intercepto.

β_1 : Sensibilidad al exceso de rendimiento de mercado.

X_1 : Diferencia entre rendimiento de mercado y tasa CETE, comúnmente conocido como exceso de rendimiento.

β_2 : Sensibilidad a la diferencia entre la variable instrumental PIB y la tasa TIIIE en el momento t .

X_2 : Variable dummy que indica el periodo de cambio estructural.

X_3 : Diferencia entre variable instrumental PIB y la tasa TIIIE en el momento t .

ε : Término del error.

Es posible notar, que es un modelo de regresión lineal múltiple. Por otro lado, el segundo modelo, en el cuál se consideró el rendimiento de cada acción perteneciente a la cartera relacionada con el valor de la tasa libre de riesgo CETES, el rendimiento del mercado financiero mexicano, la variable instrumental PIB, la TIIIE y la inclusión de una variable Dummy bajo las mismas condiciones del modelo anterior, es decir:

$$(R_{it}-R_{ft}) = \alpha_1 + \beta_{i1}(R_{mt} - R_{ft}) + \beta_{i2}d_t(R_{mt}^f - R_{ft}^f) + \varepsilon, \quad (4.3)$$

siendo $i \in \{1, \dots, 28\}$ las acciones pertenecientes a la cartera estudiada.

Para que el modelo de regresión múltiple sea válido, se realiza la comprobación de las siguientes hipótesis:

- Linealidad: A manera de comprobación, la relación existente entre las variables independientes y la variable dependiente es lineal. Verificamos esta hipótesis mediante la inspección de los gráficos de dispersión de los residuos contra los valores ajustados, buscando patrones no lineales, es posible notar que dichos residuos se distribuyen aleatoriamente alrededor de la línea horizontal ($y = 0$) sin formar patrones sistemáticos.
- Homogeneidad: Para comprobar que el valor promedio del error sea cero, es decir, $E(\varepsilon) = 0$ se utilizó la observación de los residuos; debido a que la media de los residuos es cercana a cero, este supuesto se cumple.
- Homocedasticidad: Con el fin de comprobar que la varianza de los errores es constante a lo largo de todas las observaciones, es decir, $Var(\varepsilon) = \sigma^2$ Se observaron los gráficos de

4.2. CASO 2: APLICACIÓN DEL MODELO CAPM CON VARIACIONES EN EL MERCADO FINANCIERO

dispersión de los residuos, se concluye que no existen patrones de aumento o disminución en la dispersión de los errores a medida que los valores predichos aumentan.

- Independencia: El término del error ε_i no está correlacionado con las variables independientes observadas $x_i, x_2, \dots, x_k$, es decir, $E[\varepsilon_i | x_i, x_2, \dots, x_k] = 0$ donde $x_1, x_2, \dots, x_k$ son las variables independientes del modelo. Esta hipótesis se comprobó mediante la verificación de la independencia temporal en los datos, asegurando que no hay autocorrelación significativa en los residuos usando la prueba de Durbin-Watson.
- Normalidad: Los errores deben seguir una distribución normal, $\varepsilon \sim N(0, \sigma)$. Para evaluar esta hipótesis, se realizó una prueba de normalidad, Shapiro-Wilk, y se inspeccionaron los gráficos $Q - Q$ de los residuos. Se verificó que se ajustan a una distribución normal. Los puntos en el gráfico $Q - Q$ siguen aproximadamente una línea recta, por lo que se puede asumir la normalidad de los errores.

Posterior a la creación de los modelos, para el primero se construyó la matriz de covarianza con el fin de analizar si estadísticamente considerar a la cartera completa representa relevancia. Como resultado, se observó que es singular, por lo que no se puede invertir correctamente, es decir, el modelo posee multicolinealidad entre dos o más variables independientes. Para solucionar este problema, se realiza una prueba de regularización LASSO, con el fin de controlar la multicolinealidad y mejorar la estabilidad del modelo.

Esta prueba penaliza algunos coeficientes del modelo con el valor 0 para elegir las variables adecuadas. Los coeficientes penalizados tras la prueba son los precios de cierre de: omab, tlevisapco, gapb, ac, bolsaa, gfnorteo, gcarsoa1, kimbera, orbia y pinfra. Los coeficientes no nulos indican que la variable es significativa para la predicción del modelo [30].

Para el segundo modelo, se recopilaron los datos desconocidos, β y α . En la Tabla 4.1 se muestra la relación de los datos obtenidos tras la ejecución del programa:

Tabla 4.1: Recopilación de los parámetros estimados: α , β_1 y β_2 de cada acción.

Acción	α	β_1	β_2
omab	3.851133000	-0.000000403	-0.294512000
gapb	3.859716000	-0.000000403	-0.294453900
tlevisacpo	3.840875000	-0.000000402	-0.294972700
asurb	3.855070000	-0.000000403	-0.294546300
grumab	3.875785000	-0.000000404	-0.294391400
cemexcpo	3.869623000	-0.000000404	-0.294481500
alfaa	3.868338000	-0.000000404	-0.294299900
alsea	3.8712990000	-0.0000004038	-0.2943129000
ac	3.8509160000	-0.0000004026	-0.2944257000
bimboa	3.8505420000	-0.0000004026	-0.2944549000
bolsaa	3.8568020000	-0.0000004029	-0.2945054000
elektra	3.8403540000	-0.0000004019	-0.2944853000
gfnorteo	3.8568040000	-0.0000004029	-0.2944667000
femsaubd	3.8325670000	-0.0000004015	-0.2945267000
labb	3.86624700000	-0.00000040359	-0.29404210000
gmexicob	3.84294700000	-0.00000040207	-0.29463860000
gcarsoa1	3.85816100000	-0.00000040297	-0.29455920000
gcc	3.85922300000	-0.00000040308	-0.29437710000
chdrauib	3.86084300000	-0.00000040320	-0.29434900000
gfinburo	3.83617600000	-0.00000040173	-0.29466270000
kimbera	3.85754800000	-0.00000040297	-0.29441960000
megacpo	3.87117500000	-0.00000040377	-0.29432950000
orbia	3.86368800000	-0.00000040332	-0.29438160000
pinfra	3.85736700000	-0.00000040300	-0.29428310000
walmex	3.84398200000	-0.00000040213	-0.29463940000
q	3.86184900000	-0.00000040324	-0.29434960000
vws	3.86251700000	-0.00000040326	-0.29433270000
volara	3.85755300000	-0.00000040297	-0.29441800000

4.2. CASO 2: APLICACIÓN DEL MODELO CAPM CON VARIACIONES EN EL MERCADO FINANCIERO

A manera de explicación, la tabla presentada resume los resultados del segundo modelo de regresión lineal que analiza el impacto de dos variables macroeconómicas, el PIB y la TIIE sobre los rendimientos del conjunto de acciones consideradas.

Respecto al primer factor a considerar, el intercepto (α) es un componente fundamental del modelo, pues es un coeficiente que representa el rendimiento esperado de la acción cuando las variables independientes (PIB y TIIE) tienen un valor de cero.

Los valores del intercepto no varían en gran magnitud entre las acciones, reflejando diferencias puntuales en los rendimientos base de cada una de ellas bajo condiciones neutrales de las variables estudiadas, esto debido a la limpieza de datos previamente realizada.

Como segundo componente, el coeficiente β_1 que posee relación con el PIB, se calcula con valores extremadamente próximos a cero (aproximadamente -0.0000004), indicando una influencia insignificante del PIB en los rendimientos de las acciones, lo cual representa un aspecto negativo debido a que es una variable instrumental y explicativa, esta afectación puede tener sus bases en la conformación del modelo, es decir, la necesaria inclusión o exclusión de otras variables que explican de manera más adecuada los rendimientos de cada acción.

Al mismo tiempo, este resultado sugiere que, dentro del contexto económico y temporal del estudio, las variaciones en el PIB no son determinantes significativas en la variabilidad de los rendimientos de las acciones analizadas.

Por otra parte, el coeficiente β_2 que posee relación con la TIIE muestra un efecto negativo considerable (aproximadamente -0.294) en los rendimientos de las acciones. Esto implica que incrementos en la variable TIIE podrían estar relacionados con disminuciones en los rendimientos, sugiriendo una sensibilidad notable de los precios de las acciones a variaciones en las tasas de interés interbancarias. Esta relación negativa es coherente a través de todas las acciones estudiadas y destaca la relevancia de la TIIE como predictor de rendimientos en el mercado de valores.

Como tercer factor, el error estándar de los coeficientes proporciona una medida de la variabilidad o dispersión en los estimados de los coeficientes a partir de la muestra. Con valores aproximadamente constantes alrededor de 0.62, estos errores indican que las estimaciones de los coeficientes son relativamente precisas y consistentes a través de las diferentes acciones.

Considerando ahora el primer valor de la tabla, para la acción omab, el intercepto se encuentra en 3.851133000, lo cual sugiere que cuando todas las variables independientes (PIB y TIIE) son cero, el rendimiento esperado de la acción es por ese valor, para el segundo componente, β_1 , el coeficiente es muy cercano a cero (-0.000000403) lo cual indica que el PIB tiene un impacto muy bajo sobre el rendimiento de omab.

El tercer valor, β_2 , es por -0.294512000, es decir, existe una relación inversa entre la tasa TIIE y el rendimiento de la acción, a medida que la TIIE aumenta, el rendimiento tiende

a disminuir, el cuarto valor, error estándar, se obtuvo por 0.62177050, lo cual representa la variabilidad de los residuos, un valor menor mostraría más estabilidad en el modelo, lo cual sería favorable.

En contraste con la metodología empleada los autores Black, Jensen y Scholes en [12], la forma utilizada para evitar el sesgo en las estimaciones, es realizar el ejercicio a partir de la construcción de portafolios.

En este caso de estudio que se realizó posterior a la revisión de la teoría, así como en [37], se adoptó la metodología basada en el supuesto de que el error de medida del parámetro estimado es aleatorio.

Para lo cual se creó el modelo de regresión de los excesos de rentabilidad en función de los parámetros estimados, con el fin de encontrar el factor de ajuste para corregir el error de medida asociado a la utilización del β estimado en las regresiones de sección cruzada, es decir, se utilizó el factor de corrección calculado para ajustar a los coeficientes estimados en el modelo completo de regresión lineal.

Este modelo únicamente relaciona a los rendimientos de las acciones de la cartera propuesta y al rendimiento del IPC de la Bolsa Mexicana de Valores, proporcionando una estimación corregida de los coeficientes del modelo CAPM.

Para mejorar la comprensión, se muestra el pseudocódigo en el algoritmo 6, el código completo se muestra en el **Apéndice A**.

Algorithm 6 Estimación del Factor de corrección

Require: *modelo_antesCOVID*, *modelo_despuesCOVID*

```

1: for  $i$  in 1 : ncol(rendimientos_acciones) do
2:    $\text{beta\_estimado\_antesCOVID} \leftarrow \text{coef}(\text{modelo\_antesCOVID})$ 
3:    $\text{beta\_estimado\_despuesCOVID} \leftarrow \text{coef}(\text{modelo\_despuesCOVID})$ 
4:    $\text{betas\_estimados}[i] \leftarrow \frac{\text{beta\_estimado\_despuesCOVID}}{\text{beta\_estimado\_antesCOVID}}$ 
5: end for
6:  $\text{factor\_correccion} \leftarrow \text{mean}(\text{betas\_estimados})$  return factor_correccion
```

Introduciendo la cuarta fase de la metodología propuesta, análisis de resultados, respecto al factor de corrección, este se obtuvo de -0.01568416, al ser un valor menor a cero, significa que los coeficientes estimados del modelo están subestimando significativamente el valor poblacional e implica que teóricamente el parámetro estimado subestima el valor poblacional en un 101.5684 %. Después de aplicar la corrección, los coeficientes ajustados son mayores en magnitud que los coeficientes originales estimados por el modelo.

Los parámetros estimados en los modelos de sección cruzada para los plazos 2014-01-01 al 2020-02-15 y 2020-02-16 al 2023-01-01 son corregidos por este factor.

Se procedió a calcular los intervalos de confianza con el objetivo de conocer el rango de valores dentro del cual se espera que estén los parámetros de interés, en este caso, intercepto y rendimiento, el intercepto se representa como el valor de la variable dependiente cuando

4.2. CASO 2: APLICACIÓN DEL MODELO CAPM CON VARIACIONES EN EL MERCADO FINANCIERO

todas las variables independientes son iguales a 0. A tal fin, se elaboró una tabla detallada que expone estos intervalos para cada acción, particularmente en los puntos correspondientes a los interceptos y los rendimientos obtenidos por el modelo. En la Tabla 4.2 se presenta una visión clara de las estimaciones y su variabilidad para cada acción analizada.

Tabla 4.2: Se muestran los intervalos de confianza calculados.

Acción	Intervalo 5 % Intercepto	Intervalo 95 % Intercepto	Intervalo 5 % Rendimientos	Intervalo 95 % Rendimientos
omab	-0.00015739	0.00077851	0.00009898	0.000289873
gapb	-0.00017541	0.00064186	0.000134398	0.00045653
tlevisacpo	-0.00106416	-0.00011005	-0.0007024867	-0.000203050
asurb	-0.00027244	0.00045966	0.000099798	0.0002330378
grumab	-0.00020126	0.00057792	0.0001140469	0.00033409
cemexcpo	-0.00063563	0.00033390	-0.000089550	-0.0000378987
alfaa	-0.00076631	0.00017089	-0.00044345	-0.000117378
alsea	-0.00049975	0.00042904	0.0000118767	0.000034437
ac	-0.00049374	0.00055190	0.0000454198	0.000070647
bimbo	-0.00050149	0.00032425	-0.0000443263	-0.000012456
bolsaa	-0.00045951	0.00043865	0.000024576	0.000054565
elektra	-0.00023238	0.00065563	0.00023231	0.0002323523
gfnorteo	-0.00044496	0.00033001	0.000015453	0.0000483180
femsaubd	-0.00042203	0.00013804	-0.00018758	-0.00006989
labb	-0.00062420	0.00055982	0.000008749	0.00001056
gmexicob	-0.00029416	0.00057360	0.000111943	0.000434334
gcarsoa1	-0.00063746	0.00028994	-0.0001005465	-0.00004982
gcc	-0.00027483	0.00021350	-0.000034543	-0.000006454
chdrauib	-0.00028828	0.00055656	0.000121323	0.000238498
gfinburo	-0.00076050	0.00006840	-0.00045989	-0.00017878
kimbera	-0.00046455	0.00031818	-0.000043445	-0.000011932
megacpo	-0.00050847	0.00025879	-0.000101233	-0.000043112
orbia	-0.00036329	0.00048742	0.00008985	0.000223435
pinfra	-0.00033402	0.00036178	0.0000542389	0.000078453
walmex	-0.00022456	0.00047365	0.000117098	0.000198989
q	-0.00019382	0.00060507	0.000198204	0.000343587
vws	-0.00030265	0.00098291	0.000137768	0.000567837
volara	-0.00067489	0.00058621	0.0000134345	0.000034345

La presente tabla de intervalos de confianza que se construyó, proporciona una herramienta crucial para evaluar la confiabilidad de las estimaciones de los interceptos y rendimientos en el contexto del modelo CAPM. Cada intervalo de confianza se calcula al nivel 5 % y 95 %, lo que significa que existe un 95 % de certeza de que el verdadero valor poblacional del parámetro está contenido dentro del intervalo.

Para los interceptos, estos intervalos representan la incertidumbre en el rendimiento esperado de cada acción cuando todas las demás variables se mantienen constantes. Por ejemplo, para la acción omab, el intervalo de confianza al 5 % varía desde -0.00015739 hasta 0.00077851.

Esto indica que con un nivel de confianza del 95 %, se espera que el verdadero rendimiento de omab se encuentre dentro de este rango.

Por otro lado, los intervalos de confianza para los rendimientos reflejan la variabilidad esperada en los rendimientos de las acciones como respuesta a cambios en las variables independientes, como el mercado en general o factores económicos específicos. Por ejemplo, para omab, el intervalo de confianza al 5 % para los rendimientos varía desde 0.00009898 hasta 0.000289873. Esto significa que con un nivel de confianza del 95 %, se espera que los rendimientos reales de omab se encuentren dentro de este rango, dadas las condiciones económicas y del mercado.

Observando la segunda acción listada, gapb, a manera de interpretación, el intervalo de confianza para el intercepto se encuentra entre -0.00017541 y 0.00064186, lo que significa que con un 95 % de confianza, el verdadero valor del intercepto para gapb, está dentro de este rango. Además, el intervalo de confianza para los rendimientos se presentó entre 0.000134398 y 0.00045653, indicando que con un 95 % de confianza, el valor real de los rendimientos de la acción está dentro de este rango. Esta analogía puede replicarse en las 28 acciones para interpretar los datos de manera óptima.

Considerando los valores del activo libre de riesgo, el rendimiento del mercado para el periodo considerado y la obtención de β para cada acción, se realizó un promedio de estos valores en el periodo de tiempo establecido con el fin de calcular el rendimiento esperado promedio del periodo inmediato siguiente mediante el modelo CAPM actualizado.

A manera de recopilar los datos necesarios para la aplicación del modelo y en consecuencia de las consideraciones realizadas como periodo de tiempo y datos disponibles, el activo libre de riesgo CETES ofreció un rendimiento promedio del 7.739371 % y el mercado, por su parte, ofreció una rentabilidad promedio de 1.8909 %. Por lo que, al substrair del rendimiento del mercado el activo libre de riesgo, se obtiene -5.848471 %, si observamos la ecuación (3.20), hasta este momento, el único valor faltante es el parámetro β .

Finalmente, con estas consideraciones, análogamente al primer caso, se procedió a comprobar el modelo construido a través del cálculo del rendimiento esperado promedio para cada acción en comparación con los rendimientos reales de las acciones en el periodo inmediato siguiente, lo cual permitió unificar estos modelos y establecer un punto de semejanza o diferencia.

Los resultados, así como la conclusión de la fase número cinco, valoración y presentación de resultados, se compilan en la Tabla 4.5.

4.3. Caso 3: Modelo Fama and French de tres factores

Para la primera fase de este último caso, se consideran tres variables extraídas de la misma forma que en la aplicación para el modelo de valoración de activos de capital. La primera variable es el conjunto de rendimientos de los precios de 28 acciones pertenecientes a la Bolsa Mexicana de Valores, la segunda variable, la tasa libre de riesgo, establecida por la tasa CETES

y, por último, la tercera variable, el IPC de la Bolsa Mexicana de Valores.

Análogo a los casos anteriores, el periodo de revisión es del 01 de enero de 2014 al 01 de enero de 2023 con el fin de comparar los resultados obtenidos y establecer las distancias con los resultados del rendimiento original, resaltando que la elección del periodo contiene al problema suscitado ante la pandemia COVID-19, que representó un punto clave en la economía global y, en particular, en la economía mexicana.

Posteriormente, para la segunda fase, segmentación y pruebas, se construyen las variables exógenas a partir del ordenamiento de las 28 acciones a razón de su relación Book/Market y su capitalización bursátil, como lo menciona la teoría del modelo a tres factores, para realizar estos cálculos, se extrajo la información de los recursos electrónicos financieros [43] y [77], el orden se estableció como se muestra en la Tabla 4.3:

Tabla 4.3: Orden decreciente de las acciones por su relación Book/Market y capitalización bursátil.

Orden decreciente Book/Market	Total Equity/ Common Stock	Orden decreciente Capitalización Bursátil	Capitalización Bursátil
alfaa	1.3778403	chdrauib	983396000.00
femsaubd	1.28635131	gmexicob	733200000.00
orbia	1.24550225	femsaubd	720700000.00
asurb	1.17767168	gfnorteo	492400000.00
pinfra	1.16420472	gcarsoa1	427900000.00
gmexicob	1.14647579	bimboa	377500000.00
tlevisacpo	1.12331235	ac	314300000.00
alsea	1.12070734	gfinburo	285000000.00
gcarsoa1	1.11552702	elektra	259200000.00
gapb	1.06380104	cemexcpo	191500000.00
bolsaa	1.04937602	gapb	149800000.00
megacpo	1.04287802	asurb	149200000.00
cemexcpo	1.03885344	walmex	124860000.00
bimboa	1.02812805	grumab	114900000.00
omab	1.0221537	kimbera	112900000.00
gfnorteo	1.01266295	pinfra	75519000.00
chdrauib	1.00750734	orbia	70414000.00
vw	1.0049554	omab	69428000.00
q	1.00104838	q	67835000.00
gfinburo	1.00068695	gcc	66265000.00
gcc	1.0006754	alfaa	65729000.00
elektra	1.00002704	alsea	51952000.00
walmex	1	megacpo	32554000.00
volara	1	tlevisacpo	31580000.00
labbb	1	bolsaa	20030000.00
kimbera	1	volara	17735000.00
AC	1	labbb	13161000.00
grumab	0.99975313	vw	216300000.00

La Tabla 4.3 presenta un ordenamiento detallado de las acciones en función de dos criterios cruciales: su relación Book/Market y su capitalización bursátil. Estos criterios proporcionan una visión fundamental de la posición y la importancia relativa de cada acción en el contexto del modelo Fama-French de tres factores.

El primer ordenamiento se realiza con base en la relación Book/Market, la cual es un indicador financiero clave que compara el valor contable de una empresa (Book Value) con su valor de mercado (Market Value). Este índice refleja la percepción del mercado sobre el valor de la empresa en relación con sus activos contables [9].

Como se observa, las acciones se organizan en orden decreciente en concordancia con la

relación, por ejemplo, alfaa se desempeña en el primer lugar en la tabla, lo que indica que tiene la relación Book/Market más alta entre todas las acciones analizadas. Esto sugiere que el mercado valora a alfaa en relación con su valor contable de manera significativa.

Por otro lado, la capitalización bursátil es otro indicador fundamental que representa el tamaño relativo de una empresa en el mercado de valores. Se calcula multiplicando el precio de mercado de una acción por el número total de acciones en circulación [31]. De igual forma, en esta columna, las acciones se compilan en orden decreciente según su capitalización bursátil. Por ejemplo, chdrauib se posiciona en el primer lugar en el ordenamiento por capitalización bursátil, lo que indica que es la acción con mayor participación o peso en el mercado de la cartera elegida.

Con el fin de interpretar los resultados, considerando la primer acción listada en cuestión de su relación Book/Market, alfaa, es posible visualizar que su valor es por 1.3778403, lo que sugiere que está relativamente infravalorada en comparación con sus activos contables, es decir, el valor de sus activos menos sus pasivos (“Book”) es alto en comparación con el valor de mercado (precio de sus acciones, “Market”) [39].

Como segunda observación, la acción chdrauib posee la mayor capitalización de mercado entre las acciones analizadas, indicando que es la empresa con mayor valor de mercado, es decir, más estable, con mayor dimensión y percepción de valor.

Por último, consideremos a la acción ac, de igual forma que walmex, volara, labb y kimbera, posee un valor de relación Book/Market de 1, lo cual conduce a que se encuentra justo en el punto en el que su valor “Book”, es igual a su valor “Market”, sugiriendo que está valorada adecuadamente respecto a sus activos contables.

A continuación se comienza con la tercera fase propuesta en la metodología, es decir, el reprocesamiento, en el cual es necesario además de realizar un ordenamiento, estimar tres diferentes percentiles, 50 % para la capitalización bursátil, 33 % y 66 % para la relación Book/Market, esto debido a que se obtienen ganancias de segmentar correctamente el universo de acciones.

Para el caso del percentil 50 %, también conocido como mediana, se obtuvo un valor de 119880000, y se utiliza un corte directo de dos grupos, alta capitalización para valores mayores a dicho percentil y baja capitalización, para valores inferiores, por otro lado, los percentiles 33 % y 66 %, con resultados 1.0008930 y 1.0612045 respectivamente, se basan en la división de la tercera parte del grupo, tomando los puntos extremos, con el fin de conseguir una amplia diferenciación entre las acciones en función de sus características de valoración (relación Book/Market), cuando es alta, es decir, el valor contable es relativamente bajo comparado con el valor de mercado, se pueden nombrar “acciones de crecimiento”, mientras que las que tienen relación Book/Market baja son consideradas “ acciones de valor”[10].

Además; para realizar la evaluación de la relación entre los rendimientos de las acciones y los factores Fama-French, se llevaron a cabo ajustes de modelos de regresión lineal para cada acción individualmente. Este proceso se realizó utilizando la función “lm()” en el entorno de

programación R. En cada ajuste del modelo, los rendimientos de las acciones se consideraron como la variable dependiente, mientras que los factores Fama-French, junto con otras variables predictoras pertinentes, se incluyeron como variables independientes.

Específicamente, se utilizaron los factores SMB y HML como variables predictoras, además de la diferencia entre el rendimiento del mercado y el rendimiento libre de riesgo, como lo señala el modelo.

Se construyeron los seis portafolios como se muestra en la presentación del modelo Fama-French, obteniendo la siguiente tabla, es importante destacar que por la composición de la tabla sólo se consideraron 24 acciones. La razón de no incluir a las 4 acciones (walmex, Q, omab y vw) es la discriminación propia del modelo para que los portafolios resultaran equitativos. Además, se realizó el cálculo de los rendimientos históricos tomando una suma de la ponderación del rendimiento histórico de cada uno, considerando para cada activo la misma significancia debido a que se clasificaron previamente.

En la Tabla 4.4 se muestra la conformación de los seis portafolios con base en la segmentación de la relación Book/Market y Capitalización bursátil.

Tabla 4.4: Conformación de portafolios con base en la relación Book/Market y capitalización bursátil.

		Tamaño	
		Pequeño	Grande
Relación Book to Market	Bajo	volara, labb, kimbera, grumab	chdrauib, bimboa, ac, gfinburo
	Medio	bolsaa, megacpo, alsea, gcc	gfnorteo, gcarsoa1, elektra, cemexcpo
	Alto	alfaa,orbia, pinfra, tlevisacpo.	gmexicob, femsaubd, asurb, gapb

A manera de interpretación, la Tabla 4.4 presenta la conformación de seis portafolios basados en dos criterios fundamentales: la relación Book/Market y la capitalización bursátil. Estos portafolios se dividen en diferentes tamaños y niveles de relación Book/Market, lo que proporciona una estructura organizada para la gestión de inversiones.

Para el aspecto del tamaño de los Portafolios, estos se clasifican en dos tamaños principales: pequeño (Small) y grande (Big). Esta distinción se basa en la capitalización bursátil de las acciones incluidas en cada portafolio, lo cual permite una diversificación eficaz de las inversiones en función del tamaño relativo de las empresas.

Por otro lado, para el nivel de relación Book/Market, dentro de cada tamaño de portafolio, las acciones se agrupan en tres niveles de relación Book/Market: bajo (low), medio (medium) y alto (high). Este criterio refleja la percepción del mercado sobre el valor de las empresas en relación con sus activos contables.

Cada nivel de relación Book/Market contiene acciones con características financieras y de mercado similares, lo que facilita la comparación y el análisis de rendimientos dentro de cada nivel.

Por ejemplo, en el portafolio pequeño (small) con una relación Book/Market baja (low), se encuentran las acciones volara, labb, kimbera y grumab. Estas empresas tienen una capitalización bursátil menor y una relación Book/Market más baja en comparación con otras acciones.

En contraste, en el portafolio grande (big) con una relación Book/Market alta (high), se encuentran empresas como gmexicob, femsaubd, asurb y gapb, lo que sugiere esta selección es una valoración más alta por parte del mercado en relación con sus activos contables.

En adición, como cuarta fase, análisis de resultados, se utilizaron las ecuaciones (3.21) y (3.22) para calcular las variables SMB y HML y se obtuvieron los siguientes resultados:

Para el caso de SMB, -0.000835982 , a manera de interpretación, SMB mide el rendimiento adicional que, en promedio, se obtiene invirtiendo en empresas de pequeña capitalización bursátil frente a empresas de gran capitalización.

Un valor negativo de SMB indica que, durante el período de estudio, las acciones de las empresas de gran capitalización (Big) tuvieron un mejor rendimiento que las acciones de las empresas de pequeña capitalización (Small) aproximadamente 0.0836% en el período analizado, en el contexto del mercado y el período específico estudiado, hubo una tendencia a preferir las inversiones en empresas más grandes, posiblemente debido a su mayor estabilidad y menores riesgos percibidos en comparación con las empresas más pequeñas.

Por otro lado, para el caso de HML, este se obtuvo por 0.000891868 , HML mide el rendimiento adicional que, en promedio, se obtiene invirtiendo en empresas con altos valores contables (High Book-to-Market) frente a empresas con bajos valores contables (Low Book-to-Market).

Un valor positivo de HML indica que, durante el período de estudio, las acciones de las empresas con altos valores contables tuvieron un mejor rendimiento que las acciones de las empresas con bajos valores contables en aproximadamente 0.0892% [47].

Posteriormente, se comprueba el modelo análogamente a los casos anteriores comparando el rendimiento real con el rendimiento esperado calculado con base en la implementación del modelo Fama y French de 3 factores.

En la quinta y última fase de la metodología de la tesis, denominada “Análisis de Resulta-

dos”, se presenta una evaluación detallada de la eficacia de los modelos financieros investigados. Para ello, se introduce la Tabla 4.5, que compila y exhibe los datos obtenidos para los dos modelos teóricos analizados, incluyendo una versión modificada CAPM.

Esta tabla es esencial para la comprensión integral del desempeño de cada modelo, ya que no solo incluye los rendimientos esperados calculados mediante el CAPM estándar y su variante modificada, sino también los resultados derivados de la aplicación del modelo de tres factores de Fama y French. Para proporcionar una medida cuantitativa de la precisión de cada modelo, la tabla integra tres columnas adicionales que documentan el error porcentual. Este error representa la diferencia relativa entre el rendimiento real observado de las acciones y el rendimiento pronosticado por cada modelo, expresado como un porcentaje del rendimiento real. Es calculado a través de la expresión,

$$\text{Error porcentual} = \left(\frac{\text{Rendimiento real} - \text{Rendimiento esperado}}{\text{Rendimiento real}} \right) \times 100$$

Los resultados se presentan en la Tabla 4.5

Tabla 4.5: Resultados del cálculo de los rendimientos esperados de las acciones por cada modelo.

Acción	Rendimiento real	Rendimiento Esperado CAPM	Error porcentual	Rendimiento CAPM modificado	Error porcentual	Rendimiento FF	Error porcentual
omab	0.012080805	-0.000165161	1.224597 %	0.0001867084	1.18940967 %	0.0119800877	0.01007174 %
gapb	0.015561945	-0.000163219	1.572516 %	0.0003082701	1.5253675 %	0.0149900822	0.05718627 %
tlevisacpo	0.0129658	-0.000165276	1.313108 %	0.0005050856	1.3470886 %	0.0082304590	0.47353412 %
asurb	0.007018273	-0.000162485	0.718076 %	0.0001550378	0.6863235 %	0.0082236520	0.12053794 %
grumab	0.000199079	-0.000165104	0.036418 %	0.0002250568	0.0025978 %	0.0001822271	0.00168518 %
cemexcpo	0.016472306	-0.000163741	1.663605 %	0.0000498099	1.6522116 %	0.0110082261	0.54640797 %
alfaa	0.012132593	-0.000163564	1.229616 %	0.0002283824	1.2360976 %	0.0100822757	0.20503177 %
alsea	0.010002007	-0.000165161	1.016717 %	0.0000228767	0.9979130 %	0.0120082249	0.20062177 %
ac	-0.005250775	-0.000163219	0.541399 %	0.0000604190	0.5190356 %	0.0008224297	0.60732051 %
bimboa	0.006594708	-0.000165276	0.675998 %	0.0000243265	0.6619035 %	0.0008225474	0.57721608 %
bolsaa	0.000311972	-0.000162485	0.047446 %	0.0000321285	0.0279844 %	0.0008224692	0.05104969 %
elektra	-0.003558111	-0.000165104	0.372321 %	0.0002323423	0.3325769 %	0.0008222472	0.27358636 %
gfnorteo	0.008978475	-0.000163741	0.914222 %	0.0000383130	0.8940162 %	0.0082251600	0.07533154 %
femsaubd	0.001087347	-0.000163564	0.125091 %	0.0000856350	0.1172982 %	0.0008226008	0.02647460 %
labb	0.001018902	-0.000165161	0.118406 %	0.0000095498	0.1009352 %	0.0006224910	0.03964110 %
gmexicob	0.017955175	-0.000163219	1.811839 %	0.0002210923	1.7734082 %	0.0145008223	0.34543523 %
gcarsoal	0.009907156	-0.000165276	1.007243 %	0.0000953733	1.0002530 %	0.0008226326	0.90845238 %
gcc	-0.003018442	-0.000162485	0.318093 %	0.0000164885	0.3034930 %	0.0008224895	0.21959520 %
chdrauib	0.004075139	-0.000165104	0.424024 %	0.0001815136	0.3893626 %	0.0008223247	0.32528146 %
gfinburo	0.00895182	-0.000163741	0.911556 %	0.0002624428	0.9214263 %	0.0082280490	0.07237709 %
kimbera	0.008716119	-0.000163564	0.887968 %	0.0000229322	0.8739051 %	0.0082253200	0.04907987 %
megacpo	0.013559504	-0.000165161	1.372466 %	0.0000911117	1.3650616 %	0.0110082258	0.25512780 %
orbia	0.015511154	-0.000163219	1.567437 %	0.0001289975	1.5382157 %	0.0108223967	0.46887577 %
pinfra	0.005198805	-0.000165276	0.536408 %	0.0000636789	0.5135126 %	0.0008224449	0.43763599 %
walmex	0.007704044	-0.000162485	0.786653 %	0.0001800980	0.7523946 %	0.0082233430	0.05192990 %
q	0.003241983	-0.000165104	0.340709 %	0.0002332932	0.3008690 %	0.0008222532	0.24197297 %
vws	-0.001263097	-0.000163741	0.142684 %	0.0003722392	0.0890857 %	0.0008221187	0.04409778 %
volara	0.017338698	0.000163564	1.750226 %	0.0000249535	1.7313745 %	0.0180082250	0.06695267 %

Respecto a la conformación de la Tabla 4.5, es posible visualizar el proceso de verificación de los modelos al comparar el rendimiento real de las 28 acciones con el rendimiento esperado

calculado usando los tres distintos modelos, CAPM, modelo que considera el riesgo sistemático representado por el β de la acción respecto al mercado, CAPM Modificado, una variante del CAPM que incluye una variable instrumental y variables añadidas y modelo Fama & French, un modelo de tres factores que además del mercado, considera el tamaño de la empresa SMB y la relación valor contable sobre valor de mercado HML.

Para cada acción, se registra el rendimiento real y los rendimientos esperados por cada modelo. Además, se incluyen tres columnas adicionales que calculan el error porcentual en comparación con el rendimiento real. Este error porcentual proporciona una medida de precisión en la predicción del rendimiento real de las acciones, expresado como un porcentaje del rendimiento real observado.

Por ejemplo, para la acción omab, el rendimiento real es 1.21 % mayor que el rendimiento esperado según el CAPM, lo que indica que este subestima significativamente el rendimiento de la acción. El CAPM modificado mejora ligeramente la estimación, reduciendo el error a 1.18 %. El modelo Fama-French de tres factores, proporciona una estimación mucho más cercana al rendimiento real, con un error de sólo 0.01 %.

Como segunda acción, para gapb, el rendimiento real es 1.57 % mayor que el rendimiento esperado según el CAPM, indicando una subestimación significativa. El modelo CAPM modificado por su parte, mejora la precisión, reduciendo el error a 1.52 %. El modelo Fama-French de tres factores, nuevamente proporciona una estimación más precisa, con un error de 0.06 %.

Para la acción bolsaa, el rendimiento real es 0.047446 %, mayor que el rendimiento esperado según el CAPM, lo que indica una subestimación por parte del mismo. El CAPM modificado ofrece una estimación más precisa, reduciendo el error a -0.0279844 %. Sin embargo, el modelo Fama-French de tres factores, sobreestima el rendimiento real en un 0.05104969 %. En este caso, el modelo que mejor se ajusta al rendimiento real es el modelo CAPM modificado.

Es posible observar que de las 28 acciones listadas, el modelo que se adecua de mejor manera en su mayoría es el Fama-French de tres factores, esto se basa en argumentos de estudios como [28], [37] y [76], los cuales mencionan que el modelo Fama-French al incorporar factores adicionales más allá del mercado, como es la consideración de las variables SMB y HML desempeña una mejor predicción en la estimación de los rendimientos esperados.

Este tipo de análisis puede resultar útil para los inversores y gestores de fondos en el aspecto de decidir qué modelo utilizar para estimar rendimientos futuros y gestionar sus carteras de manera más efectiva, alineando mejor las predicciones con los rendimientos reales observados [11].

Los resultados obtenidos revelan una relación significativa y una correspondencia estrecha con los rendimientos reales del periodo inmediato siguiente. Aunque los resultados más próximos y, por ende, más precisos en la predicción son aquellos derivados del modelo de tres factores de Fama-French, se observa que el modelo de valoración de activos CAPM conserva su valor explicativo y sigue siendo útil para realizar estimaciones.

Esto es notable a pesar de las críticas que se han suscitado, por ejemplo del autor Hasan en [37]. Una segunda conclusión del análisis de resultados de los tres casos respalda la afirmación de los autores Eugene Fama y Kenneth French respecto a que considerar el tamaño de la empresa y la relación Book/Market aporta una explicación sustancial sobre los rendimientos esperados de los activos. Como resultado, se observa un error porcentual menor al comparar los rendimientos esperados con los rendimientos reales.

Por otro lado, el error porcentual más pronunciado corresponde al modelo CAPM original y en segundo lugar al CAPM modificado, y esto posee distintas bases. En primer lugar, las nuevas variables consideradas no ofrecen un valor explicativo suficiente. Es importante destacar que autores como Hasan en [37], han empleado metodologías similares para modelar el rendimiento esperado en mercados de tamaño medio y bajo, como el mercado financiero colombiano.

Al enfocarnos únicamente en la evaluación del modelo CAPM, observamos que el más preciso es el modelo modificado aunque el modelo original, que utiliza el IPC como rendimiento de mercado no carece de valor explicativo.

A manera de interpretación adicional, los resultados del modelo CAPM original, indican una influencia marginal del PIB y una relación significativa y negativa con la tasa TIIE en los rendimientos de las acciones. Este hallazgo es crucial para los inversionistas y analistas financieros, ya que subraya la importancia de monitorear las tasas de interés como indicador de los cambios potenciales en los rendimientos del mercado de valores [59].

Esta perspectiva puede ser fundamental para la formulación de estrategias de inversión y manejo de portafolios en contextos de volatilidad económica o mercados emergentes [11].

Secundariamente, de acuerdo con Hasan en [37], en el caso del modelo CAPM modificado, los intervalos de confianza proporcionan una medida hacia la incertidumbre asociada con las estimaciones del modelo CAPM, permitiendo a los analistas evaluar la precisión y fiabilidad de las predicciones de rendimiento de las acciones en función de los datos disponibles y las condiciones del mercado.

Por otra parte, es importante destacar la diferencia entre el intervalo de confianza del intercepto y del rendimiento, el primero evalúa la precisión de la estimación del rendimiento esperado de la acción en condiciones donde las demás variables son cero, mientras que el segundo, evalúa la variabilidad esperada en el rendimiento de la acción en respuesta a cambios en las variables independientes.

Esta clasificación facilita la gestión de inversiones al permitir una distribución estratégica de los activos en función del tamaño y las características financieras de las empresas [39], proporcionando así una herramienta valiosa para inversores y gestores de cartera en la toma de decisiones basada en evidencia empírica.

Capítulo 5

Conclusiones

En conjunto, es importante señalar que la aplicación de los modelos financieros en los mercados de latinoamérica es relevante, particularmente, la aplicación de modelos de valuación de activos como el Modelo CAPM y modelo Fama-French de tres factores en el mercado financiero mexicano. Dicha aplicación se ejecutó de manera completa a través de la revisión de la literatura existente, que permitió empaparse de los estudios realizados, de la metodología de ambos modelos y del extenso, pero no tan exhaustivamente estudiado mundo aplicativo del modelo financiero CAPM, CAPM con modificaciones y Fama-French de tres factores.

Por esta última razón, se realizó un análisis estadístico en el software de programación estadística RStudio y sus herramientas, el cual involucra a una cartera de acciones de la Bolsa Mexicana de Valores en tres casos prácticos, uno de ellos modificando al modelo CAPM con un número relevante de variables añadidas, incluso una variable Dummy que refleja la afectación que sufrió el mercado financiero con el problema ambiental, COVID 19, esto se realizó a través de las cinco fases de la metodología planteadas.

Como primer punto, la recopilación y limpieza de los datos fue efectuada correctamente al realizar un análisis de los precios de las acciones y compilar la elección de las acciones óptimas para desarrollar los casos de estudio, a continuación, la segmentación y pruebas se lograron a través del lenguaje de programación mencionado con anterioridad.

Como tercer punto, resultó necesario el reprocesamiento del modelo CAPM con modificaciones, para asegurar la correcta interpretación de resultados, lo cual conlleva a las dos últimas fases, el análisis general de los resultados y la valoración y presentación de los mismos, las cuáles fueron satisfactorias de igual manera, realizando una comparación entre los casos de estudio desarrollados en la presente tesis con los rendimientos esperados obtenidos para el periodo inmediato futuro.

Se concluyó que el modelo Fama-French de tres factores se adecuó de mejor manera a los rendimientos reales en dicho periodo, sin embargo, los tres modelos demuestran valor explicativo y de comprensión, ya que los resultados se obtuvieron a partir de pruebas estadísticas,

parámetros y análisis, y otorgan resultados cercanos a los rendimientos reales, esto de acuerdo con el error porcentual calculado.

Por otra parte, se seleccionó el mercado financiero mexicano debido a que existe una gran oportunidad de estudio en el país, además, existen diversos factores a considerar al momento de analizar los datos como precios de acciones, índices, tasas y datos financieros en general, uno de ellos, es la afectación mundial de la pandemia COVID 19 que repercutió en fluctuaciones en las acciones, variabilidad, término de empresas, bajas en índices importantes, etc.

En cuanto a las fortalezas, se destaca el fácil acceso a la información financiera a través de recursos electrónicos como Investing y Yahoo Finance, así como el uso del entorno de codificación RStudio. Además, la practicidad de obtener modelos, pruebas, parámetros y resultados a través de este entorno permitió una apreciación, interpretación y comprensión óptimas de los datos.

Respecto a las limitaciones, se identifica una escasez de información sobre el estudio y la aplicación de los modelos de valoración de activos CAPM y Fama-French en el mercado financiero mexicano. Sin embargo, esto también representa una oportunidad para futuros estudios, análisis y propuestas financieras. Por ejemplo, podría desarrollarse un modelo basado en los principios del método de valoración de activos CAPM, al que se añadirían supuestos pertinentes para la integración de mercados financieros latinoamericanos. Estos mercados a menudo enfrentan fluctuaciones, integraciones y desintegraciones periódicas, así como falta de información, entre otros factores, lo que afecta la continuidad del análisis. Además, estos mercados suelen tener un menor número de acciones cotizando en la Bolsa de Valores.

Otra de las limitaciones a considerar es la aplicación únicamente del modelo Fama-French de tres factores en el mercado financiero mexicano y no del modelo de cinco factores Fama y French, lo cual, al mismo tiempo, libera una rama de aplicación para un trabajo futuro. Otro punto futuro a considerar es la migración a un mercado Latinoamericano paralelo como el de Colombia o Brasil y la creación del modelo de Fama-French de tres factores con la inclusión de variables del entorno local [39].

Bibliografía

- [1] P. Agrrawal. “The Gibbons, Ross, and Shanken Test for Portfolio Efficiency: A Note Based on Its Trigonometric Properties”. En: *Mathematics* 11 (2023). ISSN: 22277390. DOI: 10.3390/math11092198.
- [2] D. A. Agudelo Rueda y J. H. Uribe Estrada. “¿Realidad o sofisma? Poniendo a prueba el análisis técnico en las acciones colombianas”. En: *Cuadernos de Administración* (2009).
- [3] N. Aguilar-Rivera. “Paradigma de la diversificación de la agroindustria azucarera de México”. En: *Convergencia, revista de Ciencias Sociales* 59 (2012), págs. 187-213.
- [4] J.J. Alas Romero. “Riesgo y rentabilidad de las inversiones en la bolsa de valores”. Tesis doct. Universidad De El Salvador, 2016.
- [5] J.C. Alonso y A Muñoz. “Interpretación de variables Dummy en modelos log-lin”. En: *Cali, Colombia- Departamento de Economía, Universidad ICESI* (2014).
- [6] J. C. C. Alonso Luis Berggrun P. *Introducción al análisis de riesgo*. ECOE ediciones, 2015.
- [7] J.L. Álvarez. “Rentabilidad Empresarial Propuesta Práctica de Análisis y Evaluación”. En: *Cámaras* (2004).
- [8] H. J. G. Amezcua y M. Á. Samano Celorio. “Teoría del riesgo - Selección de un portafolios de inversión”. En: *UDLAP* (dic. de 2004).
- [9] Y. Arismendy, S. Hincapié y L. Marín. “Aplicación del modelo de los tres factores de Fama French al mercado accionario colombiano y su comparación con el CAPM”. 2018.
- [10] A. Atehortúa Montoya et al. *Disimilitudes existentes entre el valor mercado y valor contable de las acciones del mercado colombiano y su relación con la rentabilidad patrimonial 1*. 2015.
- [11] R. Barrezueta, Altamirano Flores y J. E. Ordóñez. “Aplicación del CAPM en Mercados Emergentes: Una revisión teórica”. En: *Podium* 39 (2021), págs. 53-70. ISSN: 2588-0969. DOI: 10.31095/podium.202.
- [12] F. Black, M. C. Jensen y M. S. Scholes. “The Capital Asset Pricing Model: Some Empirical Tests”. En: *Preager Publishers Inc* (2017).
- [13] Z. Bodie, A. Kane y A.J. Marcus. *Essentials of Investments*. McGraw-Hill Irwin, 2011.
- [14] Z. Bodie, A. Kane y A.J. Marcus. *Essentials of investments*. McGraw-Hill/Irwin, 2013, pág. 22. ISBN: 0078034698.

- [15] J. A. Bonilla Alarcón. “Análisis del crecimiento económico a través del indicador global de la actividad económica (IGAE) bajo la metodología de Box-Jenkins y redes neuronales artificiales”. Tesis doct. Benemérita Universidad Autónoma de Puebla, 2022.
- [16] H. A. Botello-Peñaloza e I. Guerrero-Rincon. “Modelo CAPM para valorar el riesgo de los inversionistas a partir de la información contable antes y después de las NIIF en los bancos de Colombia”. En: *Entramado* 17.1 (2021), págs. 122-135.
- [17] C.N. Bouza. *Modelos de regresión y sus aplicaciones*. 2018.
- [18] J.J. Cambra Fierro. “Gestión estratégica de una cartera de clientes industriales en un contexto relacional”. En: *Revista de Dirección y Administración de Empresas* (2005).
- [19] A. Carrasquilla-Batista et al. “Regresión lineal simple y múltiple: aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal”. En: *Revista Tecnología en Marcha* 29 (8 dic. de 2016), pág. 33. ISSN: 0379-3982.
- [20] J.F. Carrizo. “La Tasa de Interés”. En: *Revista economía y estadística* (1978).
- [21] L. P. Comun Tamariz y P. M. Huaman Ojeda. “Adaptación del modelo CAPM en mercados emergentes”. En: *Universidad Peruana de Ciencias Aplicadas (UPC)* (2019).
- [22] J. Dagnino S. “Regresión lineal”. En: *Rev Chil Anest* (2014).
- [23] A. Damodaran. “Equity Risk Premiums (ERP): Determinants, Estimation and Implications- A post-crisis Update”. En: *Stern School of Business* (2009).
- [24] F De Sousa Santana. “Modelo de valoración de activos financieros (CAPM) y teoría de valoración por arbitraje (APT): Un test empírico en las empresas del sector eléctrico brasileño”. En: *Cuadernos de contabilidad* 14.35 (2013), págs. 731-746.
- [25] M.P. Durango-Gutiérrez y L.D. Delgado-Vélez. “Diseño metodológico para la estructuración de portafolios de inversión según el perfil de riesgo del inversionista”. En: *Chio América* 11.22 (2017).
- [26] J. M. En, G. R. Contreras y M. Cañadas. *La relación entre la variable aleatoria y la variable estadística: un análisis epistemológico disciplinar*. 2013.
- [27] J. M. En, G. R. Contreras y M. Cañadas. *La relación entre la variable aleatoria y la variable estadística: un análisis epistemológico disciplinar*.
- [28] E. F. Fama y K. R. French. “The Cross-Section of Expected Stock Returns”. En: *The Journal of Finance* 47 (2 1992), págs. 427-465. ISSN: 15406261. DOI: 10.1111/j.1540-6261.1992.tb04398.x.
- [29] P. Fernández y S.P. Díaz. “Pruebas diagnósticas”. En: *Cad Aten PriMaría* 10.1 (2003), págs. 120-4.
- [30] J. I. Fernández Villafanez. “Diagnóstico de fallos en rodamientos de motores eléctricos mediante técnicas LASSO”. En: *Universidad De Valladolid* (2018).
- [31] B. R. R. Fija y O. F. M. Otros. “Capitalización Bursátil”. En: *BME* 5.2.586 (2017), pág. 556.
- [32] J.A. Forero, L.E. Bohórquez y E. Lozano. “Impacto de la calidad en la rentabilidad”. En: *Ingeniería* (2008).

- [33] C. Giménez y F.R. Silvente. “Estrategias de diversificación en las exportaciones manufactureras”. En: *Revista de Economía Aplicada* (2003).
- [34] C.A. González Rodríguez. “Portafolios y estrategias de inversión”. Tesis doct. Universidad De El Salvador, 2008.
- [35] R.M. Granados. *Modelos de regresión lineal múltiple*. 2016.
- [36] E.B. Gup. *Principios básicos sobre inversiones*. Compañía editorial continental, S.A. de C.V., 1982.
- [37] A.R. Hassan y M.S. Rodríguez. “Validación empírica del modelo CAPM para Colombia Empirical validation of CAPM for Colombia 2003-2010”. En: *Ecos de Economía* (2003). ISSN: 1657-4206.
- [38] J.D. Hernández Lalinde, C.R. Espinosa Castro y J. Gerardo. “Sobre el uso adecuado del coeficiente de correlación de Pearson, definición, propiedades y suposiciones”. En: *AVFT* (2018).
- [39] Y. A. Hincapié y S. L. Marín. “Aplicación del modelo de los tres factores de Fama French al mercado accionario colombiano y su comparación con el CAPM”. En: (2014).
- [40] A. C. Hornero y D. C. Hornero. “Oferta Monetaria: El Tiempo de la Moneda.” En: *XI Congreso Nacional de Numismática: actas, Zaragoza, 2002*. Real Casa de la Moneda, Fabrica Nacional de Moneda y Timbre. 2003, págs. 363-368.
- [41] P. Huerta R y J.E. Navas López. “Análisis de la relación entre la diversificación y los resultados empresariales- una revisión teórica”. Espanol. En: *Análisis Económico* (2007). ISSN: 0185-3937.
- [42] J.C. Hull. *Risk Management and Financial Institutions*. Wiley, 2018.
- [43] *Investing México - Finanzas*. 2024. URL: <https://mx.investing.com/>.
- [44] J. Jaramillo. “Propuesta para la implementación de nuevos productos con derivados financieros en el mercado colombiano (Tesis de pregrado)”. Tesis doct. Universidad De El Salvador, 2009.
- [45] Á. Lazaro Vega. “Ampliaciones del CAPM”. En: *Universidad del País Vasco, Facultad de Economía y Empresa* (2022).
- [46] L.E. López Martínez. “Una prueba de razón de verosimilitudes para discriminar entre la distribución Poisson, Binomial y Binomial Negativa”. En: *Colegio de Postgraduados-Institución de enseñanza e investigación en ciencias agrícolas* (2010).
- [47] J. G. López Vera. “El modelo de tres factores de Fama & French: aplicación en el mercado de valores peruano”. En: *Observatorio de la Economía Latinoamericana* 210 (2015).
- [48] H. Markowitz. “Portfolio Selection”. En: *The Journal of Finance* 7 (1952), págs. 77-91.
- [49] H. M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. Yale University, 1959.
- [50] M. Mateo Guillen. “Análisis del comportamiento de activos en el sistema financiero español basado en la aplicación del modelo CAPM en periodo 2014-2018.” En: *Universidad Pontificia Comillas, Facultad de Ciencias Económicas y Empresariales* (2019).

- [51] M. A. Mejía, A. D. T. Carrasco y R. G. H. Contreras. “Análisis bursátil sectorial con relación a la tasa de interés en México”. En: *Horizontes de la Contaduría en las Ciencias Sociales* 12 (2020).
- [52] F.O. Mendivelso e I.M. Rodríguez. *Revista médica sanitas 193 definition and appropriate interpretation of p-value and confidence intervals in clinical research*. 2018.
- [53] P. L. Meyer. “Estimación de los parámetros”. En: *Probabilidad y aplicaciones estadísticas* (1992), págs. 290-319. URL: <https://es.slideshare.net/LuisAngelMezaQuispe/probabilidad-y-aplicaciones-estadisticaspaul-meyerpdf>.
- [54] M. Miguel Lanseros. “Aplicación del Modelo de 3 Factores de Fama y French a las empresas del IBEX 35”. Tesis doct. Universidad de Valladolid-Facultad de Derecho, 2018.
- [55] D. C. Montgomery, E. A. Peck y G. G. Vining. *Introduction to Linear Regression Analysis*. WILEY, 2012.
- [56] I. Moral Peláez. “Modelos de regresión: lineal simple y regresión logística”. En: *Revistas Den* (2006).
- [57] M. A. Moreno. *El CAPM, un modelo de valoración de activos financieros*. 2021. URL: <https://www.elblogsalmon.com/conceptos-de-economia/el-capm-un-modelo-de-valoraci%C3%B3n-de-activos-financieros>.
- [58] A. Novales. “Series temporales. Estacionariedad, raíces unitarias”. En: *Madrid, España* (2016).
- [59] M. C. R. Núñez, L. M. G. Briseño y G. R. Núñez. “Importancia de las bolsas de valores en la economía”. En: *Diagnóstico FACIL Empresarial Finanzas Auditoria Contabilidad Impuestos Legal* 4 (2021), págs. 55-64.
- [60] R. A. Pacheco-Robles et al. “Modelo LASSO para comparar indicadores de desarrollo social y bienestar en Perú y la región suramericana”. En: *UNAAACIENCIA-PERÚ* 2.2 (2023), págs. 10-20.
- [61] L. Palich, L. Cardinal y C. Miller. “Curvilinearity in the diversification-performance linkage An examination of over three decades of research”. En: *Strategic Management Journal* 21 (feb. de 2000), págs. 155-174.
- [62] M. P. G. Plata, S. L. López y A. Cárdenas. “La relación de causalidad entre el índice bursátil mexicano y el tipo de cambio spot”. En: *Análisis Económico* 22.51 (2007), págs. 81-105.
- [63] V. Ramanujam y P. Varadarajan. “Research on corporate diversification: A synthesis”. En: *Strategic Management Journal* 10.6 (1989), págs. 523-551.
- [64] P. Ricardo. “Decisiones financieras”. En: *Revista CE* (1998).
- [65] L. Rincon. “Curso elemental de probabilidad y estadística”. En: (2007).
- [66] R. Roca. “La tasa de interés y sus principales determinantes”. En: *Instituto de Investigaciones de Economía - Facultad de Ciencias Económicas Universidad Nacional Mayor de San Marcos* (2002). URL: <http://economia.unmsm.edu.pe/prof/rroca>.
- [67] D. Rodríguez Benavides y F. López Herrera. “Desarrollo financiero y crecimiento económico en México”. En: *Problemas del desarrollo* 40.159 (2009), págs. 39-60.

- [68] M.J. Rodríguez-Jaume y R. Mora Catalá. “Análisis de regresión simple”. En: (2001).
- [69] D. Rodrik. “Políticas de diversificación económica”. En: *Revista de la CEPAL* (2005).
- [70] I. Roy-García et al. “Correlación: no toda correlación implica causalidad”. En: *Revista Alergia México* 66.3 (2019), págs. 354-360.
- [71] D. Salsburg. *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*. Henry Holt y Company, 2011.
- [72] E. Saucedo y J. Gonzalez. “The effect of macroeconomic variables on the robustness of the traditional Fama–French model. A study for Mexico using different portfolios”. En: *Journal of Economics, Finance and Administrative Science* 26.52 (2021), págs. 252-267.
- [73] J.H. Spíndola Del Castillo. “Una Prueba del CAPM Condicional para México con Técnicas No-paramétricas”. En: (2005).
- [74] R. Tibshirani. “Regression Shrinkage and Selection via The Lasso - A Retrospective”. En: *Journal of the Royal Statistical Society Series B- Statistical Methodology* 73.3 (abr. de 2011), págs. 273-282. ISSN: 1369-7412.
- [75] V. Torres et al. “Criterios de bondad de ajuste en la selección de modelos no lineales en la descripción de comportamientos biológicos”. En: *Revista Cubana de Ciencia Agrícola* 46.4 (2012), págs. 345-350.
- [76] L. F. Torres Gutiérrez y O. Losada Salazar. *Alternativas al beta de mercado en el modelo CAPM para mercados emergentes*. spa. 2021-11-12. URL: <http://hdl.handle.net/10726/4367>.
- [77] *Yahoo Finance - Stock Market Live, Quotes, Business and Finance News*. 2024.

Apéndice A

```
1
2 library(quantmod)
3 library(tseries)
4 library(ggplot2)
5 library(dynlm)
6 library(FinTS)
7 library(fGarch)
8 library(rugarch)
9 library(PerformanceAnalytics)
10 library(rugarch)
11 library(forecast)
12 library(zoo)
13 library(qcc)
14 #library(terrUtils)
15 library(MFT)
16 library(readxl)
17 library(fpp2)
18 library(siebanxicor)
19 library(tidyverse)
20 library(tidyquant)
21 library(lubridate)
22 library(TTR)
23 library(dplyr)
24 library(BatchGetSymbols)
25 library(rvest)
26 library(tidyr)
27 library(imputeTS)
28 library(AER)
29 library(readxl)
30
31 #C digo del Algoritmo 1
32
33 fechas <- seq(as.Date("2013-12-31"), as.Date("2023-01-01"), by = "days"
34 )
35 cartera <- c("OMAB.MX", "GAPB.MX", "TLEVISACPO.MX", "ASURB.MX",
36             "GRUMAB.MX", "CEMEXCPO.MX", "ALFAA.MX", "ALSEA.MX",
37             "AC", "BIMBOA.MX", "BOLSAA.MX", "ELEKTRA.MX",
```

```

37     "GFNORTEO.MX", "FEMSAUBD.MX", "LABB.MX", "GMEXICOB.MX", "
38     GCARSOA1.MX",
39     "GCC", "CHDRAUIB.MX", "GFINBURO.MX", "KIMBERA.MX", "MEGACPO.
    MX",
    "ORBIA.MX", "PINFRA.MX", "WALMEX.MX", "Q.MX", "VWS.CO", "
    VOLARA.MX")
40 lista_precios_cierre <- list()
41 for (accion in cartera) {
42     getSymbols(accion, src = "yahoo", from = "2014-01-01", to = "
    2023-01-01", env = globalenv())
43     lista_precios_cierre[[accion]] <- Cl(get(accion))
44 }
45 precios_cierre_combinados <- do.call(merge, lista_precios_cierre)
46
47 #C digo del Algoritmo 2
48
49 precios_cierre_combinados <- merge(xts(, order.by = fechas), precios_
    cierre_combinados)
50 rendimientos_combinados <- ROC(precios_cierre_combinados, type = "
    continuous", na.pad = FALSE)
51 rendimientos_combinados[is.na(rendimientos_combinados)] <- 0
52 head(rendimientos_combinados)
53 promedio_acciones <- colMeans(rendimientos_combinados, na.rm = TRUE)
54 print(promedio_acciones)
55 promedio_acciones <- data.frame(promedio_acciones)
56 fecha_inicio <- "2014-01-01"
57 fecha_fin <- "2023-01-01"
58 sp500_data <- getSymbols("^GSPC", from = fecha_inicio, to = fecha_fin,
    src = "yahoo", auto.assign = FALSE)
59 sp500_cierre_ajustado <- sp500_data$GSPC.Adjusted
60 fechas <- seq(as.Date(fecha_inicio), as.Date(fecha_fin), by = "days")
61 sp500_series_tiempo <- xts(rep(NA_real_, length(fechas)), order.by =
    fechas)
62 sp500_combinado <- merge(sp500_series_tiempo, sp500_cierre_ajustado,
    fill = 0)
63 rendimientos_sp500 <- diff(log(sp500_combinado))
64 rendimientos_sp500[is.nan(rendimientos_sp500) | is.infinite(
    rendimientos_sp500)] <- 0
65 head(rendimientos_sp500)
66 rendimientos_sp500$GSPC.Adjusted[1]=0
67 rendimientos_combinados$Fecha <- as.Date(index(rendimientos_combinados)
    )
68 df_fechas <- data.frame(Fecha = fechas)
69 sp500_ajustado <- rendimientos_sp500$GSPC.Adjusted
70 calcular_exceso_rendimiento <- function(accion) {
71     # Obtener los rendimientos de la acci n y el S&P 500
72     rendimientos_accion <- Cl(get(accion))
73     exceso_rendimiento <- Return.calculate(rendimientos_accion) - sp500_
    ajustado
74     return(exceso_rendimiento)
75 }

```

```

76 excesos_rendimiento_por_accion <- lapply(cartera, calcular_exceso_
    rendimiento)
77 df_excesos_rendimiento_por_accion <- do.call(cbind, excesos_rendimiento_
    _por_accion)
78 df_excesos_rendimiento_por_accion$Fecha <- df_fechas$Fecha
79 df_excesos_rendimiento_por_accion$Cartera_Exceso <- rowSums(df_excesos_
    rendimiento_por_accion[, -ncol(df_excesos_rendimiento_por_accion)])
80 df_excesos_rendimiento_por_accion[is.na(df_excesos_rendimiento_por_
    accion)] <- 0
81 head(df_excesos_rendimiento_por_accion)
82 if (!requireNamespace("urca", quietly = TRUE)) {
83   install.packages("urca")
84 }
85 library(urca)
86 resultado_prueba <- ur.df(df_excesos_rendimiento_por_accion$Cartera_
    Exceso, type = "trend", lags = 0)
87 summary(resultado_prueba)
88 colnames(CETE_bien) <- c("Fecha_Ori", "Valor_Ori", "Fecha", "Valor")
89 CETE <- CETE_bien[, c("Fecha", "Valor")]
90 rendimientos_CETE <- diff(CETE$Valor) / lag(CETE$Valor)
91 print(rendimientos_CETE)
92 promedio_rendimientos_CETE <- mean(rendimientos_CETE, na.rm = TRUE)
93 dias_habiles_anio <- 252
94 promedio_rendimientos_anualizado_CETE <- promedio_rendimientos_CETE *
    dias_habiles_anio
95 print(promedio_rendimientos_anualizado_CETE)
96 setToken("5
    e7481a95594bc8db4751ff4cfe121bb80545d0091e42004c25da98752bdb85a")
97 idSeries <- "SR16573"
98 metadata <- getSeriesMetadata(idSeries)
99 #view(metadata)
100 pib_mexico <- getSeriesData(idSeries, '2014-01-01','2024-01-02')
101 print(pib_mexico)
102 pib_mexico_df <- data.frame(pib_mexico)
103 colnames(pib_mexico_df) <- c('Fecha','Valor')
104 pib_mexico_df
105 todas_fechas <- data.frame(Fecha = seq(min(pib_mexico_df$Fecha), max(
    pib_mexico_df$Fecha), by = "1_day"))
106 pib_completo <- merge(todas_fechas, pib_mexico_df, by = "Fecha", all.x
    = TRUE)
107 pib_completo <- pib_completo %>% fill(everything())
108 print(pib_completo)
109 simbolo_bmv <- "~MXX"
110 fecha_inicio <- "2014-01-01"
111 fecha_fin <- "2023-01-01"
112 ipc_data <- tq_get(simbolo_bmv, from = fecha_inicio, to = fecha_fin)
113 fechas_completas <- seq(from = as.Date(fecha_inicio), to = as.Date(
    fecha_fin), by = "days")
114 ipc_data_completo <- data.frame(date = fechas_completas)
115 for (col in colnames(ipc_data)[-1]) {

```

```

116     ipc_data_completo[[col]] <- approx(x = ipc_data$date, y = ipc_data[[
      col]], xout = fechas_completas)$y
117 }
118 ipc_data_completo$date <- fechas_completas
119 print(ipc_data_completo)
120 resultado <- ipc_data_completo[c("date", "close")]
121 resultado$close[1] <- 41185.45
122 resultado$close[3287] <- 48462.86
123 resultado$close[3288] <- 48459.80
124 resultado
125 diferencias_precios <- diff(resultado$close)
126 lag_precios <- lag(resultado$close)
127 longitud_minima <- min(length(diferencias_precios), length(lag_precios)
      )
128
129 rendimientos <- c(NA, diferencias_precios[1:longitud_minima] / lag_
      precios[1:longitud_minima])
130 resultado$rendimientos <- rendimientos
131 print(resultado)
132 resultado$rendimientos[1] <- -2.998158e-03
133 resultado$rendimientos[2] <- -2.998158e-03
134 rendimiento_mercado <- resultado$rendimientos
135
136 promedio_rendimientos_ipc <- mean(resultado$rendimientos)
137 dias_habiles_anio <- 252
138 promedio_rendimientos_anualizado_ipc <- promedio_rendimientos_ipc *
      dias_habiles_anio
139 print(promedio_rendimientos_anualizado_ipc)
140 setToken("5
      e7481a95594bc8db4751ff4cfe121bb80545d0091e42004c25da98752bdb85a")
141 idSeries <- "SF43718"
142 metadata <- getSeriesMetadata(idSeries)
143 #view(metadata)
144 Tipo_Cambio <- getSeriesData(idSeries, '2014-01-01', '2023-01-01')
145 print(Tipo_Cambio)
146 Tipo_Cambio_df <- data.frame(Tipo_Cambio)
147 colnames(Tipo_Cambio_df) <- c('Fecha', 'Valor')
148 Tipo_Cambio_df
149 todas_fechas <- data.frame(Fecha = seq(as.Date("2014-01-01"), as.Date("
      2023-01-01"), by = "1␣day"))
150 TC_completo <- merge(todas_fechas, Tipo_Cambio_df, by = "Fecha", all.x
      = TRUE)
151 Tc_completo_sinNA <- TC_completo %>% fill(everything())
152 Tc_completo_sinNA$Valor[1] <- 13.1011
153 Tc_completo_sinNA
154 setToken("5
      e7481a95594bc8db4751ff4cfe121bb80545d0091e42004c25da98752bdb85a")
155 idSeries <- "SF331451"
156 metadata <- getSeriesMetadata(idSeries)
157 TIIE <- getSeriesData(idSeries, '2013-12-31', '2023-01-01')
158 print(TIIE)

```

```

159 TIIE_df <- data.frame(TIIE)
160
161 colnames(TIIE_df) <- c('Fecha', 'Valor')
162 TIIE_df
163 todas_fechas <- data.frame(Fecha = seq(as.Date("2014-01-01"), as.Date("
164 2023-01-01"), by = "1_day"))
165 TIIE_completo <- merge(todas_fechas, TIIE_df, by = "Fecha", all.x =
166 TRUE)
167 TIIE_sinNA <- TIIE_completo %>% fill(everything())
168 TIIE_sinNA$Valor[1] <- 3.55
169 print(TIIE_sinNA)
170 setToken("5
171 e7481a95594bc8db4751ff4cfe121bb80545d0091e42004c25da98752bdb85a")
172 idSeries <- "SM16"
173 metadata <- getSeriesMetadata(idSeries)
174 #view(metadata)
175 CMM10 <- getSeriesData(idSeries, '2014-01-01', '2023-01-01')
176 print(CMM10)
177 CMM10_df <- data.frame(CMM10)
178 colnames(CMM10_df) <- c('Fecha', 'Valor')
179 CMM10_df
180 todas_fechas <- data.frame(Fecha = seq(as.Date("2014-01-01"), as.Date("
181 2023-01-01"), by = "1_day"))
182 CMM10_completo <- merge(todas_fechas, CMM10_df, by = "Fecha", all.x =
183 TRUE)
184 CMM10_sinNA <- CMM10_completo %>% fill(everything())
185 print(CMM10_sinNA)
186 datos_mexico <- data.frame(PIB = pib_completo, Rendimiento = resultado)
187
188 #C digo del Algoritmo 4
189
190 resultado_cor_test <- cor.test(datos_mexico$PIB.Valor, datos_mexico$
191 Rendimiento.close)
192 datos_financieros = data.frame(Fecha = pib_completo$Fecha, PIB=pib_
193 completo$Valor,
194 Rendimiento = resultado$close,
195 Tipo_Cambio = Tc_completo_sinNA$Valor,
196 TIIE= TIIE_sinNA$Valor,
197 Ciculaci n_Moneda_10 =CMM10_sinNA$Valor)
198 PIB=pib_completo$Valor
199 Rendimiento = resultado$close
200 Tipo_Cambio = Tc_completo_sinNA$Valor
201 TIIE= TIIE_sinNA$Valor
202 Circulaci n_Moneda_10 =CMM10_sinNA$Valor
203 modelo <- lm(Rendimiento ~ PIB + Tipo_Cambio + TIIE + Circulaci n_
204 Moneda_10 , data = datos_financieros)
205 modelo_adicional <- lm(PIB ~ Tipo_Cambio + TIIE + Circulaci n_Moneda_
206 10, data = datos_financieros)
207 residuos_principal <- residuals(modelo)
208 residuos_adicional <- residuals(modelo_adicional)
209 shapiro.test(residuos_principal)

```

```

201 shapiro.test(residuos_adicional)
202 correlacion_residuos <- cor(residuos_principal, residuos_adicional)
203
204 prueba_correlacion <- cor.test(residuos_principal, residuos_adicional)
205 print(correlacion_residuos)
206 print(prueba_correlacion)
207 modelo_restringido <- lm(Rendimiento ~ Tipo_Cambio + TIIE+ Circulaci n
    _Moneda_10, data = datos_financieros)
208 modelo_completo <- lm(Rendimiento ~ Tipo_Cambio + TIIE+ Circulaci n_
    Moneda_10 + PIB, data = datos_financieros)
209
210 #C digo del Algoritmo 5
211
212 prueba_lr <- lrtest(modelo_restringido, modelo_completo)
213 print(prueba_lr)
214 datos_financieros$Fecha <- as.Date(datos_financieros$Fecha)
215 Tc_completo_sinNA$Fecha <- as.Date(Tc_completo_sinNA$Fecha)
216 datos_financieros$d_t <- ifelse(datos_financieros$Fecha > as.Date("
    2020-02-19"), 1, 0)
217 datos_combinados <- merge(datos_financieros, Tc_completo_sinNA, by = "
    Fecha")
218 datos_combinados$d_t <- ifelse(datos_combinados$Fecha > as.Date("
    2020-02-19"), 1, 0)
219 modelo_actualizado <- modelo <- lm((rendimientos_combinados - CETE$
    Valor) ~(resultado$rendimientos - CETE$Valor) +
220 datos_financieros$d_t * (resultado$rendimientos - CETE$Valor) + PIB +
    TIIE)
221 summary(modelo)
222 datos_modelo <- data.frame(Fecha = pib_completo$Fecha, Rentabilidad=
    rendimientos_combinados, Tasa_CETE = CETE$Valor, Rent_mer_mex=
    resultado$rendimientos,
223 IB=pib_completo$Valor, TIIE= TIIE_sinNA$Valor)
224 datos_modelo$Fecha <- as.Date(datos_modelo$Fecha)
225 datos_modelo$d_t <- as.numeric(datos_modelo$Fecha > as.Date("2020-02-19
    "))
226 modelo <- lm((rendimientos_combinados$VOLARA.MX.Close - CETE$Valor) ~ (
    resultado$rendimientos - CETE$Valor) +
227     d_t * (resultado$rendimientos - CETE$Valor) + PIB + TIIE
    + rendimientos_combinados$OMAB.MX.Close
228     + rendimientos_combinados$GAPB.MX.Close + rendimientos_
    combinados$TLEVISACPO.MX.Close
229     + rendimientos_combinados$ASURB.MX.Close + rendimientos_
    combinados$GRUMAB.MX.Close
230     + rendimientos_combinados$CEMEXCPO.MX.Close + rendimientos_
    combinados$ALFAA.MX.Close
231     + rendimientos_combinados$ALSEA.MX.Close + rendimientos_
    combinados$AC.Close
232     + rendimientos_combinados$BIMBOA.MX.Close + rendimientos_
    combinados$BOLSAA.MX.Close
233     + rendimientos_combinados$ELEKTRA.MX.Close + rendimientos_
    combinados$GFNORTEO.MX.Close

```

```

234 + rendimientos_combinados$FEMSAUBD.MX.Close + rendimientos
      _combinados$LABB.MX.Close
235 + rendimientos_combinados$GMEXICOB.MX.Close + rendimientos
      _combinados$GCARSOA1.MX.Close
236 + rendimientos_combinados$GCC.Close + rendimientos_
      combinados$CHDRAUIB.MX.Close
237 + rendimientos_combinados$GFINBURO.MX.Close + rendimientos
      _combinados$KIMBERA.MX.Close
238 + rendimientos_combinados$MEGACPO.MX.Close + rendimientos_
      combinados$ORBIA.MX.Close
239 + rendimientos_combinados$PINFRA.MX.Close + rendimientos_
      combinados$WALMEX.MX.Close
240 + rendimientos_combinados$Q.MX.Close + rendimientos_
      combinados$VWS.CO.Close, data = datos_modelo)
241 resumen_modelo <-summary(modelo)
242 coeficientes <- resumen_modelo$coefficients
243 intercepto <- coeficientes["(Intercept)", "Estimate"]
244 betas <- coeficientes[, "Estimate"]
245 error_std <- sqrt(resumen_modelo$sigma)
246 print(intercepto)
247 print(betas)
248 print(error_std)
249 # 1. Linealidad
250
251 plot(modelo$fitted.values, modelo$residuals)
252 abline(h = 0, col = "red")
253 title("Residuos vs Valores Ajustados")
254 # 2. Homogeneidad (Media cero del error)
255 mean_residuos <- mean(modelo$residuals)
256 print(paste("Media de los residuos:", mean_residuos))
257
258 # 3. Homocedasticidad
259 plot(modelo$fitted.values, modelo$residuals)
260 abline(h = 0, col = "red")
261 title("Residuos vs Valores Ajustados")
262
263 library(lmtest)
264 bptest_result <- bptest(modelo)
265 print(bptest_result)
266
267 # 4. Independencia
268 dwtest_result <- dwtest(modelo)
269 print(dwtest_result)
270
271 # 5. Normalidad
272 qqnorm(modelo$residuals)
273 qqline(modelo$residuals, col = "red")
274 title("Gráfico Q-Q de los Residuos")
275
276
277 if (!requireNamespace("glmnet", quietly = TRUE)) {

```

```

278   install.packages("glmnet")
279 }
280 library(glmnet)
281 X <- model.matrix(~ (resultado$rendimientos - CETE$Valor) + d_t * (
      resultado$rendimientos - CETE$Valor) + PIB + TIIE +
282     rendimientos_combinados$OMAB.MX.Close +
      rendimientos_combinados$GAPB.MX.Close +
283     rendimientos_combinados$TLEVISACPO.MX.Close +
      rendimientos_combinados$ASURB.MX.Close +
284     rendimientos_combinados$GRUMAB.MX.Close +
      rendimientos_combinados$CEMEXCPO.MX.Close +
285     rendimientos_combinados$ALFAA.MX.Close +
      rendimientos_combinados$ALSEA.MX.Close +
286     rendimientos_combinados$AC.Close + rendimientos_
      combinados$BIMBOA.MX.Close +
287     rendimientos_combinados$BOLSAA.MX.Close +
      rendimientos_combinados$ELEKTRA.MX.Close +
288     rendimientos_combinados$GFNORTEO.MX.Close +
      rendimientos_combinados$FEMSAUBD.MX.Close +
289     rendimientos_combinados$LABB.MX.Close +
      rendimientos_combinados$GMEXICOB.MX.Close +
290     rendimientos_combinados$GCARSOA1.MX.Close +
      rendimientos_combinados$GCC.Close +
291     rendimientos_combinados$CHDRAUIB.MX.Close +
      rendimientos_combinados$GFINBURO.MX.Close +
292     rendimientos_combinados$KIMBERA.MX.Close +
      rendimientos_combinados$MEGACPO.MX.Close +
293     rendimientos_combinados$ORBIA.MX.Close +
      rendimientos_combinados$PINFRA.MX.Close +
294     rendimientos_combinados$WALMEX.MX.Close +
      rendimientos_combinados$Q.MX.Close +
295     rendimientos_combinados$VWS.CO.Close, data = datos_
      modelo)
296
297 y <- rendimientos_combinados$VOLARA.MX.Close - CETE$Valor
298 lasso_model <- glmnet(X, y, alpha = 1) # alpha = 1 para LASSO
299 cv_lasso <- cv.glmnet(X, y, alpha = 1)
300 plot(cv_lasso)
301 lambda_optimo <- cv_lasso$lambda.min
302 lambda_optimo
303 coeficientes_lasso <- predict(lasso_model, s = lambda_optimo, type = "
      coefficients")
304 coeficientes_lasso <- coeficientes_lasso[-1, ] # Excluir el intercepto
305 print(coeficientes_lasso)
306 modelo62 <- lm(rendimientos_combinados ~ resultado$rendimientos)
307 coeficientes <- coef(modelo62)
308 beta_estimado <- coeficientes["returns_mercado"]
309 residuos <- residuals(modelo62)
310 betas_corregidos <- numeric(length = ncol(rendimientos_combinados))
311 for (i in 1:ncol(rendimientos_combinados)) {
312     rendimientos_accion <- rendimientos_combinados[, i]

```

```

313 rendimientos_mercado <- resultado$rendimientos
314 modelo <- lm(rendimientos_accion ~ rendimientos_mercado)
315
316 coeficientes <- coef(modelo)
317 beta_estimado <- coeficientes["rendimientos_mercado"]
318
319 residuos <- residuals(modelo)
320 var_error_residual <- var(residuos)
321
322 beta_corregido <- beta_estimado / (1 + var_error_residual)
323
324 betas_corregidos[i] <- beta_corregido
325 }
326 inicio_subperiodo1 <- as.Date("2014-01-01")
327 fin_subperiodo1 <- as.Date("2020-02-15") # Por ejemplo, un a o
      despu s del inicio
328 inicio_subperiodo2 <- as.Date("2020-02-16") # El d a siguiente al
      final del primer subperiodo
329 fin_subperiodo2 <- as.Date("2023-01-01") # Fecha final del periodo
      completo
330 rendimientos_mercado_sub1 <- resultado[resultado$date >= inicio_
      subperiodo1 & resultado$date < fin_subperiodo1, ]
331 rendimientos_mercado_sub2 <- resultado[resultado$date >= inicio_
      subperiodo2 & resultado$date <= fin_subperiodo2, ]
332 betas_estimados <- numeric(length = ncol(rendimientos_combinados))
333
334 #C digo del Algoritmo 6
335
336 for (i in 1:ncol(rendimientos_combinados)) {
337   rendimientos_accion <- rendimientos_combinados[, i]
338
339   rendimientos_mercado_sub1_subset <- rendimientos_mercado_sub1$
      rendimientos[1:length(rendimientos_accion)]
340   rendimientos_mercado_sub2_subset <- rendimientos_mercado_sub2$
      rendimientos[1:length(rendimientos_accion)]
341
342   modelo_sub1 <- lm(rendimientos_accion ~ rendimientos_mercado_sub1_
      subset)
343   modelo_sub2 <- lm(rendimientos_accion ~ rendimientos_mercado_sub2_
      subset)
344
345   beta_estimado_sub1 <- coef(modelo_sub1)[2]
346   beta_estimado_sub2 <- coef(modelo_sub2)[2]
347
348   betas_estimados[i] <- beta_estimado_sub2 / beta_estimado_sub1
349 }
350 factor_correccion <- mean(betas_estimados)
351
352 # 1. Linealidad
353
354 plot(modelo62$fitted.values, modelo$residuals)

```

```

355 abline(h = 0, col = "red")
356 title("Residuos vs Valores Ajustados")
357 # 2. Homogeneidad (Media cero del error)
358 mean_residuos <- mean(modelo62$residuals)
359 print(paste("Media de los residuos:", mean_residuos))
360
361 # 3. Homocedasticidad
362 plot(modelo62$fitted.values, modelo$residuals)
363 abline(h = 0, col = "red")
364 title("Residuos vs Valores Ajustados")
365
366 library(lmtest)
367 bptest_result <- bptest(modelo)
368 print(bptest_result)
369
370 # 4. Independencia
371 dwtest_result <- dwtest(modelo)
372 print(dwtest_result)
373
374 # 5. Normalidad
375 qqnorm(modelo62$residuals)
376 qqline(modelo$residuals, col = "red")
377 title("Gráfico Q-Q de los Residuos")
378 modelo_completo <- lm(rendimientos_combinados ~ resultado$rendimientos,
379   data = rendimientos_combinados)
380 betas_estimados_completos <- coef(modelo_completo)[-1] * factor_
381   correccion
382 subestimacion_porcentual <- (1 - factor_correccion) * 100
383 cat("El factor de corrección encontrado es", factor_correccion, ", lo
384   cual implica que teóricamente el parámetro estimado subestima el
385   valor poblacional en un", subestimacion_porcentual, "%.\n")
386 plazo_1 <- "2014-01-01 al 2020-02-15"
387 plazo_2 <- "2020-02-16 al 2023-01-01"
388 cat("Los parámetros estimados en los modelos de sección cruzada para
389   los plazos", plazo_1, "y", plazo_2, "son corregidos por este factor
390   .\n")
391 coeficientes_modelo_completo <- coef(modelo_completo)
392 print(coeficientes_modelo_completo)
393 coeficientes_ajustados <- coeficientes_modelo_completo * factor_
394   correccion
395 print(coeficientes_ajustados)
396 intervalos_confianza <- confint(modelo_completo)
397 print(intervalos_confianza)
398 intervalos_confianza <- data.frame(intervalos_confianza)
399 R_Mt <- resultado$rendimientos
400 R_ft <- CETE$Valor
401 SMB <- -0.000835982
402 HML <- 0.000891868
403 rendimientos_esperados <- list()
404 for (i in 1:ncol(coeficientes)) {
405   coeficientes_accion <- coeficientes[, i]

```

```

399 alpha_i <- coeficientes_accion["(Intercept)"]
400 beta_i1 <- coeficientes_accion["R_Mt"]
401
402 rendimiento_esperado <- alpha_i + beta_i1 * (R_Mt - R_ft) + SMB *
403     0.000835982 + HML * 0.000891868
404
405 rendimientos_esperados[[i]] <- rendimiento_esperado
406 }
407 nombres_acciones <- c("OMAB.MX", "GAPB.MX", "TLEVISACPO.MX", "ASURB.MX"
408     , "GRUMAB.MX", "CEMEXCPO.MX", "ALFAA.MX", "ALSEA.MX", "AC", "
409     BIMBOA.MX", "BOLSAA.MX", "ELEKTRA.MX", "GFNORTEO.MX", "FEMSAUBD.MX",
410     "LABB.MX", "GMEXICOB.MX", "GCARSOA1.MX", "GCC", "CHDRAUIB.MX", "
411     GFINBURO.MX", "KIMBERA.MX", "MEGACPO.MX",
412     "ORBIA.MX", "PINFRA.MX", "WALMEX.MX", "Q.MX", "VWS.CO", "VOLARA.MX")
413 df_rendimientos_esperados <- data.frame(Fecha = seq(as.Date("2014-01-01"),
414     as.Date("2023-01-01"), by = "day"), stringsAsFactors = FALSE)
415 colnames(df_rendimientos_esperados) <- nombres_acciones
416 library(forecast)
417 resultados <- list()
418 resta <- -0.05848471 # Retorno del mercado menos el libre de riesgo
419 for (i in 1:ncol(rendimientos_combinados)) {
420     R_it <- rendimientos_combinados[, i]
421
422     variables_predictoras <- data.frame(resultado = resultado$
423         rendimientos, resta = resta, SMB = SMB, HML = HML)
424
425     modelo <- lm(R_it ~ CETE_bien$Valor - 1, data = variables_
426         predictoras)
427
428     resultados[[i]] <- summary(modelo)
429 }
430 for (i in 1:length(resultados)) {
431     cat("Resultados para la acci n", i, ":\n")
432     print(resultados[[i]])
433     cat("\n")
434 }
435 tasa_rf_matrix <- matrix(rep(CETE$Valor, ncol(rendimientos_combinados))
436     , nrow = nrow(CETE), ncol = ncol(rendimientos_combinados), byrow =
437     FALSE)
438 rendimiento_mercado$rendimiento_excesivo_mercado <- resultado$
439     rendimientos - CETE$Valor
440 str(resultado)
441 str(CETE)
442 CETE$Fecha <- as.Date(CETE$Fecha)
443 rendimiento_mercado$date <- as.Date(rendimiento_mercado$date)
444 rendimiento_mercado$rendimiento_excesivo_mercado <- resultado$
445     rendimientos - CETE$Valor
446
447 rendimientos_excesivos <- as.data.frame(rendimientos_combinados - tasa_
448     rf_matrix)

```

```

437
438 betas <- sapply(rendimientos_excesivos, function(rend_excesivo) {
439   cov(rend_excesivo, rendimiento_mercado$rendimiento_excesivo_mercado,
        use = "complete.obs") / var(rendimiento_mercado$rendimiento_
        excesivo_mercado, use = "complete.obs")
440 })
441 tasa_rf_representativa <- mean(CETE$Valor) # 0 puedes elegir otra
        forma de seleccionar una tasa_rf
442
443 rendimiento_esperado_mercado <- mean(rendimiento_mercado$rendimiento_
        excesivo_mercado) + tasa_rf_representativa
444
445 rendimiento_esperado_CAPM <- tasa_rf_representativa + betas * (
        rendimiento_esperado_mercado - tasa_rf_representativa)
446 resultados_df <- data.frame(
447   Accion = names(rendimiento_esperado_CAPM),
448   RendimientoEsperado = rendimiento_esperado_CAPM
449 )
450 print(resultados_df, row.names = FALSE)
451 library(tseries)
452
453 #C digo de Algoritmo 3
454
455 for (col in colnames(rendimientos_excesivos)) {
456   print(paste("Serie:", col))
457
458   adf_test <- adf.test(rendimientos_excesivos[, col], alternative = "
        stationary")
459
460   print(adf_test)
461
462   if (adf_test$p.value > 0.05) {
463     rendimientos_excesivos[, col] <- diff(rendimientos_excesivos[, col
        ], differences = 1)
464     print("Serie no estacionaria. Aplicada diferenciación.")
465   } else {
466     print("Serie estacionaria.")
467   }
468 }
469 if (!requireNamespace("forecast", quietly = TRUE)) {
470   install.packages("forecast")
471 }
472 library(forecast)
473 resultado$prima_mercado <- resultado$rendimientos - CETE$Valor
474 modelo_arima_prima_mercado <- auto.arima(resultado$prima_mercado)
475 summary(modelo_arima_prima_mercado)
476 prediccion_prima_mercado_Tplus1 <- forecast(modelo_arima_prima_mercado,
        h = 1)
477 print(prediccion_prima_mercado_Tplus1)
478
479 valor_predicho <- prediccion_prima_mercado_Tplus1$mean

```

```
480 print(valor_predicho)
```

Listing 1: Código de R