



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Estudio de Sistemas Complejos a través de Series de
Tiempo.

Tesis presentada al

Colegio de Física

como requisito parcial para la obtención del grado de

LICENCIADO EN FÍSICA

por

Miguel Labra Lorenzana

Asesorado por

Dr. Jorge Velázquez Castro

Puebla Pue.
Mayo de 2024



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Estudio de Sistemas Complejos a través de Series de
Tiempo.

Tesis presentada al

Colegio de Física

como requisito parcial para la obtención del grado de

LICENCIADO EN FÍSICA

por

Miguel Labra Lorenzana

Asesorado por

Dr. Jorge Velázquez Castro

Puebla Pue.
Mayo de 2024

Título: Estudio de Sistemas Complejos a través de Series de Tiempo.
Estudiante: MIGUEL LABRA LORENZANA

COMITÉ

Dr. Roberto Ramírez Sánchez
Presidente

Dr. José Juan Castro Alva
Secretario

Dr. Ivan Fuentesilla Carcamo
Vocal

Dr. Jorge Velázquez Castro
Asesor

Índice general

Introducción	XI
1. Introducción a la Econofísica	1
1.1. Historia de la Econofísica	1
1.2. Mercados Financieros como Sistemas Complejos	2
1.3. Oferta, Demanda y Arbitraje	6
1.4. Procesos Estocásticos	9
1.4.1. Variables Aleatorias.	9
1.4.2. Momentos	10
1.4.3. Covarianza y Coeficiente de Correlación	10
1.4.4. Caminata Aleatoria	11
1.4.5. Movimiento Browniano	13
1.4.6. Proceso Markoviano	15
1.5. Hipótesis de Mercado Eficiente	16
1.5.1. Conceptos	16
1.5.2. Hipótesis de Mercado Eficiente	17
1.6. Series de Tiempo Financieras	17
1.6.1. Series de Tiempo.	18
1.6.2. Autocorrelación y Autocorrelación Parcial.	19
1.6.3. Estacionariedad y Estacionalidad.	20
1.6.4. Ruido Blanco.	21
2. Herramientas	23
2.0.1. Aglomeramiento k-medias	23
2.0.2. Descomposición en Valores Singulares	24
2.0.3. Análisis de Componentes Principales	25
2.0.4. Regresión LASSO	26
3. Antecedentes	29
3.1. Trabajos Previos	29
4. Metodología	35
4.1. Procedimiento	35
5. Resultados	39
6. Conclusiones	47
Bibliografía	49

Índice de figuras

1.1. Cotización de algunos mercados financieros del mundo. El eje x representa la fecha de cotización y el eje y corresponde al valor de cada índice en dólares.	3
1.2. Dinámica de una población en biología; modelada por la ecuación logística.	5
1.3. Visualización del diagrama de bifurcación para un mapeo logístico.	6
1.4. Diagrama de telaraña que muestra el caso convergente.	7
1.5. Interfaz para orden de compra proveída por la plataforma de la casa de bolsa Bursanet.	8
1.6. Gráfica de una caminata aleatoria.	12
1.7. Gráfica del precio de las acciones de la empresa Amazon.	12
1.8. 10 simulaciones diferentes de movimiento browniano.	14
1.9. 10 simulaciones diferentes de movimiento browniano geométrico.	15
1.10. Gráficas de funciones de autocorrelación para diferentes tipos de series de tiempo. (a) y (c) sugieren que es posible hacer predicciones ya que muestran correlaciones entre sus instancias de tiempo, a diferencia de (b) que no es posible ver algún patrón.	19
1.11. Gráfica que ejemplifica la característica de estacionalidad dentro de una serie de tiempo.	20
1.12. Gráfico de ruido blanco; esencialmente representa una secuencia de valores aleatorios.	21
2.1. Visualización del funcionamiento del algoritmo aglomeramiento k-medias	23
2.2. Estos métodos ayudan a definir cuál es el número óptimo de aglomeramientos para determinado conjunto de datos.	24
3.1. Matrices de correlación obtenidas por Munnix et al. Donde se muestran los 8 estados del mercado.	31
4.1. Construcción de las series de tiempo cortas, el marco rojo corresponde a una ventana de 20 días que va recorriendo un día a la serie de tiempo total.	36
5.1. Estados de Mercado.	40
5.2. Evolución dinámica de los estados de mercado.	41
5.3. Izquierda: Matriz de transición de estados. Derecha: Matriz de probabilidad de transición.	41
5.4. Descomposición truncada en valores singulares del estado 1.	41
5.5. Descomposición truncada en valores singulares del estado 2.	42
5.6. Descomposición truncada en valores singulares del estado 3.	42
5.7. Descomposición truncada en valores singulares del estado 4.	43
5.8. Descomposición truncada en valores singulares del estado 5.	43
5.9. (a) Proyección de los componentes principales empleando los datos de los precios estandarizados. (b) Proyección de las componentes principales empleando la matriz de covarianza.	44

- 5.10. Representación 3-dimensional de la segmentación de los estados de mercado, resultado del análisis de componentes principales vía SVD. El aglomeramiento rojo corresponde al estado 1 el cuál es un estado de calma, mientras que el aglomeramiento negro corresponde al estado 5 el cuál representa cambios drásticos y crisis financieras. 44
- 5.11. Representación 2-dimensional de la segmentación de estados de mercado; a la derecha se muestran los ejes $x-z$ y a la izquierda los ejes $x-y$. Ambas gráficas incluyendo la gráfica en la figura 5.10, representan los mismo resultado, sin embargo, cada una ofrece diferentes perspectivas. 45

Índice de tablas

1.1. Ejemplo de una colección de datos normales.	18
1.2. Ejemplo de una colección de datos correspondientes a una serie de tiempo.	18
3.1. Abreviaciones para los diferentes sectores del SP500 y el Nikkei 225 (las siglas están en inglés).	32
5.1. Correlación promedio y eigenvalor más grande de cada estado de mercado.	40
5.2. 50 acciones seleccionadas de acuerdo a los resultados obtenidos tras llevar a cabo la regresión lasso sobre los datos del precio de las acciones de cada empresa.	45
5.3. 196 pares de acciones seleccionadas tras llevar a cabo la regresión lasso sobre los datos de las correlaciones únicas entre cada par de acciones.	46

Introducción

La complejidad es una característica que poseen muchos entornos de la naturaleza que nos rodea; se puede tratar desde una galaxia, el clima del planeta Tierra, un tornado, una parvada, una población de hormigas, el cerebro humano hasta el núcleo de los átomos. Tratar de entenderlos y describirlos no es una tarea fácil. Afortunadamente, la física posee ciertos campos interdisciplinarios a través de los cuales es posible abordar estos fenómenos.

La econofísica es una novedosa línea de investigación que busca aplicar las herramientas y técnicas de la física estadística, teoría de sistemas complejos y dinámica no lineal, para estudiar los sistemas financieros [8].

Este campo surgió en el siglo XX [1, 9], con la incorporación de la era digital, millones de datos de información financiera empezaron a quedar registrados en las bases de datos de las casas de bolsa. Prontamente, un grupo de físicos empezaría a explorar e investigar los comportamientos y patrones presentes en dichos datos, aplicando los métodos conocidos en la física y como consecuencia, se generó una vasta cantidad de investigaciones sobre el tema.

Fue en una conferencia en Kalkota, en el año 1995 que el físico y economista E. H. Stanley acuñó por primera vez el término de econofísica, que culminó con la publicación de su libro *An Introduction to Econophysics Correlations and complexity in Finance* [1], donde recopilaba todas las investigaciones referentes al tema. Cabe mencionar que, entre los aportes más importantes, se encuentra la tesis de L. Bachelier hecha bajo la dirección de Henry Poincaré con el título *Theorié de la Spéculation* [17] donde analiza la bolsa de valores de París y en el proceso formula gran parte de la teoría del movimiento Browniano¹ anticipando mucho del trabajo central para la física estadística de Einstein del año 1905 [21].

En esta tesis, el objeto de estudio son los mercados financieros como un ejemplo de sistema complejo, en particular los mercados de acciones. Esto en parte por la disponibilidad de datos de buena calidad sin costo o a costos limitados. Se sigue la línea de investigación de los artículos publicados primero por Münnix et al. en 2008 [13], donde establece que es posible caracterizar con estados a los mercados financieros a través de matrices de correlación donde los estados con valores más altos de correlación representan estados de cambio drástico. Posteriormente este tema lo retoman Pharasi et al. en 2018 [14], en 2020 [15] y en 2021 [16], quienes profundizan en esta premisa e incluso proponen el desarrollo de un indicador que ayude a mejorar la gestión de un portafolio de inversión protegiéndose de las crisis financieras.

Uno de los objetivos centrales de esta tesis es identificar las principales empresas y correlaciones entre empresas que afectan a dichos estados del mercado, para ello, es necesario determinar el número óptimo de estados de mercado que caracterizan al índice SP500 utilizando los precios de apertura. Una vez determinado el número óptimo de estados

¹Esto anticipa una estrecha conexión entre los mercados de acciones y otros sistemas complejos que observó el Dr. Brown años antes.

del mercado, se llevará a cabo un análisis de su dinámica. Posteriormente, se implementará una *textbf*descomposición en valores singulares buscando reconstruir de forma óptima los patrones formados en las estructuras de las matrices de correlación de cada estado. Esta descomposición también permitirá realizar un **análisis de componentes principales** que funcionará para visualizar la segmentación de los estados determinados. Finalmente se realiza una **regresión lasso**, esta buscará predecir en qué estado se encuentra el mercado con base en los precios de las acciones o con base en los valores de las correlaciones de determinado par de acciones. Además, esta regresión también nos permitirá **identificar qué acciones o par de acciones son quienes más afectan a la ocurrencia de estos estados de mercado**. Por lo que puede ser de gran utilidad dar seguimiento a algunas acciones con el fin de prevenir o anticipar pérdidas derivado de algún cambio drástico.

Capítulo 1

Introducción a la Econofísica

1.1. Historia de la Econofísica

De acuerdo con la definición de la economía como la ciencia que estudia la distribución y administración de los recursos dentro de una sociedad con el fin de satisfacer las necesidades humanas, es posible llegar a pensar que la física no puede influir o aportar algo a esta disciplina. Sin embargo, en el presente capítulo se ofrece un breve resumen histórico de cómo estas dos ciencias se han complementado, en específico como la física ha aportado ideas y métodos para la creación de lo que hoy conocemos como econofísica, ver [3, 8, 9].

Podemos establecer el punto de partida en los tiempos de Aristóteles quien es famoso por haber sentado las bases de la física, entre otras de sus aportaciones encontramos sus reflexiones sobre el intercambio de bienes. El planteó que mientras el trueque funcionaba bien en la mayoría de los casos, uno podía encontrarse en problemas cuando una persona quería algo, pero no tenía nada que le interesara a la otra parte. Con la introducción del dinero este problema se solucionaba ya que el dinero (oro y plata en ese tiempo) ya tenía un valor intrínseco.

Nicholas Copérnico es conocido dentro de la física por plantear la teoría heliocéntrica, sin embargo, Copérnico también hizo contribuciones muy acertadas al área de la economía. Durante un tiempo se desempeñó como consejero del parlamento del estado de Prusia y del rey de Polonia, mejoró el enfoque que se tenía a temas de economía y finanzas usando un enfoque empírico y práctico. Incluso desarrollo una serie de reglas para evitar la depreciación de la moneda.

Si seguimos avanzando en el tiempo llegamos a la publicación de los trabajos de Isaac Newton, sobre cálculo, mecánica y especialmente la teoría de gravitación universal. Estos trabajos son de tal magnitud que quien es considerado hoy padre de la economía, Adam Smith, se vio influenciado por ellos al tratar de buscar leyes generales que explicaran la economía. No solo eso, Newton también fue maestro de la casa de la moneda real en Londres desde 1699 hasta el día de su muerte en marzo de 1772.

Con grandes aportaciones a la astronomía, óptica y contemporáneo a Newton, Christian Huygens, escribió varios trabajos sobre probabilidad y teoría de juegos. También estuvo en contacto con Blaise Pascal y Pierre de Fermat con quienes intercambio ideas sobre conceptos de probabilidad. Estos últimos dos fueron quienes fundaron las bases de la teoría de probabilidad.

Edmond Halley, famoso fisico que calculo la órbita del cometa que lleva su nombre también llevo a cabo un estudio sobre un pueblo alemán, hoy llamado Wroclaw, donde en su trabajo

plantea una ecuación que calcula el valor anual de una pensión tomando como trasfondo las tasas de intereses, mortalidad y la edad de quien pretendía adquirir una pensión. Su trabajo tuvo éxito y logro aportar al desarrollo de las ciencias actuariales.

El matemático Adolph Quetlet en conjunto con el filósofo Auguste Comte consideraron que las leyes de la física también tenían la capacidad para gobernar el comportamiento humano y la economía. Debido a una discrepancia entre ambos, descarto el termino física social por sociología.

Uno de los aportes más importantes tomo lugar a inicios del siglo XX en la tesis [17] de doctorado de Louis Bachelier, publicada en 1900, donde desarrolla un modelo de caminata aleatoria para describir la dinámica de los precios. Más tarde estas ideas serían formalizadas por Einstein al tratar de calcular el número de Avogadro.

La línea de pensamiento que uso Bachelier en su tesis terminaría desembocando en el origen de la hipótesis de mercado eficiente y la ecuación de Black-Scholes. Fue en 1973 cuando Fischer Black y Myron Scholes [38] publicaron el primer artículo donde establecían una ecuación (1.1) que definía el precio más justo a una opción.

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} + \frac{\partial V}{\partial t} = 0. \quad (1.1)$$

Más tarde en los años 80's, las divisas se empezaban a comerciar de forma digital [1] como ya lo hacían los mercados de acciones y otros activos, dando origen al mercado de divisas, que a diferencia de los mercados de acciones, este estaría deslocalizado y abierto 24 horas, de ahí su nombre coloquial del mercado que nunca duerme. Esto originaría que todos los mercados financieros se digitalizarán y toda la información de cada una de las transacciones quedarán almacenada.

Para tener una idea clara del volumen de estos mercados, tan solo el mercado de divisas mueve alrededor de 6.6 billones de dólares al día. Con toda esta cantidad de datos disponibles, creció un fuerte interés en crear modelos para describir con más exactitud los movimientos del mercado. Wall Street comenzó a contratar físicos y matemáticos para llevar a cabo esta tarea.

En 1990 Eugene Stanley y Rosario Mantenga se dieron a la tarea de juntar todos estos trabajos y 5 años después en una conferencia en Kalkota [8], Stanley acuñaría el neologismo de econofísica. La econofísica es un campo interdisciplinario que emplea conceptos y métodos de la física para estudiar sistemas económicos. Tiene como objetivo entender, modelar y explicar fenómenos económicos al aplicar técnicas y herramientas de física estadística, sistemas complejos y dinámica no lineal.

1.2. Mercados Financieros como Sistemas Complejos

Hay registros de que los primeros mercados financieros empezaron en el año 1400 en la localidad de Antwerp hoy Bélgica [10], después en 1611 el comercio moderno de acciones es creado en Ámsterdam, aunque solamente una compañía cotizaba en esa época. Posteriormente se fueron originando el New York Exchange Market, el Dow Jones, el SP500 y como se ha mencionado, con la llegada de la era digital estos mercados han ido creciendo.

Prueba de lo anterior es el siguiente caso hipotético, si en 1957 se hubieran invertido 20,000 MXN (1,600 USD considerando la tasa de cambio de ese año) en el SP500, tomando en cuenta que el índice tiene un porcentaje anual promedio de 10.26 % y aplicando la fórmula de interés compuesto, en el año 2023 esta inversión valdría aproximadamente 19 millones de pesos mexicanos. Lo que significaría una ganancia bastante conveniente.

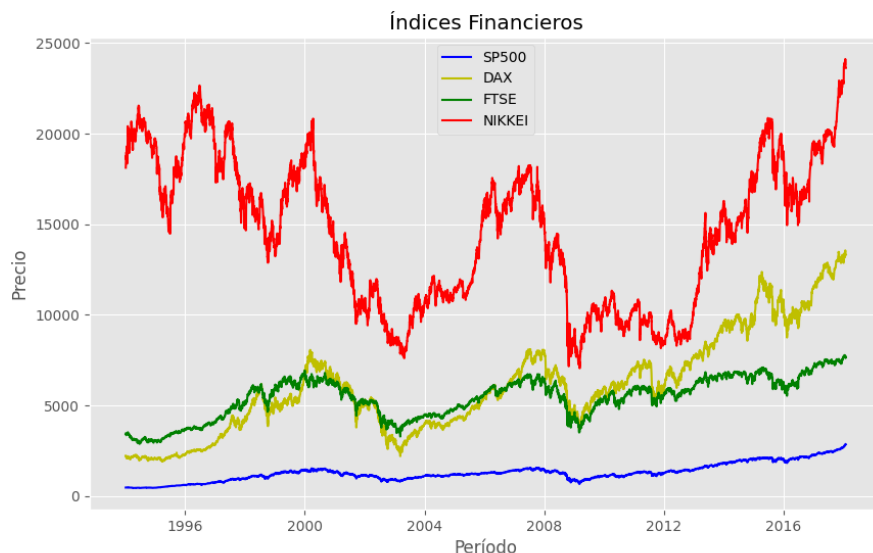


Figura 1.1: Cotización de algunos mercados financieros del mundo. El eje x representa la fecha de cotización y el eje y corresponde al valor de cada índice en dólares.

Sin embargo, invertir en algún mercado financiero no garantiza una ganancia segura e incluso se puede correr el riesgo de perder todo el patrimonio si no se es precavido. Las crisis financieras son las protagonistas de estas grandes pérdidas, por esto es que prevenirlas o tratar de predecirlas ha tomado mucha relevancia [11].

Los mercados financieros comprenden una gran gama de instrumentos financieros, dentro de los más conocidos podemos encontrar a las commodities, que son materias primas, algunos ejemplos comunes de estos son productos energéticos como el gas natural y el petróleo, también pueden ser productos de la agricultura como algodón, maíz, y soya, del sector de la minería como el oro, la plata, el cobre entre algunos otros minerales. Recientemente, en el año 2020 se empezó a cotizar futuros del agua como un commodity debido a la escasez sobre la que se empieza a especular en este recurso [28, 29].

Ya se ha hablado del mercado de divisas, conocido como Forex (por sus siglas en inglés foreign exchange) dentro de este mercado se compran y se venden divisas llegando a tener trillones de transacciones por día. Por otro lado, se pueden encontrar otro tipo de mercado negociando instrumentos financieros aún más complejos como lo son las opciones, este mercado tuvo un crecimiento masivo después de la publicación de la ecuación de Black-Scholes, en ocasiones se culpa a esta ecuación de ser la responsable de las crisis financieras actuales como la del 2008 [24].

Estos instrumentos son derivados financieros y fueron creados con el propósito principal de reducir el riesgo que tienen los inversionistas y comerciantes al momento de comprar o vender acciones [1, 2]. Una vez que un participante se vuelve titular de uno de estos instrumentos, tiene el derecho de vender o comprar un activo subyacente dentro de un plazo específico a un precio determinado. La complejidad asociada a estos instrumentos produjo, en consecuencia, que una gran cantidad de matemáticos y economistas se involucraran para desarrollar estrategias y maximizar las ganancias a través de estos instrumentos.

Los mercados financieros han alcanzado tal desarrollo, que recientemente se ha sumado un nuevo tipo de activo financiero [30], las llamadas criptomonedas que basan su funcionamiento en la tecnología criptográfica y a diferencia de las divisas normales, estas no tienen un intermediario como el banco y no se pueden encontrar en formato físico ya sea moneda o papel. Este mercado se ha ganado, entre los inversores, una reputación por ser extremadamente volátil.

Este trabajo de tesis se enfoca únicamente en las acciones de las empresas que cotizan dentro del índice más popular de Estados Unidos, el SP500 (por sus siglas en inglés: standard and poors 500). Este índice está compuesto de acciones, en específico de las 503 empresas mejor calificadas de Estados Unidos. Teniendo como referencia su capitalización de mercado, liquidez y representación en la industria. Así como existe este índice, podemos encontrar otros índices similares como el Dow Jones Industrial Average (DJIA), MSCI World Index, Nikkei 225, DAX Index entre otros.

En la Figura 1.1 se muestra una gráfica comparativa entre el precio de cotización de algunos índices del mundo desde 1997 hasta 2018, a pesar de que cada uno tiene un valor distinto en su precio de cotización, es evidente que en el paso del tiempo estos índices han sufrido movimientos grandes ya sea hacia bajo o hacia arriba casi de forma sincronizada.

Por otro lado, una acción es una fracción de una empresa [2], lo que quiere decir que si un individuo adquiere una acción, ahora se vuelve propietario de una fracción de la empresa. Las empresas empiezan a emitir acciones con el objetivo de financiarse a ellas y sus planes a futuro, esta práctica en finanzas se le llama oferta pública inicial (IPO).

Veamos el siguiente ejemplo: Supongamos que una empresa se dedica a la venta de maguey de agave y hasta el momento con inversiones propias ha logrado generar cierto margen de ganancia. Sin embargo, para crecer esta empresa necesita más inversión por lo que decide vender títulos que avalan que la persona que los compre es dueña parcial de la empresa y no solo eso, sino que estos títulos a medida que la empresa crezca y siga generando ingresos, aumentarán su valor.

Lo anterior es el funcionamiento básico de una acción y esta, dependiendo del estado de la empresa aumentará o disminuirá su valor. La parte interesante entra cuando son millones de personas las que compran y venden acciones y no solamente de una empresa sino de todas las que cotizan en el índice y el mercado.

Para fines de esta tesis, a cada uno de los individuos participantes del mercado se le considerará dentro del marco de la **hipótesis de agente racional** [1, 8, 34]. Esta hipótesis sirve para modelar el comportamiento humano y asume que cada individuo, al momento de tomar una decisión, lo hará racionalmente de manera que sus decisiones puedan maximizar su bienestar o utilidad. Por lo que, si miramos el panorama general, nos damos cuenta que los mercados financieros están formados por agentes independientes que interactúan entre sí dentro de un sistema el cual se ve afectado por sucesos internos como externos. Es por esto que los mercados financieros son considerados como un buen ejemplo de un sistema complejo. Para tener un concepto base sobre lo que es un sistema complejo, nos quedaremos con la proporcionada por Kwapien J. y Drozd, S. [6]:

“Un sistema complejo es un sistema formado por un gran número de constituyentes interactuando de forma no lineal entre sí, el cuál exhibe comportamiento colectivo y, debido a un intercambio de energía o información con el entorno, puede cambiar fácilmente su estructura interna y patrones de actividad.”

En el ámbito financiero los mercados son un sistema formado por una gran cantidad de agentes que interactúan entre sí de forma simultánea, los cuales pueden verse afectados por factores internos y factores externos. Como consecuencia de estas interacciones, el precio del mercado se moverá hacia arriba o hacia abajo de forma aparentemente aleatoria.

Dentro de la física, el ejemplo de sistema complejo más común son los fluidos [2], ya que sus moléculas interactúan entre sí a través de fuerzas electromagnéticas y cuánticas, lo que provoca comportamientos complejos a diferente escala, por ejemplo, la estructura dentro de un tornado, considerando la perspectiva de una partícula dentro, pareciera tener una estructura caótica, totalmente diferente en comparación con la estructura vista desde un punto lejano, donde el tornado se presenta como un fenómeno sumamente ordenado. Otros ejemplos de sistemas complejos son los siguientes:

- El cerebro.
- Ecosistemas.
- La sociedad.
- El clima.

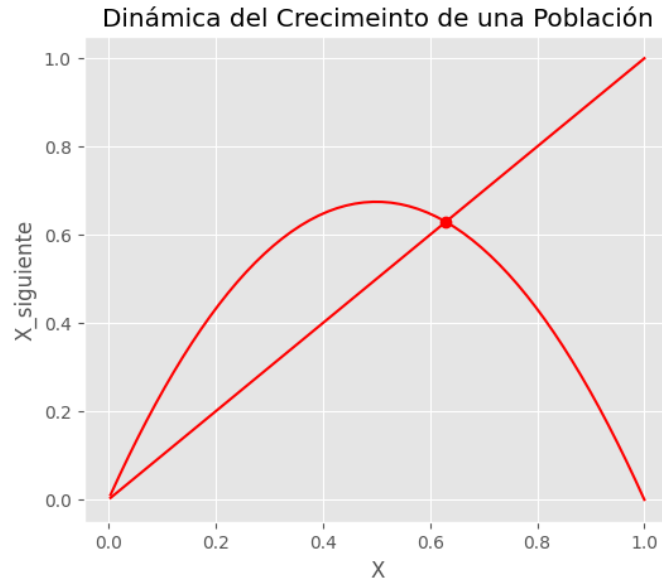


Figura 1.2: Dinámica de una población en biología; modelada por la ecuación logística.

En seguida presento un mapeo logístico [31], considerado como un ejemplo arquetipo dentro de los sistemas complejos. Este a pesar de ser simple, bajo ciertas condiciones puede presentar comportamiento complejo conocido también como caos determinista y se expresa como la siguiente ecuación:

$$x_{n+1} = rx_n(1 - x_n) \quad (1.2)$$

Donde x_n representa la población a determinado tiempo n , x_{n+1} representa la población posterior al tiempo n y r corresponde a la tasa de crecimiento de la población. La expresión anterior guarda cierta relación con la ecuación logística; mientras que ambas describen el crecimiento poblacional de una especie en determinado tiempo, la ecuación logística lo hace considerando un tiempo continuo

(ver ecuación 1.3), por otro lado, el mapeo logístico lo hace considerando un tiempo discreto.

$$\frac{dx}{dt} = rx(1 - \frac{x}{k}) \quad (1.3)$$

Un mapeo logístico demuestra como un proceso simple puede evolucionar a un proceso complejo. La ecuación 1.2 funciona como un modelo que permite calcular el valor de x en un instante posterior siempre y cuando se conozca su valor en determinado momento. Si tomamos un valor inicial de $x = 0.04$ y $r = 2.7$, al iterar varias veces esta ecuación se obtendría una gráfica como en la Figura 1.2, que muestra, como lo describió P. Verhulst, el crecimiento de una población en biología y otros campos. Si en principio se tiene una densidad de población baja, con el paso del tiempo la población iría creciendo hasta cierto límite para después decrecer, debido a la sobrepoblación, aumento de mortalidad y otros factores. Además, la

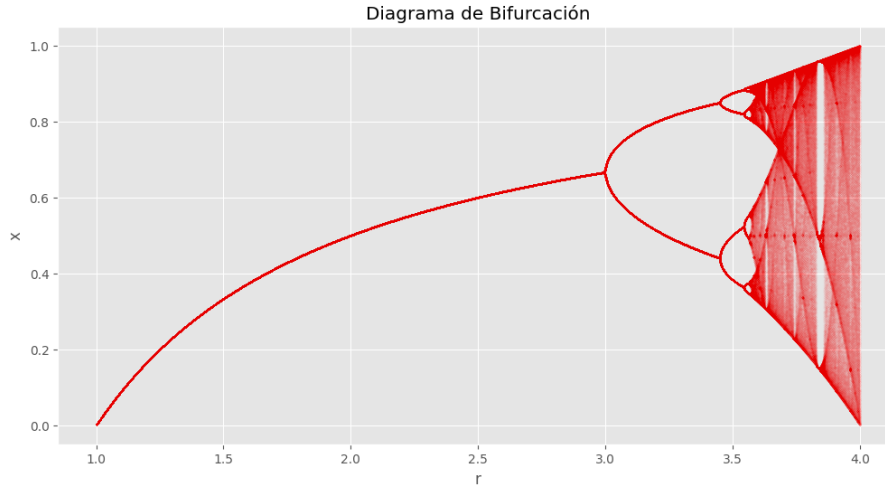


Figura 1.3: Visualización del diagrama de bifurcación para un mapeo logístico.

Figura 1.2, presenta una región de atracción en el punto $(0.6296, 0.6296)$, este punto es un estado en el cual el sistema tiende a evolucionar a través del tiempo, también conocido como un **atractor**.

Una vista general del modelo se puede ver en la Figura 1.3, este diagrama de bifurcación¹ representa el comportamiento del mapeo logístico para diferentes valores del parámetro r . Como se puede ver en la gráfica, para valores de r entre uno y tres no existe complejidad en el comportamiento de la gráfica. Mientras que a partir de $r = 3.0$ el mapeo logístico presenta bifurcaciones que se duplican hasta llegar a un comportamiento caótico.

1.3. Oferta, Demanda y Arbitraje

Como se ha explicado con anterioridad, los mercados financieros están compuestos por agentes independientes que interactúan entre ellos [6]. La forma que interactúan entre ellos es a través de la compra y venta de activos. Aquí es importante definir el concepto de **oferta y demanda**. Este concepto es un principio fundamental en economía que se encarga de asignar el precio más justo disponible para los agentes del mercado.

La oferta se refiere a la cantidad de un bien o servicio que los productores están dispuestos a ofrecer a determinados precios a lo largo de un periodo de tiempo [2, 34]. Esta puede

¹Referencia: https://en.wikipedia.org/wiki/Logistic_map

verse influenciada por factores como el costo de producción, competencia en el mercado, factores externos como políticas regulatorias, pandemias o desastres naturales.

Por otro lado, la demanda se refiere a la cantidad de bienes y servicios que los consumidores están dispuestos y son capaces de adquirir a los diferentes precios ofertados en un período de tiempo. Los factores que afectan a la demanda son el nivel de ingresos, la preferencia de los consumidores, el tamaño de la población, entre otros [2, 34].

Tanto la oferta y la demanda pueden llegar a un equilibrio cuando la cantidad de oferta iguala a la cantidad de demanda, resultando en un precio de mercado estable conocido como el precio de equilibrio.

Tomemos el ejemplo de un emprendedor. Esta persona busca generar ganancias a través de la venta de un producto, ahora, dicho producto para ser creado necesita una inversión inicial, considerando el material, la mano de obra, el lugar de elaboración, entre otros factores. Por ello, el emprendedor busca fijar el precio de su producto con el objetivo de cubrir esos gastos y además generar un margen de ganancia.

Fijémonos en el caso hipotético donde existe una tasa de producción baja (oferta), pero demanda de producto alto. Dicho emprendedor al darse cuenta de este fenómeno, aumentaría el precio de su producto. El emprendedor buscaría seguir creciendo por lo que mantendría el precio y aumentaría su producción. Esta acción generaría una competición en el mercado (dentro de los otros emprendedores con un producto similar o igual) debido a la gran cantidad de productos a elegir, por lo que los clientes, de entre toda la gama de productos a elegir terminarían decidiéndose por el mejor precio. Si dicho emprendedor estuviera al tanto de este hecho, bajaría el precio de su producto y esto ocasionaría que no se pudiera costar gran cantidad de productos por lo que se bajaría el nivel de producción y este proceso se reproduciría hasta encontrar un equilibrio entre el precio de producto y su producción.

La situación anterior ejemplifica a lo que se conoce como un caso convergente y se muestra en la Figura 1.4. Donde podemos notar que el punto de equilibrio se encuentra en la intersección entre la curva de oferta y la curva de demanda [2].

Es importante notar que el papel del precio del producto, sirve como información funda-

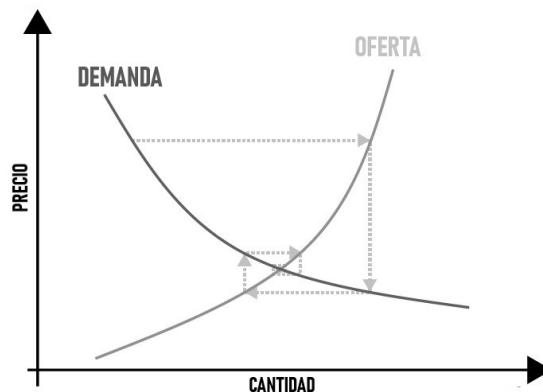


Figura 1.4: Diagrama de telaraña que muestra el caso convergente.

mental sobre el estado del mercado, y que el emprendedor interpreta para tomar decisiones futuras

sobre su empresa. De igual forma, este precio lo interpreta el consumidor para tomar decisiones sobre si seguir comprando un producto o no.

La oferta y la demanda son conceptos relacionados con el **arbitraje** [1], que es otro fenómeno presente dentro de la economía. Este básicamente permite obtener una ganancia libre de riesgo sobre una operación de compra y venta de cualquier producto o activo, tomando en cuenta que existe una diferencia en el tipo de cambio de moneda de donde se compra y vende.

Por ejemplo, supongamos que el precio de compra de un kilo de aguacate en México es de 58 pesos y el precio de venta del mismo kilo en estados unidos es de 4 dólares, si decide comprar 100,000 kilos de aguacate para vender en estados unidos, considerando que transportar y almacenar el cargamento costaría 3.39 pesos y el tipo de cambio entre el peso y el dólares de 16.93 entonces:

$$100,000[58 - (16.93 \times 4) - 3.39] = 5,437,400 \text{ pesos}$$

Se tendría una ganancia libre de riesgo de 5,437,000 pesos, si esta transacción se realiza de forma sucesiva, se vería un aumento considerable en la compra de aguacate en México, lo que se traduce como una demanda de aguacate alta, mientras que en estados Unidos esta demanda se vería reducida, por lo tanto, el precio del aguacate en México aumentaría y en Estados Unidos reduciría. Después cierto tiempo el precio llegaría a un punto de equilibrio o precio más justo, esta última consecuencia generaría que los mercados no ofrezcan diferentes precios para activos y bienes idénticos.

Este ejemplo se puede extrapolar al caso de los mercados financieros, sustituyendo el kilo de aguacate por un activo financiero y la tarifa por almacenar y transportar el producto por la comisión para el intermediario o casa de bolsa por proveer el servicio de acceder al mercado. Las oportunidades de arbitraje no son fenómenos raros, sino que más bien, ocurren espontánea y repetitivamente dentro de la dinámica de los mercados financieros, tan pronto una de estas oportunidades aparece los agentes querrán explotarla hasta que el mismo mercado elimine esta oportunidad [1].

The screenshot shows a trading interface for AAPL (Apple Computer Inc.). It displays the current bid price at \$3,025.380 and the ask price at \$3,034.990. The volume for both is 7,100. The order entry section has radio buttons for 'Compra' (selected) and 'Venta'. Under 'Nivel de operación', 'Básica' is selected. The order type is set to 'Limitada' (limited) with a validity of 'un día (DAY)'. The order is set to 'Libro' (book) rather than 'Mesa' (floor). The 'Títulos' (shares) field is set to 0, and the 'Precio límite' (limit price) is set to 3,034.990. The 'Importe total' is \$0.00 and the 'Comisión' is 0.25%. There are 'Cancelar' and 'Continuar' buttons at the bottom.

Figura 1.5: Interfaz para orden de compra proveída por la plataforma de la casa de bolsa Bursanet.

Dentro de los mercados financieros esta dinámica es fundamental. Con la llega de la era digital, vender y comprar acciones se ha vuelto más automático. Las casas de bolsa, que son intermediarios entre los agentes y el mercado, proveen a los participantes de una plataforma que muestra en tiempo real el precio de las acciones, este precio es el llamado precio de mercado, cuando una persona compra una acción puede hacerlo al precio de mercado o puede mandar una operación de límite de compra que consiste en comprar cuando el precio del activo llegue a un cierto nivel. De igual forma, el agente que busca vender una acción puede hacerlo al precio del mercado o a un precio que el considere beneficioso, a este tipo de operación se le llama límite de venta.

Estas operaciones se irán poniendo en la cola y cuando llegue la mejor oferta la operación se ejecutará. Este es un proceso que ocurre casi de manera automática y rápida por lo que a lo largo de un día, como se ha mencionado con anterioridad, la cantidad de operaciones que se ejecutan son bastantes. Y todas las operaciones tanto de venta como de compra se verán reflejadas en lo que se conoce como volumen. La Figura 1.5, muestra la interface para realizar una operación sobre la acción de Apple.

1.4. Procesos Estocásticos

Al momento de tratar de describir un fenómeno de la naturaleza, un acercamiento valido es a través del determinismo. Por ejemplo, considerando un objeto en caída libre, es posible estar seguros en que tiempo llegará al suelo, siempre y cuando tengamos conocimiento de ciertas condiciones iniciales como la altura desde donde empieza la caída [19]. Sin embargo, para describir los fenómenos más complejos de la realidad, es necesario tomar un punto de vista probabilístico. En esta sección, se describe este acercamiento estocástico.

1.4.1. Variables Aleatorias.

Dentro de los fenómenos de la naturaleza la incertidumbre siempre está presente y por ello es importante saber distinguir en qué casos analizar el problema a través de variables deterministas o a través de variables aleatorias. Una variable aleatoria cuantifica el nivel de incertidumbre de un modelo [19, 20]. De forma general podemos decir que

Definición 1.4.1. *Una variable aleatoria, X , es una función que asocia un número real con cada elemento del espacio muestral.*

En el ejemplo del objeto en caída libre se hacen asumpciones del fenómeno real donde se está ignorando la fricción del aire, para así tratar con variables deterministas. Por otro lado, tratar con variables aleatorias nos permite asignarle a la variable X un rango de posibles valores x cada uno con una probabilidad de ocurrir. En general se puede considerar que todos los procesos físicos son aleatorios siempre que se consideren a las variables deterministas como un tipo de variable aleatoria con un rango de variación cero [19].

Definición 1.4.2. *La evolución temporal de una variable aleatoria es llamada un proceso estocástico, $X(t)$.*

El número de puntos en la cara de un dado representa una variable aleatoria, X , especificada por el rango de valores $\{1, 2, 3, 4, 5, 6\}$, y cada uno de estos valores x , asume una probabilidad $P(x)$. Por lo tanto, para cada valor del rango de X , tenemos que

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6} \quad (1.4)$$

Considerando, ahora, dos dados, el evento de tirar el primer dado y obtener un 3, no afectará la probabilidad de que el resultado de la tirada del segundo dado sea igual a 3. Se dice entonces que ambos resultados son estadísticamente independientes y la probabilidad de que ambas tiradas arrojen un resultado de 3 es el producto $(\frac{1}{6})(\frac{1}{6}) = \frac{1}{36}$.

Si la ocurrencia de $X = x$ no cambia la probabilidad $P(y)$ de que el resultado $Y = y$ se obtenga y viceversa, los resultados $X = x$ y $Y = y$ son independientes y la probabilidad de que ocurran conjuntamente es el producto de su propia probabilidad

$$P(x, y) = P(x)P(y)$$

De otra forma, si el evento $X = x$ afecta la probabilidad de que ocurra $Y = y$, se dice que las variables X y Y son estadísticamente dependientes [19, 20], por lo que, tratar de obtener la probabilidad $P(y)$ dado que ya ocurrió $X = x$, se trataría de una probabilidad condicional determinada como

$$P(y|x) = \frac{P(x, y)}{P(x)}$$

1.4.2. Momentos

Los momentos, en el contexto de una variable aleatoria, consisten en una colección de medidas numéricas diseñadas para capturar diferentes atributos de la distribución de probabilidad subyacente. Sirven como medio para cuantificar y representar con precisión características específicas de la distribución como su forma, ubicación de su centro, el alcance de sus extensión, entre otras [2, 7, 19, 20]. Los momentos más comunes son los siguientes:

- Valor esperado (primer momento).
- Varianza (segundo momento).
- Oblicuidad (tercer momento).
- Curtosis (cuarto momento).

Para fines de esta tesis, solamente se definirán los primeros dos momentos.

El **valor esperado** de una variable aleatoria, es una operación matemática que transforma la probabilidad $P(x)$ en una cantidad determinística conocida como el promedio de X , [19, 20]. Este promedio es el número que mejor caracteriza todos los posibles valores de una variable aleatoria. Para una variable aleatoria discreta se define como

$$E(X) = \sum_{i=0}^{\infty} x_i P(x_i) \quad (1.5)$$

El valor esperado de una variable aleatoria se puede expresar mediante la notación de corchetes triangulares. En esta tesis, se adoptará esta notación, por lo tanto, podemos expresar la ecuación 1.5 como:

$$\langle X \rangle = \sum_{i=0}^{\infty} x_i P(x_i) \quad (1.6)$$

La **varianza** de una variable aleatoria, X , es el segundo momento alrededor del promedio denotado por $V(X)$ o $var\{X\}$, [19, 20]. Esta da una medida de variación o dispersión de la distribución de probabilidad de X y se define como

$$V(X) = var\{X\} = \langle (X - \langle X \rangle)^2 \rangle \quad (1.7)$$

1.4.3. Covarianza y Coeficiente de Correlación

Dado que la suma, así como cualquier otra función, de dos variables aleatorias es una variable aleatoria, es posible definir el valor esperado como

$$\langle X + Y \rangle = \langle X \rangle + \langle Y \rangle. \quad (1.8)$$

Definición que se puede extender considerando que el desplazamiento total, X , de una partícula en un intervalo de tiempo, es la suma de los desplazamientos sucesivos X_i dentro de intervalos sucesivos

$$\langle X \rangle = \langle X_1 \rangle + \langle X_2 \rangle + \dots + \langle X_n \rangle. \quad (1.9)$$

Es decir, el promedio de la suma es igual a la suma de promedios, la ecuación 1.9 es conocida como el teorema del promedio de la suma.

Por otro lado, tratar de definir los momentos del producto XY no es tan directo. En general podemos decir que

$$\langle XY \rangle = \sum_i \sum_j x_i x_j p(x_i, y_j). \quad (1.10)$$

Cuando se trata de dos variables estadísticamente independientes, la ecuación 1.10 se reduce a

$$\begin{aligned} \langle XY \rangle &= \sum_i x_i p(x_i) \sum_j y_j p(y_j) \\ &= \langle X \rangle \langle Y \rangle. \end{aligned} \quad (1.11)$$

Sin embargo, cuando las variables aleatorias X y Y son estadísticamente dependientes no es posible aplicar la ecuación 1.11. Y por ello se define la covarianza:

$$\text{cov}\{X, Y\} = \langle XY \rangle - \langle X \rangle \langle Y \rangle \quad (1.12)$$

y el coeficiente de correlación:

$$\text{corr}\{X, Y\} = \frac{\text{cov}\{X, Y\}}{\sqrt{\text{var}\{X\} \text{var}\{Y\}}}. \quad (1.13)$$

Ecuación 1.12 y 1.13 son mediciones de dependencia estadística entre X y Y . El coeficiente de correlación establece una escala adimensional definida en el rango $-1 \leq \text{corr}\{X, Y\} \leq 1$ que nos permite saber cuan relacionadas están las variables aleatorias [19, 20]. Considerando tres casos; se dice que X y Y están completamente correlacionadas cuando $\text{corr}\{X, Y\} = 1$. Cuando X y Y están completamente anticorrelacionadas ocurre que $\text{corr}\{X, Y\} = -1$. Y el tercer caso se refiere a cuando X y Y son estadísticamente independientes, por lo tanto $\text{cov}\{X, Y\} = 0$ y $\text{corr}\{X, Y\} = 0$.

1.4.4. Caminata Aleatoria

La caminata aleatoria tiene un papel fundamental en las ciencias naturales. Una descripción teórica se encuentra plasmada en el artículo publicado por Einstein en 1905, *Sobre el movimiento de pequeñas partículas suspendidas en un líquido estacionario, según lo exige la teoría cinética molecular del calor* [21], sin embargo, la formalización de la teoría de una caminata aleatoria ya se había abordado, en la tesis doctoral de Bachelier, *Teoría de la Especulación* [17], cinco años antes. En ella, Bachelier propone que el movimiento de los precios dentro de los mercados financieros siguen un patrón aleatorio e impredecible. Es decir, el precio de las acciones evoluciona en el tiempo de acuerdo a una caminata aleatoria.

Una caminata aleatoria es un proceso en el cual una variable aleatoria se mueve desde un valor inicial hacia otro. La forma más sencilla de representar este movimiento es a través de una dimensión, es decir, la variable aleatoria se moverá cierto número de pasos consecutivos a lo largo del eje x [2].

Consideremos la variable aleatoria X , cuya posición al tiempo t se representa por $X(t)$, donde t es una variable de tiempo discreta no negativa. Suponiendo que $X(0) = 0$, podemos determinar la posición de la variable con

$$X(t) = X(t-1) + \Delta X(t) \quad (1.14)$$

Donde $\Delta X(t)$, representa el desplazamiento de la variable aleatoria y obedece una función de probabilidad con dos posibles resultados: $+1$ con probabilidad p y -1 con probabilidad $q = 1 - p$,

donde $0 < p < 1$, [7, 18, 19]. En general, dada una caminata aleatoria, la posición de la variable después de n pasos, es conocida como **estado** mientras que a los pasos se refiere como **etapas** [7]. Se dice que una caminata es aleatoria es simétrica, siempre que $p = q = \frac{1}{2}$. En la Figura 1.6 se muestra una gráfica de una caminata aleatoria. En la Figura 1.7 se muestra una gráfica del precio de la

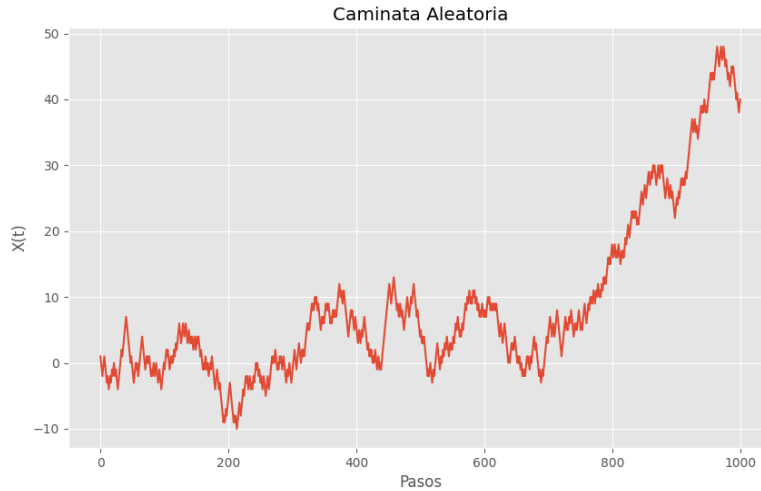


Figura 1.6: Gráfica de una caminata aleatoria.

acción de Amazon desde el año 2006 hasta el año 2022, como es posible apreciar, el movimiento del precio de las acciones se asemeja al movimiento reproducido por la caminata aleatoria. Esta fue la idea que precisamente Bachelier planteo en su tesis, asegurando que el cambio en los precios de las acciones tenían una distribución gaussiana [1, 2]. Sin embargo, Bachelier paso por alto la existencia de precios negativos. Problema que se abordó considerando que el logaritmo de los precios están gaussianamente distribuidos, resultando en el movimiento Browniano geométrico descrito en la siguiente sección.



Figura 1.7: Gráfica del precio de las acciones de la empresa Amazon.

1.4.5. Movimiento Browniano

De acuerdo con la primera ley de Newton, si una partícula se encuentra en reposo en un determinado instante y tiene velocidad cero, permanecerá en reposo en la misma posición siempre y cuando ninguna fuerza actué sobre dicha partícula.

Incluso cuando se procura aislar a una partícula de forma que ninguna fuerza actué sobre ella, aún están presente las moléculas del aire que pueden afectar el estado de la partícula [19]. El efecto de estas moléculas de aire pueden ser observadas a través de un microscopio, siempre que se trate de partículas muy pequeñas ($\leq 50 \times 10^{-6}\text{m}$). Este fenómeno fue visto por primera vez, por el botánico escocés Robert Brown² en el año 1827.

Cuando Robert Brown observaba a través del microscopio, granos de polen suspendidos en agua, se dio cuenta que cierto grupo de granos siempre se dispersaba y movía continua e irregularmente. En 1905 Einstein publicó su artículo [21] donde mostraba que el movimiento térmico de las moléculas resultaba en un movimiento aleatorio de partículas suspendidas posiblemente parecido al movimiento browniano.

Sabemos que el movimiento browniano es consecuencia de la teoría atómica de la materia. Que establece que una partícula suspendida en un fluido llámese agua, aire o gas, se ve afectada por las moléculas o átomos que componen dicho fluido [18, 19]. Golpeando a la partícula desde varias direcciones en un intervalo de tiempo.

Es posible modelar al movimiento browniano a través de la caminata aleatoria definida en la sección anterior. Considerando el modelo como la suma de desplazamientos aleatorios, donde la partícula browniana empieza en $x=0$. Después de cierto intervalo de tiempo Δt , el efecto neto de los muchos impactos individuales de moléculas sobre la partícula se verá reflejado en un desplazamiento de la partícula una cantidad aleatoria X_i . Cada desplazamiento X_i , tiene dos posibilidades, $X_i = +\Delta X$ o $X_i = -\Delta X$, cada una con igual probabilidad $\frac{1}{2}$ de ocurrencia. Después de n intervalos, el desplazamiento neto X es

$$X = X_1 + X_2 + \dots + X_n. \quad (1.15)$$

Dado que cualquier partícula browniana pueda desplazarse a una posición diferente a su punto de partida, el promedio del desplazamiento, $\langle X \rangle$, mantiene su valor cero inicial, esto es de acuerdo al modelo

$$\langle X_1 \rangle = \langle X_2 \rangle = \dots \langle X_n \rangle = 0 \quad (1.16)$$

ya que

$$\langle X_i \rangle = \left(\frac{1}{2}\right)(+\Delta X) + \left(\frac{1}{2}\right)(-\Delta X) = 0 \quad (1.17)$$

para cada $i = 1, 2, \dots, n$. Entonces

$$\begin{aligned} \langle X \rangle &= \langle X_1 \rangle + \langle X_2 \rangle + \dots + \langle X_n \rangle \\ &= 0. \end{aligned} \quad (1.18)$$

Por otro lado, la varianza es

$$\begin{aligned} \text{var}\{X_i\} &= \langle X_i^2 \rangle - \langle X_i \rangle^2 \\ &= \Delta x^2. \end{aligned} \quad (1.19)$$

Entonces

$$\langle X^2 \rangle = \sum_i \text{var}\{X_i\} = n\Delta x^2. \quad (1.20)$$

²De ahí el nombre movimiento browniano.

Dado que la duración total de la caminata aleatoria es $t = n\Delta t$, la ecuación 1.20 es equivalente a

$$\langle X^2 \rangle = \left(\frac{\Delta x^2}{\Delta t} \right) t. \quad (1.21)$$

La ecuación 1.21 anterior nos dice que la varianza del desplazamiento neto, es proporcional al tiempo t durante el cual el desplazamiento es realizado [19]. Esta es una propiedad fundamental del movimiento browniano. El movimiento Browniano o el movimiento Browniano aritmético, la

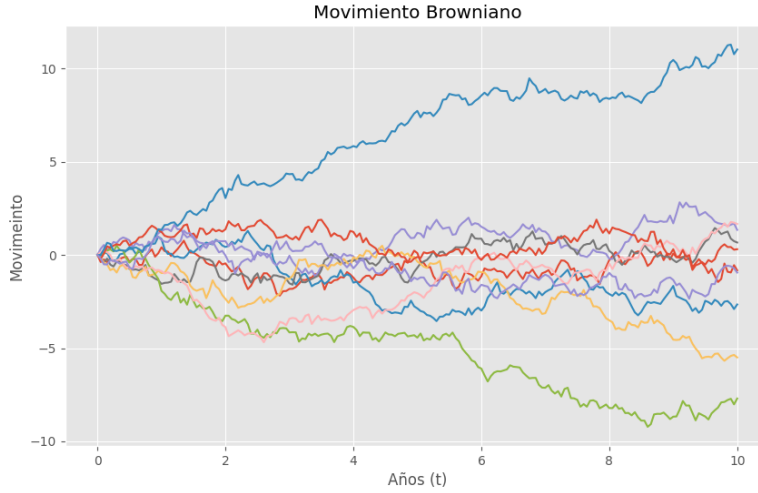


Figura 1.8: 10 simulaciones diferentes de movimiento browniano.

variable cambia una cantidad fija cada paso. Este cambio es aditivo lo que quiere decir que el valor absoluto de la variable cambia aleatoriamente, pero el cambio porcentual se mantiene constante y la dinámica del movimiento esta descrita por la siguiente ecuación:

$$dX_t = \mu dt + \sigma dW_t. \quad (1.22)$$

La expresión anterior es una ecuación diferencial estocástica [3, 32], donde el termino dX_t representa el cambio infinitesimal de la variable aleatoria X_t sobre un pequeño intervalo de tiempo dt . El término μ se refiere a la componente no aleatoria que describe como X_t evoluciona, σ es el coeficiente de volatilidad de la variable X_t . Y el termino dW_t es un proceso diferencial de Wiener, y representa una componente aleatoria o ruido.

En la Figura 1.8 se muestran diez caminos diferentes para un modelo de caminata aleatoria aritmético. Es importante apuntar que, la variable en un movimiento Browniano aritmético sigue una distribución normal con promedio y varianza constantes sobre el tiempo. Por otro lado, la variable en el caso del movimiento Browniano geométrico sigue una distribución normal logarítmica [1] y su dinámica está definida por la siguiente ecuación

$$dS_t = \mu S_t dt + \sigma S_t dW_t \quad (1.23)$$

donde en contraste con la ecuación 1.22 se ha añadido el termino s_T a la componente determinística y a la componente de ruido. Que representa el valor de la variable aleatoria al tiempo t . En el contexto de finanzas, S_t , se refiere al precio de un activo [32]. En la Figura 1.9 se muestran diez caminos diferentes para un modelo de caminata aleatoria geométrica.

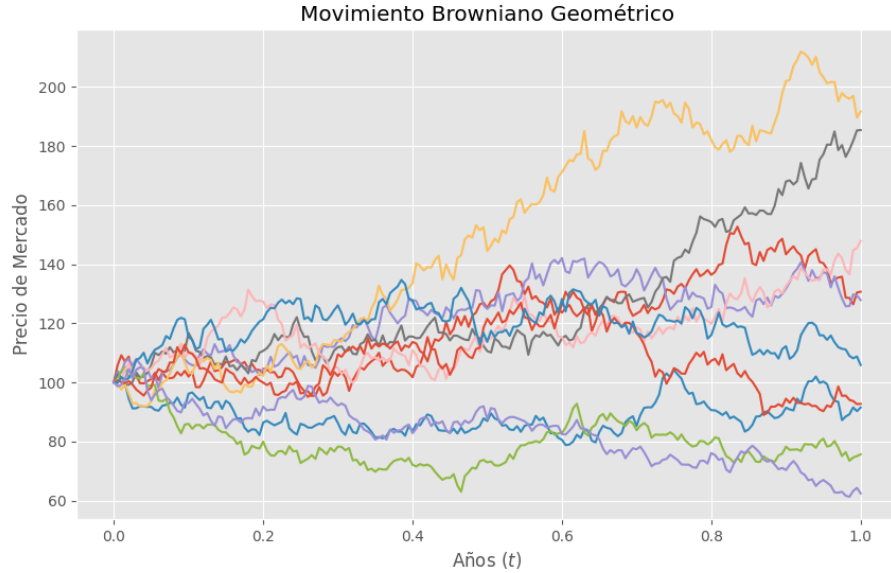


Figura 1.9: 10 simulaciones diferentes de movimiento browniano geométrico.

1.4.6. Proceso Markoviano

Andrey Markov fue un matemático ruso, reconocido principalmente por afirmar que no solo los eventos independientes pueden converger a distribuciones predecibles, sino que también era posible modelar una secuencia de eventos aleatorios usando estados y transiciones de estados. Más tarde este tipo de modelo se conoció como una cadena de markov [33].

Dicho de otra forma, una cadena de Markov es un modelo que describe una secuencia de estados donde el estado futuro del modelo depende únicamente del estado presente ignorando los estados pasados del modelo.

Por ejemplo, podemos considerar un proceso aleatorio que atraviesa una secuencia de pasos enumerados $n = 0, 1, 2, \dots$, donde la variable aleatoria X_n es el resultado después de n pasos y X_0 es la posición inicial del proceso [7]. Esta variable aleatoria puede tomar valores $i = 1, 2, \dots, m$. Donde los resultados reales son llamados los estados del sistema y están denotados por $E_i (i = 1, 2, \dots, m)$.

Se dice que el sistema ha hecho una transición del estado $E_i \rightarrow E_j$ si las variables aleatorias $X_{n-1} = i$ y $X_n = j$. Las cadenas de Markov tienen la propiedad de que la probabilidad de que $X_n = j$ únicamente depende en el estado previo al sistema, se puede decir que, las cadenas de markov se caracterizan por ser un modelo de memoria de corto plazo [2, 3, 7].

Definición 1.4.3 (Probabilidades de transición.). *Para una cadena de markov finita con m estados $E_1, E_2, \dots, E_m$, donde $i, j = 1, 2, \dots, m$.*

$$p_{i,j} = \mathbf{P}\{X_n = j | X_{n-1} = i\} \quad (1.24)$$

son las probabilidades de transición de la cadena y representan la probabilidad de una transición del estado E_i al estado E_j .

Estas probabilidades de transición deben ser mayores a 0 y la suma de todas estas sobre los

m estados debe ser igual a 1. Igualmente, es importante apuntar que para cada subíndice i fijo, la lista $\{p_{i,j}\}$ es una distribución de probabilidad.

Definición 1.4.4 (Matriz de transición.). *Las transiciones de probabilidad forman un arreglo $m \times m$, el cual se puede reordenar en la matriz de transición T , donde*

$$t = [p_{i,j}] = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix} \quad (1.25)$$

1.5. Hipótesis de Mercado Eficiente

En esta sección se introduce el concepto de hipótesis de mercado eficiente que es una idea fundamental dentro del área de economía y finanzas. Esta hipótesis se vio motivado por las investigaciones de L. Bachelier sobre caminata aleatoria y procesos estocásticos que ya se han introducido en las secciones anteriores.

1.5.1. Conceptos

A continuación, se introducen conceptos que ayudaran a facilitar el entendimiento sobre la relación entre caminata aleatoria y la hipótesis de mercado eficiente.

Rendimientos Logarítmicos. Las tasas de rendimiento esperadas se refieren a las ganancias o pérdidas respecto a una determinada inversión dentro de un periodo de tiempo y sirven como indicador para los inversores para saber si sus inversiones tienen buenos o malos desempeños [3]. Estas tasas de rendimiento esperadas normalmente se expresan mediante porcentajes a través de la siguiente expresión

$$\begin{aligned} r(t) &= \frac{p(t + \Delta) - p(t)}{p(t)} \\ &= \frac{p(t + \Delta)}{p(t)} - 1. \end{aligned} \quad (1.26)$$

En la ecuación 1.26 [2], $p(t)$ representa el precio de un activo a determinado tiempo t y Δ es algún periodo de tiempo. Sin embargo, si se tratase de comparar dos inversiones, se estarían comparando las diferencias de dos órdenes de magnitud. Por lo tanto, es preferible usar los rendimientos logarítmicos

$$r_n \approx \log(p_{n+1}) - \log(p_n) \quad (1.27)$$

donde la ecuación 1.27 es una aproximación de primer orden, por lo que aplica mejor para rendimientos pequeños. Dentro de la ecuación p_n es una secuencia de valores positivos que representa los precios en pasos de tiempo discreto.

Distribuciones de ley de potencias [1, 8] Estas son un tipo específico de distribución de probabilidad que describe una relación de escala entre la frecuencia de un evento y su magnitud. Por lo que la frecuencia de un evento es inversamente proporcional a la magnitud de dicho evento

$$P(x) \propto x^{-\alpha}. \quad (1.28)$$

En la ecuación 1.28 [8], $P(x)$ es la probabilidad de un evento de magnitud x y α es el exponente de una distribución de ley de potencias. Estas distribuciones también son conocidas como distribuciones de pareto [2, 3] debido a los trabajos pioneros del italiano Vilfredo Pareto que

trataban de calcular la distribución de la riqueza dentro de una población, el "principio de pareto" o la "regla 80-20" fue una consecuencia de estos trabajos y establece que el 80 % de los efectos proviene del 20 % de las causas. En el contexto de economía esta regla quiere decir que alrededor del 20 % de la población posee aproximadamente el 80 % de la riqueza.

Colas de Distribución [2]. Las distribuciones normalmente se clasifican de acuerdo al comportamiento de sus colas, específicamente cuando las probabilidades decrecen a medida que la variable se aleja del centro de la distribución. La distribución de ley de potencias se caracteriza por tener una cola de distribución pesada. Esto quiere decir que los eventos extremos tienen una probabilidad alta de suceder.

1.5.2. Hipótesis de Mercado Eficiente

Esta hipótesis fue formulada por primera vez por Eugene Fama en su artículo "Efficient Capital Markets: A Review of Theory and Empirical Work" [22] donde la principal idea establece que un inversionista es incapaz de superar al mercado, es decir, de sacar alguna ganancia significativa comprando una acción o armando un portafolio con acciones seleccionadas de forma aleatoria.

De manera más formal se dice que un mercado es eficiente si toda la información relevante al precio de las acciones está libremente disponible y esta información tiene un efecto inmediato en el precio de las acciones [1, 2, 8].

Esta hipótesis tiene tres formulaciones diferentes:

1. Débil.
2. Semi-fuerte.
3. Fuerte.

El enfoque débil, solo toma en cuenta que la información pública del mercado está disponible para todos los participantes, mientras que el enfoque semi-fuerte que dicha información se ve inmediatamente reflejada en el precio de los activos. Y, por último, la formulación fuerte considera que tanto la información pública como privada actualiza el mercado de forma inmediata [34].

Esta hipótesis tiene algunas consecuencias; dado que se cumpla que un mercado es eficiente entonces el precio de los activos se comportaran de acuerdo a una caminata aleatoria, y como se ha mencionado anteriormente, esto lleva a que el precio de los activos no dependa del precio en tiempos anteriores, sino que únicamente dependa del precio presente (es decir, los precios siguen una cadena de Markov) [1].

Es importante resaltar que esta hipótesis solamente sirve para modelar el mercado, los mercados reales no son perfectamente eficientes ya que, si lo fueran, las burbujas económicas no existirían e incluso tampoco existiría el trading, puesto que toda la información disponible sobre los activos se actualizaría y entonces dejaría de existir alguna forma de generar ganancias a través de estas actividades.

1.6. Series de Tiempo Financieras

Al momento de estudiar los mercados financieros, en específico los precios de cada una de las acciones que se encuentran cotizando dentro de ellos, es necesario hacer el tratamiento a través de las series de tiempo. En este capítulo se hace una introducción a dicho concepto y algunas características relevantes para este proyecto.

1.6.1. Series de Tiempo.

A diferencia de otros tipos de datos que se pueden estudiar de mejor manera desde la teoría estadística estándar y ciertos métodos basados en muestras aleatorias que asumen independencia entre sus elementos. Los elementos por los que están compuestas las series de tiempo son dependientes del tiempo [35].

Por ejemplo, al momento de llevar a cabo un experimento y hacer las observaciones apropiadas, la recolección de los datos se vería como en la Tabla 1.1. Por otro lado, si consideramos

Observación	Valor
1	7
2	27
3	5
$\vdots$	$\vdots$
n	n

Tabla 1.1: Ejemplo de una colección de datos normales.

una serie de tiempo como una secuencia de observaciones registradas de forma secuencial en el tiempo, estamos añadiendo de forma explícita una dependencia del tiempo a cada una de las observaciones como se muestra en la Tabla 1.2. Existen muchos fenómenos y casos de estudio que

Observación	Valor
Tiempo 1	7
Tiempo 2	27
Tiempo 3	5
$\vdots$	$\vdots$
Tiempo n	n

Tabla 1.2: Ejemplo de una colección de datos correspondientes a una serie de tiempo.

pueden considerarse como series de tiempo, en seguida se enlistan algunos ejemplos [27, 35]:

- Un electroencefalograma medido a segundos para determinar qué tipo de actividad realiza un paciente.
- Taza de nacimientos de un hospital cada año.
- Número de pasajeros de un aeropuerto a lo largo del año.
- La temperatura de una oficina a lo largo del día.
- Datos meteorológicos.

Como se puede ver, es posible caracterizar cualquier fenómeno a través de series de tiempo siempre y cuando pueda ser medible y asociarse a determinada período de tiempo. En el caso de esta tesis el objeto de estudio son los precios de apertura diarios de las acciones cotizadas dentro del índice SP500.

El intervalo de tiempo entre un punto en el tiempo y el siguiente, se le llama *período de tiempo* y la periodicidad con la que los valores son registrados se le conoce como frecuencia [36]. Esta frecuencia puede variar desde unos milisegundos hasta varios años. En el caso de los mercados financieros es posible acceder a datos con una frecuencia de hasta un minuto. Sin embargo, estas

ventanas de tiempo tan corto en los mercados se consideran de alta volatilidad.

En términos más formales [4], una serie de tiempo se define como una secuencia de observaciones, donde cada observación $Z_{i,t}$ está asociada a cierto orden de tiempo t :

$$\mathbf{Z}_t = [Z_{1,t}, Z_{2,t}, \dots, Z_{n,t}] \quad (1.29)$$

Para $i = 1, 2, \dots, m$ y $t = 0, \pm 1, \pm 2, \dots, T$, donde el primer subíndice corresponde a la componente y el segundo subíndice corresponde al tiempo, la T indica la última observación. En este caso, y por practicidad, se ha considerado un intervalo de tiempo discreto.

La relaciones entre las observaciones a diferentes puntos de tiempo, $Z_{i,s}$ y $Z_{j,t}$, pueden exhibir correlaciones sin importar si las instancias de tiempo, s y t coinciden o no. Dichas observaciones son variables aleatorias valuadas vectorialmente y como se ha dicho antes, la teoría estadística estandar para su tratamiento no es aplicable.

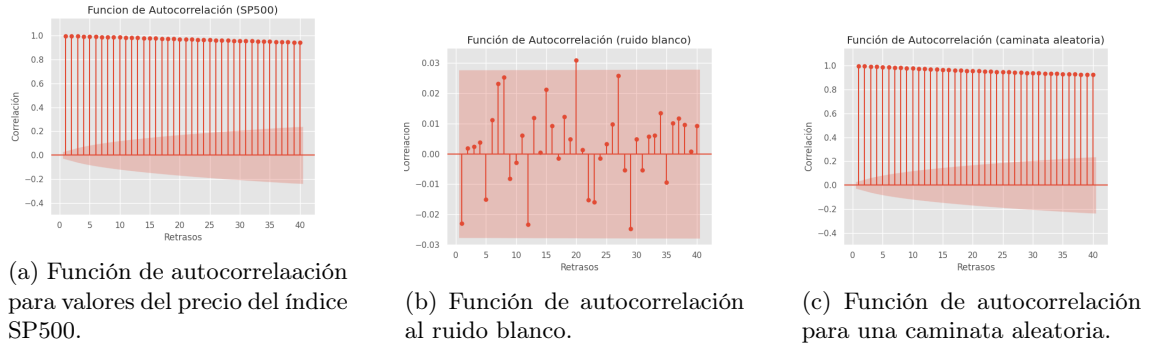


Figura 1.10: Gráficas de funciones de autocorrelación para diferentes tipos de series de tiempo. (a) y (c) sugieren que es posible hacer predicciones ya que muestran correlaciones entre sus instancias de tiempo, a diferencia de (b) que no es posible ver algún patrón.

1.6.2. Autocorrelación y Autocorrelación Parcial.

La autocorrelación dentro de una serie de tiempo univariada se refiere a cuan relacionada está dicha serie con una versión de ella misma retrasada [37]. Es decir, mide el nivel de semejanza entre una secuencia de hace varios periodos y los datos reales. El retraso o lag determinará los periodos previos de la secuencia a comparar. Por ejemplo, si se decide calcular la autocorrelación para una serie de tiempo con frecuencia diaria, se estaría midiendo el nivel de similitud entre los datos de ayer con los datos de hoy.

Existen 3 tipos de autocorrelaciones:

- **Autocorrelación positiva:** Si existe una autocorrelación positiva con respecto a la secuencia con retraso k , los valores altos tenderán a seguir a los valores altos y los valores bajos tenderán a seguir a los valores bajos, con un retraso de k periodos de tiempo.
- **Autocorrelación negativa:** Inversamente a la autocorrelación positiva, los valores altos tenderán a seguir a los valores bajos y los valores bajos tenderán a seguir a los valores altos, con un retraso de k periodos de tiempo.
- **Autocorrelación cero:** Este caso sugiere que no existe relación alguna entre valores a diferentes instancias de tiempo.

En la Figura 1.10, se muestra un ejemplo de la autocorrelación presente para diferentes series de tiempo, en la gráfica (a) se muestra para el precio del índice SP500, en (b) para ruido blanco y la gráfica (c) se muestra para una caminata aleatoria.

Por otro lado, tenemos la función de autocorrelación parcial que sirve para capturar los efectos de los momentos anteriores de manera indirecta, es decir, a todos los demás canales a través de los cuales los datos del pasado afectan a los datos actuales.

1.6.3. Estacionariedad y Estacionalidad.

Una serie de tiempo se considera estacionaria cuando se mantiene estable a lo largo del tiempo. Dicho de otra forma, su media y varianza se mantienen constantes y no cambian con el tiempo y además, no presentan tendencias [35, 37]. Esta característica es visualmente evidente cuando los valores de las series de tiempo tienden a fluctuar alrededor de una media fija, y el grado de variabilidad alrededor de dicha media se mantiene constante con el tiempo.

Definición 1.6.1. Se dice que, el siguiente proceso de serie de tiempo vectorial m -dimensional

$$\mathbf{Z}_t = [Z_{1,t}, Z_{2,t}, \dots, Z_{m,t}]' \quad (1.30)$$

es un proceso estacionario si cada uno de sus series componentes es un proceso estacionario univariado y sus primeros dos momentos son invariantes del tiempo.

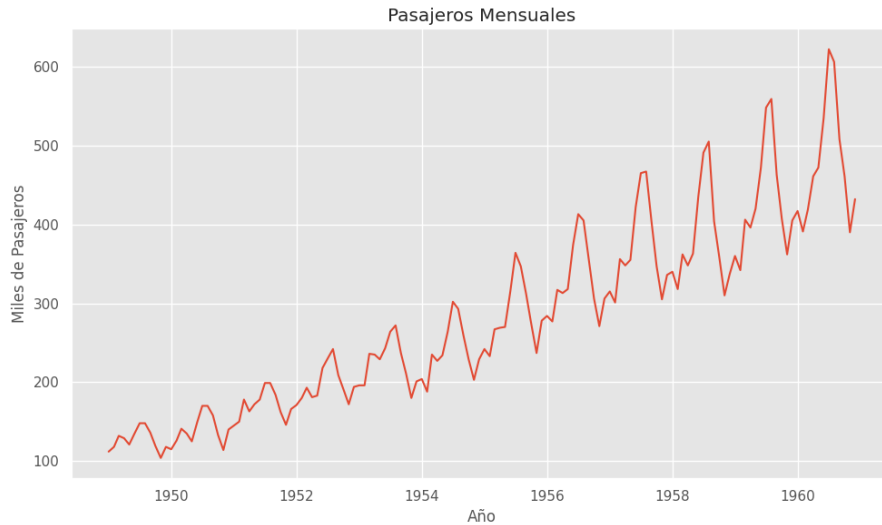


Figura 1.11: Gráfica que ejemplifica la característica de estacionalidad dentro de una serie de tiempo.

Por otro lado, una serie de tiempo no estacionaria se refiere a una serie en la cual se muestra efectos estacionales, tendencias, y otras estructuras que dependen del tiempo. Los cambios en la media determinarán una tendencia a crecer o decrecer a largo plazo, por lo que no oscilaría alrededor de un valor constante.

Un proceso estacionario de series de tiempo vectoriales es caracterizado por su vector de medias, función de matriz de correlación y su función de matriz de correlación parcial.

Por otro lado, la estacionalidad de una serie de tiempo se refiere a que ciertas tendencias

aparecerán cíclicamente de forma regular dentro de un período de tiempo. Una estructura de ciclo en una serie de tiempo puede o no ser estacional [36, 37]. Para que la estructura se considere estacional, es necesario que se repita consistentemente a la misma frecuencia.

Por ejemplo, si se estudian las temperaturas a lo largo del año se puede encontrar un patrón repetitivo en ciertos meses del año. Igualmente, si en lugar de estudiar las temperaturas a lo largo del año, se estudian a lo largo del día se encontrará que las temperaturas suben y bajan en determinadas horas del día. Existen distintos tipos de estacionalidad, algunos se presentan en seguida:

- Diarios.
- Semanales.
- Mensuales.
- Anuales.

En la Figura 1.11, se muestran la cantidad de pasajeros a lo largo de los años con una estacionalidad mensual, como se puede ver existen estructuras de patrones que se repiten a lo largo de los años, esto se puede explicar debido a que la mayoría de las personas viaja en temporada de vacaciones, provocando los picos en la gráfica.

1.6.4. Ruido Blanco.

El ruido blanco [35] es un tipo especial de serie temporal donde los datos no siguen un patrón y cada uno de estos puntos de datos están distribuidos independiente e idénticamente, con promedio y varianzas constantes en el tiempo. Esto quiere decir que no existe autocorrelación significativa en ningún período de la serie, lo que genera que no existan patrones aparentes en las series y no sea posible hacer predicciones. En la Figura 1.12, se muestra una gráfica que ilustra como se ve el ruido blanco.

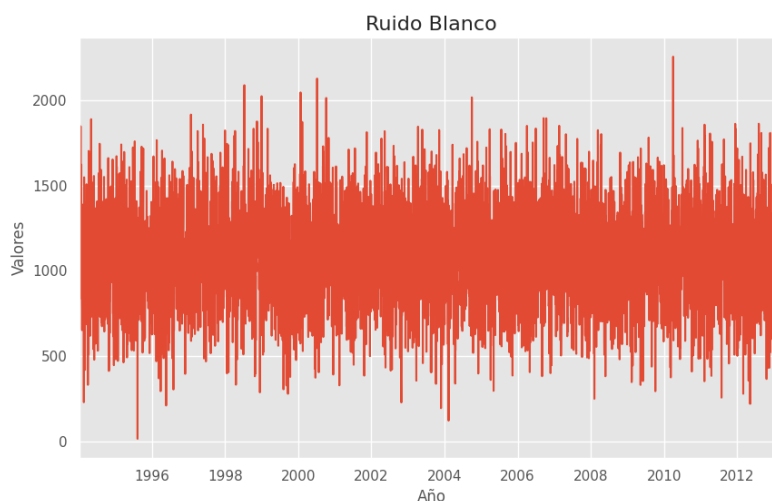


Figura 1.12: Gráfico de ruido blanco; esencialmente representa una secuencia de valores aleatorios.

Capítulo 2

Herramientas

Dentro de este capítulo se describen las herramientas principales para poder llevar a cabo el análisis de los datos.

2.0.1. Aglomeramiento k-medias

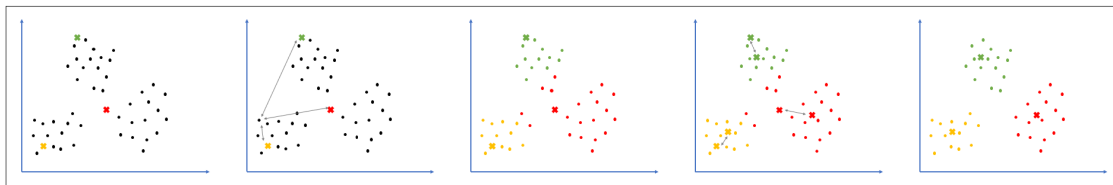


Figura 2.1: Visualización del funcionamiento del algoritmo aglomeramiento k-medias

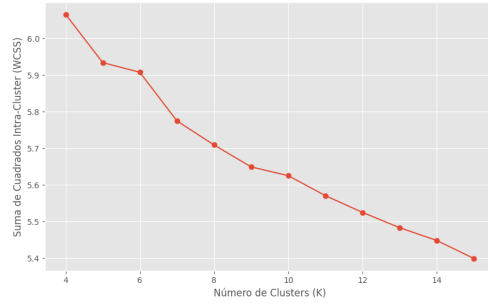
Desarrollado por Stuart Lloyd y Edward W. Forgy, este algoritmo funciona para realizar tareas de clasificación, es decir, dado un conjunto de datos, este algoritmo, puede asignar a cada elemento a un grupo donde sobresale una característica en específico [25]. Es importante señalar que este algoritmo pertenece a las técnicas de aprendizaje de máquina no supervisados, lo que quiere decir que, los datos de entrada no contienen ninguna etiqueta que los identifique por grupos, por esta razón es preciso especificar el número de grupos en los que se desea dividir determinado conjunto de datos.

La forma en la que funciona este algoritmo [23] se muestra en la Figura 2.1. Primero se asigna de forma aleatoria los centroides de cada aglomeramiento, procurando que estos centroides estén separados considerablemente, el número de aglomeramientos a considerar se especifica antes de forma arbitraria. En seguida se mide la distancia del primer punto a cada uno de los centroides, dicho punto será asignado al centroide del aglomeramiento más cercano. Lo anterior se repite para cada uno de los puntos de los datos por lo que cada punto será asignado a un aglomeramiento. Se calcula el centro de cada aglomeramiento, y se mueve el centroide de acuerdo al nuevo centro, una vez más se vuelven a asignar los puntos de acuerdo al nuevo centroide y esto se repite hasta que la distribución de los centroides ya no se mueva.

Una gran desventaja de este algoritmo reside en el hecho de que la definición del número de grupos es arbitraria. Por ello, al implementar este algoritmo en un conjunto de datos resulta útil seleccionar el número óptimo de aglomeramientos a través de técnicas como el método del codo, el cual consiste en definir el número de grupos cuando la figura de un codo se dibuje como

en la Figura 2.2a, igualmente se puede tener una confirmación de este número con el método de la silueta presentado en la Figura 2.2b. También es necesario tener en cuenta que la selección del número de aglomeramientos depende mucho de los datos presentes y el objetivo que se busque [25].

Como se ve en la gráfica de la Figura 2.2a, el eje y representa la suma del cuadrado de



(a) Método del codo.



(b) Método de la silueta.

Figura 2.2: Estos métodos ayudan a definir cuál es el número óptimo de aglomeramientos para determinado conjunto de datos.

las distancias dentro de los aglomeramientos, por lo qué si estos valores reducen drásticamente para después estabilizarse, eso significaría que ya se ha llegado al número óptimo de aglomeramientos. En este caso, 5 grupos o 7 grupos podrían ser una buena elección según la gráfica, pero para poder definir solo un número, este primer resultado se complementaría con lo que muestra la Figura 2.2b, en ella el eje y representa el coeficiente de silueta, el cual oscila entre -1 y +1, donde un valor cercano a +1 representaría aglomeramientos bien segmentados, si este coeficiente es cercano a 0; esto significaría que los aglomeramientos están cercanos a sus fronteras y un valor de -1 representaría una mala asignación de los grupos. Por lo qué en este caso, se confirmaría 5 como el número óptimo de aglomeramientos, a pesar de que estos tengan sus fronteras cercanas.

2.0.2. Descomposición en Valores Singulares

La descomposición en valores singulares¹ de una matriz es una técnica que se emplea principalmente para obtener aproximaciones óptimas de bajo rango de matrices [23]. Para fines de esta tesis, estas aproximaciones servirán para identificar aquellos patrones dominantes que podrían explicar las estructuras de mayor dimensión dentro de los datos. Esencialmente, SVD, descompone una matriz $X_{n \times m}$ en el producto de tres matrices de la siguiente manera:

$$X = U \Sigma V^T \quad (2.1)$$

donde U pertenece al espacio de matrices de dimensión $n \times n$, V al espacio de matrices de dimensión $m \times m$ y la matriz Σ pertenece al espacio de matrices de dimensión $n \times m$. Cabe resaltar que tanto U como V son matrices unitarias con columnas ortonormales, mientras que Σ es una matriz con entradas no negativas a lo largo de su diagonal y ceros en el resto de sus valores.

Además, a las columnas de U se les conoce como valores singulares izquierdos de X y a las columnas de V se les conoce como valores singulares derechos de X . Por otro lado, los elementos de la diagonal de Σ , son llamados valores singulares y representan la importancia de cada uno de los vectores singulares al momento de capturar la variación dentro de los datos, es por ello que están ordenados de mayor a menor. El rango de X es igual al número de valores singulares diferentes de cero.

¹SVD por sus siglas en inglés, singular value decomposition.

Como se ha mencionado antes esta técnica provee una manera de aproximar una matriz determinada X manteniendo sus características principales. Y debido a que Σ es diagonal, es posible expresar dicha matriz como:

$$X = U\Sigma V^T = \sum_{k=1}^m \sigma_k u_k v_k^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_m u_m v_m^T \quad (2.2)$$

Dado que, para muchos sistemas, los valores singulares σ_k reducen su valor de forma rápida, es posible aproximar X de forma efectiva al momento de truncarla en determinado rango r , tal como se muestra en la siguiente ecuación:

$$X \approx \tilde{X} = \sum_{k=1}^r \sigma_k u_k v_k^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_m u_m v_m^T \quad (2.3)$$

donde $\tilde{X}$ representa la matriz truncada y sus matrices descompuestas de la siguiente manera:

$$\tilde{X} = \tilde{U} \tilde{\Sigma} \tilde{V}^T \quad (2.4)$$

Por lo tanto, para datos de una dimensión grande es posible describirlos a través de ciertos patrones dominantes dados por las columnas $\tilde{U}$ y $\tilde{V}^T$. Esto tiene el beneficio de reducir el tamaño de la dimensión de un conjunto grande de datos. Igualmente, si las columnas de nuestros datos consisten en medidas espaciales en el tiempo, U representaría patrones espaciales mientras que V contendría los patrones temporales.

Una de las aplicaciones más importantes y directas derivada de una descomposición de valores singulares se describe a continuación.

2.0.3. Análisis de Componentes Principales

El análisis de componentes principales² es una interpretación estadística sobre las coordenadas jerárquicas basadas en datos que provee una representación de los datos correlacionados de alta dimensión.

La unión entre una descomposición en valores singulares y un análisis de componentes principales reside en las columnas de los eigenvectores de la matriz V presentada en la ecuación 2.1, que son los componentes principales. En estos componentes principales, cada una de las características son linealmente no correlacionadas entre sí.

Para llevar a cabo un PCA vía SVD [23], es recomendable centralizar los datos, esto quiere decir que es necesario sustraer el promedio a cada uno de los valores que componen los datos (ver ecuación 2.5). En seguida, se ajusta la varianza a una varianza unitaria; lo anterior asegura que los datos tengan una media igual cero y una desviación estándar igual a uno, por lo que al momento de calcular la matriz de covarianza, esta representaría correlaciones y no covarianzas [23, 25, 26]. Además, el resultado representaría un sistema de coordenadas determinado por los componentes principales, los cuales son ortonormales entre sí, pero tienen una correlación máxima con los valores de las medidas, es decir, representarían las direcciones de mayor varianza de los datos.

$$B = X - \bar{X} \quad (2.5)$$

²PCA por sus siglas en inglés, principal component analysis.

En la ecuación 2.5, B representa el conjunto de datos centralizado, mientras que $\bar{X}$ corresponde a la matriz promedio definida como sigue:

$$\bar{X} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \bar{x} \quad (2.6)$$

Por otro lado, la matriz de covarianza de B esta definida como:

$$C = \frac{1}{n-1} B * B \quad (2.7)$$

Los componentes principales son los eigenvectores de C, los cuales definen un cambio de coordenadas en el cual la matriz de covarianza es diagonal:

$$CV = VD \rightarrow C = VD^T \rightarrow D = V^T CV \quad (2.8)$$

Los elementos de la matriz diagonal D en la ecuación 2.8 representan las varianzas de los datos a lo largo de las direcciones de los componentes principales.

Si se desea hacer PCA vía SVD, es necesario proyectar las componentes principales que se encuentran en la matriz V sobre los vectores de los datos originales, $X_{n \times m}$. Para la coordenada x sería:

$$x = \sum_{i=1}^n (VT_{0,i} \times X_{i,j}) \quad (2.9)$$

coordenada y :

$$y = \sum_{i=1}^n (VT_{1,i} \times X_{i,j}) \quad (2.10)$$

coordenada z :

$$z = \sum_{i=1}^n (VT_{2,i} \times X_{i,j}) \quad (2.11)$$

Por lo tanto, SVD ofrece un acercamiento numéricamente robusto para calcular los componentes principales. Para fines de esta tesis, el análisis de componentes principales se llevará a cabo, con la finalidad de poder visualizar dichos patrones y correlaciones en nuestro conjunto de datos y contrastar si dicha formación de los patrones genera aglomeramientos visiblemente segmentados para los diferentes estados del mercado.

2.0.4. Regresión LASSO

Para entender que es exactamente una regresión lasso, primero es necesario revisar lo que es una regresión lineal.

Una regresión lineal [26] es un método estadístico que trata de modelar la relación entre una variable y una o más variables independientes mediante el ajuste de una ecuación lineal. Dependiendo del contexto, a la variable modelada se le conoce como variable dependiente, y a las variables independientes se les conoce como regresores, predictores o features. De forma general, este modelo lineal hace predicciones simplemente calculando la suma ponderada de las variables independientes de entrada, más una constante llamada término de sesgo como se muestra en la ecuación 2.12:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (2.12)$$

donde $\hat{y}$ es la variable dependiente, n es el número de variables independientes, x_i es el i -ésimo valor independiente y β_j es el j -ésimo parámetro del modelo (incluyendo el valor de sesgo β_0 y los pesos de cada variable independiente $\beta_1 + \beta_2 + \dots + \beta_n$). La ecuación 2.12 se puede escribir de forma vectorial de la siguiente manera:

$$\hat{y} = \boldsymbol{\beta} \cdot \mathbf{x} \quad (2.13)$$

Para hacer buenas predicciones es necesario ajustar los parámetros de forma que el modelo mejor se ajuste a los valores del conjunto de entrenamiento. Para saber que tan bien el modelo se ajusta al conjunto de entrenamiento es necesaria una medida de desempeño. Para fines de esta tesis se considerará el error cuadrático medio³ descrita por la ecuación 2.14:

$$MSE(\boldsymbol{\beta}) = \frac{1}{m} \sum_i^m (\boldsymbol{\beta}^T \mathbf{x}^{(i)} - y^{(i)})^2 \quad (2.14)$$

La ecuación 2.14 también se conoce como la función costo y específicamente promedia las distancias cuadradas entre los puntos del conjunto de entrenamiento y las predicciones.

Teniendo en cuenta lo anterior, la regresión lasso [25] (por sus siglas en inglés, Least Absolute Shrinkage and Selection Operator Regression) es una versión regularizada de la regresión lineal, es decir, agrega un término de regularización a la función de costo con la finalidad de restringir los pesos del modelo. Este término de regularización usa la norma L1 sobre cada uno de los pesos de las variables independientes, por lo que la función costo se vería de la siguiente manera:

$$J(\boldsymbol{\beta}) = MSE(\boldsymbol{\beta}) + \beta \sum_{i=1}^n |\beta_i| \quad (2.15)$$

Esta restricción provoca que los pesos menos importantes a la formación de patrones dentro del modelo los elimine igualándolos a cero. Por lo tanto, la regresión lasso funciona como un modelo para predecir una variable dependiente pero también para seleccionar aquellas variables independientes más significativas para llevar a cabo la predicción. En el contexto de esta tesis, cada una de las instancias de tiempo están etiquetadas de acuerdo al estado de mercado al que pertenecen, siendo esta la variable dependiente dentro de los datos, por otro lado, las variables independientes serán el valor de las correlaciones entre los precios de las acciones o directamente el precio de las acciones.

³Mean Squared Error en inglés (MSE).

Capítulo 3

Antecedentes

En el presente capítulo se busca dar a conocer las bases que soportan esta tesis, por ello, se introducen algunos conceptos y se detallan los métodos que previamente se han usado para analizar el problema.

3.1. Trabajos Previos

En el año 2001 Eugene et al. [12] publicaron un artículo donde mostraban como abordaban las correlaciones cruzadas en finanzas a través de teoría de matrices aleatorias. Una vez que obtuvieron la matriz de correlación C , evaluaron la estadística de sus eigenvalores contra los de una matriz de correlación construida con series de tiempo mutuamente no correlacionadas. Para ello, emplearon 3 conjuntos de datos diferentes:

1. El primer conjunto constaba de los rendimientos de 1000 acciones estadounidenses en un rango de 30 minutos sobre un período de 2 años: 1994-1995.
2. El segundo conjunto estaba construido similarmente con respecto a 881 acciones estadounidenses y en un período de 2 años: 1996-1997.
3. El tercer conjunto correspondía a los retornos diarios de 442 acciones estadounidenses sobre un periodo de 35 años desde 1962 hasta 1996.

Stanley et al. establecen que debido a que las interacciones subyacentes dentro de los mercados financieros no son conocidas, es necesario abordar el problema desde la teoría de matrices aleatoria.

Los rendimientos de las acciones las obtuvieron a través de la siguiente ecuación:

$$G_i(t) \equiv \ln S_i(t + \Delta t) - \ln S_i(t) \quad (3.1)$$

Donde S_i corresponde al precio de la acción i y Δt corresponde a la escala de tiempo. Para el tratamiento de la desviación estándar que cada acción tiene, se define un rendimiento normalizado como sigue:

$$g_i(t) \equiv \frac{G_i(t) - \langle G_i \rangle}{\sigma_i} \quad (3.2)$$

$$\sigma_i \equiv \sqrt{\langle (G_i)^2 \rangle - (\langle G_i \rangle)^2} \quad (3.3)$$

Donde σ_i es la desviación estándar. Finalmente se procede a calcular la matriz de correlación:

$$C_{ij} \equiv \langle g_i(t) g_j(t) \rangle \quad (3.4)$$

Los elementos de la matriz C_{ij} tendrán valores definidos dentro del rango $[-1,1]$, donde 1 corresponde a una correlación positiva perfecta, -1 corresponde a una correlación negativa perfecta y se entenderá que no existe correlación cuando el elemento sea 0.

Stanley et al. concluyeron en 2002 que es posible distinguir correlaciones genuinas de las correlaciones aparentes. De igual forma establecieron dos cosas, primero, que el eigenvalor más grande junto con su eigenvector correspondiente representan la influencia del mercado sobre todas las acciones. Y segundo, es posible caracterizar al mercado en sectores de negocio previamente establecidos a través de los eigenvalores desviados.

Más adelante en 2012 Munnix et al. [13] propusieron una forma de determinar estados de cambio drástico dentro del mercado a través de los patrones estructurales de las matrices de correlación. Estos estados serían característicos de ciertas situaciones del mercado, es decir, el mercado estaría oscilando entre algunos de estos estados y dependiendo el valor de las correlaciones dentro del estado se consideraría un mercado volátil o calmado. Para ello, consideraron los datos financieros del índice SP500 en un período de 19 años desde 1992 hasta 2010.

Se consideraron períodos de tiempo corto para analizar el estado actual del mercado y así obtener resultados significativos. El acercamiento a la construcción de la matriz de correlación fue a través del coeficiente de correlación de Pearson definido de la siguiente manera

$$C_{ij} = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sigma_i \sigma_j}. \quad (3.5)$$

Este coeficiente de correlación se realizó sobre los rendimientos de las acciones, a las cuales se les aplicó una normalización local, esto para tratar el problema de los cambios drásticos en la volatilidad que pudieran provocar mediciones equivocadas en las correlaciones

$$\bar{r}(t) \equiv \frac{r(t) - \langle r(t) \rangle_n}{\sqrt{\langle r^2(t) \rangle_n - \langle r(t) \rangle_n^2}}. \quad (3.6)$$

Para cuantificar las diferencias entre las estructuras de correlación para dos épocas diferentes, introdujeron el concepto de distancia entre dos estados, que define una medida de similitud

$$\zeta(t_1, t_2) \equiv \langle |C_{ij}(t_1) - C_{ij}(t_2)| \rangle_{i,j}. \quad (3.7)$$

Y finalmente para definir el número de estados del mercado emplearon el algoritmo de aglomeramiento k-medias llegando al resultado de que el mercado estudiado tenía 8 estados, en los cuales se podía ver una alta correlación cuando el mercado tenía caídas considerables como la crisis crediticia en 2008. La conclusión principal en este artículo fue que **las matrices de correlación pueden representar estados de los mercados financieros. En el caso de este artículo, se encontraron 8 estados característicos.** En la Figura 3.1 se muestran las matrices de correlación obtenidas por Munnix et al. El estado 1 representa un estado calmado mientras que el estado 8 representa un estado de crisis.

Tomando como pauta el trabajo realizado por Munnix et al. en 2012, investigaciones más exhaustivas llegaron en 2018 con Pharasi et al. [14] donde nuevamente desarrollan un análisis sobre la caracterización y dinámica de estados del mercado, en específico de los índices SP500 y Nikkei 255. Al igual que en las investigaciones anteriores, se estudiaron los patrones de correlación estructural de épocas de tiempo corto sobre un período de 32 años desde 1985 hasta 2016. En esta ocasión se utilizó el método de power mapping para atacar el problema del ruido que surge en las series de tiempo al considerar épocas de tiempo corto. Para definir el número óptimo de estados de mercado se consideraron las distancias intra-cluster concluyendo que el índice SP500 presume tener 4 estados de mercado mientras que el índice japonés Nikkei 255 tiene 5 estados.

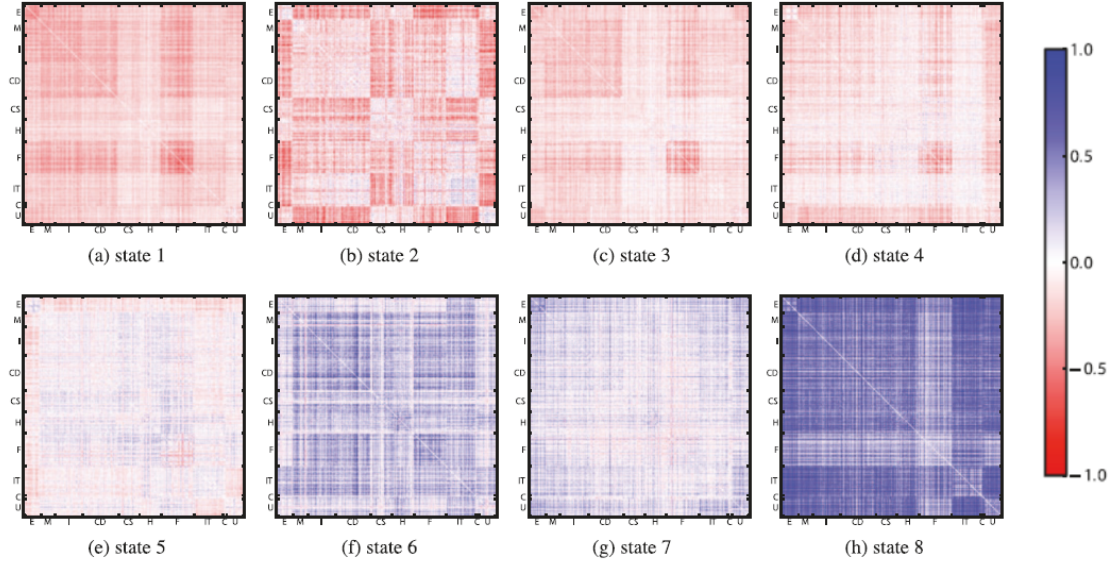


Figura 3.1: Matrices de correlación obtenidas por Munnix et al. Donde se muestran los 8 estados del mercado.

De igual forma en este trabajo se consideraron los rendimientos diarios de las series de tiempo como

$$r_k(t) = \ln P_k(t) - \ln P_k(t-1) \quad (3.8)$$

donde P_k es el **precio ajustado al cierre** de la k -ésima acción al tiempo t . Para posteriormente obtener los coeficientes de correlación cruzada

$$C_{ij}(\tau) = (\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle) / \sigma_i \sigma_j. \quad (3.9)$$

Cabe mencionar qué para el mercado estadounidense, el tamaño de las series de los rendimientos es de $T = 8060d$ con un total de $N = 194$ acciones. Mientras que, el análisis al mercado japonés se llevó a cabo con $T = 7990$ series, $N = 165$ acciones. Dichas series de rendimientos se decidieron dividir en épocas, τ , de tamaño $M=20d$ para romper la falta de estacionaridad de las series de tiempo largas, con lo que se obtuvieron $n = 403$ épocas para el mercado estadounidense y $n = 399$ épocas para el mercado japonés.

Posteriormente, trabajaron en la supresión del ruido aplicando la técnica de power mapping a cada uno de los coeficientes

$$C_{ij} = (\text{signo} C_{ij}) |C_{ij}|^{1+\epsilon} \quad (3.10)$$

donde ϵ es el parámetro de supresión de ruido.

Desarrollaron una escalada multidimensional, a dos y tres dimensiones, cuyos valores de entrada se consideraron de la distancia de la matriz, usando una transformación no lineal:

$$d_{ij} = \sqrt{2(1 - C_{ij})} \quad (3.11)$$

Una vez con el proceso anterior realizado, se llevó a cabo el k-means clustering para determinar el número óptimo de estados de mercado basado en las similitudes de cada matriz de correlación. Consiguiendo un resultado de $k=4$ estados para el mercado estadounidense y $k=5$ estados para el

mercado japonés.

Dentro de las conclusiones a las que llegaron, también establecen que el estado adyacente al estado crítico del mercado puede servir como un indicador para evitar posibles grandes pérdidas dentro de un portafolio de inversión. Además, dentro del estudio de transiciones entre estados del mercado se concluye que:

- La probabilidad de permanecer en un mismo estado es mayor que la transición hacia otro estado. Esto sugiere una inercia a permanecer en un mismo estado por períodos de tiempo largo.
- Es más probable que ocurra una transición hacia un estado adyacente.
- Las transiciones hacia estados remotos son raras, pero existen.

En 2020, el problema fue abordado nuevamente, y como se establece en el artículo, Dynamics of Market States and Risk Assessment [15], se hizo un enfoque en la evolución dinámica temporal de las matrices de correlación.

Dentro de este estudio se consideró un período de tiempo de 14 años, desde Enero de 2006 hasta Diciembre de 2019. Nuevamente, se emplearon series de tiempo del precio ajustado al cierre de los índices SP500 y el Nikkei 225, considerando únicamente las acciones presentes durante la duración del período de 14 años. La disposición de las acciones se llevó a cabo de acuerdo al sector perteneciente de la empresa que cotiza, misma disposición de sectores fue usada en los trabajos anteriores. La Tabla 3.1 muestra los sectores del mercado.

Los métodos empleados, fueron similares. Para obtener la matriz de correlación para los

Abreviación	Sector	Abreviación	Sector
CD	Consumidor discrecional	ID	Industrias
CS	Productos básicos de consumo	IT	Tecnologías de la información
CP	Bienes de equipo	MT	Materiales
CN	Bienes de consumo	PR	Farmacéuticos
EG	Energía	TC	Tecnología
FN	Finanzas	UT	Materias primas
HC	Salud		

Tabla 3.1: Abreviaciones para los diferentes sectores del SP500 y el Nikkei 225 (las siglas están en inglés).

incrementos logarítmicos $r_i(\tau)$ de cada acción se utilizó el coeficiente de correlación de pearson usado en el trabajo de 2018, ecuación 1.13. Sin embargo, para obtener la matriz de similitud entre dos matrices de correlación $C(\tau_1)$ y $C(\tau_2)$ evaluadas a diferentes tiempos τ_1 y τ_2 , se empleó la ecuación 3.7 abordada por Münnix et al. Como resultado se obtuvo que la mejor opción para el índice SP500, son **cinco estados de mercado** con un parámetro de supresión de ruido, $\epsilon = 0.9$., en tanto para el índice Nikkei 225, se obtuvo como mejor opción **7 estados de mercado** con $\epsilon = 0$.

Las conclusiones que obtuvieron se pueden resumir en los siguientes puntos:

- Cuando un índice financiero entra en crisis, se entraría rápidamente en las regiones de alta correlación, esperando que las matrices de correlación salten de un estado a un estado de crisis.
- Se sugiere empezar adoptar acciones contra la protección de riesgo, cuando el mercado se encuentra en el segundo mayor estado de correlación.

- Las transiciones más probables pueden ocurrir hacia estados cercanos o a mantenerse en el mismo estado (e excepción del punto uno).
- De acuerdo a que el mercado japonés obtuvo 7 estados de mercado con $\epsilon = 0.0$, se dice que es más complejo e incluso que el concepto de estado de mercado tiene más sentido que con el índice estadounidense.

Capítulo 4

Metodología

Como se ha mencionado, el objetivo de este trabajo de tesis toma como punto de partida el problema presentado en el capítulo anterior, recordando el hecho de que las matrices de correlación representan estados de los mercados financieros, esta vez, los datos iniciales son **los precios de apertura de las acciones cotizando en el índice SP500** y se realiza un análisis más exhaustivo en las estructuras de los patrones presentes en dichas matrices de correlación, aunado a eso, también se implementa un modelo de regresión para determinar futuros estados de mercado para determinadas correlaciones futuras entre acciones o futuros precios de acciones. El análisis completo se puede dividir en 3 partes:

1. La primera parte consiste en seguir la metodología establecida en los artículos presentados; preprocesar los datos, obtener una clasificación de estados de mercado y analizar su dinámica.
2. La segunda parte consiste en analizar los datos a través de una descomposición de valores singulares y posteriormente implementar un análisis de componentes principales.
3. En la parte tres se implementa un modelo de regresión lasso a los datos del precio de las acciones con cada instancia ya etiquetados con sus correspondiente clasificación, de igual forma este modelo se aplica sobre las correlaciones entre precio de acciones.

A continuación, se describe el procedimiento, el cual se realizó en Python. (<https://github.com/Miguellabra/Study-of-Complex-Systems>)

4.1. Procedimiento

Para poder manipular los datos de forma correcta es necesario usar los rendimientos logarítmicos, para ello se transformaron el precio de apertura de las acciones usando la siguiente ecuación:

$$r_k(t) = \log P_k(t + \Delta t) - \log P_k(t) \quad (4.1)$$

En este caso $P_k(t)$ representa el precio de un activo al tiempo t , mientras que Δt es un período de tiempo correspondiente a un día. Por lo tanto, el término $\log P_k(t + \Delta t)$ se refiere al precio de la acción k un día posterior a t , para $k = 326$ acciones.

Antes de trabajar directamente con los datos, hay un aspecto a tomar en cuenta; considerar la serie de tiempo completa supone trabajar con una serie de tiempo no estacionaria, lo que sería un problema para el análisis. Por ello, se divide la serie de tiempo original en series de tiempo más pequeñas, con esto estaríamos trabajando con series de tiempo aproximadamente estacionarias. Este último paso se ejemplifica en la Figura 4.1, donde la serie de tiempo completa

se representa con los cuadros verdes, cada uno de estos cuadros representa una ventana de tiempo de 20 días. Nuestras épocas serán de tamaño $M = 20$ días y tendrán un traslape de 1 día, esto último quiere decir que la época se formará recorriéndose un día. En la Figura 4.1 el marco rojo representa como las épocas se van recorriendo; la primera época 2 consiste del día 1 al día 20, la época 2 consiste del día 2 al día 21, la época 3 del día 3 al día 22 y así sucesivamente.

Con el procedimiento anterior se generaron 4191 épocas diferentes cada una de un tamaño de 20 días (filas) y 326 acciones (columnas). Ahora, para cada una de las épocas se calcula el coeficiente de correlación de pearson definido de la siguiente manera:

$$C_{ij}(\tau) = (\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle) / \sigma_i \sigma_j. \quad (4.2)$$

De acuerdo con los procedimientos presentados en los antecedentes, ahora se implementa la técnica de power mapping, en la cual se incorpora un coeficiente con valores entre 0 y 1 como potencia a cada uno de los valores de las correlaciones. Este coeficiente reduce el ruido dentro de las correlaciones así como rompe la degeneración en los eigenvalores, por lo que las matrices de correlación dejan de ser singulares. En la ecuación 4.3, se muestra como se implementa la técnica de power mapping.

$$C_{ij} = (\text{signo} C_{ij}) |C_{ij}|^{1+\epsilon} \quad (4.3)$$

En el caso de esta tesis, se optó por usar un parámetro $\epsilon = 0.6$ para reducir el ruido dentro de las

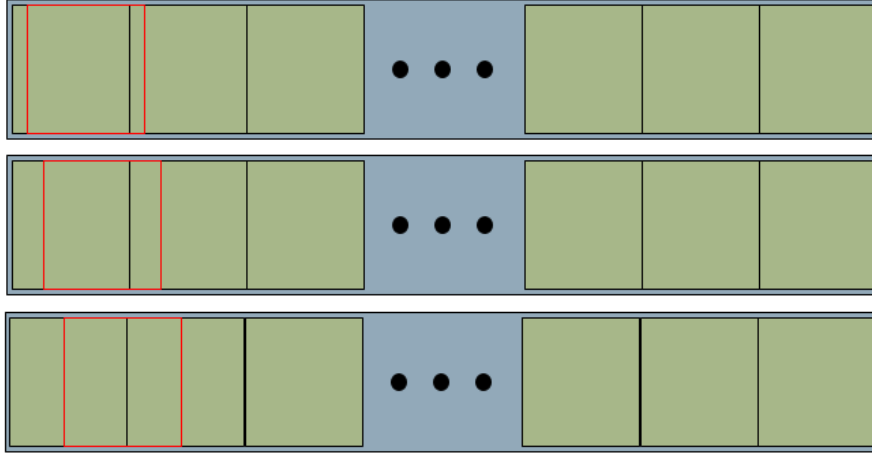


Figura 4.1: Construcción de las series de tiempo cortas, el marco rojo corresponde a una ventana de 20 días que va recorriendo un día a la serie de tiempo total.

series de tiempo. Para tener una manipulación manejable de los datos en términos de memoria, fue preciso obtener los valores de cada una de las matrices triangulares superiores; consolidándolos en una única matriz de tamaño (4191x52975), donde cada fila corresponde a una época, el número de columnas es definido por los valores únicos de correlaciones entre las 326 acciones.

Esta matriz de datos se introdujo como los valores de entrada para el algoritmo de k-medias; para definir el número óptimo de estados de mercado, se evaluó usando la técnica del codo y puntaje de silueta. Con base en los resultados, se agruparon las 4191 matrices de correlación. Para cada grupo se calculó la matriz de correlación promedio, y sobre esa matriz se obtuvo el eigenvalor más grande; con ello se pudo etiquetar a cada matriz con su respectivo aglomeramiento, donde la correlación promedio más alta representa el estado crítico del mercado y la correlación promedio más baja al estado más tranquilo. Finalmente, para la primera parte, se

calculó la matriz de transición entre estados de mercado y la matriz de probabilidades de transición.

Para la parte 2, se implementó un SVD a cada una de las matrices de correlación promedio que representan a cada estado del mercado, esto con el objetivo de determinar si es posible reconstruir los patrones dentro de las matrices de correlación a partir de cierto número de valores singulares. Para cada una de las matrices de correlación promedio por estado del mercado, se aplicó una descomposición truncada, para determinado rango; $r = 5$, $r = 10$, $r = 20$, $r = 30$, $r = 50$, $r = 100$ y $r = 150$.

Para validar los resultados obtenidos en esta parte, se graficó la suma cumulativa contra el número de los valores singulares; esta gráfica nos explica que porcentaje de la matriz original hemos aproximado dependiendo el número de valores singulares usados. Y la gráfica de los valores contra el número de valores singulares, esta última nos ayuda a visualizar el nivel de significancia de cada valor singular, se dice que el valor singular más grande nos indica la dirección de mayor variabilidad o significancia.

En seguida se implementó un análisis de componentes principales vía SVD; para ello se usaron los datos del precio de apertura de las acciones. Se estandarizaron los datos y posteriormente se aplicó descomposición en valores singulares, para poder proyectar las componentes principales sobre cada uno de los vectores de los datos originales, como se mostró en la ecuación 2.9, 2.10 y 2.11. Esto se realizó tanto para los valores singulares derechos como para los valores singulares izquierdos con la finalidad de obtener una visualización de la segmentación de los estados del mercado; al usar los valores singulares izquierdos se estarían tomando en cuenta los componentes principales, mientras que al usar los valores singulares derechos se tomarían en cuenta las correlaciones temporales.

El mismo procedimiento se llevó a cabo, para la matriz de datos correspondiente a los valores de correlaciones únicas para cada par de acciones. Esta ocasión no se estandarizaron debido a que se trataban de valores de correlación, acotados entre -1 y 1 .

En la parte final del procedimiento, se aplicó una regresión lasso al conjunto de datos etiquetados con cada uno de sus estados de mercado correspondiente. La regresión lasso se implementó de manera correcta; es decir, el conjunto de los datos se separó en conjunto de entrenamiento y en conjunto de evaluación. El tamaño del conjunto de entrenamiento consistió en el 80 % de los datos totales, mientras que el conjunto de evaluación al restante 20 %.

Se estableció nuestra variable objetivo, en este caso, las etiquetas de los estados del mercado. El precio de las acciones fueron nuestros predictores. Estos últimos se estandarizaron con el fin de re-escalar los valores a una escala más procesable para el algoritmo y que los resultados fueran más precisos. Para los parámetros de regularización del modelo, se ajustó alfa en un valor de 0.01; este parámetro es el responsable de seleccionar que predictores son de mayor importancia.

Este mismo procedimiento, se realizó sobre el conjunto de los datos correspondiente a cada una de las correlaciones únicas entre cada acción. Nuevamente, no se realizó el proceso de estandarización, pues como los valores de las correlaciones ya se encontraban acotados entre -1 y 1 , hubiera sido redundante aplicar la estandarización. Es importante señalar que, para ambos conjuntos de datos, el modelo se validó usando validación cruzada usando 4 folds, y el desempeño del modelo se evaluó mediante el error cuadrático medio. En el siguiente capítulo se presentan los resultados que se obtuvieron.

Capítulo 5

Resultados

Las matrices de correlación promedio correspondientes a cada estado del mercado las podemos visualizar en la Figura 5.1. Donde las leyendas sobre los ejes representan el nombre de las acciones de forma resumida. El estado 1, tiene una correlación promedio de **0.15** y representa un estado de calma. A medida que la correlación promedio se eleva el mercado se volverá más volátil; el estado 5 tiene una correlación promedio de **0.59** y representa un estado de cambio drástico o estado crítico. En la Tabla 5.1 se muestra la correlación promedio y el eigenvalor más grande de cada uno de los estados.

Derivado de los estados de mercado, en la Figura 5.2, se muestra la dinámica que siguen estos estados a lo largo del período de tiempo de estudio. Las líneas verticales representan las crisis financieras más importantes durante ese tiempo. En la Figura 5.3 se muestra la matriz de transición de los estados de mercado y la matriz de probabilidad de transición, en ambas el eje x corresponde al estado adyacente mientras que el eje y corresponde al estado actual.

En la Figura 5.4, 5.5, 5.6, 5.7 y 5.8 se muestra la descomposición truncada en valores singulares para cada uno de los estados. Para cada uno de ellos se muestra la aproximación a la matriz original de acuerdo al rango al que fueron truncadas para los diferentes valores de r . Además, cada figura muestra dos gráficas; la primera muestra el nivel de significancia respecto al número de valores singulares y la segunda muestra que porcentaje de la estructura original de la matriz de correlación se ha aproximado con base al número de los valores singulares considerados.

En la Figura 5.9a se muestran la proyección en 3 dimensiones de las componentes principales empleando el conjunto de datos estandarizados de los precios de apertura de cada acción. En la figura 5.9b se muestra la proyección en 3 dimensiones de las componentes principales, esta vez, usando la matriz de covarianza obtenida de los datos estandarizados.

En la Figura 5.10 se muestra el resultado obtenido de realizar el análisis de componentes principales vía SVD, usando los datos de correlaciones únicas para cada par de acciones. Este resultado es importante debido a que es posible ver la segmentación de estados de mercado. Una visualización en 2 dimensiones se puede ver en la Figura 5.11, la primera muestra los ejes xy y la segunda los ejes xz .

Al aplicar la regresión lasso para los datos del precio de las acciones, como resultado, el modelo consideró únicamente **50** acciones del total de **326** acciones. Las acciones consideradas se pueden ver en la Tabla 5.2. El error cuadrático medio a través de 4 folds fue de **3.61** y las predicciones fueron correctas en un **57.56 %** de las veces con relación al conjunto de evaluación.

Por otro lado, al aplicar la regresión lasso a los datos que comprenden cada par de correlaciones únicas entre acciones, el modelo consideró **197** pares de acciones del total de **52,975**, cada par de acciones se encuentran listadas en la Tabla 5.3. El error cuadrático medio a través de 4 folds fue de **0.21** y las predicciones fueron correctas en un **92.13 %** de las veces con relación al conjunto de evaluación.

Estado de Mercado	Correlación Promedio	Eigenvalor más Grande
Estado 1	0.15	48.48
Estado 2	0.22	74.38
Estado 3	0.31	103.89
Estado 4	0.41	134.62
Estado 5	0.59	194.43

Tabla 5.1: Correlación promedio y eigenvalor más grande de cada estado de mercado.

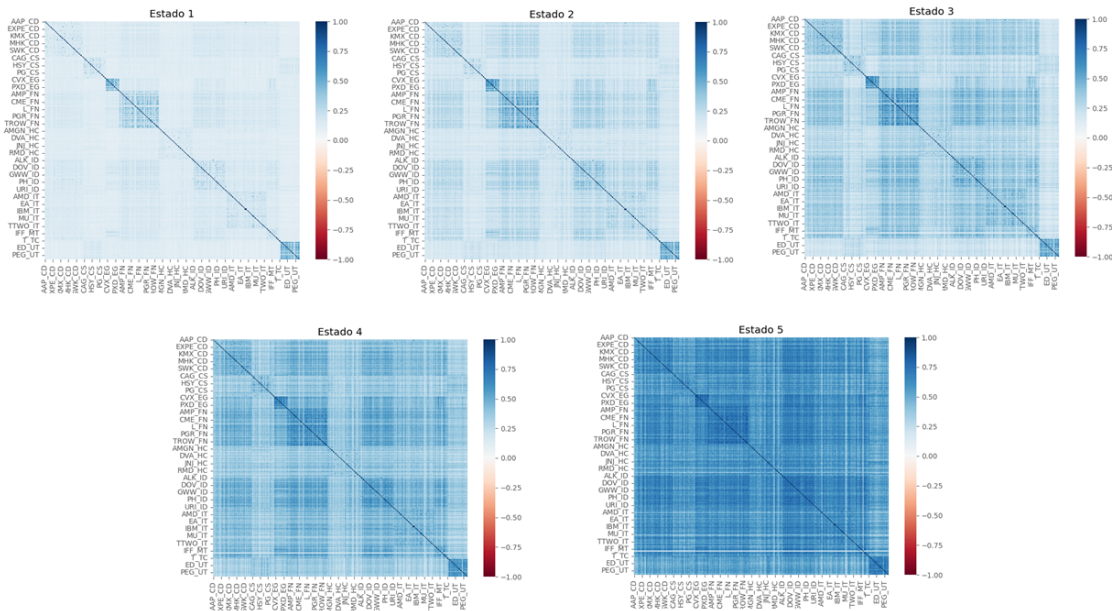


Figura 5.1: Estados de Mercado.

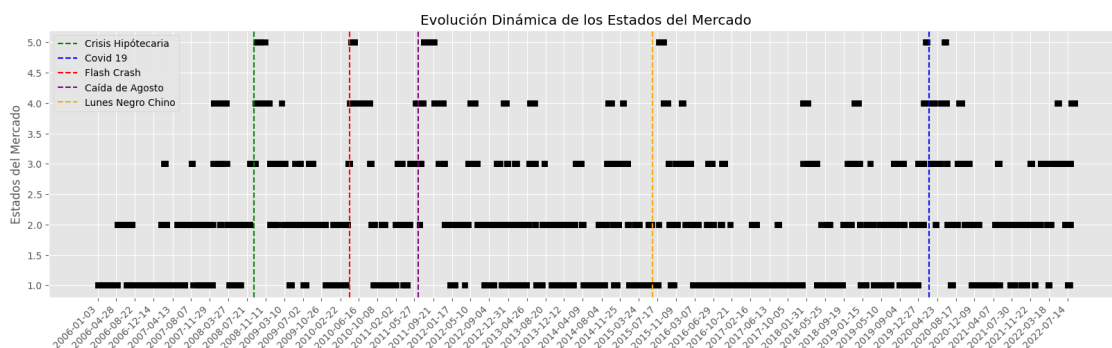


Figura 5.2: Evolución dinámica de los estados de mercado.

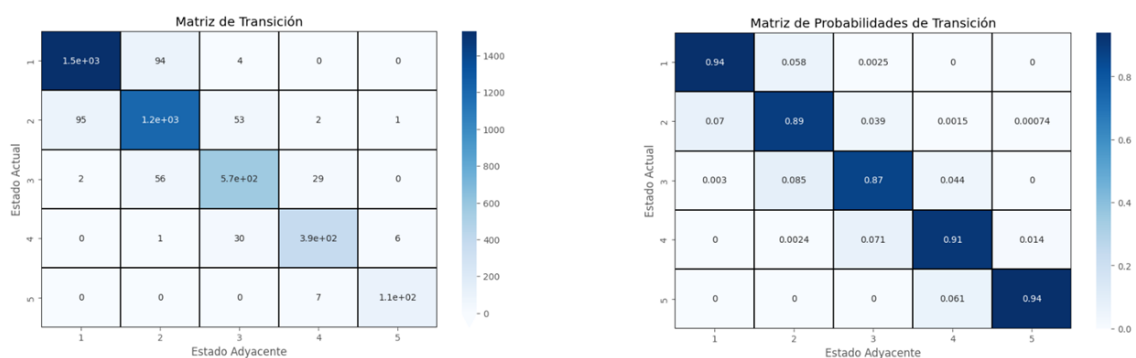


Figura 5.3: Izquierda: Matriz de transición de estados. Derecha: Matriz de probabilidad de transición.

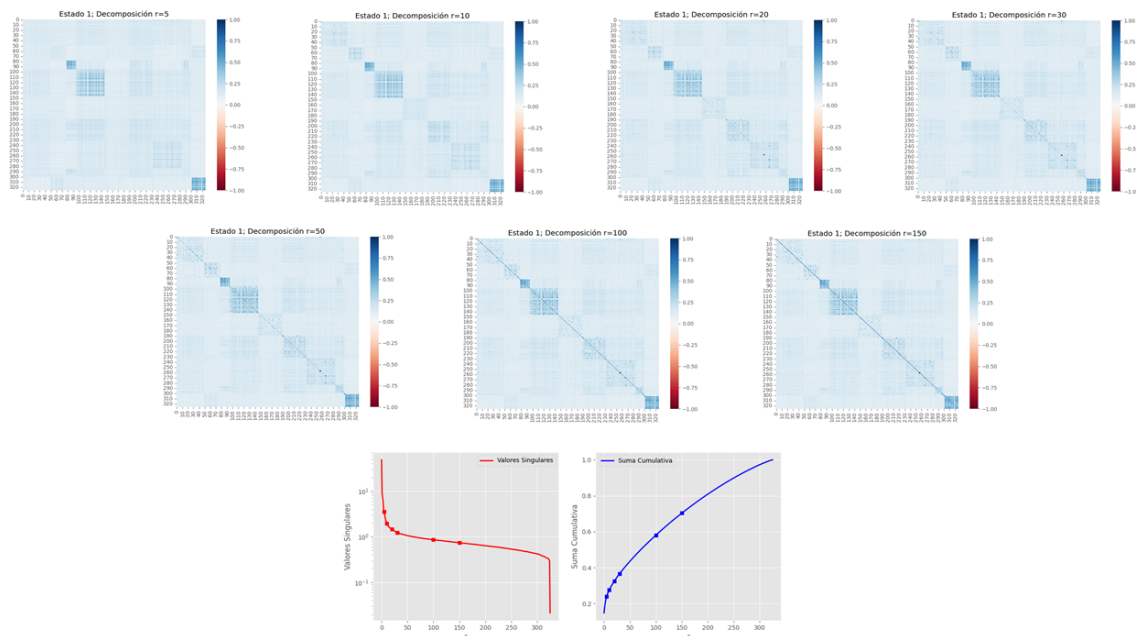


Figura 5.4: Descomposición truncada en valores singulares del estado 1.

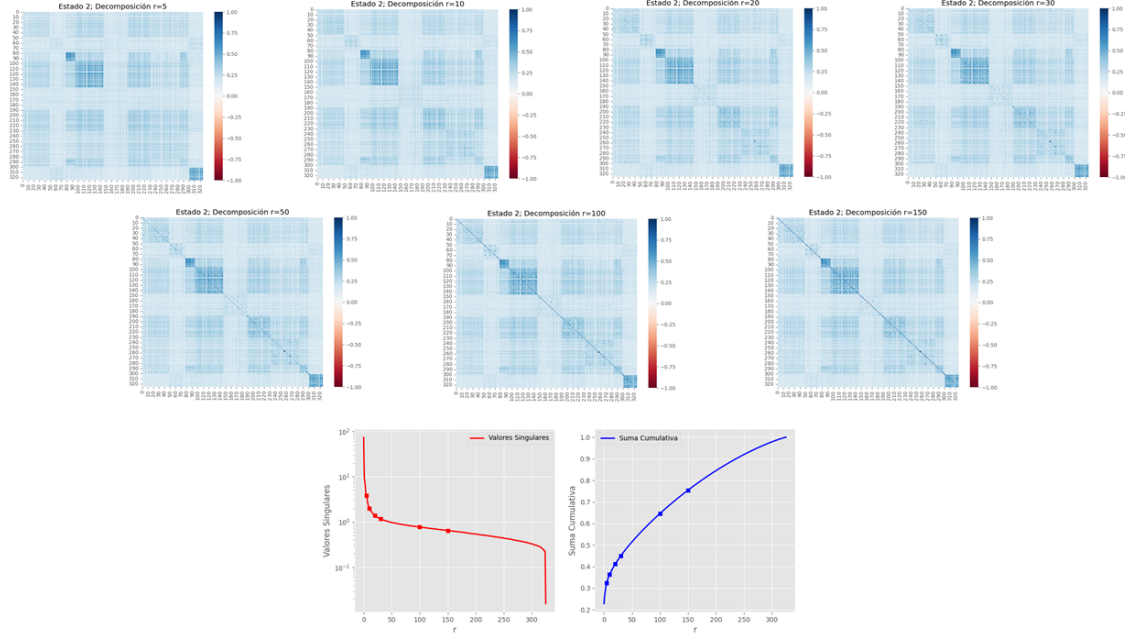


Figura 5.5: Descomposición truncada en valores singulares del estado 2.

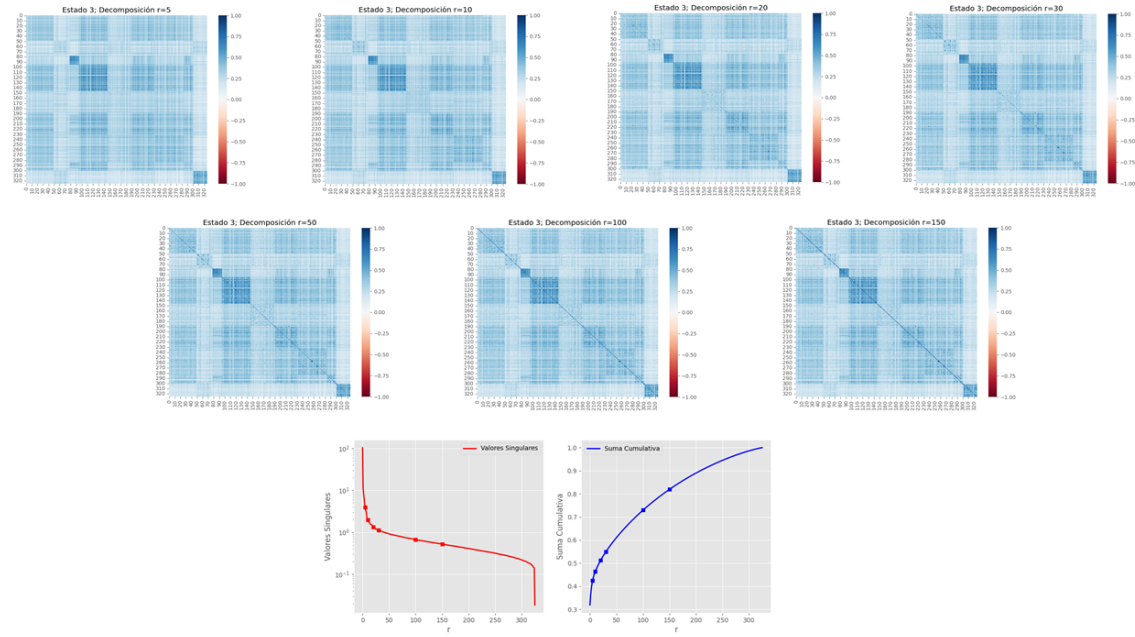


Figura 5.6: Descomposición truncada en valores singulares del estado 3.

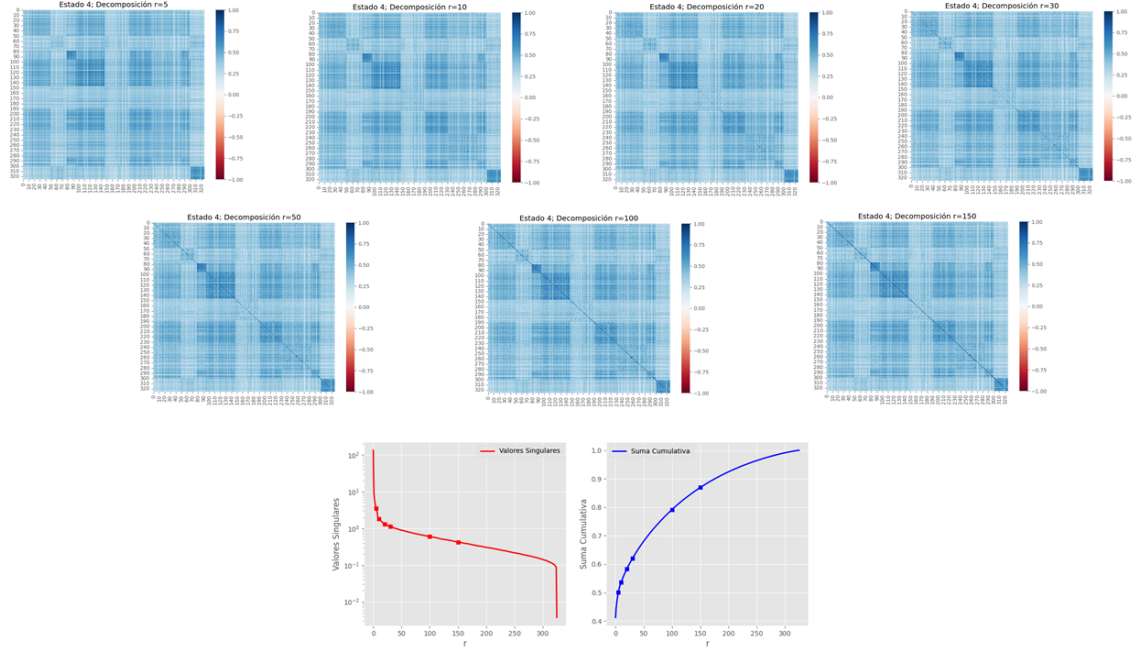


Figura 5.7: Descomposición truncada en valores singulares del estado 4.

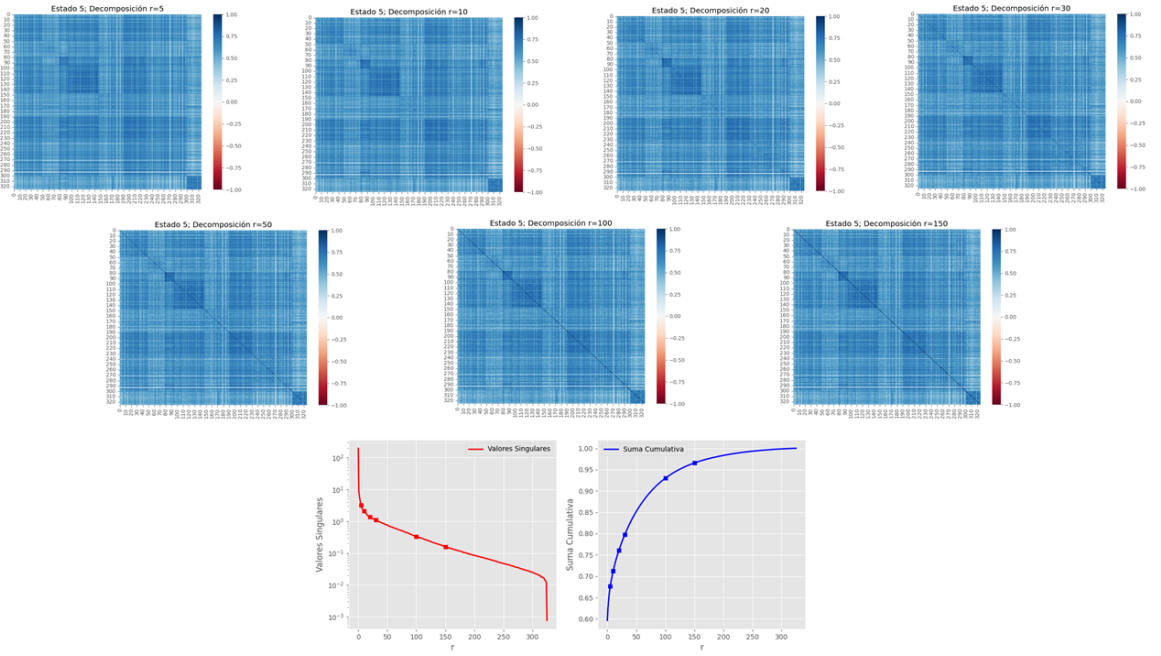


Figura 5.8: Descomposición truncada en valores singulares del estado 5.

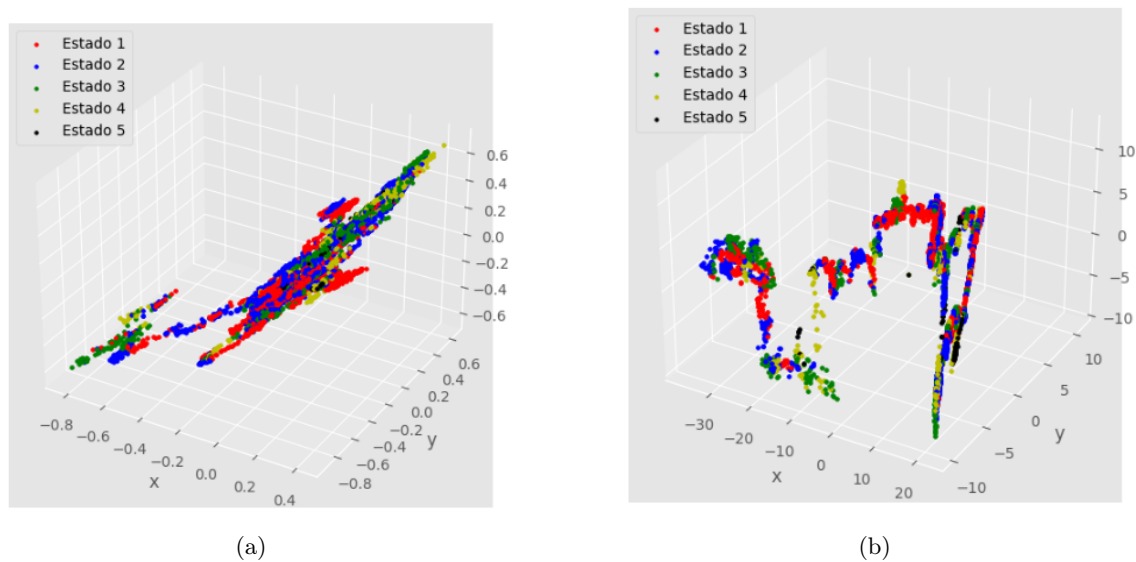


Figura 5.9: (a) Proyección de los componentes principales empleando los datos de los precios estandarizados. (b) Proyección de las componentes principales empleando la matriz de covarianza.

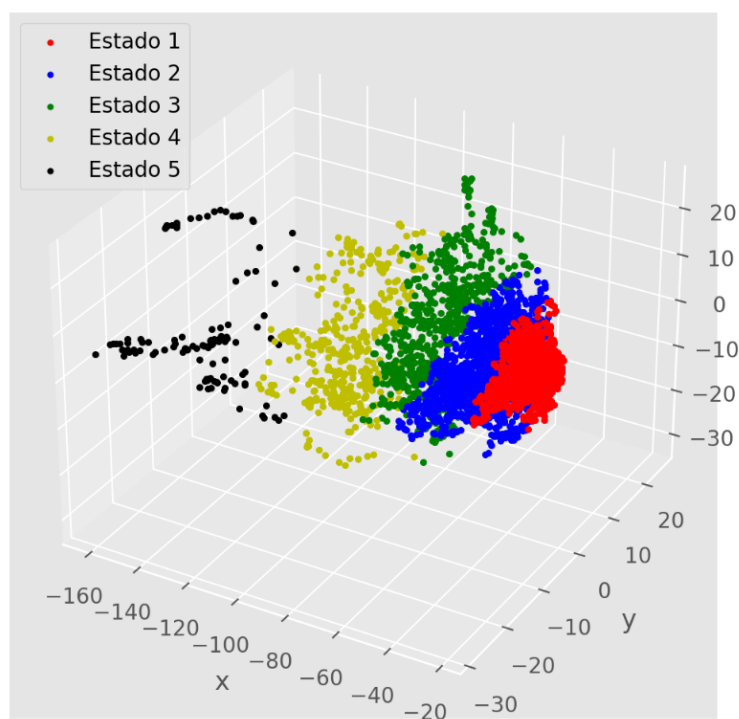


Figura 5.10: Representación 3-dimensional de la segmentación de los estados de mercado, resultado del análisis de componentes principales vía SVD. El aglomeramiento rojo corresponde al estado 1 el cuál es un estado de calma, mientras que el aglomeramiento negro corresponde al estado 5 el cuál representa cambios drásticos y crisis financieras.

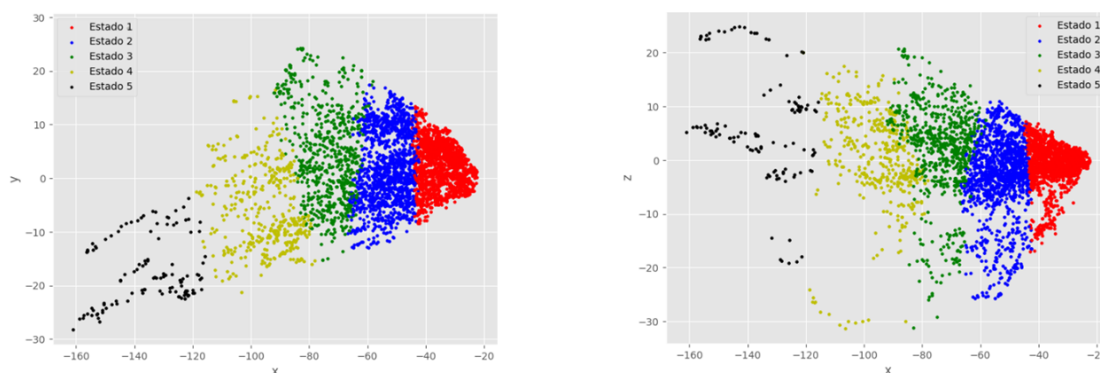


Figura 5.11: Representación 2-dimensional de la segmentación de estados de mercado; a la derecha se muestran los ejes $x - z$ y a la izquierda los ejes $x - y$. Ambas gráficas incluyendo la gráfica en la figura 5.10, representan los mismo resultado, sin embargo, cada una ofrece diferentes perspectivas.

BBY_CD	DLTR_CD	F_CD	KMX_CD	KSS_CD
MGM_CD	MHK_CD	RL_CD	TPR_CD	WYNN_CD
K_CS	MDLZ_CS	MO_CS	FTI_EG	HAL_EG
VLO_EG	WMB_EG	XOM_EG	AMG_FN	AXP_FN
GS_FN	MS_FN	MTB_FN	PRU_FN	UNM_FN
DGX_HC	INCY_HC	PRGO_HC	UHS_HC	AAL_ID
CHRW_ID	EMR_ID	FLS_ID	GE_ID	JCI_ID
DXC_IT	EA_IT	HPQ_IT	IBM_IT	NVDA_IT
WDC_IT	FCX_MT	MOS_MT	SEE_MT	T_TC
VZ_TC	AES_UT	EIX_UT	FE_UT	NRG_UT

Tabla 5.2: 50 acciones seleccionadas de acuerdo a los resultados obtenidos tras llevar a cabo la regresión lasso sobre los datos del precio de las acciones de cada empresa.

F_CD_KLAC_IT	F_CD_TXN_IT	GPC_CD_DE_ID
HD_CD_HOG_CD	HD_CD_YUM_CD	HD_CD_COST_CS
HD_CD_OKE_EG	HD_CD_ISRG_HC	HD_CD_AME_ID
HD_CD_GE_ID	HD_CD_NSC_ID	HOG_CD_PHM_CD
IPG_CD_L_FN	LEN_CD_LNC_FN	LEN_CD_WFC_FN
LOW_CD_MAR_CD	LOW_CD_PNR_ID	MAR_CD_PHM_CD
MAR_CD_TJX_CD	MAR_CD_ISRG_HC	MAR_CD_APH_IT
MAR_CD_VMC_MT	MCD_CD_ROST_CD	MCD_CD_AME_ID
MCD_CD_NUE_MT	MGM_CD_ROP_ID	OMC_CD_SWK_CD
OMC_CD_CVX_EG	OMC_CD_MS_FN	OMC_CD_SIVB_FN
OMC_CD_A_HC	OMC_CD_ETN_ID	OMC_CD_PNR_ID
OMC_CD_UNP_ID	OMC_CD_FMC_MT	PHM_CD_WFC_FN
PHM_CD_ANSS_IT	RL_CD_SBUX_CD	SBUX_CD_SWK_CD
SBUX_CD_YUM_CD	SBUX_CD_C_FN	SBUX_CD_ICE_FN
SBUX_CD_EFX_ID	SBUX_CD_FLS_ID	SBUX_CD_URI_ID
SWK_CD_BK_FN	SWK_CD_PRU_FN	SWK_CD_PAYX_IT
SWK_CD_VZ_TC	TPR_CD_PNR_ID	VFC_CD_MS_FN
WYNN_CD_AMG_FN	YUM_CD_AFL_FN	YUM_CD_NDAQ_FN
YUM_CD_WFC_FN	ADM_CS_COP_EG	ADM_CS_ROK_ID
KO_CS_UNM_FN	MDLZ_CS_PEP_CS	MO_CS_PRU_FN
PG_CD_JCI_ID	STZ_CS_CVX_EG	STZ_CS_AMP_FN
STZ_CS_PRU_FN	SYN_CS_T_TC	COP_EG_NSC_ID
CVX_EG_MS_FN	CVX_EG_TRV_FN	CVX_EG_FMC_MT
DVN_EG_BEN_FN	DVN_EG_EMR_ID	HAL_EG_NUE_MT
HES_EG_A_HC	HES_EG_FMC_MT	MRO_EG_TROW_FN
NOV_EG_PRU_FN	NOV_EG_ORCL_IT	NOV_EG_NUE_MT
VLO_EG_UNP_ID	VLO_EG_MSFT_IT	XOM_EG_AIG_FN
XOM_EG_GE_ID	XOM_EG_MSFT_IT	XOM_EG_CNP_UT
AFL_FN_EFX_ID	AFL_FN_PH_ID	AMG_FN_RSG_ID
AMG_FN_ANSS_IT	AMP_FN_L_FN	AMP_FN_SYK_HC
AMP_FN_EXPD_ID	AMP_FN_RSG_ID	AMP_FN_CSCO_IT
AMP_FN_MSFT_IT	AMP_FN_PAYX_IT	AXP_FN_WM_ID
AXP_FN_ADP_IT	AXP_FN_INTU_IT	BAC_FN_ACN_IT
BAC_FN_IBM_IT	BAC_FN_MSFT_IT	BEN_FN_KEY_FN
BEN_FN_ZION_FN	BEN_FN_TXT_ID	BEN_FN_CSCO_IT
BK_FN_EFX_ID	C_FN_MCO_FN	C_FN_EFX_ID
C_FN_JCI_ID	C_FN_INTU_IT	C_FN_CNP_UT
CINF_FN_IBM_IT	CINF_FN_CNP_UT	CMA_FN_RSG_ID
HRB_FN_PPG_MT	IVZ_FN_INTU_IT	KEY_FN_TROW_FN
L_FN_RMD_HC	LNC_FN_SYK_HC	MCO_FN_IP_MT
MET_FN_AVY_MT	MMC_FN_CDNS_IT	MS_FN_NSC_ID
MS_FN_APD_MT	NDAQ_FN_ANSS_IT	PGR_FN_PRU_FN
PGR_FN_WM_ID	PNC_FN_WM_ID	PRU_FN_JCI_ID
PRU_FN_IP_MT	RJF_FN_ROK_ID	TROW_FN_ADI_IT
TRV_FN_UNP_ID	UNM_FN_RSG_ID	USB_FN_VZ_TC
WFC_FN_APH_IT	WFC_FN_MSFT_IT	A_HC_BAX_HC
A_HC_GD_ID	A_HC_ANSS_IT	A_HC_CRM_IT
A_HC_MSFT_IT	A_HC_FMC_MT	A_HC_VZ_TC
ABT_HC_UPS_ID	ABT_HC_AVY_MT	ISRG_HC_ROP_ID
ISRG_HC_CRM_IT	JNJ_HC_VZ_TC	MTD_HC_GLW_IT
MTD_HC_IT_IT	MTD_HC_PPG_MT	SYK_HC_PPG_MT
TMO_HC_EFX_ID	TMO_HC_ETN_ID	TMO_HC_PNR_ID
WAT_HC_SNPS_IT	AME_ID_CSCO_IT	CMI_ID_APH_IT
CTAS_ID_GD_ID	CTAS_ID_IBM_IT	EFX_ID_CTSH_IT
EFX_ID_VZ_TC	EMR_ID_ROK_ID	EMR_ID_UNP_ID
EMR_ID_UPS_ID	EMR_ID_IP_MT	ETN_ID_IFT_MT
ETN_ID_VMC_MT	FDX_ID_ADP_IT	GD_ID_RSG_ID
GE_ID_UNP_ID	GE_ID_AMAT_IT	GWV_ID_IBM_IT
GWV_ID_PAYX_IT	MAS_ID_UNP_ID	NSC_ID_TXT_ID
NSC_ID_CSCO_IT	NSC_ID_NUE_MT	PH_ID_AMAT_IT
PNR_ID_FISV_IT	PNR_ID_IFT_MT	ROP_ID_FMC_MT
RSG_ID_PKG_MT	TXT_ID_FMC_MT	UNP_ID_FMC_MT
UNP_ID_IFT_MT	AMAT_IT_CSCO_IT	AMAT_IT_PPG_MT
ANSS_IT_CSCO_IT	APH_IT_IFT_MT	CSCO_IT_NUE_MT
CTSH_IT_MSFT_IT	GLW_IT_IBM_IT	INTU_IT_JNPR_IT
INTU_IT_KLAC_IT	MSFT_IT_QCOM_IT	

Tabla 5.3: 196 pares de acciones seleccionadas tras llevar a cabo la regresión lasso sobre los datos de las correlaciones únicas entre cada par de acciones.

Capítulo 6

Conclusiones

Se clasificó el índice financiero SP500 en 5 estados de mercado a través de matrices de correlación usando el precio de apertura de las acciones. En las matrices de correlación, Figura 5.1, que representan a cada estado del mercado, se pueden observar ciertas estructuras constantes. Estas estructuras corresponden a tres sectores; energético, financiero y de materias primas. A medida que cambia el estado de mercado, estas estructuras se acentúan y se acompañan de nuevas estructuras de correlación emergentes, que se pueden observar en forma de cuadrados.

Lo anterior demuestra que usar los precios de apertura de las acciones es una opción para llevar a cabo este análisis. Aún queda comprobar usar otras métricas financieras derivadas de los mercados, como puede ser la diferencia entre el precio de cierre y apertura, la diferencia entre la apertura y el cierre del día anterior, el precio máximo alcanzado, el precio mínimo, el volumen, entre otras.

A través del tiempo, como se puede ver en la Figura 5.2, el cambio de estados parece realizarse hacia los estados adyacentes, sin embargo, la línea amarilla que indica la crisis del lunes negro chino; el mercado parece realizar un comportamiento inusual, es decir, realiza un salto del estado 2 al estado 5. Este salto lo podemos confirmar en la matriz de transición en la Figura 5.3, que muestra dos saltos del estado 2 al 4 y un salto del estado 2 al 5. De igual forma, derivado de la misma figura, se puede decir que es más probable que un estado conserve su estado actual a que cambie a un estado adyacente.

Una posible explicación al comportamiento anterior reside en el hecho de que se están usando los precios de apertura del mercado, esto quiere decir que, terminada una jornada de mercado, los participantes ya no reciben una retroalimentación directa del mercado, en su lugar la retroalimentación proviene principalmente de las noticias, reportes, anuncios que pudieran hacer las empresas e incluso eventos geopolíticos que se puedan desarrollar fuera de los horarios del mercado. Esto por lo general genera una diferencia en el precio de apertura la jornada siguiente. Para el caso específico del lunes negro chino, dicho salto en la Figura 5.2 se da de 2015-07-24 a 2015-07-27, siendo el 27 día lunes, por lo que hay dos días entre el salto lo que pudo originar un salto muy repentino cuando la crisis en el mercado chino ya estaba consolidada el día viernes 24 cuando el mercado ya estaba cerrado.

La descomposición en valores singulares de cada estado permite darnos cuenta que la formación de patrones en las estructuras de correlación se deben principalmente a unas cuantas acciones, por ejemplo, en la Figura 5.4, se puede ver que casi el 80 % de la estructura original se puede aproximar usando solo 150 valores singulares, incluso desde un rango $r=20$ ya es posible ver los patrones estructurales de la matriz de correlación original, esto puede ayudar a optimizar

el uso de recursos computacionales para llevar a cabo el análisis.

Es importante destacar el resultado obtenido por el análisis de componentes principales ejecutado sobre los datos de las correlaciones únicas entre cada par de acciones. Este resultado nos permite confirmar que las correlaciones entre el precio de apertura de las acciones permiten caracterizar al mercado a través de estados. Se capturó con éxito las fuentes más significativas de variación; el espacio de dimensiones reducidas que se muestra en la Figura 5.10 conserva la información esencial necesaria para distinguir entre diferentes segmentos.

En la Figura 5.11 la gráfica del lado izquierdo muestra la proyección de los datos de las correlaciones sobre la primera y segunda componente principal, por lo que se puede observar una representación 2-dimensional de la Figura 5.10, sobre los ejes x e y . De igual forma, la gráfica a la derecha en la Figura 5.11 es una representación 2-dimensional de la Figura 5.10, esta vez proyectando los datos de las correlaciones sobre la primera y la tercera componente principal, es decir, sobre los ejes x y z . De esta forma la Figura 5.11, nos muestra otra perspectiva de cómo se visualiza la segmentación de los estados de mercado.

Desafortunadamente, la segmentación de estados de mercado no se pudo visualizar usando los datos estandarizados o sobre la matriz de covarianza. Una futura investigación puede implementar un procedimiento que permita generar esta segmentación de estados empleado la matriz de covarianza, ya que esta, en principio, también representa correlaciones.

El modelo de regresión lasso obtuvo mejores resultados al ser aplicado sobre las correlaciones únicas entre cada par de acciones que sobre el precio de las acciones. Este resultado sugiere que el comportamiento de los mercados financieros se debe a la interconexión entre los elementos que lo componen, en este caso las correlaciones entre las acciones. Este modelo puede ser una propuesta para el desarrollo de un sistema que funcione como un indicador para los agentes del mercado, que disponga de información que permita al agente tomar mejores decisiones.

Complementariamente, este modelo se puede comparar con algún otro modelo de aprendizaje de máquina como por ejemplo una regresión logística, bosques aleatorios o incluso algo más sofisticado como redes neuronales. Además de eso, sería eficiente, complementarlo cuantificando el efecto positivo o negativo de las noticias, eventos y posts sobre los mercados a través de un análisis sentimental. Esta última propuesta implicaría salirse del concepto de "agente" en el contexto de economía pues se estaría intentando interpretar la perspectiva y sentimientos de los participantes del mercado.

En cuanto a las limitaciones que se presentaron; llevar a cabo este análisis implica obtener datos de calidad que funcionen para el objetivo que se quiera lograr. En primer lugar, es necesario obtener los datos con la misma calidad y homogeneidad. Y, en segundo lugar, están los recursos computacionales que se tenga a disposición. Si se desea hacer un análisis parecido al que se llevó a cabo en este trabajo de tesis, los puntos anteriores se necesitan tomar en cuenta.

Finalmente, se puede concluir que se lograron los objetivos centrales planteados. Se clasificó el índice financiero en estados de mercado y **se identificaron las acciones de las empresas, así como los pares de correlaciones únicas entre empresas**, que más influyen en la formación de dichos estados de mercado. La **implementación de la descomposición en valores singulares sobre las correlaciones** representó un enfoque innovador en este estudio, los resultados obtenidos a través de esta metodología ofrecieron una perspectiva única, aportando una gran cantidad de información relevante, permitiendo lograr una **visualización clara de la segmentación de los estados de mercado**.

Bibliografía

- [1] MANTEGNA, R. N. & STANLEY, H. E. (2000). *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press.
- [2] DA CUNHA, C. R. (2021). *Introduction to Econophysics: Contemporary Approaches with Python Simulations*. CRC Press.
- [3] RICHMOND, P., MIMKES, J., HUTZLER, S. (2013). *Econophysics and Physical Economics*, Oxford University Press.
- [4] WEI, W. W. S. (2019). *Multivariate Time Series Analysis and Applications*, John Wiley & Sons.
- [5] JOVANOVIĆ F. & SCHINCKUS C.. (2017). *Econophysics and Financial Economics, an Emerging Dialogue*. Oxford University Press.
- [6] KWAPIEŃ, J. & DROZDZ, S.. (2012). *Physical Approach to Complex Systems*. Physics Reports. v. 515. pages 115-226. 10.1016/j.physrep.2012.01.007.
- [7] JONES, P. W. & SMITH, P.. *Stochastic Processes, An Introduction*. (2018). CRC Press.
- [8] FERNANDES, M., AREA LEAO, E. & PEREIRA H.B.B. (2017). *Econophysics: Past and Present*. Physica A: Statistical Mechanics and its Applications, Volume 473, Pages 251-261, <https://doi.org/10.1016/j.physa.2017.01.007>.
- [9] DASH, C. K.. (2018). *The Story of Econophysics*. Cambridge Scholars Publishing.
- [10] MOSSELAAR, S. J.. (2018). *A Concise Financial History of Europe*. Robeco.
- [11] MIRAMONTES, O. & VOLKE, K.. (2013). *Fronteras de la Física en el Siglo XXI*. Copit-arXives.
- [12] PLEROU, V., GOPIKRISHNAN, P., ROSENOW, B., NUNES, L.A., GUHR, T. AND STANLEY, H.E. (2001) *A Random Matrix Approach to Cross-Correlations in Financial Data*, Physical Review, 65, 066126-1-18.
- [13] MÜNNIX, M., SHIMADA, T., SCHÄFER, R. ET AL., *Identifying States of a Financial Market*, Sci Rep 2, 644 (2012). <https://doi.org/10.1038/srep00644>
- [14] PHARASI, HIRDESH & SHARMA, KIRAN & CHATTERJEE, RAKESH & CHAKRABORTI, ANIRBAN & LEYVRAZ, F. & SELIGMAN, T.. (2018). *Identifying Long-term Precursors of Financial Market Crashes using Correlation Patterns*, New Journal of Physics, 20, 10304, 10.1088/1367-2630/aae7e0.
- [15] PHARASI, H. & SELIGMAN, E. & SADHUKHAN, S. & SELIGMAN, T.. (2020). *Dynamics of Market States and Risk Assessment*.

- [16] PHARASI, H. & SADHUKHAN, S. & MAJARI, P. & CHAKRABORTI, A. & SELIGMAN, T.. (2021). *Dynamics of the Market States in the Space of Correlation Matrices with Applications to Financial Markets*.
- [17] BACHELIER L.. (1900). *Théorie de la spéculation*. PhD Thesis. Ecole Normale Supérieure de Paris, Paris.
- [18] SCHILLING, L. R. & PARTZZCH, L.. *Brownian Motion, An Introduction to Stochastic Processes*. (2012). Walter de Gruyter.
- [19] LEMONS, S. D.. (2002). *An Introduction to Stochastic Processes in Physics*. The Johns Hopkins University Press.
- [20] WALPOLE, E. R. MYERS, H. R. MYERS, L. S. & YE K.. (2012). *Probabilidad y Estadística para Ingeniería y Ciencias*. Pearson.
- [21] EINSTEIN A.. (1905). *On the Movement of Small Particles Suspended in Stationary Liquids Required by the Molecular-Kinetic Theory of Heat*. Annals of Physics, 322, 549-560.
- [22] FAMA, E. F. (1970). *Efficient Capital Markets: A Review of Theory and Empirical Work*. The Journal of Finance, 25(2), 383-417.
- [23] BRUNTON L. STEVE. & KUTZ NATHAN, J. (2021). *Data-Driven Science and Engineering, Machine Learning, Dynamical Systems, and Control*. Cambridge University Press.
- [24] STEWART, I. (2013). *In Pursuit of the Unknown: 17 Equations That Changed the World*. Basic Books.
- [25] GÉRON, A. (2019). *Hands-on Machine Learning with Scikit-learn, Keras & TensorFlow, Concepts, Tools, and Techniques to build Intelligent Systems*. O'Reilly Media.
- [26] DEISENROTH, P. M. & FAISAL A. A. & ONG S. C. (2019). *Mathematics for Machine Learning*. Cambridge University Press.
- [27] NIELSEN, A. (2020). *Practical Time Series Analysis, Prediction with Statistics and Machine Learning*. O'Reilly Media.
- [28] SHUKLA, N. (2021). *Water is Now Being Traded as a Commodity Amid Fear of Scarcity*. Earth.Org. <https://earth.org/water-trade/>.
- [29] CHIPMAN, K. (2020). *California Water Futures Begin Trading Amid Fear of Scarcity* Bloomberg. <https://www.bloomberg.com/news/articles/2020-12-06/water-futures-to-start-trading-amid-growing-fears-of-scarcity?embedded-checkout=true>.
- [30] NAKAMOTO, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*.
- [31] SCHIFTER, I. (2018). *La Ciencia del Caos*. Fondo de Cultura Económica.
- [32] SHREVE, E. S. (2004). *Stochastic Calculus for Finance II, Continuous-Time Models*. Springer.
- [33] BASHARIN, P. G. LANGVILLE, N. A. & NAUMOV, A.V. (2004). *The Life and Work of A.A.Markov*. Elsevier, Linear Algebra and Its Applications 386, p.3-26.
- [34] MANKIW, G. N. (2018). *Principles of Economics*. Cengage Learning.
- [35] BROWNLEE, J. (2020). *Introduction to Time Series Forecasting with Python*. Machine Learning Mastery.

- [36] MILLS, C. T. (2019). *Applied Time Series Analysis, A Practical Guide to Modeling and Forecasting*. Elsevier, Academic Press.
- [37] PAL, A. & PRAKASH, PKS. (2017). *Practical Time Series Analysis*. Packt.
- [38] BLACK, F. & SCHOLES, M. (1973). *The Pricing of Optiuonsand Corporate Liabilities*. Journal of Political Economy, 81(3).