



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

CORRELACIÓN Y AGRUPACIONES DE
SERIES DE TIEMPO FINANCIERAS

T E S I S
QUE PARA OPTAR POR EL GRADO DE:
FÍSICO

P R E S E N T A :
Jesús Eduardo Salgado Hernández

DIRECTOR DE TESIS:
DRA. MANAN VYAS



CIUDAD UNIVERSITARIA, CDMX, 2022



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

1. Datos del alumno
Salgado
Hernández
Jesús
Eduardo
Universidad Nacional Autónoma de México
Facultad de Ciencias
Física
417029676
2. Datos del tutor
Dra.
Manan
Vyas
3. Datos del sinodal 1
Dr.
Francois
Alain
Leyvraz
Waltz
4. Datos del sinodal 2
M. en C.
Edgar
Vázquez
Luis
5. Datos del sinodal 3
Dr.
Ruben
Yvan
Maarten
Fossion
6. Datos del sinodal 4
Dr.
Thomas
Werner
Stegmann
7. Datos del trabajo escrito
Correlación y agrupaciones de series de tiempo financieras
70 pág.
2022

Agradecimientos

A mi familia, cuya fe ciega en mi determinación por cumplir las cosas me dio la motivación necesaria para siempre salir adelante. A mi madre, que me dio los impulsos que necesitaba para dar ese primer paso, y aunque el miedo me dominara, hacer las cosas con determinación. A mi padre, que, a pesar de no siempre estar de acuerdo en mis decisiones, siempre ha sabido cómo ser mi soporte. A mi asesora Manan Vyas, por el apoyo que me ha brindado, por haber compartido su conocimiento y el haberme instruido en la rama de la física que más me apasiona. A mis amigos que fueron el soporte durante los tiempos difíciles y me ayudaron a entender, y sobre todo a ver siempre bajo distintos enfoques. A ustedes, que ha llegado el momento de despedirnos y seguir nuestros propios caminos espero volverlos a encontrar y sentarnos a charlar. A las personas que me han querido incondicionalmente y que por ellas se dieron las oportunidades que me llevaron a donde estoy. Y en especial, a esas personas que su tiempo pasó y ya no se encuentran conmigo; y sin embargo su recuerdo fue el combustible de mi día a día.

La vida no alcanza para devolver lo que me todos me han concedido, espero poder retribuirlo como se debe de un modo u otro en algún futuro. Gracias por su apoyo y confianza en mi persona.

Por otra parte, agradezco a los proyecto Conacyt Fronteras 10872 por el apoyo para realizar esta investigación.

Declaración de autenticidad

Por la presente declaro que, salvo cuando se haga referencia específica al trabajo de otras personas, el contenido de esta tesis es original y no se ha presentado total o parcialmente para su consideración para cualquier otro título o grado en esta o cualquier otra Universidad.

Esta tesis es resultado de mi propio trabajo y no incluye nada que sea el resultado de algún trabajo realizado en colaboración, salvo que se indique específicamente en el texto.

Jesús Eduardo Salgado Hernández. Ciudad Universitaria, CDMX, 2022

Resumen

Nuestro objetivo es analizar correlaciones de distinta índole en series de tiempo del mercado financiero. El coeficiente de correlación de Pearson se ha utilizado ampliamente para analizar relaciones entre las componentes de mercados financieros, este coeficiente cuenta con un alcance únicamente lineal. Sin embargo, debido a la naturaleza compleja de los mercados, sus componentes tienen relaciones de carácter más allá del lineal. Es por ello que se requieren de distintas alternativas para medir este tipo de relaciones, por tanto, planeamos analizar datos del mercado con distintos tipos de correlación.

Existe cierta incertidumbre en el comportamiento de los mercados financieros, cómo se autorregulan y cómo pueden presentar cierto comportamiento azaroso bajo distintas circunstancias que acontecen en el mundo; asimismo, hay mucho que asimilar sobre modelos que sean capaces de predecir parte del comportamiento de las series de tiempo a partir de sus propiedades.

Habiendo distintas maneras de estudiar los mercados de distintas regiones del mundo, se desarrollará un análisis en base con la teoría de matrices aleatorias y con herramientas estadísticas tomando como grupo de series de tiempo al índice más representativo del mercado estadounidense, el índice Standards & Poors 500 (S&P500), contemplando los datos históricos de enero 2000 a septiembre 2020, dando una limpieza y purificación de los datos, y otorgando a la información resultante una estructura matricial. A partir de ello, se dividirá la matriz en periodos de 40 días hábiles (épocas), obteniendo lotes de distintos periodos de tiempo, algunos periodos con crisis entre las fechas que abarcan, y utilizando los eigenvalores y eigenvectores de las matrices de correlación obtenidas por época se seleccionarán las épocas más relevantes, así como propiedades como lo son sus cuatro primeros momentos, sus razones de participación, haciendo una comparativa con lo que arroja la teoría de matrices aleatorias para analizar diferencias al usar distintas definiciones de correlación. Con el conjunto de resultados para cada caso se empleará a un algoritmo de agrupación a las 130 épocas donde cada grupo represente los distintos estados que el mercado presenta en distintos periodos de tiempo. Todo el código será desarrollado en Python.

Índice general

Agradecimientos	3
Resumen	7
1. Introducción	11
2. Fundamentos	15
2.1. Sistema Financiero	15
2.2. Índices Bursátiles: S&P 500	16
2.3. Series de Tiempo	17
2.4. Matriz de Correlación	18
2.5. Coeficiente de Correlación	19
2.6. Momentos	21
2.7. Razones de Participación	22
2.8. Teoría de Matrices Aleatorias (RMT)	23
3. Resultados Numéricos y Análisis	27
3.1. Descripción de los Datos	27
3.2. Correlación: eigenvalores y razones de participación	29
3.3. Correlación: épocas	31
4. Algoritmo de Clustering	45
5. Conclusiones	53
A. Datos del Mercado	55
A.1. Constitución de la Base de Datos	56
A.2. Tabla de Compañías	57
A.3. Periodo Temporal por Época	65
A.4. Datos de Clustering	66

Capítulo 1

Introducción

La multidisciplinariedad permite abarcar problemas desde perspectivas tan distintas que, incógnitas de un grupo de investigadores de una rama del conocimiento pueden ser resueltas por otro colectivo ajeno a la rama gracias a conocimientos propios de su campo. La física, por su parte, ha sido capaz de contribuir en la construcción de modelos de sistemas complejos, y en recientes años, físicos de todo el mundo se han enfocado en el análisis de sistemas económicos y financieros. Con investigaciones donde se conectan las teorías físicas con la practicidad de los mercados, con resultados muy beneficiosos debido al aporte multidisciplinario [1].

El mundo de las finanzas maneja variación en precios, relaciones de ganancias o pérdidas, riesgo de inversión, etc. El entendimiento de los mercados financieros con el tiempo ha incrementado la complejidad que estos tienen, requiriendo de mejores modelos capaces de describir los datos de periodos pasados o de predecir resultados futuros con un alto grado de confianza. Con el tiempo se llegó a la construcción de las series de tiempo financieras, que dotan de características a los datos históricos del mercado. Como consecuencia, se implementó el concepto de correlación de Pearson a las series de tiempo financieras, aunque esto presentaba un nuevo problema ya que la definición de Pearson se restringe a relaciones lineales. Las propias limitantes requerían de análisis más profundos.

Muchos estudios comenzaron a enfocarse en cómo la correlación de los mercados financieros cambiaba con el paso del tiempo, y para esta clase de sistemas, la teoría física tenía algo que aportar. Surgida en la física nuclear, la teoría de matrices aleatorias permitió comprender de mejor manera la interconexión entre elementos de una matriz meramente aleatoria y que, comparándola con elementos de matrices de correlación de mercados, diversos grupos interdisciplinarios destacaron que las propiedades universales de las matrices aleatorias pueden aplicarse a las matrices provenientes de sistemas reales [2].

El análisis de matrices de correlación ha sido de bastante utilidad para discernir la estructura del mercado, así como su dinámica en variedad de periodos de tiempo. Se ha encontrado que la alta volatilidad de los mercados está fuertemente relacionada con un alta correlación entre ellos, siguiendo las mismas tendencias durante crisis financieras [2]; así como posibles clasificaciones de estados que puede tener el mercado al considerar la correlación general promedio de sus ac-

tivos que la componen al dividir las series de tiempo financieras y conservando épocas con una ventana de tiempo corta [3, 4].

Grandes avances se han alcanzado gracias al empleo de matrices de correlación a los mercados financieros, aunque hay un hecho que es muy importante mencionar, que puede ser considerado como una limitante bajo ciertos casos, y está presente desde cómo se definen las matrices. Usualmente, ya sea por convención o por ser el más habitual, en análisis de series de tiempo financieras se usan matrices cuyos elementos emplean la definición de coeficiente de correlación de Pearson. Un coeficiente que mide la dependencia lineal. Esto presenta un problema en los casos de estudio donde se tratan de manera lineal a los datos cuyas relaciones existen, mas no son forzosamente lineales. Es en estos casos que, a pesar de la vasta investigación existente, aún se desconocen las distintas interpretaciones del uso de coeficientes de correlación ajenos al de Pearson. Coeficientes monótonos como el de Spearman [5] o no lineales como el de Distancia [6] muestran resultados satisfactorios al momento de capturar características del mercado donde Pearson no.

Con esto último en mente, en este trabajo se busca mostrar las principales diferencias de emplear los coeficientes de Pearson, Spearman y Distancia a un mismo grupo de activos financieros dentro de un mismo intervalo de tiempo; así como el empleo de la teoría de matrices aleatorias (RMT) para recabar qué diferencias existen al emplear distinta definición y cómo es que se adecua a las predicciones de la teoría. Y finalmente, agrupar al mercado en estados dentro del intervalo de tiempo estudiado, a partir de la correlación promedio de sus activos que componen las matrices de correlación.

Los datos empleados en este trabajo son los históricos de cierre del índice bursátil Standard & Poor's 500, mejor conocido como S&P500; considerado por muchos como el índice más representativo del mercado estadounidense, con las 500 empresas más importantes de dicho mercado, distribuidas en distintas bolsas de valores como NASDAQ y NYSE. Se tomaron las acciones de compañías constituyentes del índice dentro de un marco temporal de más de 20 años, que va de enero 2000 a septiembre 2020. Al eliminar compañías con valores nulos en alguno de los días estudiados se quedó con un total de 374 acciones pertenecientes a 11 sectores en los que se subdivide el índice. Con esto se buscó medir los rendimientos de acciones, esto es, el cambio del valor diario por empresa para poder así analizar las relaciones tanto de acciones individuales como de sectores.

Enfocándose en la construcción de matrices de correlación, se parcelaron los datos, rendimientos de precios de cierre de un periodo de más de 20 años, en épocas de 40 días hábiles, esto debido a las propiedades que cuentan las series de tiempo cortas, y se aplicaron las respectivas definiciones de coeficiente de correlación. Terminando así de construir tres grupos, uno por cada definición de correlación empleada, de 130 matrices de correlación. En cada grupo, a cada matriz por separado se calcularon sus valores de media (llamada más adelante como coeficiente de correlación promedio), desviación estándar, asimetría y kurtosis; así como su distribución de eigenvalores y sus razones de participación, estos dos últimos para comparar resultados entre las

tres definiciones de correlación y lo que arroja la teoría de matrices aleatorias con el ensamble ortogonal de Wishart (WOE). Para la metodología anteriormente descrita, se decidió seleccionar un grupo selecto de épocas para hacer la comparativa de distribuciones, matrices y resultados complementarios.

Por último, para cada grupo de matrices se empleó el método de clustering conocido como jerárquico aglomerante o *bottom-up* para seleccionar los 130 coeficientes de correlación promedio de cada grupo y agruparlos en distintos clusters o agrupaciones que representan los posibles estados que presenta el mercado que dependen de qué tan relacionados están los activos entre sí durante un periodo de tiempo. Cada agrupación contó con un mapa de calor de la matriz de correlación promedio de cada época que la constituye, además de valores de media y número de elementos; y así descubrir qué similitudes o diferencias presentan los tres casos de correlación en cuanto a las épocas dentro de las agrupaciones más importantes. Por último, se generó una matriz de transición, que contabiliza los saltos entre estados del mercado de épocas, al ordenarse cronológicamente; por lo tanto, se analizaron las transiciones más abruptas entre estados de correlación baja a estados críticos y vicerversa.

Capítulo 2

Fundamentos

2.1. Sistema Financiero

Un sistema financiero es el conjunto de entidades, ya sean financieras o gubernamentales, que permiten el intercambio de fondos. Es por medio de los mercados financieros que dichos fondos pasan de un agente a otro. Dentro de los mercados, los fondos se intercambian por distintos instrumentos financieros como lo pueden ser acciones, bonos, futuros, opciones, entre otros, sus precios están bien definidos dentro del mercado, regidos por la ley de oferta y demanda; donde normalmente hay un número finito de ese instrumento y cierta cantidad de vendedores y compradores [7, 8]. Los mercados se ocupan de los flujos de efectivo a lo largo del tiempo, así como el valor de cada instrumento, ya que los precios de los mismos no suelen permanecer estáticos a lo largo del tiempo.

El sistema financiero representa una gran importancia en la economía, pues es el encargado de generar riqueza a las dos partes de cualquier transacción financiera. Una sana economía requiere de un sistema financiero capaz de mover fondos a los agentes económicos que tienen las oportunidades de inversión más productivas. De manera más general, en un sistema financiero, un largo número de agentes interactúan entre sí, reaccionando a la información externa que reciban para poder determinar el mejor valor o precio para un instrumento dado, a este paradigma mayormente aceptado por las escuelas en finanzas de finales del siglo XX es conocido como la Hipótesis del Mercado Eficiente. Sin embargo, a raíz de lo acontecido con la crisis financiera global del 2008 [9], este panorama ya no es el más aceptado y se considera que no es una representación completa de cómo el mercado se comporta, sino como un sistema ideal. Con el pensamiento actual, los mercados siempre conllevan ineficiencias residuales, capaces de llevar a un sistema financiero, en ciertos casos, a entrar en crisis [1, 10, 11].

Crisis Financiera

Una crisis financiera se da al momento en que una alta cantidad de activos pierden repentinamente gran parte de su valor, teniendo influencia contundente sobre las decisiones de negocios y otros agentes económicos, causando un problema en el sistema financiero que, pasando de un estado de correcto desempeño a uno en el que los instrumentos se reducen drásticamente, provoca un desequilibrio entre la oferta y demanda de los instrumentos, así como en la dis-

tribución de fondos hacia quienes tienen las mejores oportunidades de inversión [12]. Pueden ser provocadas por el sistema bancario, los distintos mercados, inclusive la deuda pública o el impago de entidades tanto empresariales como gubernamentales. Las crisis suelen originarse en un solo país, no obstante, el carácter globalizado del mundo moderno produjo una liberación de los movimientos de capital, y por consiguiente, una alta correlación financiera entre los distintos mercados globales, por lo que el alcance inicial de cualquier crisis puede ser mayor al del país de origen, adquiriendo un carácter global en su punto de mayor extensión [13].

Burbuja Financiera

En el contexto histórico de las crisis financieras, entre toda la variedad de causas, muchas de ellas se han dado por un fenómeno conocido como burbuja de especulación o financiera. A causa de la especulación del mercado, el sobreoptimismo del estado futuro de un activo o un grupo de ellos genera una subida o sobrevaloración de los mismos por un periodo de tiempo sin que llegue a restituirse su valor real o intrínseco, y ya que estos valores intrínsecos son difíciles de determinar en un mercado, es complicado identificar una burbuja hasta el momento en que los precios presentan una caída sustancial en un corto periodo de tiempo, tal caída se conoce como explosión o reventar de la burbuja [14]. Después de que los precios eventualmente caigan a su valor real (o inclusive se deprecien aún más hasta que se estabilice en su valor intrínseco), los inversores recibirán grandes pérdidas, y dependiendo de la magnitud de la burbuja y de su alcance, podría terminar desencadenando una crisis. Como ejemplos en años recientes se tiene la burbuja del Dot-com en 2002 [15] y la burbuja inmobiliaria estadounidense del 2008 [16], donde esta última fue de carácter global.

2.2. Índices Bursátiles: S&P 500

Un índice bursátil es una medida del cambio en un grupo de distintos instrumentos financieros que generalmente presentan características comunes [8]. Los índices son de gran ayuda para comparar los precios actuales con los del pasado, y así poder representar el desempeño del mercado o una parte de él en una ventana de tiempo. De igual manera, una de las características más importantes de los índices bursátiles es su capacidad de diversificación; al tener una gran cantidad de activos van disminuyendo los riesgos a los que pueden estar expuestos los inversores al tomar una estrategia de inversión que siga estos índices. Cada índice rastrea distintos activos como acciones, divisas y futuros, por mencionar algunos; y cada uno de estos activos tiene una ponderación individual dentro del índice, el método de ponderación puede variar en cada índice.

De entre los índices bursátiles más conocidos, es importante resaltar el Índice Standard & Poor's 500, o simplemente S&P 500. El S&P 500 se considera una representación efectiva de la economía estadounidense (e incluso del mercado global) debido a su inclusión de 500 empresas líderes de todas las industrias de los Estados Unidos que cotizan en las bolsas de valores NYSE o NASDAQ¹, y que abarcan una captura de más de 38 billones de dólares, aproximadamente un

¹A modo de contexto, se entiende a la bolsa de valores como un mercado donde se permite tanto compra como venta de valores.

80 % de toda la capitalización del mercado de ese país [17].

2.3. Series de Tiempo

Las mediciones del mercado se registran con intervalos de tiempo regulares: pueden ser diarios, semanales, mensuales, anuales, entre otros. Al conjunto de datos registrados de manera periódica se le denomina serie de tiempo. Una serie de tiempo Y con una colección T de valores, ordenados cronológicamente, se expresa de la siguiente manera

$$Y = (Y_t : t \in T)$$

Las mediciones de esta naturaleza suelen tomarse en lapsos de tiempo igualmente espaciados. Las series de tiempo pueden contener cierto grado de aleatoriedad, así como fluctuaciones en sus valores; para esto, se cuentan con herramientas estadísticas que permiten, ya sea describir propiedades de la serie, o bien, predecir resultados [18, 19]. Existen procesos para clasificar series de tiempo a partir de sus propiedades, una de ellas es el concepto de estacionariedad: una serie de tiempo se dice que es estacionaria si sus propiedades estadísticas no cambian a lo largo del tiempo.

El término serie de tiempo puede aplicarse a los datos históricos de cualquier instrumento de inversión. Esto es muy útil ya que el hecho que cada operación efectuada se registra en tiempo real, con la serie temporal se puede monitorear la evolución del activo cada cierta cantidad de tiempo, ya sea en un corto plazo como lo puede ser minutos u horas, o en el largo plazo con días, meses o años. El índice S&P 500 no es la excepción a la regla. Al estar constituido por 500 empresas existen distintos instrumentos que replican en su portafolio de inversión la misma ponderación del índice, obteniendo una serie de tiempo que sigue al índice original, acercándose a su valor a partir de las transacciones de los agentes del mercado.

En el análisis de series de tiempo financieras se suele emplear como intervalo de tiempo el valor de cierre de cada día hábil del mercado, ya que con ellos pueden definirse a los rendimientos como la ganancia o pérdida del capital de inversión sobre un periodo de tiempo [20, 21]. Entonces el rendimiento bajo el tiempo t se representa de la siguiente manera:

$$S(t) = \frac{P(t) - P(t-1)}{P(t-1)} \quad (2.1)$$

donde P_t es el precio del activo en el tiempo t . Ahora bien, es común asumir que la serie de tiempo del retorno de un activo es débilmente estacionaria [21]. Se dice que una serie de tiempo Y es débilmente estacionaria si cumple las siguientes condiciones:

- La media es finita y constante.
- Su varianza es finita y constante.
- La autocovarianza $\text{cov}(Y_t, Y_{t+\tau}) = \gamma(\tau)$, con $\tau \in \mathbb{Z}$ y $\gamma(\tau)$ no depende del tiempo.

Gráficamente esto se ve reflejado en que los valores de una serie de tiempo oscilan alrededor de

una media constante y su variabilidad también permanece siendo constante a lo largo del tiempo.

2.4. Matriz de Correlación

Para trasladar el concepto de correlación al análisis multivariado de series de tiempo se requiere rescatar el concepto de rendimientos previamente mencionado, no obstante, es necesario trabajar en los mismos a causa de la alta variación que puedan tener, normalizando los datos para que tengan media 0 y una misma desviación estándar. Esto puede ser hecho con la siguiente fórmula:

$$A_k(t) = \frac{S_k(t) - \langle S_k \rangle}{\sigma_k} \quad (2.2)$$

con $\sigma_k \equiv \sqrt{\langle S_k^2 \rangle - \langle S_k \rangle^2}$ como desviación estándar, y $\langle \dots \rangle$ denota el promedio temporal sobre el periodo estudiado [1, 22]. Es con estas nuevas definiciones que se puede construir la matriz de correlación con coeficientes de correlación.

Escrito en notación matricial, la matriz de correlación de Pearson para N series de tiempo con longitud de tiempo T se define como:

$$C = \frac{1}{T} AA^t \quad (2.3)$$

donde A^t es la transpuesta de la matriz de datos A de las series de tiempo², con dimensión $T \times N$. Existe el caso de contar con series de tiempo cortas donde el periodo de tiempo T es mucho más pequeño que el número N de series de tiempo, en consecuencia, la matriz C contará con solamente $T - 1$ eigenvalores distintos de cero [23].

Es importante mencionar que la anterior medida de correlación entre variables no es única y tiene sus limitantes. En particular, la definición del coeficiente de Pearson tiene dos principales problemas: depende de la distribución de los datos y puede ser influenciada por valores atípicos, capaces de dar conclusiones erróneas sobre la correlación real; y por otro lado, el coeficiente de Pearson mide la relación lineal entre variables, esto ya es un problema porque existen relaciones donde se concluya que no existe correlación alguna, cuando de hecho la haya, solo que de manera no necesariamente lineal. A raíz de esta problemática, a continuación se proporcionarán definiciones del coeficiente de Pearson y de otras alternativas de coeficientes de correlación, se indagará más acerca de la construcción de matrices con un mejor enfoque a la correlación de los datos.

²Por otro lado, los elementos C_{ij} que también pueden ser denotados como $C_{ij}(t) = \langle A_i(t) A_j(t) \rangle$, se restringen al mismo dominio que el coeficiente de Pearson [3].

2.5. Coeficiente de Correlación

Correlación de Pearson

Un resultado interesante de las componentes de un índice bursátil es encontrar instrumentos que se muevan conjuntamente, como lo es en el aumento o caída de los precios, la inflación, etc.; en otras palabras, están correlacionados de distintas maneras. Para medir el grado de correlación entre dos variables se tiene el coeficiente de correlación, y el más empleado y conocido es el de Pearson [24, 25],

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (2.4)$$

La correlación de Pearson muestra para cualquiera dos valores se tiene un valor de correlación dentro del intervalo $[-1, 1]$, donde $\rho_{X,Y} = 1$ indica una correlación positiva perfecta y existe dependencia total entre los dos valores, mientras que $\rho_{X,Y} = -1$, corresponde a una relación inversa o anticorrelación perfecta, y el caso de $\rho_{X,Y} = 0$ muestra a los dos valores sin ninguna correlación.

Correlación de Rango de Spearman

Una alternativa para coeficiente de correlación no necesariamente lineal es la correlación de rango de Spearman, pero para definirlo es necesario primero hablar de las correlaciones de rango en el contexto de conjuntos de datos. De manera conceptual, las correlaciones de rango comparan dos clasificaciones dadas de objetos para probar si estas clasificaciones son independientes [5]. Con esta idea, el coeficiente de correlación de rango de Spearman se define como una estadística no paramétrica que mide la fuerza de la asociación monótona³ entre dos series de tiempo, si la relación no es lineal, la correlación de Spearman da un valor de asociación más alto que la correlación de Pearson. Esto sucede dado que este método no asume ninguna distribución de probabilidad de los datos originales [26].

El coeficiente de correlación de Spearman ρ se define como el coeficiente de correlación de Pearson entre variables clasificadas. Teniendo N series de tiempo con T rendimientos $\{S_k(t), t = 1, 2, \dots, T\}$, para cada una de ellas se van a ordenar sus elementos y se definirá una nueva variable de rango R_k , donde $R_k(t)$ representa el orden del $S_k(t)$ correspondiente, y a las nuevas variables se les aplican la definición de correlación de Pearson. Los coeficientes de Spearman de la matriz de correlación serán de la siguiente forma:

$$C_{ij} = \frac{\sum_{t=1}^N (R_i(t) - \langle R_i \rangle)(R_j(t) - \langle R_j \rangle)}{\sqrt{\sum_{t=1}^N (R_i(t) - \langle R_i \rangle)^2} \sqrt{\sum_{t=1}^N (R_j(t) - \langle R_j \rangle)^2}} \quad (2.5)$$

³En una relación monótona, las variables tienden a moverse en la misma dirección relativa, pero no necesariamente a un ritmo constante.

Si no hay datos idénticos, entonces la ecuación puede ser simplificada, siendo expresada como:

$$C_{ij} = 1 - 6 \frac{\sum_{t=1}^N D_{ij}(t)}{N(N^2 - 1)} \quad (2.6)$$

donde $D_{ij}(t) = R_i(t) - R_j(t)$. La interpretación de coeficiente de Spearman es la misma que la correlación de Pearson. Oscila entre -1 y +1, indicándonos asociaciones negativas o positivas respectivamente; 0 significa no correlación mas no independencia.

Correlación de Distancia

La otra alternativa presentada en este trabajo es la correlación de distancia. La correlación de distancia es una medida de la fuerza de asociación entre variables aleatorias no lineales, no necesariamente con las mismas dimensiones. Va más allá de la correlación de Pearson porque puede detectar más que asociaciones lineales y puede funcionar de forma multidimensional [27, 28]. La correlación de distancia varía de 0 a 1, donde 0 implica total independencia entre las variables aleatorias, si $0 < \mathcal{R}(\mathbf{X}, \mathbf{Y}) < 1$ entonces existe correlación pero no se explica si es positiva o negativa, mientras que 1 implica que los subespacios lineales de las variables aleatorias son iguales, es decir, $\mathcal{R}(\mathbf{X}, \mathbf{Y}) = 1 \Leftrightarrow \mathbf{Y} = a\mathbf{X} + b$, $a \neq 0, b \in \mathbb{R}$ [29].

Sea (X_k, Y_k) , $k = 1, 2, \dots, n$ una muestra estadística de un par de variables aleatorias con valores aleatorios (\mathbf{X}, \mathbf{Y}) . Primero, se calculan las matrices de distancias de orden $n \times n$, $a_{l,r}$ y $b_{l,r}$, que contienen todas las distancias por pares,

$$\begin{aligned} a_{l,r} &= \|X_l - X_r\|, & l, r &= 1, 2, \dots, n \\ b_{l,r} &= \|Y_l - Y_r\|, & l, r &= 1, 2, \dots, n \end{aligned}$$

con $\|\cdot\|$ denotando una norma. Se toman las distancias doblemente centradas:

$$\begin{aligned} A_{l,r} &= a_{l,r} - \bar{a}_{l\bullet} - \bar{a}_{\bullet r} + \bar{a}_{\bullet\bullet} \\ B_{l,r} &= b_{l,r} - \bar{b}_{l\bullet} - \bar{b}_{\bullet r} + \bar{b}_{\bullet\bullet} \end{aligned}$$

donde $\bar{a}_{l\bullet}$ es la media de la fila l -ésima, $\bar{a}_{\bullet r}$ es la media de la columna r -ésima, y $\bar{a}_{\bullet\bullet}$ es la media de la matriz de distancia de la muestra \mathbf{X} ; la construcción para valores de \mathbf{Y} es recíproca. La covarianza de la distancia de la muestra al cuadrado es el promedio aritmético de los productos de las distancias doblemente centradas:

$$\mathcal{V}_n^2(\mathbf{X}, \mathbf{Y}) = \frac{1}{n^2} \sum_{l,r=1}^n A_{l,r} B_{l,r} \quad (2.7)$$

Por último, la varianza de la distancia de la muestra no es más que la covarianza de distancia de dos variables idénticas, la raíz cuadrada de lo siguiente:

$$\mathcal{V}_n^2(\mathbf{X}) = \mathcal{V}_n^2(\mathbf{X}, \mathbf{X}) = \frac{1}{n^2} \sum_{l,r=1}^n A_{l,r}^2 \quad (2.8)$$

Con todo esto, la correlación de distancia se define como sigue:

$$\mathcal{R}_n(\mathbf{X}, \mathbf{Y}) = \frac{\mathcal{V}_n(\mathbf{X}, \mathbf{Y})}{\sqrt{\mathcal{V}_n(\mathbf{X})\mathcal{V}_n(\mathbf{Y})}} \quad (2.9)$$

El caso generalizado con una muestra de N variables aleatorias, que en el caso de estudio que a este trabajo concierne, series de tiempo, es análogo.

2.6. Momentos

Los momentos de una función son las medidas cuantitativas relacionadas a la forma de la distribución de una función. Si se tiene una función de distribución, los momentos de una distribución son los valores esperados de las potencias de una variable aleatoria que tiene dicha distribución [30, 31]. El n -ésimo momento de una variable aleatoria \mathbf{X} se define como:

$$\mu'_n = E[\mathbf{X}^n] = \begin{cases} \int_{-\infty}^{\infty} x^n f(x) dx & \text{distribución continua} \\ \sum_{i=1}^{\infty} x_i^n p_i & \text{distribución discreta} \end{cases}$$

A partir de aquí se hará énfasis sobre el caso discreto, ya que al manejar series de tiempo se tienen conjuntos de puntos, es decir, una distribución de tipo discreta. El primer momento, la media, habitualmente denotado por $\mu'_1 = \mu$ es

$$\mu \equiv E[\mathbf{X}] = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.10)$$

Momentos centrales

Si el momento es tomado a partir de un punto arbitrario c , se definen como

$$\mu'_n(c) = E[(\mathbf{X} - c)^n] = \sum_{i=1}^{\infty} (x_i - c)^n p_i$$

y en particular, cuando c es igual a la media μ , se tienen los llamados momentos centrales o momentos alrededor de la media, usualmente denotados por μ_n

$$\mu_n = E[(\mathbf{X} - \mu)^n] = \sum_{i=1}^{\infty} (x_i - \mu)^n p_i$$

Al segundo momento central se le conoce como varianza, comúnmente denotado por $\mu_2 = \sigma^2$

$$\sigma^2 \equiv E[(\mathbf{X} - \mu)^2] = E[\mathbf{X}^2] - E^2[\mathbf{X}] = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (2.11)$$

donde su raíz cuadrada positiva es la desviación estándar σ ,

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (2.12)$$

Momentos estandarizados

Los momentos estandarizados $\tilde{\mu}_n$ son los momentos centrales divididos por la desviación estándar a la misma n -ésima potencia:

$$\tilde{\mu}_n = \frac{\mu_n}{\sigma^n} = E \left[\left(\frac{\mathbf{X} - \mu}{\sigma} \right)^n \right]$$

El tercer momento estandarizado es el conocido como coeficiente de asimetría, sesgo, o *skewness*, mide que la curva de la distribución del lado izquierdo y derecho sean similares con respecto al valor central,

$$\tilde{\mu}_3 = E \left[\left(\frac{\mathbf{X} - \mu}{\sigma} \right)^3 \right] = \frac{1}{N} \frac{\sum_{i=1}^N (x_i - \mu)^3}{\sigma^3} \quad (2.13)$$

El cuarto momento estandarizado es el coeficiente de curtosis o kurtosis, que mide lo concentrado que se encuentran los valores tanto en la zona central de una distribución de datos como en las zonas más alejadas (colas de la distribución) [32],

$$\tilde{\mu}_4 = E \left[\left(\frac{\mathbf{X} - \mu}{\sigma} \right)^4 \right] = \frac{1}{N} \frac{\sum_{i=1}^N (x_i - \mu)^4}{\sigma^4} \quad (2.14)$$

Al basarse en la distribución normal se puede obtener otra definición de coeficiente de kurtosis llamada exceso de kurtosis, donde se subtrae el valor de kurtosis de la distribución normal (igual a 3) a la definición anterior:

$$g = \tilde{\mu}_4 - 3 = \frac{1}{N} \frac{\sum_{i=1}^N (x_i - \mu)^4}{\sigma^4} - 3$$

2.7. Razones de Participación

Los eigenvectores o vectores propios brindan información útil en cuanto a cómo se distribuyen las componentes que tienen mayor peso en una matriz de correlación. Contemplando N series de tiempo financieras, su matriz de correlación de $N \times N$ tiene N valores propios y N vectores propios, estos últimos a su vez cuentan con N componentes, donde cada componente está ligada a la contribución de una única compañía [2, 22].

Para cuantificar el número de componentes que contribuyen significativamente por vector propio, se tiene el cociente o razón de participación. La razón de participación de un vector propio k se define como

$$PR^k = \frac{1}{\sum_{l=1}^N (e_l^k)^4} \quad (2.15)$$

donde N es el número de componentes. Se tienen resultados importantes con respecto al comportamiento de las componentes de un eigenvector:

- Si todas las componentes son idénticas, $PR^k = N$
- Si solo hay una componente distinta de cero, $PR^k = 1$

Con todo esto, se puede decir que PR^k está relacionado con el número de componentes del vector propio significativamente diferentes de cero [22].

2.8. Teoría de Matrices Aleatorias (RMT)

Existen dificultades en analizar la significancia de los coeficientes de correlación cruzados obtenidos empíricamente. Entre las más importantes se encuentran las condiciones del mercado que, en constante cambio, hacen que las correlaciones cruzadas entre acciones puedan no ser estacionarias; y la longitud finita de las series de tiempo, al ser analizadas para la estimación de las correlaciones introduce ruido de medición. Si se utiliza una serie de tiempo larga para resolver el problema del ruido, las estimaciones obtenidas se verán afectadas por la no estacionariedad de las correlaciones. Es por esto que las correlaciones cruzadas medidas empíricamente contendrán siempre contribuciones aleatorias [33]. Para identificar los coeficientes que, en promedio, permanecen correlacionados a lo largo del tiempo se requiere una hipótesis nula, dicha hipótesis la encontramos en el concepto de matriz aleatoria.

Una matriz aleatoria, según la Teoría de Matrices Aleatorias (**RMT**, por sus siglas en inglés), es una matriz donde cada entrada es una variable aleatoria [34, 35]. Estudios han permitido saber que los resultados de esta teoría son capaces de coincidir con las matrices de correlación de datos empíricos con un alto grado de precisión [22]. La teoría de matrices aleatorias representa un promedio sobre todas las posibles interacciones, desviaciones a sus predicciones indican propiedades no aleatorias. A continuación se indagará sobre el ensamble ortogonal de Wishart, uno de los modelos más empleados para describir las propiedades universales de las matrices de correlación.

Ensamble Ortogonal de Wishart (WOE)

Una matriz de Wishart (en forma matricial) es de la forma

$$\mathbf{C} = \frac{1}{T} \mathbf{G} \mathbf{G}^t \quad (2.16)$$

donde \mathbf{G} es una matriz aleatoria de dimensiones $N \times T$ cuyas entradas son variables gaussianas reales independientes, es decir, $G_{ij} \in N(0, 1)$. Con estas condiciones se tiene el Ensamble Ortogonal de Wishart (**WOE**)⁴ [23, 33].

⁴Para obtener resultados similares entre los ensambles de Wishart con los datos empíricos, es necesario normalizarlos para tener media $\mu = 0$ y varianza $\sigma^2 = 1$.

Las propiedades de las matrices de Wishart son bien conocidas. Por construcción, y viéndose en la figura 2.1, la matriz \mathbf{C} no tendrá correlaciones cruzadas, sin embargo, las matrices individuales presentan fluctuaciones, aunque para T grandes este efecto es despreciable. Para $T < N$, \mathbf{C} es singular y tiene exactamente $(N - T - 1)$ eigenvalores iguales a cero. Particularmente, para grandes N y T tal que se denota $Q \equiv T/N$ finito y mayor a 1, la distribución $\rho(\lambda)$ de los eigenvalores λ de la matriz de correlación aleatoria \mathbf{C} está dada por la Ley de Marchenko-Pastur,

$$\rho(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \quad (2.17)$$

para cada eigenvalor λ dentro de los límites $[\lambda_-, \lambda_+]$, donde λ_- y λ_+ son los eigenvalores mínimos y máximos respectivamente de \mathbf{C} , dados por:

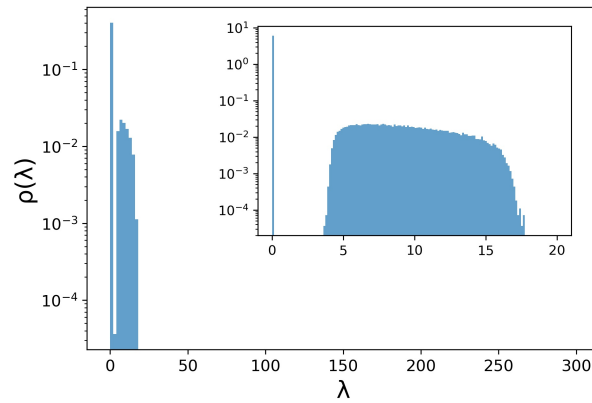
$$\lambda_{\pm} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}} \quad (2.18)$$

Resultados de la teoría RMT

Exceptuando el pico en la distribución de valores nulos de la figura 2.1a, la mayoría de eigenvalores de una matriz de correlación real cae en un abultamiento delimitado por la distribución de Marchenko-Pastur, no obstante, lo que es el eigenvalor más largo no suele caer dentro de este intervalo, implicando que gran parte de los activos presentan correlaciones entre sí [22].

Importantes resultados de sobre la teoría RMT se han dado a partir de matrices aleatorias simétricas con valores reales, cuyos elementos están distribuidos acorde al ensamble ortogonal Gaussiano (**GOE**, por sus siglas en inglés). En otras palabras, sea X_{ij} real y simétrica tal que $X_{ij} \sim N(0, 1)$ y $X_{ii} \sim N(0, 2)$; la matriz resultante es llamada ensamble ortogonal gaussiano. En lo que respecta a los eigenvectores, la hipótesis nula se da en los resultados del ensamble GOE y el promedio de las razones de participación $\langle PR \rangle$, que mide la cantidad de contribuyentes mayoritarios dentro del promedio de los eigenvectores [22, 36]. La hipótesis dice que el promedio de las razones de participación debe tender a $N/3$, donde N es el número de componentes de cada eigenvector, por tanto, se tienen dos casos (véase figura 2.1b):

- Al tener valores alrededor de $\langle PR \rangle \approx N/3$ o mayores, se tienen eigenvectores o **estados extendidos**, indican que gran parte de las compañías contribuyen de igual manera en los eigenvectores, dejando a cada vector propio con sutiles diferencias en la ponderación de los activos.
- Si, en cambio, hay desviación considerable en el promedio, siendo menor al promedio teórico ($\langle PR \rangle < N/3$), significa la existencia de eigenvectores o **estados localizados**, sugiriendo que solo un número de compañías contribuyen en los eigenvectores, siendo a su vez no correlacionadas entre sí.



(a) Distribución de Eigenvalores

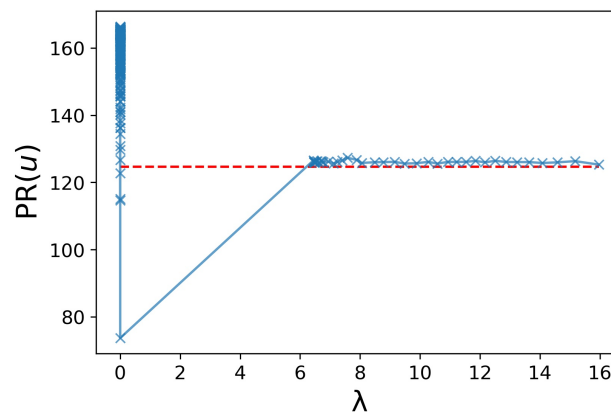
(b) Eigenvalores vs PR

Figura 2.1: Distribución ejemplo de eigenvalores y razones de participación en función a los eigenvalores de un Ensemble Ortogonal de Wishart (WOE). a) Pueden verse dos tipos de patrones, el pico de valores nulos y el abultamiento que forma la distribución de Marchenko-Pastur. b) Las razones de participación correspondientes a eigenvalores distintos de cero muestran estados extendidos, es decir, presentan contribuciones de gran parte de los agentes.

Capítulo 3

Resultados Numéricos y Análisis

En este capítulo se exponen los pasos al filtrado y construcción de la matriz de datos con los precios históricos de cierre de cada acción constituyente del índice S&P500 en el periodo de enero del 2000 a septiembre del 2020; así como su fragmentación en épocas o lotes de matrices de mismo tamaño y así obtener para cada una de ellas las matrices de correlación con los distintos tipos de coeficientes previamente estudiados, incluyendo matrices que representan la hipótesis nula de la teoría RMT. Posteriormente, se desglosarán las propiedades para cada matriz, que son los primeros cuatro momentos de media, varianza, asimetría y kurtosis; además de mostrar el método de obtención de las razones de participación para cada empresa dentro de cada periodo de tiempo.

3.1. Descripción de los Datos

A partir de una base de datos adquirida del portal Quandl [37] con los valores históricos entre 1970 y septiembre 2020 de los instrumentos constituyentes de los índices más importantes del mercado estadounidense¹, se filtraron los datos correspondientes a las acciones de empresas que conformaron el índice S&P 500 durante el mes de octubre del 2020 [38], es muy importante mencionar esto ya que el índice cambia respecto a las empresas que lo constituyen, quitando y agregando empresas².

El índice contaba con 505 acciones de 500 empresas distintas (algunas empresas contando con más de una acción) que fueron organizadas en grupos dependiendo a qué sector pertenecían [39, 40], sin embargo, por diversas circunstancias no todas las empresas registraron precios de cierre en todos los días hábiles o laborales, que es lo que concierne a esta investigación, por tanto, se eliminó de la base de datos todo lo relacionado a códigos bursátiles que tuvieran datos nulos entre las fechas de enero 2000 y septiembre 2020³. En consecuencia, se obtuvo una matriz de datos con los precios de cierre de 374 empresas entre el 3 de enero del 2000 y el 1 de septiembre

¹Para una mayor información en la constitución de la base de datos inicial, consultar la sección Apéndice B.1.

²La última actualización tomada en cuenta en la investigación de este trabajo fue el día 9 de octubre del 2020. Consultada el día 11 del mismo mes.

³La razón de que la ventana de tiempo no fuese más allá del año 2000 se debe a que gran parte de las empresas del índice no estaban constituidas antes de ese año, y con el criterio de eliminación de empresas, el número de empresas restantes hubiera descendido de manera considerable.

del 2020; en otras palabras, una matriz con dimensiones de 374×5200 , donde 374 representan las acciones y 5200 los días laborales del mercado⁴.

Sector	Símbolo	Compañías	Ponderación
Consumo Discrecional	CD	47	0.126
Consumo Básico	CS	24	0.064
Salud	HC	44	0.118
Industria	IN	48	0.128
Tecnologías de la Información	IT	52	0.139
Materiales	MA	21	0.056
Inmobiliario	RE	24	0.064
Servicios de Comunicación	TS	20	0.053
Servicios Públicos	UT	27	0.072
Financiero	FI	49	0.131
Energía	EN	18	0.048
Total		374	1.000

Tabla 3.1: Distribución de sectores constituyentes del índice S&P 500.

CD: Bienes de consumo discrecional, CS: Bienes de consumo básico, HC: Salud, IN: Industria, IT: Información y tecnología, MA: Materiales, RE: Inmobiliario, TS: Servicios de comunicación, UT: Servicios públicos, FI: Financieras, EN: Energía.

Crisis y Burbujas Financieras

El conocer el contexto del mercado es fundamental, es por ello que se hizo una investigación de las crisis financieras que se dieron en el marco 2000-2020. A continuación se muestran las crisis y burbujas financieras más importantes que afectaron al mercado estadounidense durante el periodo de investigación⁵.

Burbujas Financieras	Fecha (dd-mm-aaaa)	Época
Burbuja Dot-Com	10-03-2000	2
Burbuja Inmobiliaria Estadounidense	01-05-2006	40
Burbuja del Mercado Chino del 2007	27-05-2007	45
Burbuja de Criptomonedas	17-12-2017	114

Tabla 3.2: Épocas con el registro de máximos históricos de burbujas financieras más importantes del mercado estadounidense de enero 2000 a septiembre 2020.

⁴En la sección Apéndice B.2. se muestran las 374 compañías empleadas, con su respectivo sector y código bursátil o *Ticker*.

⁵Durante el proceso de investigación fue complicado encontrar el día exacto donde la burbuja inmobiliaria estadounidense marcó su máximo histórico antes de su desplome. Basándose en el S&P CoreLogic Case-Shiller 10-City Composite Home Price NSA Index (SPCS10) [41], cuya frecuencia es mensual, la fecha elegida para el registro de la tabla 3.2 fue mayo 1 de 2006.

Crisis	Fecha (dd-mm-aaaa)	Época
Ataques del 11 de Septiembre	11-09-2001	11
Caída del Mercado Estadounidense en 2002	09-10-2002	18
Mercado Bajista Estadounidense de 2007-2009	11-10-2007	49
Gran Recesión	16-09-2008	56
Flash Crash	06-05-2010	66
Caída del Mercado Estadounidense en 2011	08-06-2011	73
Lunes Negro Chino del 2015	24-08-2015	99
Recesión por COVID-19	09-03-2020	127

Tabla 3.3: Épocas con las crisis más relevantes para el mercado estadounidense de enero 2000 a septiembre 2020.

Como se observa en las tablas 3.2 y 3.3, se contemplaron cuatro burbujas financieras y ocho crisis. No es necesario describir si fueron muy grandes o pequeñas, salvo la más importante, la crisis conocida como la Gran Recesión, ocasionada, entre otras cosas, por la burbuja inmobiliaria que se vivió en Estados Unidos. Durante los periodos que abarcaron se esperan ver correlaciones altas entre los sectores del mercado [2]. Estas fechas serán consideradas como puntos de referencia durante el proceso de investigación a los que se van a someter los datos.

3.2. Correlación: eigenvalores y razones de participación

Dentro de la matriz de datos construida, cada fila representa la serie de tiempo de una acción y cada columna representa un día laboral. Es con esta estructura que se calculan los rendimientos para obtener una matriz de 374×5199 . Se elaboraron mapas de calor con las matrices de correlación provenientes de los coeficientes de Pearson, Spearman y Distancia.

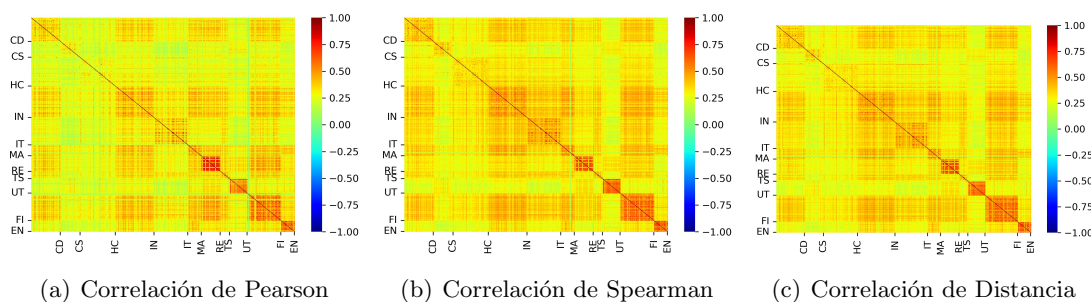


Figura 3.1: Matrices de correlación (según su definición) de 374 acciones del índice S&P500, ordenada por sector, del 3 de enero del 2000 al 1 de septiembre del 2020.

Como se puede ver en la figura 3.1, sin importar la definición, existe una fuerte correlación dentro de los sectores de energía (EN), finanzas (FI), servicios públicos (UT) e inmobiliario (RE). Más aún, los mapas correspondientes a los tres coeficientes se asemejan bastante, en especial los de Pearson y Spearman. Sin embargo, al manejar la matriz de datos completa (con $T = 5200$ días) no se alcanza a percibir ninguna clase de fenómeno que llegue a afectar al mercado a corto plazo, como lo son las crisis.

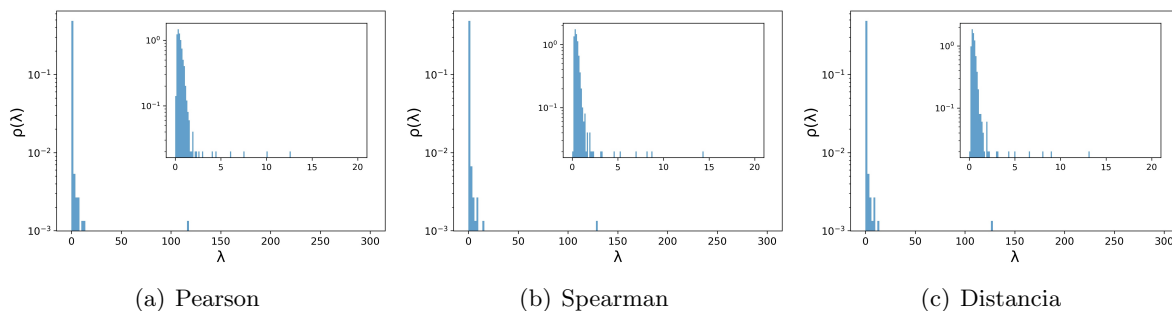


Figura 3.2: Distribución de densidad probabilística (a escala logarítmica) de eigenvalores para la matriz de correlación de series de tiempo a 5200 días.

Después, se analizaron las distribuciones de eigenvalores, cuyos resultados son presentados en la figura 3.2, donde para ningún caso se aprecia la distribución de Marchenko-Pastur. Al haber manejado series de tiempo con $T \gg N$ se cuenta con un abultamiento en la distribución de eigenvalores, y viendo con más detenimiento dicha zona se puede observar que los eigenvalores son distintos de cero; asimismo, se llegan a contar con valores atípicos bastante alejados de esta concentración, siendo el más lejano el mayor de estos el máximo eigenvalor: para Pearson es 116.93, mientras que para Spearman y Distancia son 129.03 y 126.63 respectivamente; esto pareciera tener una relación con el coeficiente de correlación promedio de cada matriz, siendo para cada matriz 0.2980, 0.3366 y 0.3000 respectivamente.

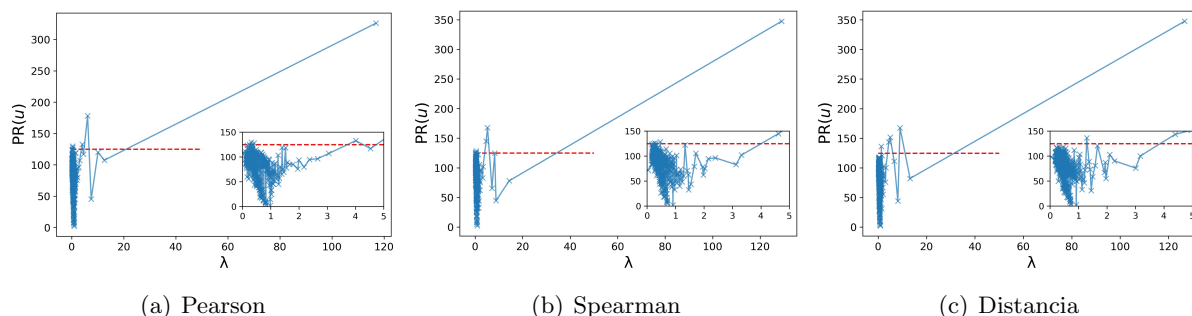


Figura 3.3: Razones de participación en función de los eigenvalores para cada tipo matriz de correlación con distinto coeficiente, siendo la línea roja punteada $N/3$ con N igual al número de acciones contempladas dentro de la matriz de correlación teórica.

En la figura 3.3 se muestran las variaciones de las razones de participación en función de los eigenvalores de cada matriz de correlación. Al contar con matrices reales, la media de las razones de participación de WOE ($\langle PR \rangle \approx N/3 \approx 124.67$) representa el umbral entre los eigenvectores para determinar la naturaleza de sus estados. La razón de participación, que representa la cantidad de compañías que participan en el eigenvector, asociada al mayor de los eigenvalores siempre se encuentra alejada del resultado teórico, y al estar alejada del cúmulo de valores se asegura que estas compañías están correlacionadas. Para Pearson, $PR_{\lambda_{max}} \approx 326.03$; para Spearman, $PR_{\lambda_{max}} \approx 347.37$; y Distancia con $PR_{\lambda_{max}} \approx 347.54$. Se cuentan con otros estados extendidos, no obstante, en promedio, las razones de participación son menores al valor teórico, sugiriendo que se tienen predominantemente estados localizados [34, 36]: para Pearson, $\langle PR \rangle \approx$

81.83; en Spearman, $\langle PR \rangle \approx 83.99$; y para Distancia, $\langle PR \rangle \approx 78.86$.

3.3. Correlación: épocas

La matriz de datos como se ha ido trabajando no es suficiente para el análisis ya que, además de no percibir periodos breves de alta correlación, las series de tiempo financieras no necesariamente son estacionarias cuando el periodo de tiempo es muy grande. Con esta justificación, a la matriz de dimensiones 374×5200 le fueron calculados rendimientos (pasando a una matriz de 374×5199); luego, esta nueva matriz fue particionada en épocas o matrices de datos de 40 días laborales⁶; quedando con 130 épocas⁷ de 374 acciones y 40 días hábiles. Aplicando a sus elementos las correlaciones de Pearson, Spearman y Distancia, se asoció a cada definición un grupo de 130 matrices de correlación de dimensiones 374×374 , un total de tres. Fueron calculados los valores de media, desviación estándar, asimetría y kurtosis a cada matriz. Al trabajar con matrices, en este caso particular, la media de la matriz puede renombrarse como coeficiente de correlación promedio (CCP).

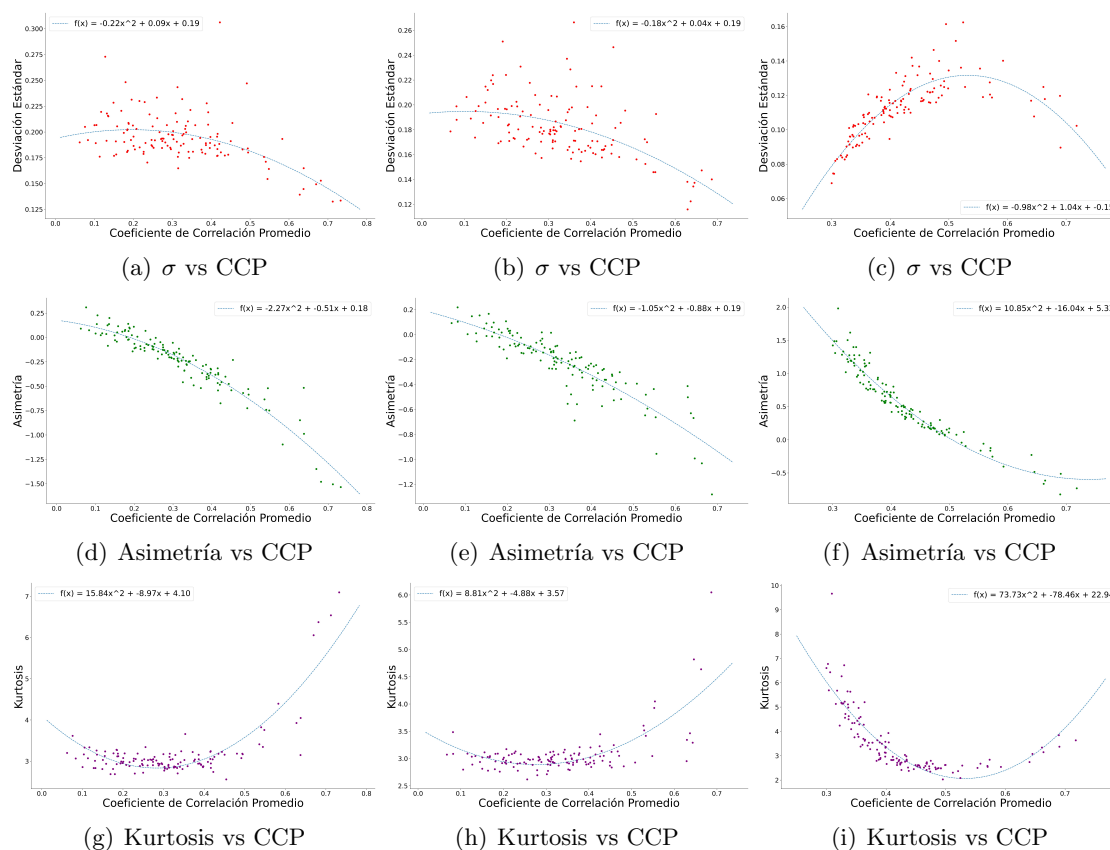


Figura 3.4: Gráficas de momentos para los coeficiente de correlación de Pearson (izquierda), Spearman (centro) y Distancia (derecha).

Para cada definición de correlación, con los cuatro momentos asociados a cada matriz se

⁶Se eligieron 40 días como ventana de tiempo por la facilidad de tomar los días hábiles dentro de 2 meses, además que en mismas investigaciones se suele emplear esta misma ventana de tiempo.

⁷Por emplear los rendimientos la última época constó de una matriz de 374×39 , no obstante, este hecho fue irrelevante al momento de analizar los datos pues esta diferencia fue imperceptible.

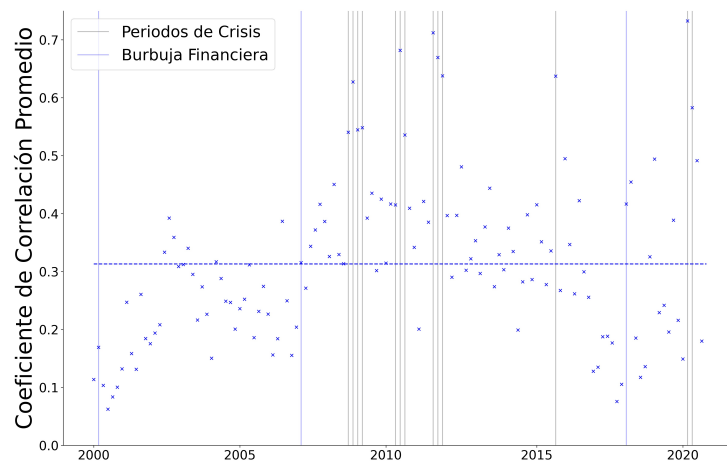
establecieron los siguientes esquemas dentro de la figura 3.4: coeficiente de correlación promedio contra desviación estándar, coeficiente de correlación promedio contra asimetría y coeficiente de correlación promedio contra kurtosis. Las curvas de ajuste son las que mejor se adaptan a la dispersión de los datos, todas ellas siendo de orden cuadrático. Esto fue corroborado con el uso del coeficiente de determinación o R^2 -score, que mide en gráficas de dispersión la calidad de la regresión obtenida, significando que en 0 el modelo no explica nada y que en 1 el modelo se ajusta perfecto a los datos [24].

	Pearson	Spearman	Distancia
σ vs CCP	0.271	0.307	0.735
Asimetría vs CCP	0.901	0.789	0.938
Kurtosis vs CCP	0.755	0.546	0.785

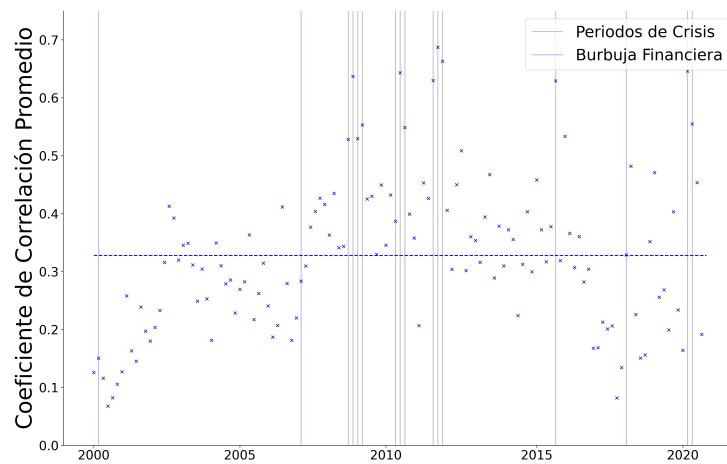
Tabla 3.4: Coeficientes de determinación (R^2 -score) asociados a la curva de ajuste para las gráficas de dispersión de los momentos.

En particular, el orden cuadrático es visto en el caso de la correlación de Distancia, manteniendo coeficientes de determinación arriba de 0.7 en cada ocasión, siendo el más alto el de asimetría con 0.938. A su vez mostrando un mejor ajuste cuando el coeficiente de correlación promedio se compara con la asimetría; lo anterior, junto con el alto coeficiente de determinación en Distancia, se puede afirmar que el coeficiente de correlación promedio y la asimetría poseen una correlación negativa. Por otro lado, existen similitudes en los patrones de ajuste de Pearson y Spearman: sus dispersiones no muestran cambios significativos; en especial, con correlaciones altas el ajuste se comporta similar sin importar la definición. Lo destacable se da en el caso de Distancia, en la desviación estándar contra correlación promedio se nota mayor variación cuando se tienen correlaciones promedio bajas, y al sobrepasar el umbral de 0.5 en correlación promedio la desviación tiende a bajar; en la figura 3.4i, en bajas correlaciones hay una alta kurtosis que baja a medida que la correlación promedio aumenta, estancándose en valores de kurtosis cercanos a 3 cuando sobrepasa una correlación promedio de 0.4, y pasando el umbral de 0.6 la kurtosis vuelve a subir.

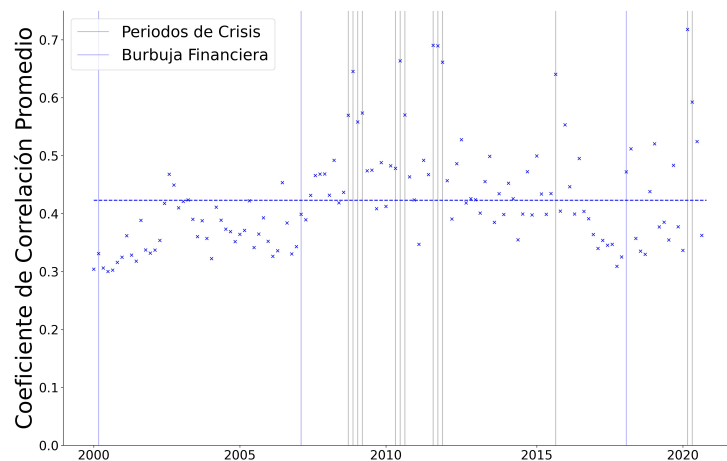
Recordando una premisa pasada, en la figura 3.5 se aprecia la evolución temporal del coeficiente de correlación promedio a lo largo de las 130 matrices de correlación, que abarcan de enero 2000 a septiembre 2020. En cada subfigura se tienen marcados los periodos de crisis más importantes de la tabla 3.3; así como de la tabla 3.2, que contemplan las épocas donde se registraron picos de burbujas financieras dentro del mercado estadounidense. Con lo anterior, no se alcanza a apreciar un patrón aparente, más allá de una tendencia de alta correlación, sin importar la definición empleada, en aquellas zonas grises que abarcan crisis [2]. Que se tenga el pico de una burbuja financiera para su inmediato reventar no parece implicar una correlación particularmente alta. Tanto Pearson como Spearman cuentan con épocas con una correlación promedio cercanos a cero, sin embargo, eso no implica que no hubiera correlación en esos periodos del mercado; también afecta el hecho que estos dos coeficientes toman valores negativos, donde el coeficiente de Distancia va de 0 a 1.



(a) Pearson



(b) Spearman



(c) Distancia

Figura 3.5: Coeficiente de correlación promedio (CCP) a lo largo de las épocas, ordenadas cronológicamente: 1 de enero del 2000 al 1 de septiembre del 2020.

Viendo la media para los coeficientes de correlación promedio, Pearson y Spearman mantienen valores cercanos, siendo 0.313 y 0.328, respectivamente; mientras que la correlación de Distancia cuenta con una media de 0.433, bastante alejada de las otras dos. Con su serie de

tiempo en la figura 3.5 se percibe que los coeficientes más bajos superan por mucho a los de otras alternativas. En otras palabras, el coeficiente de Distancia puede ser más sensible al medir relaciones que no detectan los otros coeficientes.

Selección de Épocas

A partir de este punto se realizó un análisis profundo para épocas selectas y así percibir sus diferencias en el corto plazo al emplear distintos tipos de correlación, notar reacciones del mercado a eventos con alta influencia y una prolongación muy corta en el tiempo como pueden ser crisis, crashes, burbujas, entre otros. Para esto, de la figura 3.5 dentro del grupo de Pearson fueron seleccionadas las épocas con los valores más altos de coeficiente de correlación promedio⁸, y de entre ellas se propuso filtrar y mantener las tres épocas con el mayor coeficiente de correlación promedio que ocurrieran en distintas crisis o burbujas; con un criterio opuesto, se seleccionó a la época con menor coeficiente de correlación promedio, que coincidentemente fue el mismo para las tres definiciones de correlación. A continuación se hace un compilado con el contexto necesario de las cuatro épocas seleccionadas:

- **Época 4 - 23 de junio al 18 de agosto del 2000.** Época con el menor coeficiente de correlación para las tres definiciones, considerado como un periodo estable.
- **Época 66 - 6 de mayo al 1 de julio del 2010.** Flash Crash.
- **Época 73 - 15 de agosto al 10 de agosto del 2011.** Caída del mercado estadounidense del 2011.
- **Época 127 - 15 de enero al 12 de marzo del 2020.** Recesión por COVID-19.

Se tomaron los mismos números de época en los grupos de matrices de Spearman y de Distancia⁹. Para cada una de estas épocas se graficaron sus mapas de calor, distribución de eigenvalores, distribución de coeficientes de correlación, razones de participación en función de sus eigenvalores, e inclusive su evolución temporal de las razones de participación asociados al mayor y menor de los eigenvalores. Los mapas de calor de las matrices de correlación, así como las distribuciones de densidad probabilística de los coeficientes de correlación de las épocas 4, 66, 73 y 127 están dados en las figuras 3.6, 3.7, 3.8 y 3.9 respectivamente. Asimismo, los valores de sus cuatro momentos están en las tablas 3.5, 3.6, 3.7 y 3.8.

Para la época 4, el coeficiente de correlación promedio de Distancia es significativamente más alto. En las distribuciones de la figura 3.6 son casi imperceptibles los coeficientes cercanos a cero, lo cual sugiere la existencia de correlaciones no lineales en los datos; sin embargo, el hecho de que los otros coeficientes tienen una cantidad cercana a la mitad con valores negativos, influye en que sus coeficientes sean más bajos. El resto de momentos muestran diferencias presentes entre la correlación de Distancia y sus contrapartes. Tanto Pearson como Spearman tienen valores de

⁸Como era de esperarse, las tres épocas seleccionadas se encontraban en distintos periodos que abarcan una crisis distinta.

⁹Puede parecer contraintuitivo para el lector no ver entre las épocas selectas alguna que abarque la burbuja inmobiliaria y/o la Gran Recesión, su justificación está en que ninguna de las épocas que abarcan estos periodos tiene un coeficiente de correlación promedio tan alto para ser de las épocas selectas. Más adelante en el trabajo se verá el argumento de esto.

asimetría cercanos a cero, una kurtosis y una desviación estándar similar, asemejan bastante a una distribución normal, es decir, acampanada y bastante simétrica; para la distribución de Distancia es lo contrario, está sesgada hacia su lado derecho, marcada por la asimetría positiva y mayor a 1, y hay una alta concentración en los datos en un rango muy pequeño, denotado por su desviación estándar menor y su kurtosis el doble de grande.

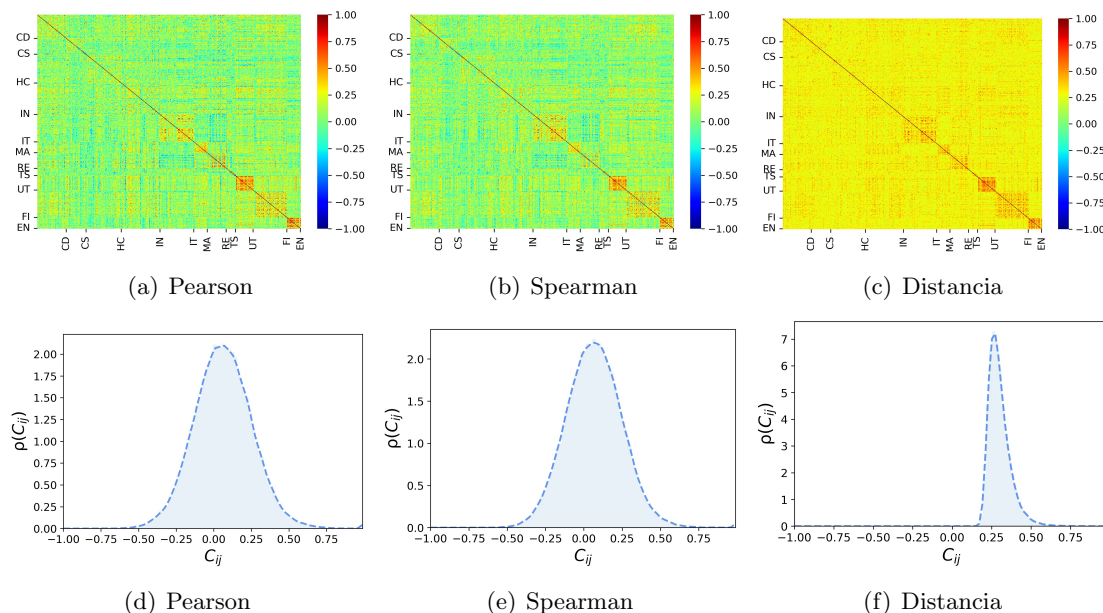


Figura 3.6: Matrices de correlación, así como distribuciones de densidad probabilística de los coeficientes de matrices de correlación de la época 4: de junio 23 a agosto 18 del 2000.

	Pearson	Spearman	Distancia
CCP	0.0623	0.0676	0.3000
σ	0.1900	0.1787	0.0690
Asimetría	0.0910	0.0927	1.4237
Kurtosis	3.1995	3.0720	6.6031

Tabla 3.5: Momentos para cada una de las definiciones de correlación empleadas en la época 4.

Para el resto de épocas alcanza a visualizarse cierto patrón. Durante periodos donde no hay crisis los coeficientes de correlación de a Pearson y Spearman pueden ser negativos; durante las crisis los coeficientes negativos son casi inexistentes. Las distribuciones tienden a sesgarse más hacia la izquierda a medida que el coeficiente de correlación promedio aumenta, esto concuerda con la inferencia dada en la tabla 3.4; lo mismo sucede con kurtosis, que en mayor o menor medida se ve reflejado qué tan altos están los picos de la distribución, dependiendo de la definición de correlación empleada. Esto desde la perspectiva financiera puede indicar que entre mayores sean las crisis los activos se comportarán como un solo agente [2], la mayoría de los activos mantendrán un coeficiente de correlación alto. Pueden notarse ligeros abultamientos en la parte izquierda para los tres tipos de distribuciones, esto implica que habrán activos, en este caso empresas, que no seguirán la corriente del resto del índice o el mercado en general, pero son tan pocos que pueden considerarse como irrelevantes.

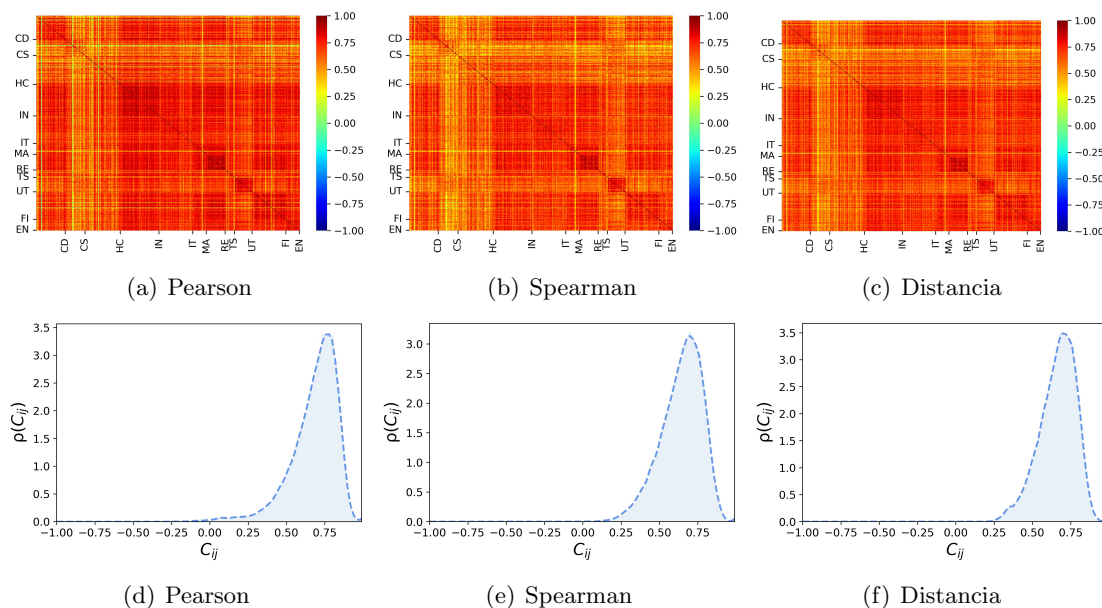


Figura 3.7: Matrices de correlación, así como distribuciones de densidad probabilística de los coeficientes de matrices de correlación de la época 66: de mayo 6 a julio 1 del 2010.

	Pearson	Spearman	Distancia
CCP	0.6817	0.6432	0.6638
σ	0.1528	0.1344	0.1179
Asimetría	-1.4794	-0.6683	-0.6142
Kurtosis	6.3732	3.2909	3.1468

Tabla 3.6: Momentos para cada definición de correlación empleada en la época 66.

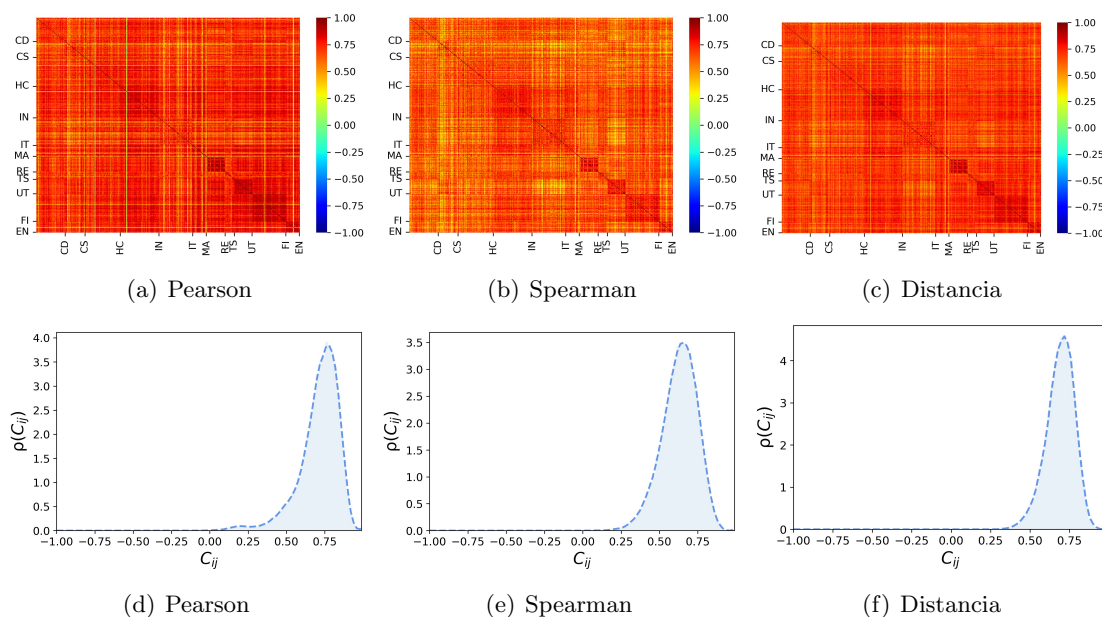


Figura 3.8: Matrices de correlación, así como distribuciones de densidad probabilística de los coeficientes de matrices de correlación de la época 73: de junio 15 a agosto 10 del 2011.

	Pearson	Spearman	Distancia
CCP	0.7123	0.6296	0.6905
σ	0.1326	0.1160	0.0896
Asimetría	-1.5058	-0.5019	-0.5125
Kurtosis	6.5401	3.3423	3.3716

Tabla 3.7: Momentos para cada definición de correlación empleada en la época 73.

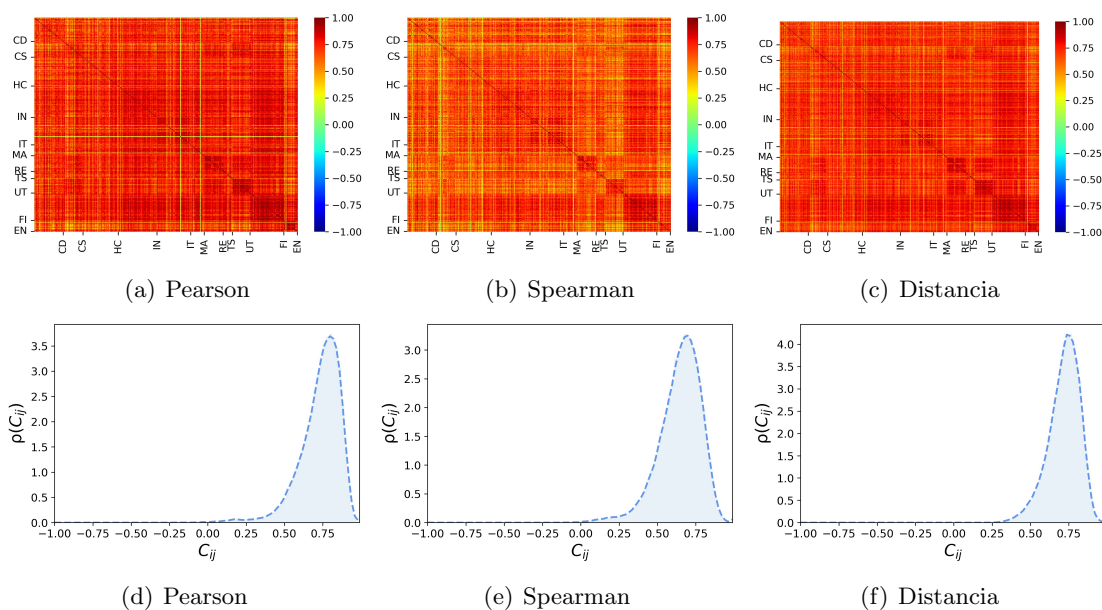


Figura 3.9: Matrices de correlación, así como distribuciones de densidad probabilística de los coeficientes de matrices de correlación de la época 127: de enero 15 a marzo 12 del 2020.

	Pearson	Spearman	Distancia
CCP	0.7326	0.6460	0.7180
σ	0.1336	0.1374	0.1023
Asimetría	-1.5345	-0.9918	-0.7320
Kurtosis	7.0978	4.8186	3.6323

Tabla 3.8: Momentos para cada definición de correlación empleada en la época 127.

El coeficiente correlación promedio mantuvo la misma tendencia en las épocas seleccionadas. En términos generales, no hay tantos cambios en correlaciones altas, donde más se encuentran diferencias son en las bajas. Para Spearman y Distancia se fue notando el incremento donde se veía en Pearson, sin embargo, el coeficiente de Spearman no siguió avanzando cuando sus contrapartes lo hacían; Pearson y de Distancia yendo de manera ascendente y superando valores mayores al 0.7, mientras que Spearman a partir de la segunda época fue descendiendo, no pasando de 0.65. Esto podría indicar que la correlación de Spearman, que mide relaciones monótonas, no es capaz de medir la correlación donde sí lo pueden hacer Pearson (de carácter lineal) y la correlación de Distancia (de carácter no lineal). Es importante recalcar que en periodos cortos de tiempo se perciben de igual modo la alta correlación entre los sectores de energía (EN), finanzas (FI), servicios públicos (UT) e inmobiliario (RE); aunque en este análisis se remarcan

nuevos sectores dependiendo la época, siendo el caso del sector industrial (IN) cuando hay alta correlación y el de investigación y tecnología (IT) cuando la correlación es baja.

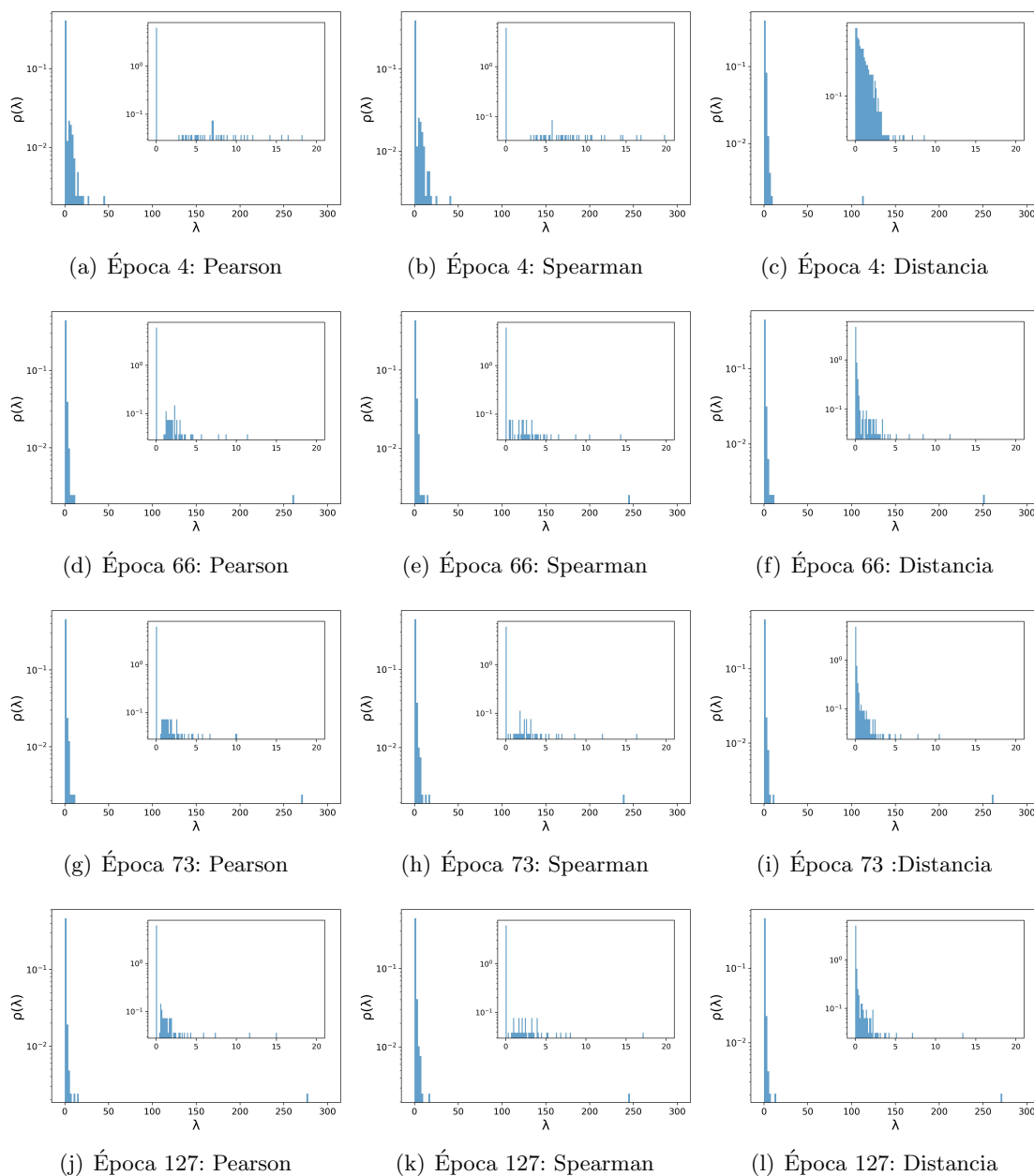


Figura 3.10: Densidad probabilística (a escala logarítmica) de eigenvalores sobre las matrices de correlación de la época 4, 66, 73 y 127, respectivamente, para los tres coeficientes de correlación. Para cada una de las distribuciones se perciben tres principales secciones: el pico de valores nulos, el abultamiento que es interpretado como ruido, y los valores atípicos que representan al máximo eigenvalor.

Las distribuciones correspondientes a los eigenvalores de las épocas 4, 66, 73 y 127 son mostradas en la figura 3.10. Como se mencionó anteriormente, para series de tiempo cortas las matrices de correlación cuentan con un número reducido de eigenvalores distintos de cero, para este análisis se tienen 39 eigenvalores distintos de cero por época [23]. Dicha cantidad no es

suficiente para obtener datos estadísticos contundentes. A pesar de ello, en cada una de las distribuciones, sin distinguir la época, puede verse el pico de la distribución en el cero, al igual que un abultamiento de los eigenvalores distintos de cero, aún más visible en los casos de correlación de Distancia, en especial en el caso de menor correlación promedio (época 4); mas se cuenta con valores atípicos, en cada distribución se aprecia el máximo eigenvalor que incrementa a medida que el coeficiente de correlación promedio lo hace.

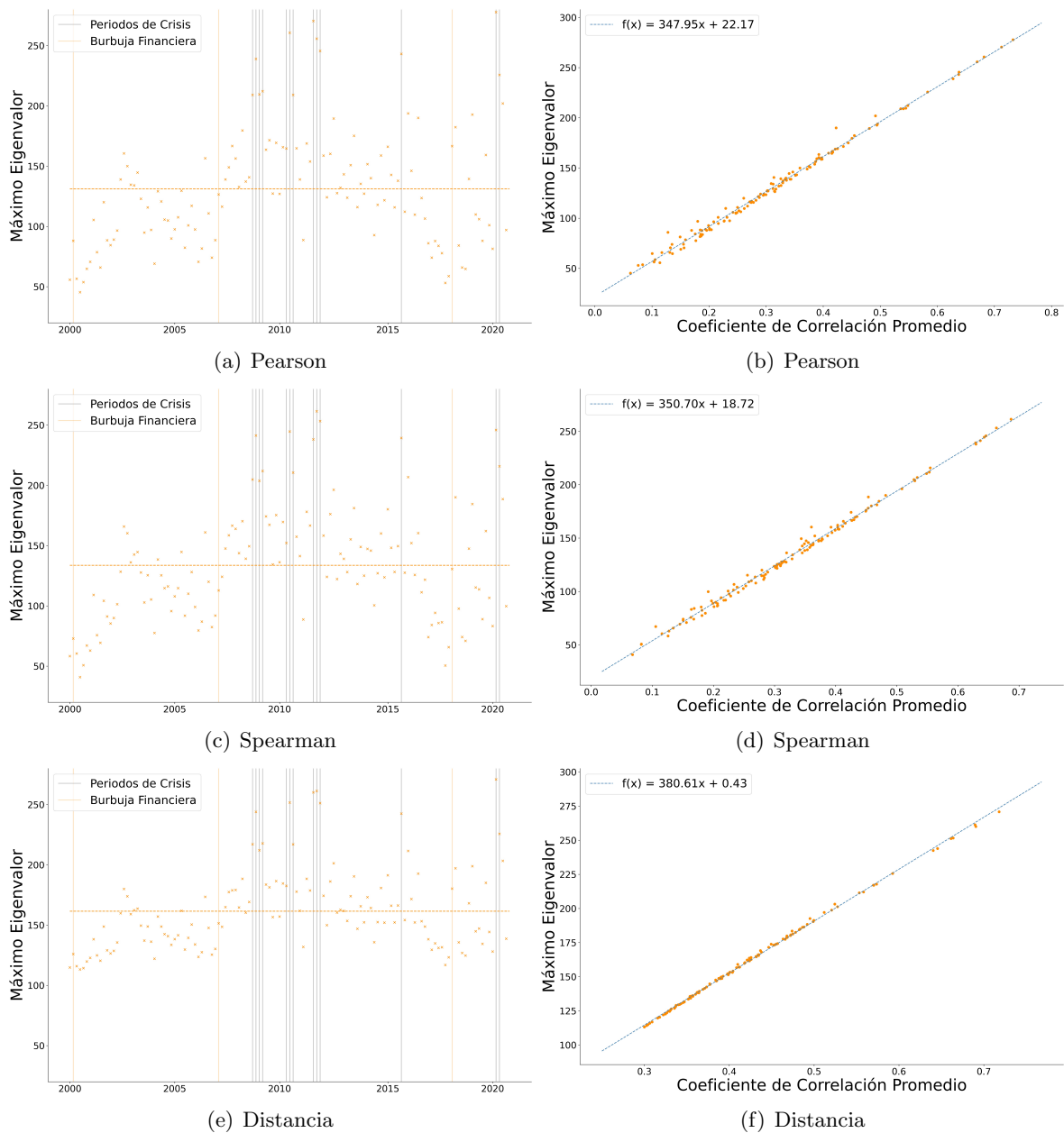


Figura 3.11: Máximo eigenvalor a lo largo de las matrices de correlación de cada época, ordenadas cronológicamente: 1 de enero del 2000 al 1 de septiembre del 2020; así como gráficas de dispersión del máximo eigenvalor por época contra su respectivo coeficiente de correlación promedio.

Lo último mencionado invitaba a analizar más a fondo la relación entre estas dos propiedades. Y así se hizo, en la figura 3.11 se visualiza la evolución temporal del máximo eigenvalor de cada matriz de correlación. Comparando estas series de tiempo con su análogo de coeficientes de correlación promedio (figura 3.5) se notan similitudes en cómo están dados sus puntos; y comparándose en una gráfica de dispersión, invariante de la definición, se nota una tendencia totalmente lineal. Esto confirma que altas correlaciones pueden indicarse con el máximo eigenvalor de la matriz de correlación. Queda clara la relación entre altas correlaciones y el máximo de los eigenvalores. A una mayor correlación, mayor el valor del máximo eigenvalor. Sin embargo, aún falta por ver si lo mismo ocurre entre correlaciones medias y los eigenvectores de las matrices de correlación en términos de sus razones de participación.

Para la época 4, las razones de participación (figura 3.12), a excepción del caso de Distancia, fluctúan alrededor del valor teórico del ensamble WOE ($N/3$). Con respecto a los eigenvectores, ignorando los eigenvalores cercanos a cero, los casos con correlaciones bajas muestran eigenvectores extendidos en mayor cantidad; mientras que en Distancia, gran parte de los eigenvectores son localizados y acumulados. El eigenvector asociado al eigenvalor máximo siempre se encuentra extendido, pero alcanzando distintos valores, y en particular, la correlación de Distancia tiene un valor muy alejado del umbral teórico¹⁰. Los casos de Pearson y Spearman, al tener correlaciones bastante bajas, su acumulación de datos se dispersa, obteniendo menor concentración de datos a medida que los eigenvalores aumentan.

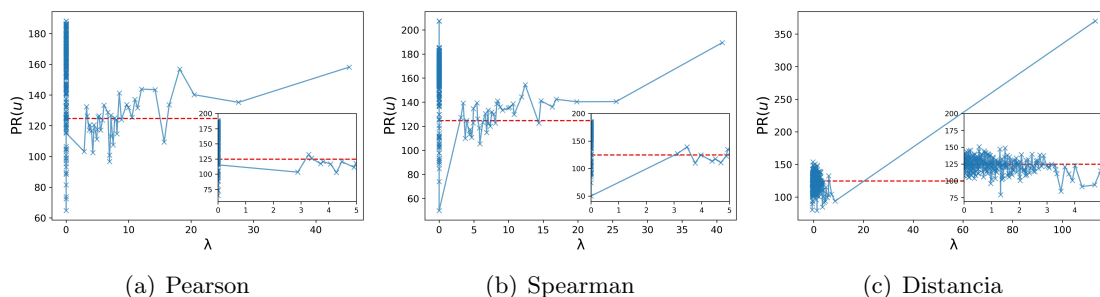


Figura 3.12: Razones de participación en función de los eigenvalores para cada tipo matriz de correlación de la época 4, siendo la línea roja punteada $N/3$, con N igual al número de acciones contempladas dentro de la matriz de correlación teórica.

Para las épocas de crisis (figura 3.13), la razón de participación correspondiente al valor propio máximo está bastante alejado del resultado WOE. Nuevamente hay una gran diferencia en cómo se distribuyen los datos de Distancia con respecto a los otros coeficientes, el abultamiento de puntos cercanos a cero mantienen un límite cercano a 150; en sus contrapartes se alcanzan cantidades mayores a 200. Por otro lado, el número máximo de activos que participan en el eigenvector asociado al eigenvalor máximo se mantiene bastante consistente, en todos superando el valor de 350, denotando un estado extendido sin excepción, se interpreta entonces que casi todos los activos están correlacionados y que participan en el eigenvector de manera similar; a diferencia del caso cuando la correlación es baja.

¹⁰Esto se explica recordando la relación lineal entre el CCP y el máximo eigenvalor.

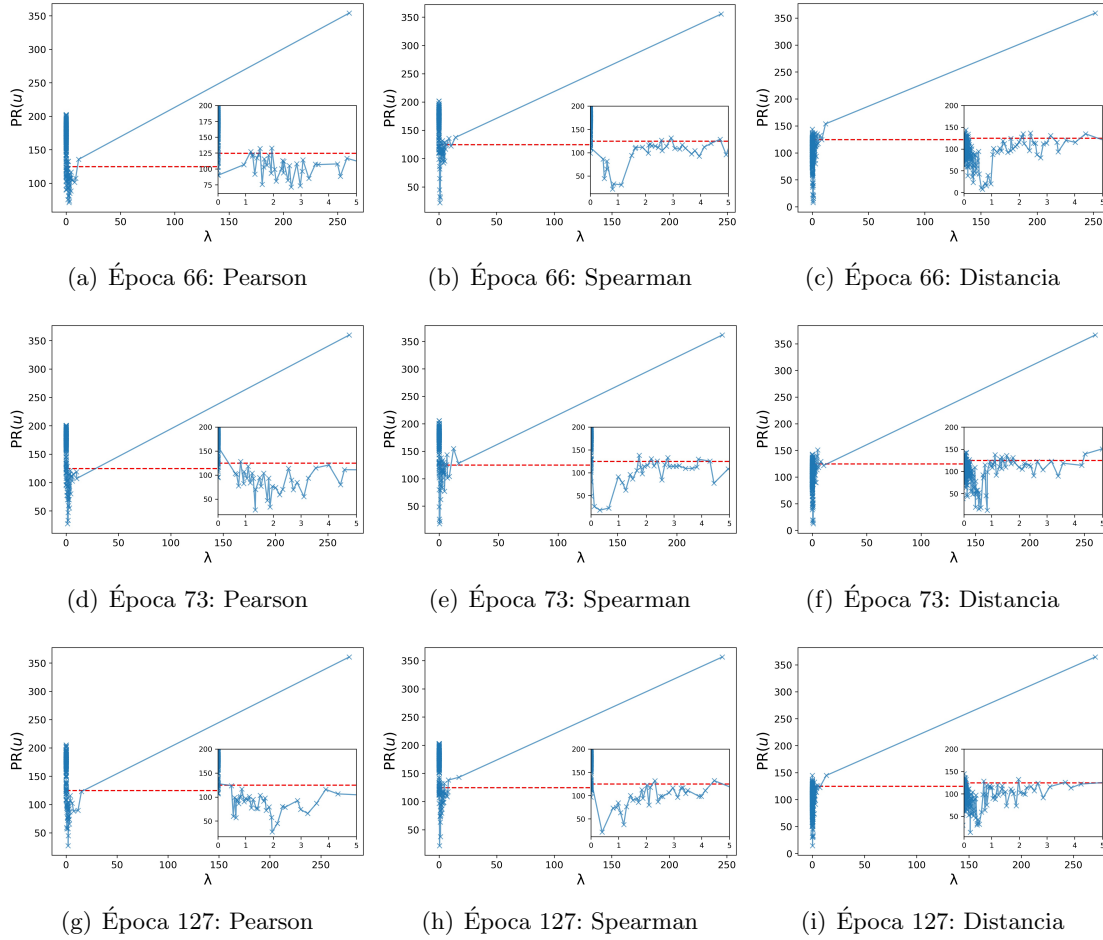


Figura 3.13: Razones de participación en función de los eigenvalores para las matrices de correlación de las épocas 66, 73 y 127. Siendo la línea roja punteada $N/3$, con N igual al número de acciones contempladas dentro de la matriz de correlación teórica.

Analizando la media $\langle PR \rangle$, con el coeficiente de Distancia no se cuenta con ninguna media que supere el umbral teórico, indicando predominancia en los eigenvectores localizados. Totalmente contrarios fueron las medias para Pearson y Spearman que siempre fueron mayores a este umbral, sugiriendo una ponderación en los estados extendidos; aunque es importante mencionar que para estos dos coeficientes el valor de razón de participación asociado al eigenvalor más alto es ligeramente menor cuando se tiene una menor correlación.

Con apenas cuatro épocas se percibe un problema que cuentan las gráficas de razones de participación con respecto a los eigenvalores, siendo este su inconsistencia en la escala vertical y la concentración de sus datos. Esto limita el análisis que se puede dar por el rango entre eigenvalores y razones de participación. Como alternativa para ampliar el análisis, se presenta la figura 3.14 que muestra las razones de participación asociados a los eigenvalores máximos y mínimos, junto con el umbral teórico de las 130 épocas. Con lo presentado se reafirma que todas las razones de participación máximos van asociados a eigenvectores extendidos. El rango entre razones de participación máximos y mínimos son considerablemente más altos en el caso de Distancia, hay una notable estacionariedad en los valores mínimos y máximos de las razones

de participación; en el caso de Pearson y Spearman, no se muestra ninguna, por otro lado, muestran valores altos en sus máximos cuando se tienen épocas con crisis¹¹. Esto expresa que al usar distintas definiciones de correlación la participación máxima de activos puede variar de manera considerable dentro del eigenvector de una misma época.

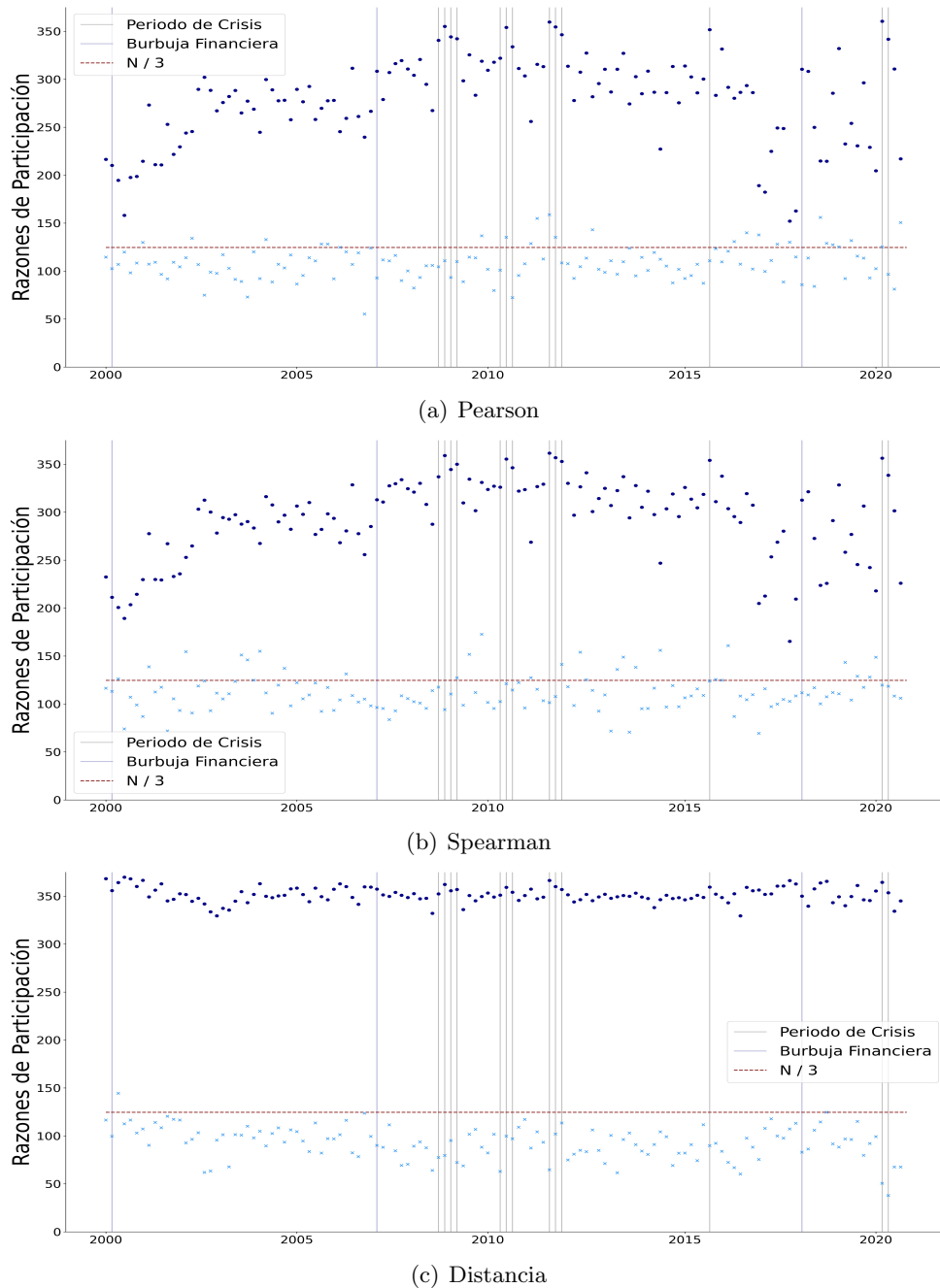


Figura 3.14: Razones de participación correspondientes a los eigenvalores máximos y mínimos para cada época. La línea horizontal corresponde al umbral teórico del ensamble de Wishart.

¹¹Ajeno a lo expuesto en este trabajo, se analizó la relación entre el máximo de razones de participación con los coeficientes de correlación promedio por época para cada uno de los coeficientes, y se obtuvieron resultados diametrales: mientras que Pearson y Spearman presentaron una correlación lineal alta con un R^2 -score mayor a 0.8 para ambos casos, en Distancia no se mostró ninguna correlación (R^2 -score = 0.0207).

Por último, en el siguiente capítulo se verá que, para cada definición de correlación empleada, con los valores de coeficiente de correlación promedio de cada época se implementó un algoritmo de agrupación jerárquica con el fin de visualizar los distintos estados del mercado en los que se encontraba el índice durante cada periodo de tiempo dependiendo de su alta o baja correlación promedio.

Capítulo 4

Algoritmo de Clustering

Habiendo analizado las matrices de correlación, sus eigenvalores, razones de participación y momentos, procedemos a ampliar el análisis empleando un algoritmo capaz de agrupar datos obtenidos del análisis previo. Dentro del grupo de algoritmos capaces de agrupar por medio de jerarquías, el agrupamiento jerárquico crea grupos o *clusters* fusionando conjuntos de manera sucesiva. Empleando un enfoque ascendente, comúnmente llamado *bottom-up*. Se comienza con muestras de un solo elemento, y en cada iteración, dado una métrica previamente definida llamada criterio de vinculación, las muestras se juntan unas con otras hasta alcanzar un único grupo [42, 43]. Comúnmente, la representación de esta jerarquía está dada por un dendrograma. Yendo de manera descendente con respecto a las iteraciones puede establecerse el número de clusters que el usuario considere más adecuado.

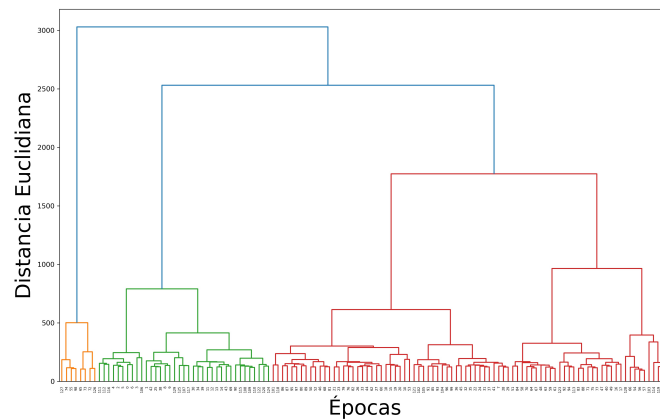


Figura 4.1: Dendrograma ejemplo.

Para la clasificación de estados del mercado se utilizó este mismo algoritmo de clustering. Recordando que por definición de correlación se tuvo un grupo de 130 matrices de correlación, cada matriz con un coeficiente de correlación promedio asociado, estos fueron los valores a agrupar, siendo un total de 130. El criterio de vinculación o *linkage* empleado fue el conocido como *ward*, el cual busca minimizar la varianza entre las muestras al ser agrupadas en cada iteración [3, 4]. La utilización del criterio seleccionado requería del cálculo de una matriz de

distancia como objeto de entrada para el algoritmo; para ello, se recuperaron una vez más los tres grupos de 130 matrices de correlación que, de manera individual por grupo, se computó una nueva matriz ξ cuyos elementos se definen como sigue:

$$\xi(t_i, t_j) = d_E |C(t_i) - C(t_j)| \quad (4.1)$$

con d_E representando norma L_2 o euclidiana y $C(t)$ como el coeficiente de correlación promedio de la época t , e $i, j = 1, \dots, 130$ [44, 45]. La figura 4.2 muestra las matrices de distancia resultantes de cada uno de los grupos de matrices. Invariante al conjunto de matrices de correlación¹, se contaba con 130 matrices representando distintos periodos de tiempo, esto implica que la matriz resultante, llamada matriz de distancia, cuenta con dimensiones de 130×130 .

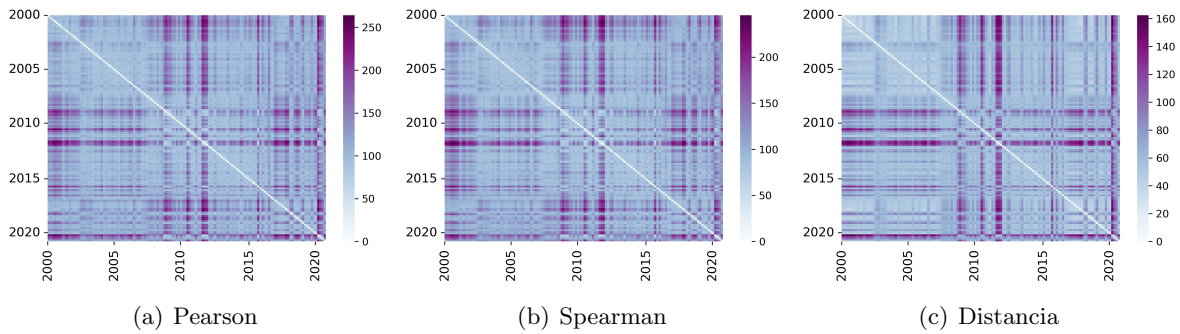


Figura 4.2: Matrices de distancia aplicando la norma L_2 a los coeficiente de correlación promedio de las 130 épocas obtenidas por los coeficiente de correlación de Pearson, Spearman y Distancia.

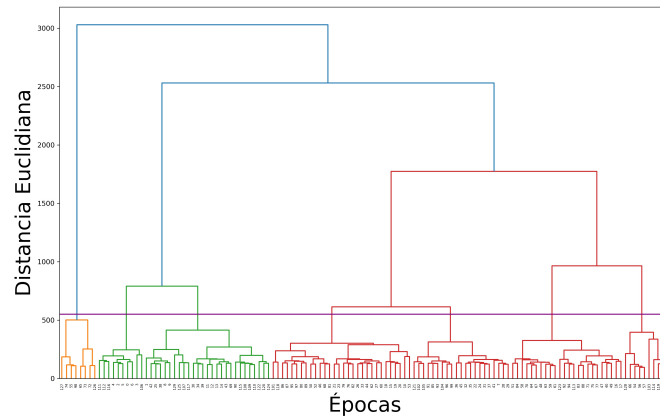
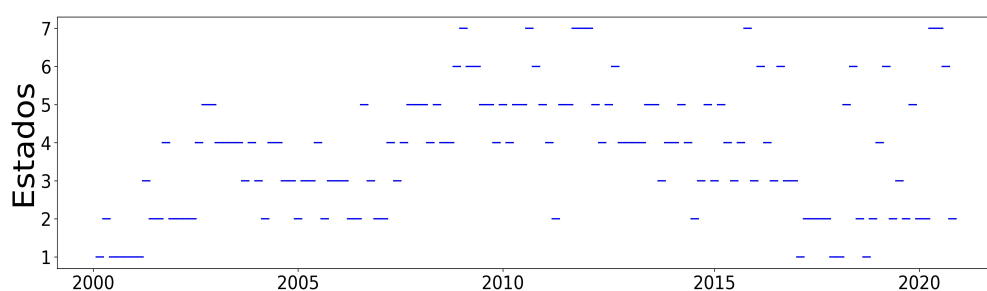


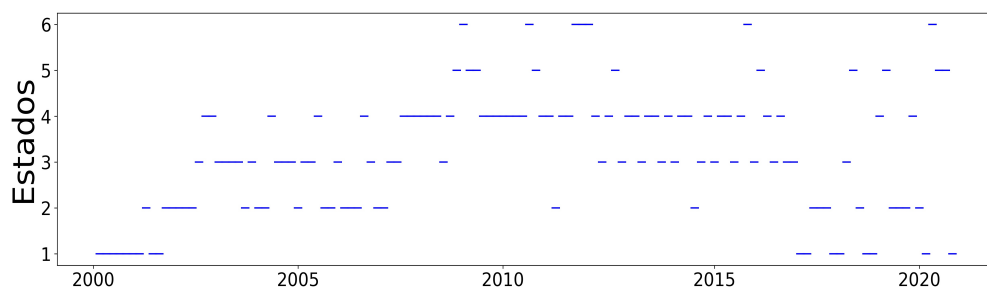
Figura 4.3: Dendrograma ejemplo, aplicando el umbral de máxima distancia (línea morada) para la obtención del número óptimo de agrupaciones.

¹Nótese que por cómo está definida la matriz de distancia, cada matriz de correlación colapsa a un solo valor escalar, es por ello que cada entrada de la matriz resultante se puede ver como la distancia entre los valores de la matriz de tiempo l con la de tiempo r .

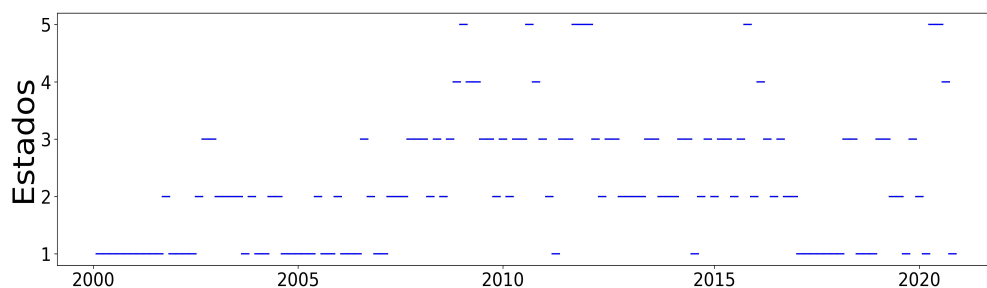
Una vez obtenida la matriz de distancia, se usaron los coeficientes de correlación promedio como entrada para poder agruparlos en n estados que se consideraran adecuados, no obstante, era necesario determinar el número óptimo de agrupaciones para los 130 coeficientes. Como criterio de selección se empleó un umbral de máxima distancia: como se ve en la figura 4.3, la máxima distancia la denota la línea azul, al realizar un corte horizontal en la zona inferior queda determinado el valor del umbral, y de manera visual las rectas verticales que atraviesen el umbral horizontal indican el número de clusters óptimos para analizar. Entonces, al utilizar el umbral (línea morada, figura 4.3) para cada grupo de coeficientes, el número de clusters resultaron de la siguiente manera: 7 para Pearson, 6 para Spearman y 5 para Distancia. Es en la figura 4.4 que se muestran los resultados de la evolución temporal de los estados del mercado a partir de su coeficiente de correlación promedio.



(a) Pearson



(b) Spearman



(c) Distancia

Figura 4.4: Agrupación de los coeficientes de correlación promedio (ordenados cronológicamente) en los posibles estados del mercado de cada definición de correlación.

Con lo contenido en la tabla 4.1 se notan los clusters de coeficientes de correlación promedio organizados de manera descendente, con sus valores de media y el conteo de sus elementos que lo componen. El primero de los estados en aparecer representa el estado de mayor correlación del mercado, significando mayores crisis, sin importar la definición, aunque para Spearman es ligeramente menor, la media se mantiene en valores similares. Contrariamente, los últimos estado figuran periodos de calma en el mercado, manifestado una correlación baja²; no obstante, la correlación de Distancia marca una media lejana al de los otros dos clusters, recordando que el más bajo de sus coeficientes es 0.3000. Por convención, se llamarán estados bajos a los clusters con una media menor a 0.31 (y el primer estado de Distancia ya que es el menor estado de este coeficiente) y al estado con mayor correlación como crítico.

Pearson			Spearman			Distancia		
Estado	Media	Elementos	Estado	Media	Elementos	Estado	Media	Elementos
7	0.6602	8	6	0.6478	7	5	0.6627	8
6	0.5007	10	5	0.5162	10	4	0.5582	6
5	0.4029	24	4	0.3939	40	3	0.4716	34
4	0.3172	30	3	0.3065	29	2	0.4043	38
3	0.2558	21	2	0.2210	27	1	0.3404	44
2	0.1791	27	1	0.1351	17			
1	0.1022	10						

Tabla 4.1: Media de los coeficientes de correlación promedio (CPP) por estado del mercado según la definición de correlación empleada, al igual que el número de elementos por estado.

Al considerar el número de estados, que es irregular, entre las definiciones de correlación, se nota una gran concentración en el último estado del caso de Distancia; algo similar se presenta en los casos de Pearson y Spearman, donde las grandes concentraciones se dan más en los niveles medios de correlación y a medida que baja la correlación también lo hace el número de componentes del cluster.

Paralelamente, se promediaron las matrices de correlación de épocas contenidas en cada cluster, y en las figuras 4.5, 4.6 y 4.7 se muestran los mapas de calor asociados a los estados del mercado. Recordando lo expuesto en la página 30 y 32 acerca de los resultados divergentes al emplear la definición de Distancia en periodos de correlaciones bajas, tiene sentido lo mostrado en los estados bajos de la figura 4.7, que cuentan con una coloración amarillenta, la cual exhibe que el coeficiente de correlación promedio de estas épocas son más altas que las de otras definiciones (estados bajos de las figuras 4.5 y 4.6).

²Para mayor información, véase la tabla A.5 de la sección Apéndice, que muestra los elementos constituyentes al estado de menor correlación para cada definición.

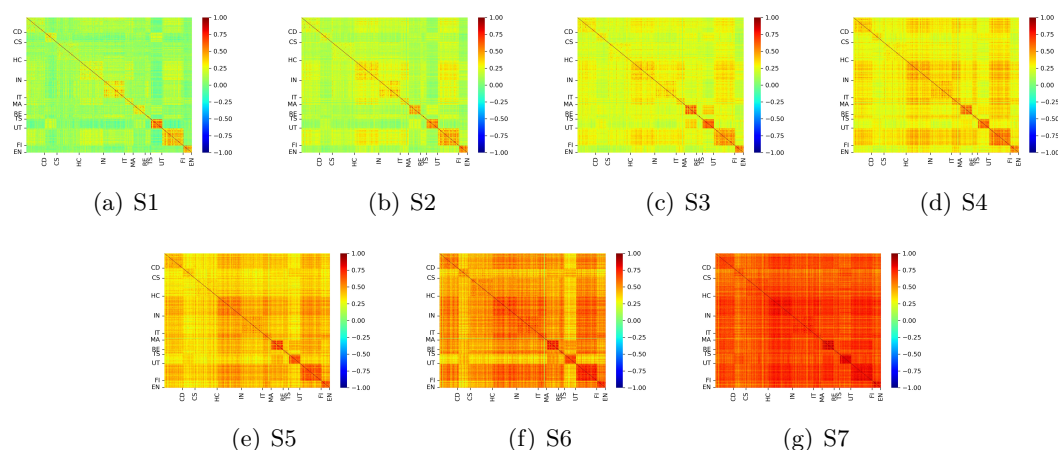


Figura 4.5: Matrices de correlación promedio por cada estado del mercado al emplear la correlación de Pearson.

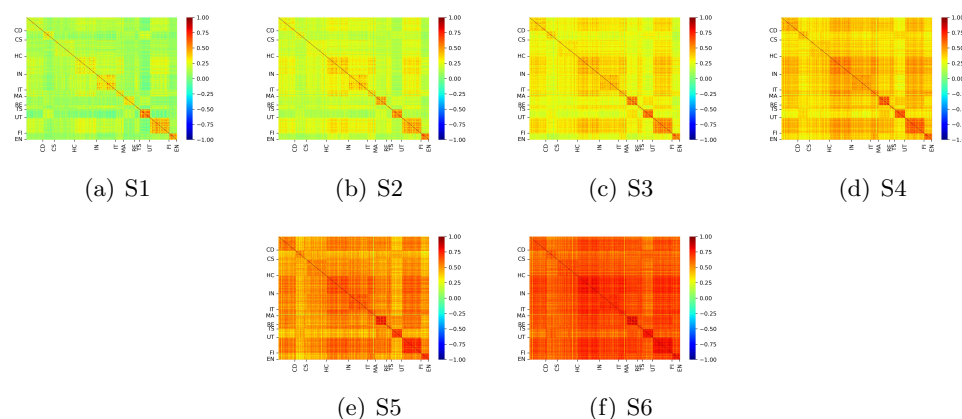


Figura 4.6: Matrices de correlación promedio por cada estado del mercado al emplear la correlación de rango de Spearman.

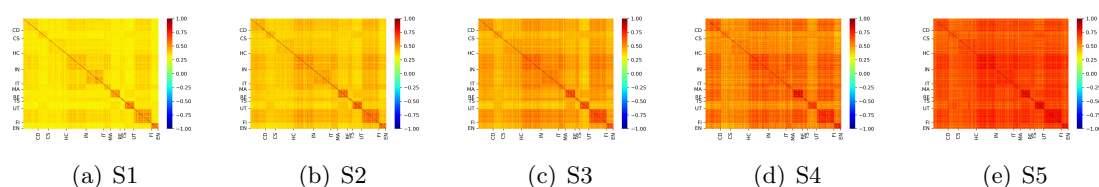


Figura 4.7: Matrices de correlación promedio por cada estado del mercado al emplear la correlación de Distancia.

Seguir la evolución de los estados por época ha mostrado ser relevante para investigaciones en años recientes. Con el uso de matrices de transición que contea las veces que se transiciona entre estados del mercado. Por cada grupo de correlación se ordenaron los coeficientes de correlación promedio en orden cronológico ascendente. Se contabilizaron las transiciones al cambiar de época, ya sea de un cluster a otro o si se mantenía en el mismo estado³. La suma de los

³La matriz de transición se lee a partir de las filas y de izquierda a derecha; por ejemplo: la entrada t_{62}

elementos por matriz es igual a $N - 1$, el número de transiciones en N épocas.

La mayor cantidad de valores se mantienen cerca de la diagonal, esto quiere decir que las transiciones se dan en transiciones o saltos pequeños hacia los estados más cercanos o en sí mismo; se evitan grandes saltos desde estados de correlación baja al crítico y viceversa [3, 46, 47]. En cada matriz se tienen estados bien definidos que preceden las crisis y cuentan con una mayor cantidad de transiciones a los estados críticos (correlación promedio alta), así como estados de cobertura donde el pico de la crisis cesa al llegar un relativo periodo de calma: en Pearson, S5 se comporta como un precursor de las caídas del mercado y como agente de cobertura se tiene mayormente a S6; para Spearman, se tiene a S4 como principal precursor y a S5 como cobertura; y para la Distancia, es S3 el precursor y S4 la cobertura.

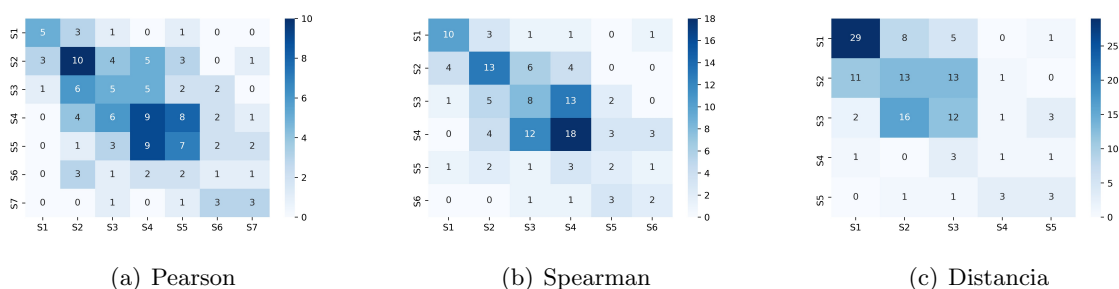


Figura 4.8: Matriz de transición de estados del mercado al aplicar los coeficientes de correlación.

Ya se mencionó que saltos entre estados bajos al crítico y viceversa se dan muy poco, cuando se tiene una disrupción a esta regla, que las hay, es necesario ver estos casos a detalle. Para Pearson se tienen saltos de S2 a S7 y de S7 a S3, en Spearman hay transiciones de S1 a S6 y de S6 a S3, y en Distancia solo de S1 a S5. Estas anomalías corresponden a los mismos dos saltos: de los estados bajos al crítico corresponde a la transición de la época 126 a la 127; y del estado crítico a los bajos es el paso de la época 99 al 100⁴.

1. **126-127.** Durante el periodo que contempla la época 127 se dio la recesión por COVID-19, pasando de correlaciones de 0.1492 (Pearson), 0.1639 (Spearman) y 0.3366 (Distancia) a valores de 0.7326 (Pearson), 0.6460 (Spearman) y 0.7180 (Distancia). La transición se explica al notar el cierre de fábricas por tiempo indefinido debido a las alzas en los contagios, llevando al mercado a tener una correlación muy alta en corto tiempo.
2. **99-100.** Durante el periodo que contempla la época 99 se dio el Lunes Negro del mercado chino, pasando de correlaciones de 0.6372 (Pearson), 0.6291 (Spearman) y 0.6403 (Distancia) a valores de 0.2673 (Pearson), 0.3189 (Spearman) y 0.4044 (Distancia). Por el crash chino se tuvo un mercado bajista con ligeras recuperaciones para el mercado.

representa el número de ocasiones que se salta en la matriz de transición de Pearson del estado 6 al 2, siendo un total de tres.

⁴Viendo la tabla 4.2 se pueden ver cuáles épocas corresponden a las crisis y épocas marcadas: Gran Recesión (56), Flash Crash (66), Caída del mercado estadounidense del 2011 (73 a 75), Lunes negro chino del 2015 (99), Recesión por COVID-19 (127 y 128).

Pearson		Spearman		Distancia	
Época	CCP	Época	CCP	Época	CCP
127	0.7326	74	0.6870	127	0.7180
73	0.7123	75	0.6630	73	0.6905
66	0.6817	127	0.6460	74	0.6896
74	0.6696	66	0.6432	66	0.6638
75	0.6378	56	0.6367	75	0.6613
99	0.6372	73	0.6296	56	0.6454
56	0.6275	99	0.6291	99	0.6403
128	0.5828			128	0.5925

Tabla 4.2: Épocas dentro del cluster con coeficientes de correlación promedio (CCP) más altos.

Capítulo 5

Conclusiones

Se realizaron análisis de correlación en series de tiempo financieras, tanto de una longitud temporal alta como corta, del índice estadounidense S&P500. La construcción del portafolio se basó únicamente en las acciones que presentaran precios de cierre en días hábiles para la ventana del 1 de enero del 2000 al 1 de septiembre del 2020. Se pudieron observar las bondades de emplear series de tiempo cortas con el uso de matrices de correlación en periodos de tiempo selectos: a pesar de sus limitantes teóricas, la descripción de los datos se enriquece de manera contundente pues soluciona el problema de estacionariedad y permite visualizar periodos de crisis a partir de una correlación alta en su promedio de los activos. Con distintas definiciones de correlación se afirma que el mercado estadounidense alcanza a tener correlaciones de carácter más allá del lineal, así que el utilizar alternativas al coeficiente de correlación de Pearson puede ser enriquecedor para este tipo de investigación. Se identifican más diferencias entre épocas de baja correlación al emplear distintas definiciones que en épocas de correlación alta, ya que en todos los casos el mercado se comporta como un solo agente con propiedades semejantes.

Buscando encontrar señales de correlación tanto en los eigenvalores como en los eigenvectores por medio de las razones de participación, se mostró que el máximo eigenvalor presenta una directa relación con la correlación promedio de las acciones; sin embargo, las razones de participación calculadas utilizando el coeficiente de correlación de Distancia no muestran ninguna diferencia entre los períodos de crisis y de los que no hay, y por otro lado, los coeficientes de correlación de Pearson y Spearman presentan variaciones que esbozan una relación directa con la correlación promedio pero no son concluyentes. Para continuar sobre esta línea de investigación se recomienda computar las acciones dentro del portafolio que son los mayores contribuidores. Asimismo, para futuras líneas de investigación se puede variar el periodo de tiempo en las series de tiempo cortas y comparar sus resultados con los actuales resultados; así como aumentar la ventana temporal de 20 años a una más amplia.

Se realizaron agrupaciones de los coeficientes de correlación promedio obtenidos del análisis de correlación, obteniendo 130 de ellos que representaron periodos de 40 días hábiles. Empleando un algoritmo de clustering jerárquico ascendente se llegaron a distintos números de estados para cada una de las definiciones, no obstante, representaban lo mismo, esto es, una cantidad de épocas muy limitadas donde se tienen las correlaciones del mercado más altas, indicando que a

mayor correlación se tiene una crisis más representativa y, a su vez, esto hizo que el mercado se comportara como un solo agente durante ese periodo; nuevamente, la correlación de Distancia parece diferir de sus análogos al mostrar menos estados y una correlación promedio más alta para los estados de menor correlación, esto es de esperarse ya que el coeficiente de correlación implica relaciones no lineales o no monótonas a diferencia de los coeficientes de correlación de Pearson o Spearman. Asimismo, se encontraron épocas que preceden a las crisis donde se transiciona de un estado de baja correlación al estado crítico o de más alta correlación, una posible continuación a este trabajo podría ser cómo es que se dan estos saltos tan abruptos.

Apéndice A

Datos del Mercado

A.1. Constitución de la Base de Datos

Símbolo	Índice
SPX	S&P 500 Index
ONE	S&P 100 Index
IDX	S&P Midcap 400
IQY	S&P Smallcap 600
ESG	S&P 500 ESG
IGX	S&P 500 Growth
IVX	S&P 500 Value
SRFI	S&P 500 Financials [Sector]
SRHC	S&P 500 Health Care [Sector]
SREN	S&P 500 Energy [Sector]
SRIT	S&P 500 Information Technology [Sector]
SRTS	S&P 500 Communication Services [Sector]
SRCD	S&P 500 Consumer Discretionary [Sector]
SRCS	S&P 500 Consumer Staples [Sector]
SRIN	S&P 500 Industrials [Sector]
SRMA	S&P 500 Materials [Sector]
SRRE	S&P 500 Real Estate [Sector]
SRUT	S&P 500 Utilities [Sector]
DOWI	Dow Jones Industrials Average
DOWT	Dow Jones Transportations Average
DOWU	Dow Jones Utilities Average
DOWC	Dow Jones Composite Average
IUXX	Nasdaq 100 Index
NASX	Nasdaq Composite
HGX	PHLX Housing Index
OSX	PHLX Oil Service
SOX	PHLX Semiconductor Index
UTY	PHLX Utility Index
XAU	PHLX Gold/Silver Index
VIX	CBOE Volatility Index
GNX	S&P GSCI Index
DXY	U.S. Dollar Index

Tabla A.1: Índices contenidos dentro de la base de datos de Quandl.

A.2. Tabla de Compañías

Ticker	Compañía	Sector
AMZN	AMAZON.COM ,INC	Consumer Discretionary
AZO	AUTOZONE, INC.	Consumer Discretionary
BBY	BEST BUY CO, INC.	Consumer Discretionary
BKNG	BOOKING HOLDINGS, INC.	Consumer Discretionary
BWA	BORG WARNER, INC.	Consumer Discretionary
CCL	CARNIVAL CORPORATION	Consumer Discretionary
DHI	D.R. HORTON, INC.	Consumer Discretionary
DLTR	DOLLAR TREE, INC.	Consumer Discretionary
DRI	DARDEN RESTAURANTS, INC.	Consumer Discretionary
EBAY	EBAY, INC.	Consumer Discretionary
F	FORD MOTOR COMPANY	Consumer Discretionary
GPC	GENUINE COMPANY	Consumer Discretionary
GPS	GAP, INC.	Consumer Discretionary
HAS	HASBRO, INC.	Consumer Discretionary
HD	HOME DEPOT, INC.	Consumer Discretionary
KMX	CARMAX, INC.	Consumer Discretionary
LB	LIMITED BRANDS, INC.	Consumer Discretionary
LEG	LEGGETT & PLATT, INC.	Consumer Discretionary
LEN	LENNAR CORPORATION	Consumer Discretionary
LOW	LOWE'S COMPANIES, INC.	Consumer Discretionary
MAR	MARRIOTT INTERNATIONAL, INC.	Consumer Discretionary
MCD	MCDONALD'S CORPORATION	Consumer Discretionary
MGM	MGM RESORTS INTERNATIONAL	Consumer Discretionary
MHK	MOHAWK INDUSTRIES	Consumer Discretionary
NKE	NIKE, INC.	Consumer Discretionary
NVR	NVR. INC.	Consumer Discretionary
NWL	NEWELL RUBBERMAID, INC.	Consumer Discretionary
ORLY	O'REILLY AUTOMOTIVE, INC.	Consumer Discretionary
PHM	PULTE GROUP, INC	Consumer Discretionary
POOL	POOL CORPORATION	Consumer Discretionary
PVH	PHILLIPS VAN HAUSEN CORPORATION	Consumer Discretionary
RCL	ROYAL CARIBBEAN GROUP	Consumer Discretionary
RL	RALPH LAUREN CORPORATION	Consumer Discretionary
ROST	ROSS STORES, INC.	Consumer Discretionary
SBUX	STARBUCKS CORPORATION	Consumer Discretionary
TGT	TARGET CORPORATION	Consumer Discretionary
TIF	TIFFANY & CO.	Consumer Discretionary
TJX	THE TJX COMPANIES, INC.	Consumer Discretionary
TSCO	TRACTOR SUPPLY COMPANY	Consumer Discretionary
VFC	V.F. CORPORATION	Consumer Discretionary
WHR	WHIRLPOOL CORPORATION	Consumer Discretionary
YUM	YUM BRANDS, INC.	Consumer Discretionary
ADM	ARCHER-DANIELS-MIDLAND COMPANY	Consumer Staples
BF.B	BROWN-FORMAN CORPORATION	Consumer Staples
CAG	CONAGRA BRANDS, INC.	Consumer Staples
CHD	CHURCH & DWIGHT	Consumer Staples
CL	COLGATE-PALMOLIVE COMPANY	Consumer Staples
CLX	CLOROX COMPANY	Consumer Staples
COST	COSTCO WHOLESALE CORPORATION	Consumer Staples
CPB	CAMPBELL SOUP COMPANY	Consumer Staples

Ticker	Compañía	Sector
EL	ESTEE LAUDER COMPANIES, INC.	Consumer Staples
GIS	GENERAL MILLS, INC.	Consumer Staples
HRL	HORMEL FOODS CORPORATIONS	Consumer Staples
HSY	HERSHEY COMPANY	Consumer Staples
K	KELLOGG COMPANY	Consumer Staples
KMB	KIMBERLY CLARK CORPORATION	Consumer Staples
KO	COCOA-COLA COMPANY	Consumer Staples
KR	KROGER COMPANY	Consumer Staples
MKC	MCCORMICK & COMPANY, INC.	Consumer Staples
MNST	MONSTER BEVERAGE CORPORATION	Consumer Staples
MO	ALTRIA GROUP, INC.	Consumer Staples
PEP	PEPSICO, INC.	Consumer Staples
PG	PROTECT & GAMBLE COMPANY	Consumer Staples
SJM	THE J. M. SMUCKER COMPANY	Consumer Staples
STZ	CONSTELATION BRANDS, INC.	Consumer Staples
SY	SYSCO CORPORATION	Consumer Staples
TAP	MOLSON COORS BEVERAGE COMPANY	Consumer Staples
TSN	TYSON FOODS, INC.	Consumer Staples
WBA	WALGREENS BOOTS ALLIANCE, INC.	Consumer Staples
WMT	WALMART, INC.	Consumer Staples
A	AGILENT TECHNOLOGIES, INC.	Health Care
ABC	AMERISOURCEBERGEN CORPORATION	Health Care
ABMD	ABIOMED, INC.	Health Care
ABT	ABBOTT LABORATORIES	Health Care
ALXN	ALEXION PHARMACEUTICAL, INC.	Health Care
AMGN	AMGEN, INC.	Health Care
BAX	BAXTER INTERNATIONAL, INC.	Health Care
BDX	BECTON DICKINSON AND COMPANY	Health Care
BIIB	BIOGEN, INC.	Health Care
BIO	BIO-RAD LABORATORIES	Health Care
BMY	BRISTOL-MYERS SQUIBB COMPANY	Health Care
BSX	BOSTON SCIENTIFIC CORPORATION	Health Care
CAH	CARDINAL HEALTH, INC.	Health Care
CERN	CERNER CORPORATION	Health Care
CI	CIGNA CORPORATION	Health Care
COO	COOPER COMPANIES, INC.	Health Care
CVS	CVS HEALTH CORPORATION	Health Care
DGX	QUEST DIAGNOSTICS, INC.	Health Care
DHR	DANAHER CORPORATION	Health Care
DVA	DAVITA, INC.	Health Care
GILD	GILEAD SCIENCES, INC.	Health Care
HOLX	HOLOGIC, INC.	Health Care
HSIC	HENRY SCHEIN, INC.	Health Care
HUM	HUMANA, INC.	Health Care
IDXX	IDEXX LABORATORIES, INC.	Health Care
INCY	INCYTE CORPORATION	Health Care
JNJ	JOHNSON & JOHNSON	Health Care
LH	LABORATORY CORPORATION OF AMERICA HOLDINGS	Health Care
LLY	ELI LILLY AND COMPANY	Health Care
MCK	MCKESSON CORPORATION	Health Care
MDT	MEDTRONIC PLC	Health Care
MRK	MERCK & COMPANY, INC.	Health Care

Ticker	Compañía	Sector
MTD	METTER-TOLEDO INTERNATIONAL, INC.	Health Care
MYL	MYLAN N.V.	Health Care
PFE	PFIZER, INC.	Health Care
PKI	PERKINELMER, INC.	Health Care
PRGO	PERRIGO COMPANY PLC	Health Care
REGN	REGENERON PHARMACEUTICALS, INC.	Health Care
RMD	RESMED, INC.	Health Care
STE	STERIS, INC.	Health Care
SYK	STRYKER CORPORATION	Health Care
TFX	TELEFLEX, INC.	Health Care
TMO	THERMO FISHER SCIENTIFIC, INC.	Health Care
UHS	UNIVERSAL HEALTH SERVICES, INC.	Health Care
UNH	UNITEDHEALTH GROUP, INC.	Health Care
VAR	VARIAN MEDICAL SYSTEMS, INC.	Health Care
VRTX	VERTEX PHARMACEUTICAL	Health Care
WAT	WATERS CORPORATION	Health Care
WST	WEST PHARMACEUTICAL SERVICES	Health Care
XRAY	DENTSPLY SIRONA, INC.	Health Care
ALK	ALASKA AIR GROUP, INC.	Industrials
AME	AMETEK, INC.	Industrials
AOS	A. O. SMITH CORPORATION	Industrials
BA	BOEING COMPANY	Industrials
CAT	CATERPILLAR, INC.	Industrials
CHRW	C.H. ROBINSON WORLDWIDE, INC.	Industrials
CMI	CUMMINS, INC.	Industrials
CPRT	COPART, INC.	Industrials
CSX	CSX CORPORATION	Industrials
CTAS	CINTAS CORPORATION	Industrials
DE	DEERE & COMPANY	Industrials
DOV	DOVER CORPORATION	Industrials
EFX	EQUIFAX, INC.	Industrials
EMR	EMERSON ELECTRIC COMPANY	Industrials
ETN	EATON CORPORATION PLC	Industrials
EXPD	EXPEDITORS INTERN OF WASHINGTON, INC.	Industrials
FAST	FASTENAL COMPANY	Industrials
FDX	FEDEX CORPORATION	Industrials
FLS	FLOWSERVE CORPORATION	Industrials
GD	GENERAL DYNAMICS	Industrials
GE	GENERAL ELECTRIC COMPANY	Industrials
GWW	W. W. GRAINGER, INC.	Industrials
HON	HONEYWELL INTERNATIONAL, INC.	Industrials
IEX	INDEX CORPORATION	Industrials
ITW	ILLINOIS TOOLS WORKS, INC.	Industrials
J	JACOBS ENGINEERING GROUP, INC.	Industrials
JBHT	J. B. HUNT TRANSPORT SERVICES, INC.	Industrials
JCI	JOHNSON CONTROLS INTERNATIONAL PLC	Industrials
KSU	KANSAS CITY SOUTHERN	Industrials
LMT	LOCKHEED MARTIN CORPORATION	Industrials
LUV	SOUTHWEST AIRLINES COMPANY	Industrials
MAS	MASCO CORPORATION	Industrials
MMM	3M COMPANY	Industrials
NOC	NORTHROP GRUMMAN CORPORATION	Industrials

Ticker	Compañía	Sector
NSC	NORFOLK SOUTHERN CORPORATION	Industrials
ODFL	OLD DOMINION FREIGHT LINE, INC.	Industrials
PCAR	PACCAR, INC.	Industrials
PH	PARKER-HANNIFIN CORPORATION	Industrials
PNR	PENTAIR PLC	Industrials
PWR	QUANTA SERVICES	Industrials
RHI	ROBERT HALF INTERNATIONAL, INC.	Industrials
ROK	ROCKWELL AUTOMATION, INC.	Industrials
ROL	ROLLINS, INC.	Industrials
ROP	ROPER TECHNOLOGIES, INC.	Industrials
RSG	REPUBLIC SERVICES, INC.	Industrials
RTX	RAYTHEON TECHNOLOGIES CORPORATION	Industrials
SNA	SNAP-ON, INC.	Industrials
SWK	STANLEY BLACK & DECKER, INC.	Industrials
TDY	TELEDYNE TECHNOLOGIES, INC.	Industrials
TXT	TEXTRON, INC.	Industrials
UNP	UNION PACIFIC CORPORATION	Industrials
UPS	UNITED PARCEL SERVICE, INC.	Industrials
URI	UNITED RENTALS, INC.	Industrials
WAB	WABTEC CORPORATION	Industrials
WM	WASTE MANAGEMENT, INC.	Industrials
AAPL	APPLE, INC.	Information Technology
ADBE	ADOBE, INC.	Information Technology
ADI	ANALOG DEVICES, INC.	Information Technology
ADP	AUTOMATIC DATA PROCESSING, INC.	Information Technology
ADSK	AUTODESK, INC.	Information Technology
AKAM	AKAMI TECHNOLOGIES	Information Technology
AMAT	APPLIED MATERIALS, INC.	Information Technology
AMD	ADVANCED MICRO DEVICES, INC.	Information Technology
ANSS	ANSYS, INC.	Information Technology
APH	AMPHENOL CORPORATION	Information Technology
CDNS	CADENCE DESIGN SYSTEMS, INC.	Information Technology
CSCO	CISCO SYSTEMS, INC.	Information Technology
CTSH	COGNIZANT TECHNOLOGY SOLUTION CORPORATION	Information Technology
CTXS	CITRIX SYSTEMS, INC.	Information Technology
FFIV	F5 NETWORKS, INC.	Information Technology
FISV	FISERV, INC.	Information Technology
FLIR	FLIR SYSTEMS, INC.	Information Technology
GLW	CORNING, INC.	Information Technology
HPQ	HP, INC.	Information Technology
IBM	INTERNATIONAL BUSINESS MACHINE CORPORATION	Information Technology
INTC	INTEL CORPORATION	Information Technology
INTU	INTUIT, INC.	Information Technology
IT	GARTNER, INC.	Information Technology
JKHY	JACK HENRY & ASSOCIATES, INC.	Information Technology
JNPR	JUNIPER NETWORKS	Information Technology
KLAC	KLA CORPORATION	Information Technology
LRCX	LAM RESEARCH CORPORATION	Information Technology
MCHP	MICROCHIP TECHNOLOGY	Information Technology
MSFT	MICROSOFT CORPORATION	Information Technology
MSI	MOTOROLA SOLUTIONS, INC.	Information Technology
MU	MICRON TECHNOLOGY, INC.	Information Technology

Ticker	Compañía	Sector
MXIM	MAXIM INTEGRATED	Information Technology
NLOK	NORTONLIFELOCK, INC.	Information Technology
NTAP	NETAPP, INC.	Information Technology
NVDA	NVIDIA CORPORATION	Information Technology
ORCL	ORACLE CORPORATION	Information Technology
PAYX	PAYCHECKX, INC.	Information Technology
QCOM	QUALCOMM, INC.	Information Technology
SNPS	SYNOPSYS, INC.	Information Technology
SWKS	SKYWORKS SOLUTIONS, INC.	Information Technology
TER	TERADYNE, INC.	Information Technology
TXN	TEXAS INSTRUMENTS, INC.	Information Technology
TYL	TYLER TECHNOLOGIES	Information Technology
VRSN	VERISIGN, INC.	Information Technology
WU	THE WESTERN UNION COMPANY	Information Technology
XLNX	XILINX, INC.	Information Technology
XRX	XEROX HOLDINGS CORPORATION	Information Technology
ZBRA	ZEBRA TECHNOLOGIES	Information Technology
AMCR	AMCOR PLC.	Materials
APD	AIR PRODUCTS & CHEMICALS, INC.	Materials
AVY	AVERY DENNISON CORPORATION	Materials
DOW	DOW, INC.	Materials
ECL	ECOLAB, INC.	Materials
EMN	EASTMAN CHEMICAL COMPANY	Materials
FCX	FREEMPORT-MCMORAN, INC.	Materials
FMC	FMC CORPORATION	Materials
IFF	INTERNATIONAL FLAVORS & FRAGRANCES, INC.	Materials
IP	INTERNATIONAL PAPER COMPANY	Materials
LYB	LYONDELLBASELL INDUSTRIES N.V.	Materials
MLM	MARTIN MARIETTA MATERIALS, INC.	Materials
MOS	MOSAIC COMPANY	Materials
NEM	NEWMONT CORPORATION	Materials
PKG	PACKAGING CORP OF AMERICA	Materials
PPG	PPG INDUSTRIES, INC.	Materials
SEE	SEALED AIR CORPORATION	Materials
SHW	SHERWIN.WILLIAMS COMPANY	Materials
WRK	WESTROCK COMPANY	Materials
AIV	APARTMENT INVESTMENT AND MANAGEMENT COMPANY	Real Estate
AMT	AMERICAN TOWER CORPORATION	Real Estate
ARE	ALEXANDRIA REAL ESTATE EQUITIES	Real Estate
AVB	AVALONBAY COMMUNITIES, INC.	Real Estate
CBRE	CBRE GROUP, INC.	Real Estate
DLR	DIGITAL REALTY TRUST, INC.	Real Estate
EQIX	EQUINIX, INC.	Real Estate
EQR	EQUITY RESIDENTIAL	Real Estate
EXR	EXTRA SPACE STORAGE, INC.	Real Estate
FRT	FEDERAL REALTY INVESTMENT TRUST	Real Estate
HST	HOST MARRIOTT FINANCIAL TRUST	Real Estate
IRM	IRON MOUNTAIN, INC.	Real Estate
KIM	KIMCO REALTY CORPORATION	Real Estate
MAA	MID-AMERICA APARTMENT COMMUNITIES	Real Estate
O	REALTY INCOME CORPORATION	Real Estate
PEAK	HEALTHPEAK PROPERTIES, INC.	Real Estate

Ticker	Compañía	Sector
PLD	PROLOGIS, INC.	Real Estate
PSA	PUBLIC STORAGE	Real Estate
REG	REGENCY CENTERS CORP	Real Estate
SBAC	SBA COMMUNICATIONS CORPORATION	Real Estate
SLG	SL GREEN REALTY CORPORATION	Real Estate
SPG	SIMON PROPERTY GROUP, INC.	Real Estate
UDR	UDR, INC.	Real Estate
VNO	VORNADO REALTY TRUST	Real Estate
VTR	VENTAS, INC.	Real Estate
WELL	WELLTOWER, INC.	Real Estate
WY	WEYERHAEUSER COMPANY	Real Estate
CHTR	CHARTER COMMUNICATIONS, INC.	Communication Services
CMCSA	COMCAST CORPORATION	Communication Services
DISCK	DISCOVERY, INC.	Communication Services
DISH	DISH NETWORK CORPORATION	Communication Services
FB	FACEBOOK, INC.	Communication Services
FOX	FOX CORPORATION	Communication Services
GOOGL	ALPHABET, INC.	Communication Services
IPG	INTERPUBLIC GROUP OF COMPANIES	Communication Services
NWSA	NEWS CORPORATION	Communication Services
OMC	OMNICOM GROUP, INC.	Communication Services
TMUS	T-MOBILE US	Communication Services
VIAC	VIACOMCBS, INC.	Communication Services
VZ	VERIZON COMMUNICATIONS, INC.	Communication Services
AEE	AMEREN CORPORATION	Utilities
AEP	AMERICAN ELECTRIC POWER COMPANY	Utilities
AES	AES CORPORATION	Utilities
AWK	AMERICAN WATER WORKS COMPANY, INC.	Utilities
CMS	CMS ENERGY CORPORATION	Utilities
CNP	CENTERPOINT ENERGY, INC.	Utilities
D	DOMINION ENERGY, INC.	Utilities
DTE	DTE ENERGY COMPANY	Utilities
DUK	DUKE ENERGY CORPORATION	Utilities
ED	CONSOLIDATED EDISON, INC.	Utilities
ES	EVERSOURCE ENERGY	Utilities
ETR	ENTERGY CORPORATION	Utilities
EVRG	EVERGY, INC.	Utilities
EXC	EXELON CORPORATION	Utilities
FE	FIRSTENERGY CORPORATION	Utilities
LNT	ALLIANT ENERGY CORPORATION	Utilities
NEE	NEXTERA ENERGY, INC.	Utilities
NRG	NRG ENERGY, INC.	Utilities
PEG	PUBLIC SERVICE ENTERPRISE GROUP, INC.	Utilities
PNW	PINACLE WEST CAPITAL CORPORATION	Utilities
PPL	PPL CORPORATION	Utilities
SO	SOUTHERN COMPANY	Utilities
SRE	SEMPRA ENERGY	Utilities
WEC	WEC ENERGY GROUP, INC.	Utilities
XEL	XCEL ENERGY, INC.	Utilities
AFL	AFLAC, INC.	Financials
AIZ	ASSURANT, INC.	Financials
AJG	ARTHUR J. GALLAGHER & CO.	Financials

Ticker	Compañía	Sector
AMP	AMERIPRISE FINANCIAL, INC.	Financials
AON	AON PLC	Financials
AXP	AMERICAN EXPRESS COMPANY	Financials
BAC	BANK OF AMERICA CORPORATION	Financials
BEN	FRANKLIN RESOURCES, INC.	Financials
BK	THE BANK OF NEW YORK MELLON CORPORATION	Financials
BLK	BLACKROCK, INC.	Financials
BRK_B	BERKSHIRE HATHAWAY, INC.	Financials
C	CITIGROUP, INC.	Financials
CFG	CITIZENS FINANTIAL GROUP, INC.	Financials
CINF	CINCINNATI FINANCIAL CORPORATION	Financials
CME	CME GROUP, INC.	Financials
DFS	DISCOVER FINANCIAL SERVICES	Financials
FRC	FIRST REPUBLIC BANK	Financials
GL	GLOBE LIFE, INC.	Financials
GS	GOLDMAN SACHS GROUP, INC.	Financials
HBAN	HUNTINGTON BANCSHARES, INC.	Financials
ICE	INTERCONTINENTAL EXCHANGE, INC.	Financials
IVZ	INVESCO PLC	Financials
JPM	JPMORGAN CHASE & CO.	Financials
KEY	KEYCORP	Financials
L	LOEWS CORPORATION	Financials
LNC	LINCOLN NATIONAL CORPORATION	Financials
MKTX	MARKETAXESS HOLDINGS	Financials
MMC	MARSH & MCLENNAN COMPANIES, INC.	Financials
MSCI	MSCI, INC.	Financials
NDAQ	NASDAQ, INC.	Financials
NTRS	NORTHERN TRUST CORPORATION	Financials
PFG	PRINCIPAL FINANCIAL GROUP	Financials
PGR	THE PROGRESSIVE CORPORATION	Financials
PRU	PRUDENTIAL FINANCIAL CORPORATION	Financials
RE	EVEREST RE GROUP	Financials
RF	REGIONS FINANCIAL CORPORATION	Financials
RJF	RAYMOND JAMES FINANCIAL	Financials
SCHW	THE CHARLESSCHWAB CORPORATION	Financials
SPGI	S&P GLOBAL, INC.	Financials
SYF	SYNCHRONY FINANCIAL	Financials
TFC	TRUIST FINANCIAL CORPORATION	Financials
TROW	PRICE T ROWE GROUPS	Financials
TRV	TRAVELERS COMPANIES, INC.	Financials
UNM	UNUM GROUP	Financials
USB	U.S. BANCORP	Financials
WLTW	WILLIS TOWERS WATSON PLC	Financials
WRB	W.R. BERKLEY CORPORATION	Financials
ZION	ZIONS BANCORPORATION N. A.	Financials
APA	APACHE GROUP	Energy
BKR	BAKER HUGHES COMPANY	Energy
COG	CABOT OIL & GAS CORPORATION	Energy
COP	CONOCOPHILLIPS	Energy
CXO	CONCHO RESOURCES, INC.	Energy
DVN	DEVON ENERGY CORPORATION	Energy
FTI	TECHNIPFMC PLC	Energy

Ticker	Compañía	Sector
HAL	HALLIBURTON COMPANY	Energy
HES	HESS CORPORATION	Energy
MPC	MARATHON PETROLEUM CORPORATION	Energy
MRO	MARATHON OIL CORPORATION	Energy
NOV	NATIONAL OILWELL VARCO, INC.	Energy
OKE	ONEOK, INC.	Energy
PSX	PHILLIPS 66	Energy
PXD	PIONEER NATURAL RESOURCES COMPANY	Energy
SLB	SCHLUMBERGER LIMITED	Energy
VLO	VALERO ENERGY CORPORATION	Energy
WMB	WILLIAMS COMPANIES, INC.	Energy
XOM	EXXON MOBIL CORPORATION	Energy

Tabla A.2: Compañías constituyentes del índice del S&P 500 empleadas, divididas por sector (en inglés). Un total de 374 acciones.

Ponderación Inicial del Índice S&P 500

Sector	Símbolo	compañías	Ponderación
Consumo Discrecional	CD	61	0.121
Consumo Básico	CS	32	0.063
Salud	HC	63	0.125
Industrial	IN	73	0.145
Tecnologías de la Información	IT	73	0.145
Materiales	MA	28	0.055
Inmobiliario	RE	31	0.061
Servicios de Comunicación	TS	26	0.051
Servicios Públicos	UT	28	0.055
Financiero	FI	65	0.129
Energía	EN	25	0.050
Total		505	1.000

Tabla A.3: Distribución entre sectores (en inglés) constituyentes del índice S&P 500.

A.3. Periodo Temporal por Época

Época	Fechas	Época	Fechas	Época	Fechas
1	01/03/00-02/29/00	45	01/04/07-03/02/07	89	12/31/13-02/27/14
2	03/01/00-04/26/00	46	03/05/07-04/30/07	90	02/28/14-04/25/14
3	04/27/00-06/22/00	47	05/01/07-06/26/07	91	04/28/14-06/23/14
4	06/23/00-08/18/00	48	06/27/07-08/22/07	92	06/24/14-08/19/14
5	08/21/00-10/16/00	49	08/23/07-10/18/07	93	08/20/14-10/15/14
6	10/17/00-12/12/00	50	10/19/07-12/14/07	94	10/16/14-12/11/14
7	12/13/00-02/09/01	51	12/17/07-02/13/08	95	12/12/14-02/10/15
8	02/12/01-04/09/01	52	02/14/08-04/11/08	96	02/11/15-04/09/15
9	04/10/01-06/06/01	53	04/14/08-06/09/08	97	04/10/15-06/05/15
10	06/07/01-08/02/01	54	06/10/08-08/05/08	98	06/08/15-08/03/15
11	08/03/01-10/04/01	55	08/06/08-10/01/08	99	08/04/15-09/29/15
12	10/05/01-11/30/01	56	10/02/08-11/26/08	100	09/30/15-11/24/15
13	12/03/01-01/30/02	57	11/28/08-01/27/09	101	11/25/15-01/25/16
14	01/31/02-03/28/02	58	01/28/09-03/25/09	102	01/26/16-03/22/16
15	04/01/02-05/24/02	59	03/26/09-05/21/09	103	03/23/16-05/18/16
16	05/28/02-07/23/02	60	05/22/09-07/20/09	104	05/19/16-07/15/16
17	07/24/02-09/18/02	61	07/21/09-09/15/09	105	07/18/16-09/12/16
18	09/19/02-11/13/02	62	09/16/09-11/10/09	106	09/13/16-11/07/16
19	11/14/02-01/13/03	63	11/11/09-01/08/10	107	11/08/16-01/05/17
20	01/14/03-03/12/03	64	01/11/10-03/09/10	108	01/06/17-03/06/17
21	03/13/03-05/08/03	65	03/10/10-05/05/10	109	03/07/17-05/02/17
22	05/09/03-07/07/03	66	05/06/10-07/01/10	110	05/03/17-06/28/17
23	07/08/03-09/02/03	67	07/02/10-08/27/10	111	06/29/17-08/24/17
24	09/03/03-10/28/03	68	08/30/10-10/25/10	112	08/25/17-10/20/17
25	10/29/03-12/24/03	69	10/26/10-12/21/10	113	10/23/17-12/18/17
26	12/26/03-02/24/04	70	12/22/10-02/17/11	114	12/19/17-02/15/18
27	02/25/04-04/21/04	71	02/18/11-04/15/11	115	02/16/18-04/16/18
28	04/22/04-06/18/04	72	04/18/11-06/14/11	116	04/17/18-06/12/18
29	06/21/04-08/16/04	73	06/15/11-08/10/11	117	06/13/18-08/08/18
30	08/17/04-10/12/04	74	08/11/11-10/06/11	118	08/09/18-10/04/18
31	10/13/04-12/08/04	75	10/07/11-12/02/11	119	10/05/18-11/30/18
32	12/09/04-02/04/05	76	12/05/11-02/01/12	120	12/03/18-01/31/19
33	02/07/05-04/05/05	77	02/02/12-03/29/12	121	02/01/19-03/29/19
34	04/06/05-06/01/05	78	03/30/12-05/25/12	122	04/01/19-05/28/19
35	06/02/05-07/28/05	79	05/29/12-07/24/12	123	05/29/19-07/24/19
36	07/29/05-09/23/05	80	07/25/12-09/19/12	124	07/25/19-09/19/19
37	09/26/05-11/18/05	81	09/20/12-11/16/12	125	09/20/19-11/14/19
38	11/21/05-01/19/06	82	11/19/12-01/16/13	126	11/15/19-01/14/20
39	01/20/06-03/17/06	83	01/17/13-03/15/13	127	01/15/20-03/12/20
40	03/20/06-05/15/06	84	03/18/13-05/13/13	128	03/13/20-05/08/20
41	05/16/06-07/12/06	85	05/14/13-07/10/13	129	05/11/20-07/07/20
42	07/13/06-09/07/06	86	07/11/13-09/05/13	130	07/08/20-09/01/20
43	09/08/06-11/02/06	87	09/06/13-10/31/13		
44	11/03/06-01/03/07	88	11/01/13-12/30/13		

Tabla A.4: Distribución de los días hábiles del mercado estadounidense por época.

A.4. Datos de Clustering

Pearson		Spearman		Distancia	
Época	CCP	Época	CCP	Época	CCP
7	0.1321	130	0.1914	29	0.3732
107	0.1279	108	0.1685	33	0.3710
117	0.1173	107	0.1678	30	0.3690
1	0.1137	126	0.1639	36	0.3647
113	0.1053	9	0.1631	32	0.3642
3	0.1035	118	0.1558	107	0.3639
6	0.1003	117	0.1505	130	0.3624
5	0.0835	2	0.1504	8	0.3618
112	0.0759	10	0.1450	23	0.3603
4	0.0622	113	0.1342	25	0.3572
		7	0.1272	116	0.3571
		1	0.1258	91	0.3549
		3	0.1158	123	0.3546
		6	0.1056	109	0.3538
		5	0.0821	15	0.3536
		112	0.0818	38	0.3523
		4	0.0675	31	0.3518
				111	0.3470
				70	0.3470
				110	0.3455
				44	0.3431
				35	0.3413
				108	0.3400
				12	0.3373
				14	0.3371
				126	0.3366
				40	0.3360
				117	0.3353
				13	0.3317
				2	0.3310
				43	0.3305
				118	0.3297
				9	0.3283
				39	0.3264
				113	0.3253
				7	0.3248
				26	0.3222
				10	0.3179
				6	0.3157
				112	0.3091
				3	0.3062
				1	0.3041
				5	0.3025
				4	0.3000

Tabla A.5: Épocas dentro de la agrupación o *cluster* con los coeficientes de correlación promedio (CPP) más bajos.

Bibliografía

- [1] Rosario N Mantegna and H Eugene Stanley. *Introduction to econophysics: correlations and complexity in finance*. Cambridge university press, 1999.
- [2] Leonidas Sandoval Junior and Italo De Paula Franca. Correlation of financial markets in times of crisis. *Physica A: Statistical Mechanics and its Applications*, 391(1-2):187–208, 2012.
- [3] Hirdesh K Pharasi, Kiran Sharma, Rakesh Chatterjee, Anirban Chakraborti, Francois Leyvraz, and Thomas H Seligman. Identifying long-term precursors of financial market crashes using correlation patterns. *New Journal of Physics*, 20(10):103041, 2018.
- [4] Michael C Münnix, Takashi Shimada, Rudi Schäfer, Francois Leyvraz, Thomas H Seligman, Thomas Guhr, and H Eugene Stanley. Identifying states of a financial market. *Scientific reports*, 2(1):1–6, 2012.
- [5] Denys Maslov and Oleg Rytchkov. Ranking stocks and returns: A non-parametric analysis of asset pricing anomalies, 2010.
- [6] Richard A Davis, Muneya Matsui, Thomas Mikosch, and Phyllis Wan. Applications of distance correlation to time series. *Bernoulli*, 24(4A):3087–3116, 2018.
- [7] Jeff Madura, Franco Modigliani, and Michael G Ferri. *Mercados e Instituciones Financieras*. Cengage Learning, 2010.
- [8] Cheng-Few Lee and Alice C Lee. *Encyclopedia of Finance*. Springer, 2013.
- [9] Mandelbrot and Richard L Hudson. *The Misbehavior of Markets: A Fractal View of Financial Turbulence*. Basic Books, 2006.
- [10] Ray Ball. The global financial crisis and the efficient market hypothesis: what have we learned? *Journal of Applied Corporate Finance*, 21(4):8–16, 2009.
- [11] Burton G Malkiel. The efficient market hypothesis and its critics. *Journal of economic perspectives*, 17(1):59–82, 2003.
- [12] Frederic S Mishkin. Anatomy of a financial crisis. *Journal of evolutionary Economics*, 2(2):115–130, 1992.
- [13] Juan Torres López. Crisis financiera. algunas reflexiones críticas sobre sus orígenes y la manera de analizarla. *Proyección: Teología y mundo actual*, (235):371–388, 2009.

- [14] Markus K Brunnermeier and Martin Oehmke. Bubbles, financial crises, and systemic risk. *Handbook of the Economics of Finance*, 2:1221–1288, 2013.
- [15] Alexander Ljungqvist and William J Wilhelm Jr. Ipo pricing in the dot-com bubble. *The Journal of Finance*, 58(2):723–752, 2003.
- [16] Peter CB Phillips and Jun Yu. Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics*, 2(3):455–491, 2011.
- [17] S&P Dow Jones Indices. S&p 500, Septiembre 2020. Consultado el 10 de octubre del 2020. <https://www.spglobal.com/spdji/es/indices/equity/sp-500/overview>.
- [18] James Douglas Hamilton. *Time series analysis*. Princeton university press, 2020.
- [19] Christoly Biely and Stefan Thurner. Random matrix ensembles of time-lagged correlation matrices: derivation of eigenvalue spectra and analysis of financial time-series. *Quantitative Finance*, 8(7):705–722, 2008.
- [20] Rudi Schäfer and Thomas Guhr. Local normalization: Uncovering correlations in non-stationary financial time series. *Physica A: Statistical Mechanics and its Applications*, 389(18):3856–3865, 2010.
- [21] Ruey S. Tsay. *Analysis of Financial Time Series*. John wiley & sons, 2005.
- [22] Vasiliki Plerou, Parameswaran Gopikrishnan, Bernd Rosenow, Luís A Nunes Amaral, and H Eugene Stanley. Universal and nonuniversal properties of cross correlations in financial time series. *Physical review letters*, 83(7):1471, 1999.
- [23] Manan Vyas, T Guhr, and TH Seligman. Multivariate analysis of short time series in terms of ensembles of correlation matrices. *Scientific reports*, 8(1):1–12, 2018.
- [24] J.L. Devore and P.S. Gómez. *Probabilidad Y Estadística Para Ingeniería Y Ciencias*. CENGAGE Learning, 2011.
- [25] M.Á.G. Álvarez. *Introducción a la teoría de la probabilidad II. Segundo curso*. Ciencia y Tecnología. FCE - Fondo de Cultura Económica, 2015.
- [26] Oleg Shirokikh, Grigory Pastukhov, Vladimir Boginski, and Sergiy Butenko. Computational study of the us stock market evolution: a rank correlation-based network model. *Computational Management Science*, 10(2-3):81–103, 2013.
- [27] Gábor J Székely, Maria L Rizzo, and Nail K Bakirov. Measuring and testing dependence by correlation of distances. *The annals of statistics*, 35(6):2769–2794, 2007.
- [28] Maria Pitsillou and Konstantinos Fokianos. dcovts: Distance covariance/correlation for time series. *R J.*, 8(2):324, 2016.
- [29] Dominic Edelman, Tamás F Móri, and Gábor J Székely. On relationships between the pearson and the distance correlation coefficients. *Statistics & Probability Letters*, 169:108960, 2021.

- [30] A. M. Mood. *Introduction to the Theory of Statistics*. McGraw-hill, 1950.
- [31] L. Rincon. *Introducción a la probabilidad*. Temas selectos de matemáticas. Universidad Nacional Autónoma de México, Facultad de Ciencias, 2014.
- [32] A. Stuart and K. Ord. *Kendall's Advanced Theory of Statistics, Distribution Theory*. Kendall's Advanced Theory of Statistics. Wiley, 2010.
- [33] Vasiliki Plerou, Parameswaran Gopikrishnan, Bernd Rosenow, Luis A Nunes Amaral, Thomas Guhr, and H Eugene Stanley. Random matrix approach to cross correlations in financial data. *Physical Review E*, 65(6):066126, 2002.
- [34] Thomas Guhr, Axel Müller-Groeling, and Hans A Weidenmüller. Random-matrix theories in quantum physics: common concepts. *Physics Reports*, 299(4-6):189–425, 1998.
- [35] Tomás A Brody, Jorge Flores, J Bruce French, PA Mello, A Pandey, and Samuel SM Wong. Random-matrix physics: spectrum and strength fluctuations. *Reviews of Modern Physics*, 53(3):385, 1981.
- [36] Yan V Fyodorov and Alexander D Mirlin. Level-to-level fluctuations of the inverse participation ratio in finite quasi 1d disordered systems. *Physical review letters*, 71(3):412, 1993.
- [37] Quandl, Octubre 2020. Consultado el 1 de octubre del 2020. <https://www.quandl.com/>.
- [38] The Globe and Mail. Canadian, world, politics and business news and analysis, Octubre 2020. Consultado el 11 de octubre del 2020. <https://www.theglobeandmail.com/>.
- [39] Tim Edwards, Craig J Lazzara, Hamish Preston, and Francesca B Pipino. Aplicaciones globales de los sectores del s&p 500, Marzo 2018. Consultado el 10 de octubre del 2020. <https://www.spglobal.com/spdji/es/documents/research/research-global-applications-of-sp-500-sectors-spa.pdf>.
- [40] Phillip Murphy and Mahavir Kaswa. Sectores del s&p 500: ponderación equitativa y por capitalización, Marzo 2016. Consultado el 10 de octubre del 2020. <https://www.spglobal.com/spdji/es/documents/education/practice-essentials-sp-500-sectors-equal-weight-and-cap-weig-sp.pdf>.
- [41] S&P Dow Jones Indices. S&p corelogic case-shiller 10-city composite home price nsa, Octubre 2020. Consultado el 21 de noviembre del 2020. <https://www.spglobal.com/spdji/en/index-family/indicators/sp-corelogic-case-shiller/sp-corelogic-case-shiller-composite/overview>.
- [42] A. Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2019.
- [43] Mark Newman. *Networks*. Oxford university press, 2018.
- [44] S.H. Friedberg, A.J. Insel, and L.E. Spence. *Linear Algebra*. Featured Titles for Linear Algebra (Advanced) Series. Pearson Education, 2003.

- [45] K.W. Gruenberg and A.J. Weir. *Linear Geometry*. Graduate Texts in Mathematics. Springer New York, 2013.
- [46] Hirdesh K Pharasi, Eduard Seligman, and Thomas H Seligman. Market states: A new understanding. *arXiv preprint arXiv:2003.07058*, 2020.
- [47] Hirdesh K Pharasi, Suchetana Sadhukhan, Parisa Majari, Anirban Chakraborti, and Thomas H Seligman. Dynamics of the market states in the space of correlation matrices with applications to financial markets. *arXiv preprint arXiv:2107.05663*, 2021.