



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS FÍSICAS  
FÍSICA ESTADÍSTICA Y SISTEMAS COMPLEJOS

SERIES DE TIEMPO FINANCIERAS EN EL MARCO DE LA FÍSICA  
ESTADÍSTICA

T E S I S

QUE PARA OPTAR POR EL GRADO DE:  
DOCTOR EN CIENCIAS (FÍSICA)

PRESENTA:  
MANUEL MIJAÍL MARTÍNEZ RAMOS

TUTORES PRINCIPALES:  
DR. THOMAS HENRY SELIGMAN SCHURCH  
INSTITUTO DE CIENCIAS FÍSICAS, UNAM  
DRA. MANAN VYAS  
INSTITUTO DE CIENCIAS FÍSICAS, UNAM

MIEMBROS DEL COMITÉ TUTOR:  
DR. CHRISTOF FRIEDRICH JUNG KOHL  
INSTITUTO DE CIENCIAS FÍSICAS, UNAM  
DR. DENIS PIERRE BOYER  
INSTITUTO DE FÍSICA, UNAM

MORELOS, MÉXICO, AGOSTO DE 2024



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**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.



# Índice general

<b>1. Introducción</b>	<b>3</b>
1.1. Antecedentes . . . . .	3
1.2. Física Estadística . . . . .	5
1.2.1. Series de tiempo en sistemas complejos . . . . .	8
1.3. Mercados financieros . . . . .	9
1.3.1. Índices . . . . .	12
<b>Objetivo General</b>	<b>13</b>
<b>2. Fundamentos teóricos</b>	<b>15</b>
2.1. Modelo de May . . . . .	15
2.2. Covarianza y correlación . . . . .	16
2.2.1. Matrices locales en el tiempo . . . . .	20
2.2.2. Análisis en componentes principales . . . . .	22
2.2.3. Matrices de correlación relativa y proyecciones de Guhr . . . . .	25
2.3. Agrupamiento y reducción dimensional . . . . .	30
2.3.1. Estados de mercado . . . . .	32
<b>3. Correlaciones relativas y proyecciones de Guhr</b>	<b>35</b>
3.1. Preliminares . . . . .	35
3.2. Correlación local en el tiempo . . . . .	38
3.3. Estados de mercado en el S&P 500 . . . . .	48

3.3.1. Estado de COVID . . . . .	57
3.3.2. Correlaciones proyectadas de Guhr . . . . .	59
<b>4. Correlaciones de grano grueso</b>	<b>67</b>
4.1. Generalidades de las matrices sectoriales de Guhr . . . . .	68
4.1.1. Estado del mercado . . . . .	68
4.1.2. Matrices de similaridad . . . . .	71
4.2. Estados de mercado . . . . .	74
4.2.1. S&P 500 . . . . .	74
4.2.2. Nikkei 225 . . . . .	75
<b>5. Conclusión</b>	<b>85</b>
5.1. Perspectivas . . . . .	86
5.1.1. Modelos de correlación con grafos . . . . .	86
5.1.2. Técnica de ensembles . . . . .	89
5.1.3. Comparación de mercados . . . . .	92
<b>Apéndices</b>	<b>93</b>
<b>A. Sobre los datos utilizados</b>	<b>95</b>
A.1. S&P 500 . . . . .	95
A.2. Nikkei 225 . . . . .	104
<b>B. Detalles metodológicos</b>	<b>111</b>
B.1. Lenguajes de programación y bibliotecas . . . . .	111
B.2. K-means y MDS . . . . .	111
B.2.1. K-means . . . . .	111
B.2.2. MDS . . . . .	113
B.3. Histogramas y estimación de densidades de probabilidad . . . . .	114
<b>C. Material suplementario</b>	<b>117</b>

C.1. Esbozo de las características de las matrices de Wishart . . . . .	117
C.2. Otras estadísticas de las matrices locales de correlación . . . . .	118
C.2.1. Proyecciones de Guhr . . . . .	118
C.3. Eigenvalores en el S&P 500 . . . . .	125
C.3.1. Proyecciones de Guhr . . . . .	125
C.4. Eigenvectores en el S&P 500 . . . . .	131
C.5. Dinámica de los estados . . . . .	133
C.5.1. Proyecciones de Guhr . . . . .	133
C.6. Figuras complementarias del MDS . . . . .	139
<b>Referencias</b>	<b>143</b>



# Agradecimientos

Investigación realizada gracias al Programa UNAM-PAPIIT AG101122: *Dinámica de sistemas complejos: desde mecánica celeste y mercados financieros hasta transporte y dispositivos cuánticos.*

Investigación realizada en el marco del proyecto CONAHCYT Fronteras 425854.

Agradezco al CONAHCYT la asignación de Beca Nacional.

Gracias a mi novia bella, Marmota del Huerto de Tecamachalco (Fany Marín), por su amor inconmensurable y su grácil perspicacia emocional.

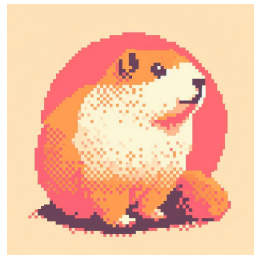


Figura 1: *Vivir nuestra individualidad es un reto que nos mantiene ocupados durante toda la vida.*  
Remo H. Largo.





# Capítulo 1

## Introducción

Les nuages ne sont pas des sphères, les montagnes ne sont pas des cônes, les côtes ne sont pas des cercles et les éclairs ne se déplacent pas en ligne droite.

Benoît Mandelbrot

En este capítulo daremos un panorama general de la influencia y utilidad de las ciencias físico-matemáticas en las finanzas. También introduciremos notación y terminología, y mencionaremos algunos marcos teóricos de referencia que nos permitirán alcanzar nuestro **Objetivo General**.

### 1.1. Antecedentes

La relación entre la economía y las *ciencias exactas* empieza hace al menos 160 años, cuando en *Calcul des chances et philosophie de la bourse* [1], el economista francés Jules Regnault, inspirado en el trabajo del matemático belga Adolphe Quételet sobre la distribución normal o *ley de los errores* [2], utilizó una caminata aleatoria simétrica para describir las fluctuaciones en precios de acciones en la Bolsa de París [3, cap. 1, págs. 2-4], cuyo registro de datos sistemático constituía una base para hacer pruebas estadísticas. Fue hasta la tesis doctoral de Louis Bachelier<sup>1</sup> de 1900, *Théorie de la spéculation* [4], que esta intuición se asentó matemáticamente [5, sec. 1.2]. En esta tesis, Bachelier estudió la probabilidad de que el precio de una acción alcanzara un valor dado a un tiempo fijo, o antes de un tiempo límite [3, cap. 1, pág. 5].

Además, si se toman en cuenta los trabajos de Adolphe Quételet y Auguste Comte sobre el *hombre promedio*, las leyes de la sociedad o la civilización, o los trabajos de Malthus y Verhulst sobre el crecimiento en sistemas económicos o sociales, a principios del siglo XIX; los estudios de Edmond Halley sobre la esperanza de vida y el cálculo subsecuente de pensiones, o la comisión de Isaac Newton en la *Royal Mint*<sup>2</sup> a finales del siglo XVII y principios del siglo XVIII; las

---

<sup>1</sup>Cabe resaltar que Bachelier consiguió su doctorado en el campo de la Física Matemática, bajo la tutela de Henri Poincaré [3, cap. I, pág. 4].

<sup>2</sup>La Royal Mint es la casa de moneda del Reino Unido: <https://www.royalmint.com/aboutus> (Accedida el 14 de

observaciones de Christiaan Huygens sobre las curvas de mortalidad (segunda mitad del siglo XVII); o incluso las contribuciones de Copérnico a la teoría del dinero (principios del siglo XVI), dicha relación entre la perspectiva física, matemática o lógica y los problemas del hombre en sociedad es mucho más antigua [3, cap. 1].

En el trabajo de Bachelier, se asume que los precios consecutivos de una acción son independientes, y que su evolución se puede separar en una parte determinista y otra aleatoria [6, cap. 3, pág. 64]. Ésta parte aleatoria estaría dada, usando una terminología actualizada, por un proceso estocástico  $\{Y_n\}_{n=1}^{\infty}$  definido mediante una secuencia de variables aleatorias  $\{X_i\}_{i=1}^{\infty}$  independientes e idénticamente distribuidas, con media  $\langle X_i \rangle = \mu$  común; por ejemplo:

1. Caminata aleatoria simple:  $Y_n := \sum_{i=1}^n (X_i - \mu)$ .
2. Caminata aleatoria geométrica:  $Y_n := \prod_{i=1}^n X_i$ , con las restricciones  $X_i > 0$  y  $\mu = 1$ .
3. Caminata aleatoria exponencial:  $Y_n := \exp[\lambda \sum_{i=1}^n X_i]$ , con las restricciones  $\lambda > 0$  y  $X_i \in \{-1, 1\}$ .

En los tres casos se satisface la siguiente relación sobre valor esperado condicional, la *propiedad de submartingala*:

$$\langle Y_{n+1} | Y_1, \dots, Y_n \rangle \geq Y_n$$

pero la igualdad siempre se cumple en los primeros dos casos (*propiedad de martingala*). Esta independencia del valor esperado futuro de una variable aleatoria, de la historia previa específica (o su limitación) por la que ha pasado, es la que se traduciría en una probabilidad igual de una variación consecutiva del precio de una acción de estar encima o debajo de su precio corriente. A su vez, esto es lo que seguramente Regnault y Bachelier habrían concebido como *la ley* de la variación de los precios de acciones bursátiles: ganancia promedio cero. En su tesis, Bachelier supuso que cada  $X_i$  es normal para tratar de calcular la probabilidad de que un precio alcanzara o rebasara un valor dado un tiempo específico, o bien, un tiempo límite [3, cap. 1, pág. 5].

El término *movimiento browniano* proviene de las observaciones detalladas, con ayuda de un microscopio, reportadas por el botánico escocés Robert Brown, entre 1827 y 1828, sobre el desplazamiento irregular de una partícula de polen inmersa en un fluido [7][8, cap. 5]. Tras casi 80 años, y utilizando los desarrollos de Termodinámica, Teoría Cinética y Física Estadística de William Thomson (Lord Kelvin), Rudolf Clausius, Josiah Willard Gibbs, James Clerk Maxwell, Ludwig Boltzmann y Joseph John Thomson, los científicos Albert Einstein (1905 [9]), Marian von Smoluchowski (1906 [10]) y Jean Perrin (1908 [11]) explicaron este fenómeno por las fuerzas moleculares<sup>3</sup> aleatorias, pequeñas pero con altas fluctuaciones, del líquido donde se suspende la partícula, cuyo desplazamiento cuadrático medio es tanto lineal en el tiempo como una función de su tamaño, la temperatura ambiental y la viscosidad del medio [13, cap. 1]. En la actualidad, el movimiento browniano es el ejemplo arquetípico para describir un fenómeno macroscópico u *observable*, como el efecto colectivo de eventos microscópicos o impredecibles [14, cap. 1]<sup>4</sup>. Como

---

noviembre de 2019).

<sup>3</sup>En ese tiempo, la estructura atómica de la materia todavía no constituía una teoría en consenso. Sin embargo, la teoría del movimiento browniano de Einstein-Smoluchowski y su verificación experimental por Perrin, con una estimación del número de Avogadro dentro del 19% del valor aceptado en la actualidad [12, sec. 4.3], más el trabajo de Joseph John Thompson sobre el electrón, fueron fundamentales para evidenciar la existencia del átomo [13, cap. 1].

<sup>4</sup>De hecho, extraer propiedades promedio de un sistema macroscópico a partir de la dinámica microscópica del sistema, está en el corazón de la Física Estadística [5].

proceso estocástico, puede caracterizarse como el límite de una caminata aleatoria, cuando la longitud del desplazamiento tiende a cero [15, cap. 3, sec. 1-3].

La teoría neoclásica de la economía, desde su formulación, ha tomado conceptos de la termodinámica en equilibrio. La idea principal es explicar el valor de un producto o servicio por la utilidad percibida por su consumidor [16], *quien racionalmente busca maximizar su satisfacción*<sup>5</sup>. Hablando en general, las teorías marginalistas asumen que se alcanza un equilibrio donde pueda igualarse el cambio de la utilidad al precio, dando lugar a un *valor fundamental*. Por ejemplo, en su tesis doctoral *Mathematical Investigations in the Theory for Value and Prices* [18], de 1891, Irving Fisher propuso que el precio en equilibrio de una mercancía está dado por su utilidad marginal, igual a su vez a la derivada de la utilidad respecto a la cantidad. Josiah Willard Gibbs, uno de los asesores doctorales y revisores de Fisher, le advirtió que tal afirmación no tenía fundamento, pues la utilidad *no es una función de estado sino un proceso* así que no tiene una diferencial exacta y, por lo tanto, no es una variable de estado [19].

En la segunda mitad del siglo pasado, en finanzas cuantitativas bajo las hipótesis neoclásicas y el uso central de la distribución normal, no siempre con respaldo empírico, se desarrollaron herramientas matemáticas sofisticadas sobre procesos estocásticos y difusión (v.gr. cálculo de Itô, movimiento browniano geométrico, martingalas, etc.) [5, 16] sobre la base del trabajo axiomático del matemático ruso Andrey Nikolaevich Kolmogórov (1933 [20])<sup>6</sup>, pero es a partir del trabajo de Benoît Mandelbrot (desde la década de 1960 hasta su fallecimiento) que se ponen en tela de juicio las bases empíricas de la teoría financiera. En la introducción de *The Misbehavior of Markets* [21], Mandelbrot escribe:

*[...] Conventional financial theory assumes that variation of prices can be modeled by random processes that, in effect, follow the simplest “mild” pattern, as if each uptick or downtick were determined by the toss of a coin. What fractals show [...] is that real prices “misbehave” very badly. A more accurate, multifractal model of wild price variation paves the way for a new, more reliable type of financial theory. Understanding fractally wild randomness, also exemplified by such diverse phenomena as turbulent flow, electrical “flicker” noise, and the track of a stock or bond price, will not bring personal wealth. But the fractal view of the market is alone in facing the high odds of catastrophic price changes.*

Justo los cambios abruptos de precios observados en los registros de las bolsas de valores, son incompatibles con la suposición de normalidad.

## 1.2. Física Estadística

De acuerdo con Ariel Caticha *et al* [22, 23], la Lógica y la Estadística comparten el objetivo de extraer conclusiones a partir de la información con la que se dispone; es decir, *inferir*. Si la

<sup>5</sup>Puede consultarse el artículo *Neoclassical economics*: <https://www.investopedia.com/terms/n/neoclassical.asp> (Accedida el 30 de noviembre de 2019). Sin embargo, cabe mencionar que esta teoría está superada desde hace al menos 50 años por la *Teoría de Juegos Evolucionaria*: asumir que los humanos actúan racionalmente en beneficio de cada individuo falla completamente, aunque funciona para explicar “la supervivencia del más apto” para especies [17].

<sup>6</sup>En el prefacio de *Foundations of the Theory of Probability* [20], Kolmogórov cita la teoría de la medida del matemático francés Henri Léon Lebesgue como fundamental para su propio trabajo.

valoración resultante de una afirmación es inequívoca (es decir, la información *es suficiente*), se trata de una inferencia deductiva; en cambio, si la información *es insuficiente*, entonces se habla de inferencia inductiva y de valoraciones con un grado de certeza.

En un artículo del *New York Times*[24], F. D. Flam escribe:

*Scientists who have learned Bayesian statistics often marvel that it propels them through a different kind of scientific reasoning than they had experienced using classical methods.*

De manera opuesta a la visión *frecuentista* de la probabilidad (dada por el número relativo de ocurrencias de un fenómeno dentro de un *ensemble* de ensayos), la interpretación bayesiana indica el grado de certeza o confianza de la validez de una afirmación, según un observador.

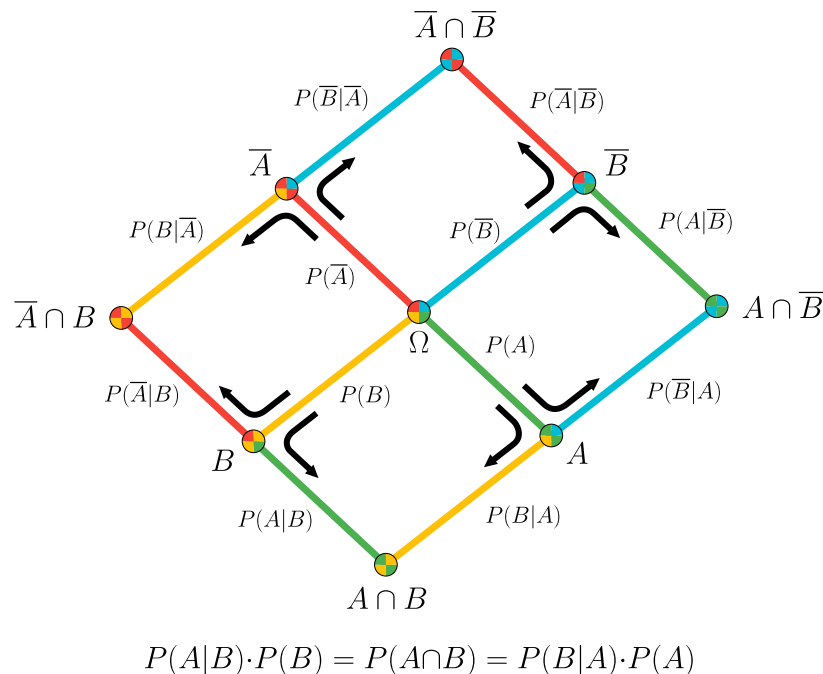


Figura 1.1: Árbol de decisión de dos eventos  $A$  y  $B$ . Tomada de: Qniemiec ([https://commons.wikimedia.org/wiki/File:Bayes'\\_Theorem\\_2D.png](https://commons.wikimedia.org/wiki/File:Bayes'_Theorem_2D.png))

Éstas interpretación pueden enmarcarse, a su vez, en las siguientes perspectivas [25, 26], respectivamente:

- Objetiva: donde la probabilidad de un evento es una propiedad inherente del mismo, y que es medible a través de la frecuencia de su ocurrencia en una experimentación aleatoria. Así, una distribución de probabilidad se hace verificable mediante la Estadística;
- Subjetiva: donde cada individuo puede asignar sus propias creencias a la luz de evidencia posiblemente razonable o de sus conocimientos previos, estableciendo así una medida de las *creencias racionales* o del *grado de ignorancia* a través de las probabilidades. Así, una

distribución de probabilidad refleja las conclusiones plausibles y acordes con la información disponible.

En ambos casos, los valores discretos de verdad  $\{0, 1\}$  de la Lógica Clásica se extienden a un continuo  $[0, 1]$  de probabilidades.

A su vez, la oposición entre la *primacía existencial* y la *primacía de la consciencia* guarda muchas semejanzas con los enfoques frecuentista y bayesiano de la probabilidad [27]; más aún: dentro del contexto de la Filosofía de la Ciencia, el falsacionismo empírico de Karl Popper, donde la validez de una afirmación es provisional mientras no se obtenga evidencia experimental que la refute (es decir: se puede poner a prueba una teoría y demostrar que es falsa, pero no se puede concluir que es correcta [28]), también es congruente con la actualización de una probabilidad *a priori*, dada una evidencia, a una probabilidad *a posteriori*:

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{k=1}^n \mathbb{P}(B|A_k)\mathbb{P}(A_k)}. \quad (1.1)$$

Cabe resaltar que esta expresión es independiente de su interpretación filosófica, y es el punto de partida teórico para expresar la *correlación parcial* o *relativa* (véase el Capítulo 2).

También la noción de actualización de probabilidades es compatible con el concepto de entropía. La idea de Edwin T. Jaynes<sup>7</sup> detrás de su presentación de la Mecánica Estadística como *una teoría de inferencia*, es plantear la distribución de probabilidad *a priori*  $\{p_i\}_{i=1}^n$ <sup>8</sup>, descriptiva de la solución de un problema, como aquella que maximiza la entropía:

$$H = - \sum_{i=1}^n p_i \ln p_i \quad (1.2)$$

sujeta a la consistencia con cualquier información inicial disponible sobre el fenómeno bajo estudio [25, 26], representada por un cierto conjunto de *valores empíricos*  $\{\langle f_r(x; \alpha_s) \rangle\}_{r=1}^m$  (posiblemente dependientes de parámetros  $\alpha_s$ ):

$$\langle f_r(x) \rangle = \sum_{i=1}^n p_i f_r(x_i) \quad r = 1, \dots, m. \quad (1.3)$$

El criterio de maximizar la entropía surge de que ésta es, módulo una constante multiplicativa, la única función no negativa, aditiva para fuentes independientes de incertidumbre, recursiva<sup>9</sup>, invariante bajo permutaciones de  $\{p_i\}_{i=1}^n$ , y monótona creciente con el número de estados del sistema  $n$ , dando entonces una estimación de la distribución de probabilidad con el menor sesgo posible acorde con la información disponible [29, cap. 1]. Citando a Jaynes [25, secs. 2 y 4]:

*[...] in making inferences on the basis of partial information we must use that probability distribution which has maximum entropy subject to whatever is known [...]*

<sup>7</sup>Cuyo trabajo precede al de Ariel Caticha.

<sup>8</sup>Tal que  $\forall i : p_i \geq 0$  y  $\sum_{i=1}^n p_i = 1$ .

<sup>9</sup> $\forall n \geq 3 : H_n(p_1, p_2, \dots, p_n) = H_{n-1}(p_1 + p_2, p_3, \dots, p_n) + (p_1 + p_2) H_2\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right)$

*Entropy as a concept may be regarded as a measure of our degree of ignorance as to the state of a system [...]*

La solución al planteamiento (1.2)-1.3 está dada por el método de los multiplicadores de Lagrange:

$$p_i = \frac{e^{-[\lambda_1 f_1(x) + \dots + \lambda_m f_m(x)]}}{Z(\lambda_1, \dots, \lambda_m)} \quad (1.4)$$

donde  $Z(\lambda_1, \dots, \lambda_m) := \sum_{i=1}^n e^{-[\lambda_1 f_1(x_i) + \dots + \lambda_m f_m(x_i)]}$  es la función de partición, y los multiplicadores  $\{\lambda_r\}_{r=1}^m$  están dados por las condiciones:

$$\begin{aligned} \langle f_r(x) \rangle &= -\frac{\partial}{\partial \lambda_r} \ln Z(\lambda_1, \dots, \lambda_m); \\ \left\langle \frac{\partial f_r(x)}{\partial \alpha_s} \right\rangle &= -\frac{1}{\lambda_r} \frac{\partial}{\partial \alpha_s} \ln Z(\lambda_1, \dots, \lambda_m), \quad r = 1, \dots, m. \end{aligned} \quad (1.5)$$

Cabe notar que la función de partición también puede considerarse como otro multiplicador, a través de la relación  $\lambda_0 := \ln(Z)$ . Con este esquema, la entropía máxima (1.2) resultante es  $H = \lambda_0 + \lambda_1 \langle f_1(x) \rangle + \dots + \lambda_m \langle f_m(x) \rangle$ .

Otra consecuencia es que cada situación que no está definitivamente excluida por la información disponible tendrá una (densidad de) probabilidad positiva, y en que las leyes de la Física sólo entran en juego para enumerar los diferentes estados posibles y mutuamente excluyentes en los que puede encontrarse el sistema [25]. Un asunto aparte, es el problema de explicar lo observado en un sistema macroscópico, como la causa de un fenómeno microscópico, para lo que uno debe valerse de principios físicos, como la conservación de la energía.

### 1.2.1. Series de tiempo en sistemas complejos

Las series de tiempo son a menudo el punto de partida para estudiar sistemas complejos, pensándolos como *cajas negras*. Ellas describen la evolución temporal de una o más variables de un sistema, y son particularmente útiles para el estudio de sus atractores, y el análisis estadístico de la sincronización [30, cap. 1].

Gracias al amplio registro de información, los mercados financieros son, como objeto de estudio, un ejemplo arquetípico de sistema complejo. La accesibilidad de los datos cambia con el tipo de activo, la frecuencia o el intervalo de tiempo de interés. Por ejemplo, en *Yahoo Finance* están disponibles de forma gratuita (a la fecha de elaboración de esta tesis) los precios de apertura, cierre (crudo y ajustado<sup>10</sup>), máximo, mínimo y volumen de cada día bursátil, para la mayoría de los mercados de valores más importantes. También los precios de criptomonedas, o de sus libros de órdenes, son de fácil acceso. Por otro lado, los precios de acciones *intradía* (con

<sup>10</sup>El precio de cierre (*crudo*) de una acción es el último precio al que se negoció durante el día de operaciones regulares. Es el punto de referencia estándar utilizado por los inversionistas para rastrear el rendimiento de una acción con el tiempo. Sin embargo, el precio de cierre no reflejará el impacto del reparto de dividendos (en efectivo o en activos), o la *divisiones de acciones* (*stock split*: la multiplicación del número de acciones en circulación con la consecuente reducción de su precio unitario) [31, cap. 7]. Los precios de cierre ajustado son una forma de abordar este problema; véase [32].

frecuencias de hasta milisegundos), o los precios de bonos o derivados, tienden a ser muy caros o difíciles de obtener.

### 1.3. Mercados financieros

Las Finanzas son la aplicación de principios de Economía, Matemáticas (particularmente de Cálculo, Estadística y Probabilidad), Psicología y Contabilidad, para la toma de decisiones que involucren la asignación y administración de recursos bajo condiciones de incertidumbre [31, cap. 1]. Pueden distinguirse tres ramas en Finanzas:

1. **Mercados de Capital:** enfocada en el estudio de los sistemas financieros, la estructura de las tasas de interés, y la fijación de precios de activos de riesgo.
2. **Administración Financiera:** dedicada a las decisiones de inversión y financiamiento en una organización, como las sociedades mercantiles (v.gr. empresas individuales, sociedades en nombre colectivo, sociedades anónimas, etc.) o las entidades públicas.
3. **Administración de Inversiones:** enfocada en el establecimiento de objetivos, políticas y estrategias de inversión; de selección de activos específicos, y la supervisión o evaluación del desempeño de las inversiones.

El sistema financiero es la plataforma en la que los recursos (fondos, dinero, financiamiento) se transfieren entre quienes los tienen (los *prestamistas*, *inversores*, *accionistas*, *partícipes*) y quienes lo necesitan (los *prestatarios*, *deudores*, *acreditados*, *emprendedores*) [31, cap. 2]. Tiene la finalidad de hacer más eficiente esa transferencia de recursos, mitigando el problema de la asimetría en la información entre prestamistas y prestatarios (los *actores*). Tiene tres componentes:

1. **Mercados:** donde tienen lugar las transacciones. Pueden ser lugares físicos o virtuales.
2. **Intermediarios:** quienes facilitan las transacciones.
3. **Reguladores:** quienes se aseguran de que las transacciones sean justas.

El sistema financiero comprende varios sectores: empresas financieras, entidades gubernamentales (federales, estatales y locales, pero también ser empresas para-estatales o patrocinadas por un gobierno), entidades empresariales no financieras, particulares (*households*) y entidades sin fines de lucro. El sector más grande es el de las entidades empresariales no financieras [31, cap. 10]. El sector financiero se compone de instituciones depositarias, no depositarias, compañías aseguradoras, y sociedades de inversión. En un sistema financiero es importante la eficiencia en la fijación de precios, el estudio del rol y comportamiento de sus actores, la mejor manera de regular su actividad, la medición del riesgo, y la fijación de precios de *activos*. Un activo es cualquier recurso del que se espera un beneficio futuro, por lo que se le asigna un valor económico presente. Puede ser tangible (v.gr. edificios, flotillas de autobuses o aviones, tierras, maquinaria, etc.) o intangible (v.gr. patentes, derechos de autor, marcas registradas o *instrumentos financieros*). En general, los activos intangibles representan un derecho legal a beneficios económicos futuros [33, cap. 1].

Los activos o instrumentos financieros constan de un emisor y un inversor. Pueden ser:



1. **De deuda:** el emisor acuerda pagar al inversor la cantidad prestada, más un interés; es decir: el inversor percibe una *renta fija*. Algunos ejemplos son los pagarés, los valores gubernamentales como los CETES<sup>11</sup>, los bonos o los préstamos (créditos).
2. **De capital:** el emisor, generalmente una empresa, da el derecho al inversor titular a parte de los beneficios (de haberlos) de su propia actividad, los *dividendos*, y a veces el derecho de voto en las decisiones de la empresa emisora. Algunos ejemplos son las acciones ordinarias y las participaciones sociales.

Los mercados financieros son el lugar donde se comercian o *cotizan* los instrumentos financieros, y tienen la utilidad de:

- Descubrir y fijar precios<sup>12</sup>, señalando cómo se distribuyen los fondos disponibles de quienes desean prestarlos o invertirlos entre quienes los necesitan, considerando que el incentivo para los que buscan fondos se relaciona con la rentabilidad exigida por los inversores. Dicho de otro modo: los precios se adaptan a la oferta y la demanda.
- Dar liquidez, por propiciar la presencia de compradores y vendedores listos para comerciar. Esto es conveniente en circunstancias donde un inversionista es forzado (o está motivado) a vender un instrumento financiero.
- Reducir los costos de transacción, asociados a la búsqueda de contrapartes (quién quiere vender y quién quiere comprar), o a la información sobre los atributos de inversión.

Los mercados financieros se pueden clasificar con diversos criterios (véase [33, cap. 1]):

- Con perspectiva nacional o territorial: Internos o externos.
- Según el plazo de vencimiento de los activos financieros que se negocian: Monetarios o de capitales.
- Por el plazo de entrega: Al contado (de efectivo, o *spot*) o de productos derivados.
- Según la *estructura de mercado*<sup>13</sup>: basada en órdenes, o basada en cotizaciones.
- Según la fase o temporada de negociación: Primarios o secundarios.

Para nosotros, es de interés la clasificación en mercados primarios o secundarios:

- **Primarios:** cuando un gobierno o una compañía emite por primera vez un instrumento financiero, éste se vende en un mercado primario para reunir capital. La oferta puede hacerse públicamente o con colocación privada. Típicamente, la *salida a bolsa* u *oferta pública inicial*

---

<sup>11</sup> *Certificados de la Tesorería* de la Federación (en México).

<sup>12</sup> La valuación de un instrumento financiero se hace con el principio de que el precio de un activo es el valor presente del flujo esperado de efectivo. En otras palabras: se estiman los flujos esperados (a futuro) de efectivo, y se ajustan con tasas de interés (o de descuento) apropiadas, dadas por la teoría de Mercados de Capital, para obtener su valor presente. Por ejemplo, para valorar una acción se estiman los dividendos futuros y su incertidumbre.

<sup>13</sup> La estructura del mercado es el mecanismo de interacción entre compradores y vendedores, con el que se determinan precio y cantidad.

de valores (IPO<sup>14</sup>) involucra bancos de inversión, quienes suscriben esas nuevas emisiones o valores a los mercados públicos. También se pueden ofrecer nuevas emisiones mediante subastas.

- **Secundarios:** los lugares centrales (físicos o virtuales) donde los instrumentos financieros son revendidos entre inversores, por lo que los emisores originales no recaudan capital adicional en estas transacciones. Se pueden distinguir por su grado de formalización:
  - **Bolsas** (o mercados organizados): los instrumentos financieros cotizados están previa y debidamente listados por una *bolsa*. Un instrumento debe satisfacer ciertos requisitos (pago de impuestos, alcanzar una cierta capitalización de mercado o un número de acciones en circulación pública, etc.) para ser aceptado y listado por una bolsa. La bolsa más grande e importante es la NYSE<sup>15</sup>, seguida por el Nasdaq (ambos en EE.UU.).
  - **Mercados secundarios no organizados** (también llamados extrabursátiles, de venta libre u OTC<sup>16</sup>): los instrumentos financieros cotizados no tienen la necesidad de incluirse en un listado. Típicamente, en un mercado OTC se cotizan bonos, créditos o acciones ordinarias.

Al día de hoy, en los mercados financieros se llevan a cabo miles de transacciones por día, coordinadas por lo general por un libro de órdenes (LOB<sup>17</sup>) electrónico de *subasta doble en tiempo continuo*, en el que se registran las ofertas de compra (con los precios máximos con que se está dispuesto a pagar) o venta (con los precios mínimos con que se está dispuesto a vender) de los comerciantes (*participantes del mercado*). Éstos se comprometen a efectuar la orden en caso de encontrar una contraparte, proveyendo liquidez [34, cap. 1, sec. 1]. La información del libro de órdenes sobre las cantidades disponibles para compraventa (el *volumen*), junto con sus precios específicos, se actualiza en tiempo real y está disponible para los participantes del mercado. Las órdenes de compra (*ask*) y venta (*bid*) cuyos precios coinciden son ejecutadas y borradas del LOB. Los detalles pueden consultarse en [34, caps. 3-8].

La *eficiencia del mercado* se refiere a la rapidez en que los precios de los activos reflejan toda la información disponible, de tal manera que los inversionistas (asumidos como racionales) no puedan obtener ganancias superiores a las definidas por el riesgo asumido (*rendimientos anómalos*). De acuerdo con Eugene F. Fama [35], hay tres niveles para la eficiencia de un mercado:

1. **Eficiencia débil:** los precios actuales de los activos reflejan todos los precios históricos.
2. **Eficiencia semi-fuerte:** los precios actuales de los activos reflejan toda la información públicamente disponible. En otras palabras: los precios de los activos reflejan esta información rápidamente<sup>18</sup>.

---

<sup>14</sup> *Initial Public Offering.*

<sup>15</sup> *New York Stock Exchange.*

<sup>16</sup> *Over the Counter.*

<sup>17</sup> *Limit Order Book.*

<sup>18</sup> En [36] se investigó la correlación temporal entre rendimientos de acciones del NYSE, encontrándose dos mecanismos principales de influencia: la debida a efectos externos y con efectos simultáneos, y la influencia de una empresa sobre otra (para la que se observó que la máxima correlación a menudo ocurría con un desfase temporal, indicativa de la dirección de la influencia), lo que es compatible con la hipótesis del mercado eficiente en su forma semi-fuerte [5].

3. **Eficiencia fuerte:** los precios actuales de los activos reflejan toda la información pública y privada<sup>19</sup>.

Una implicación de la hipótesis de los mercados eficientes, es que el precio que pagan los inversionistas por los activos debe reflejar su valor, de modo que cualquier información nueva sobre una compañía debe cambiar el precio de sus acciones. Otra consecuencia es que el precio futuro de un activo no puede deducirse a partir de la observación de sus precios pasados [37, sec. 1.1.2]. En el caso de la hipótesis de eficiencia fuerte, además, se deduce que la distribución de cambios de precio es normal (si éstos tienen segundo momento finito), lo que contrasta con los datos [5]. Benoît Mandelbrot [38]

### 1.3.1. Índices

Un índice de mercado es un *portafolio*<sup>20</sup> hipotético que representa un segmento del mercado financiero, cuyo valor se calcula a partir de los precios de los activos subyacentes (acciones, bonos, etc.). Algunos índices se ponderan según la capitalización de mercado, los ingresos, las acciones en circulación, entre otros. Los inversionistas siguen diferentes índices para evaluar los movimientos del mercado.

Un índice de mercado o bursátil es una combinación ponderada de los precios de varias acciones, que trata de representar su valor conjunto como una serie de tiempo [39]. Los tres índices bursátiles más populares para rastrear el rendimiento del mercado estadounidense son el *S&P 500* y el *Nasdaq Composite*. En el mercado de bonos, *Bloomberg* es un proveedor líder de índices, cuyo *Bloomberg U.S. Aggregate Bond Index* es uno de los más populares para los bonos estadounidenses[39].

En este trabajo, aprovecharemos los índices de mercado para obtener los listados de empresas representativas o importantes de una o varias bolsas en un país. Nos enfocaremos en:

- EE.UU.: Nasdaq y NYSE, de acuerdo con el índice S&P 500.
- Japón: JPX<sup>21</sup>, representada por las compañías del índice *Nikkei 225*.

Aunque también vale la pena mencionar aquí otras bolsas e índices de importancia mundial:

- Inglaterra: LSE<sup>22</sup>, representada por las compañías del *FTSE 100*.
- Alemania: FSE<sup>23</sup>, representada por las compañías del índice *DAX*.
- Francia: *Bourse de Paris* o *Euronext*, representadas por las compañías del índice *CAC 40*.

---

<sup>19</sup>La evidencia empírica de acontecimientos recientes sugiere que el mercado de valores de EE.UU. no es eficiente en su forma fuerte: se pueden obtener beneficios de la negociación con información privilegiada [31, cap. 2, sec. 3].

<sup>20</sup>Entenderemos como portafolio cualquier agrupación de activos financieros, particularmente acciones.

<sup>21</sup>*Japan Exchange Group*.

<sup>22</sup>*London Stock Exchange*.

<sup>23</sup>*Frankfurt Stock Exchange*.

# Objetivo General

El objetivo de nuestro proyecto de investigación consiste en estudiar la dinámica y la física estadística de los mercados financieros como sistemas complejos. Partiendo de los registros históricos de precios de distintas acciones cotizadas en bolsas de valores importantes, calcularemos las matrices de correlación de las fluctuaciones diarias de esos precios, para poder identificar parámetros que caractericen los estados por los que transitan los mercados de valores.

Seguiremos la línea de los trabajos previos sobre estados de mercado [40-42], pero en lugar de enfatizar el análisis de riesgos, nos enfocaremos en la dinámica de las matrices de correlación de Pearson, relativas y reducidas.



## Capítulo 2

# Fundamentos teóricos

El objetivo de este capítulo es revisar algunas nociones de la Estadística Multivariada (correlaciones de Pearson, correlaciones parciales y la distribución de Wishart) y de los Procesos Estocásticos (espacios de estados discretos, con una dinámica no determinista pero que satisfacen una markovianidad débil<sup>1</sup>), para tener un marco conceptual preliminar que sustente el resto de este trabajo. Comenzaremos revisando el Modelo de May [43], donde no solo se observa la relación entre el mayor eigenvalor de una matriz *de interacción* y la estabilidad del sistema que modela, sino también motiva la utilización de la Teoría de Matrices Aleatorias en el estudio de sistemas complejos.

### 2.1. Modelo de May

En 1972, R. May describió un modelo relativamente simple de un sistema complejo con  $N \gg 1$  componentes, cuyas interacciones (conexiones) aleatorias pueden pasar un nivel crítico y hacer al sistema inestable. Si este sistema se describiera por un sistema de ecuaciones no-lineales, en el análisis de estabilidad con desplazamientos  $\mathbf{x} = (x_1, \dots, x_N)^\top$  alrededor de un punto fijo  $\mathbf{x}^*$  May propuso:

$$\frac{d\mathbf{x}}{dt} = (\alpha\mathbf{J} - \mathbf{I})\mathbf{x} \quad (2.1)$$

bajo un *ensemble* de matrices aleatorias  $\mathbf{J}$  con entradas  $J_{kl} = J_{lk}$  reales, idénticamente distribuidas, simétricas respecto a su media<sup>2</sup>  $\langle J_{kl} \rangle = 0$  y con varianza  $\langle J_{kl}^2 \rangle = \sigma^2 > 0$ , y para un parámetro  $\alpha > 0$  representativo de las interacciones promedio entre las  $N$  componentes del sistema. En este modelo, la matriz identidad opuesta  $-\mathbf{I}$  modela, en el caso de ausencia de interacciones ( $J_{kl} = 0$ , o bien  $\alpha \rightarrow 0$ ), un régimen estable [44, Cap. 1, sec. 5][45, Cap. 5]. Los eigenvalores de  $\alpha\mathbf{J} - \mathbf{I}$  en la ecuación (2.1) tienen la forma  $\alpha\lambda - 1$ , donde  $\lambda$  es un eigenvalor de  $\mathbf{J}$  que sigue la *ley del semicírculo de Wigner* [46][47, cap. 4, sec. 3][48]: si  $S$  es el número de

---

<sup>1</sup>Aunque los mercados financieros son fuertemente no-markovianos, la naturaleza de las transiciones en estados de mercado se ha mostrado consistente con la hipótesis de markovianidad [41].

<sup>2</sup>En este trabajo, utilizaremos  $\langle \dots \rangle$  para denotar el valor esperado de una variable aleatoria real discreta ( $\langle X \rangle = \sum_{x \in S} p_x x$ ,  $|S| \leq |\mathbb{N}|$ ) o continua ( $\langle X \rangle = \int_S x f(x)$ ,  $|S| = |\mathbb{R}|$ ). En el caso de procesos estocásticos  $\{X_t : t \in \mathbb{R}\}$ , omitiremos la dependencia temporal cuando resulte clara a partir del contexto.

eigenvalores de  $\mathbf{J}$  en el intervalo  $(a\sqrt{N}, b\sqrt{N})$  (donde  $a < b$ ), entonces

$$\lim_{N \rightarrow \infty} \frac{\langle S \rangle}{N} = \frac{1}{2\pi\sigma^2} \int_a^b \sqrt{4\sigma^2 - x^2} \mathbb{1}_{\{|x| < 2\sigma\}} dx \quad (2.2)$$

Intuitivamente, esto significa que para  $N \gg 1$ , los  $N$  eigenvalores de  $\mathbf{J}$  tienen una densidad  $\rho(\lambda) = \frac{1}{\pi N \sigma^2} \sqrt{2N\sigma^2 - \lambda^2}$  de forma semicircular, de donde se deduce que  $\langle \lambda_{\text{máx}} \rangle = \sigma\sqrt{2N}$ .

La estabilidad del sistema (2.1) se establece por la condición  $\alpha\lambda_i - 1 < 0$  para cada  $i \in \{1, \dots, N\}$ , o de forma equivalente:

$$\lambda_{\text{máx}} < \frac{1}{\alpha} \quad (2.3)$$

Si además suponemos que  $\sigma^2 \sim 1/N$ , entonces en el límite termodinámico  $N \rightarrow \infty$  tendríamos  $\lambda_{\text{máx}} \approx \langle \lambda_{\text{máx}} \rangle = \sigma\sqrt{2N} \sim \sqrt{2}$ , que en conjunto con la condición (2.3) daría la estabilidad de (2.1) con probabilidad 1 [49] si:

$$\alpha < \frac{1}{\sqrt{2}} \quad (2.4)$$

Esto significa que  $\alpha_c := \frac{1}{\sqrt{2}}$  es un valor crítico para la estabilidad de (2.1) en el límite termodinámico.

La razón detrás de la mención de este modelo, es que ejemplifica la relevancia de los sistemas complejos en el estudio de mercados financieros: la alta conectividad de las componentes de un sistema se traduce en algún tipo de inestabilidad, o en la existencia de comunidades (los sectores, los países o conglomerados) y de grandes desviaciones (desplomes de los índices); la de *nodos influyentes* en la comunidad, y en cómo éstos arrastran a los demás miembros a crisis o estados extremos. Además, relaciona la estabilidad del sistema con un valor crítico del mayor eigenvalor de la matriz de las interacciones.

## 2.2. Covarianza y correlación

Si  $X$  y  $Y$  son variables aleatorias con medias  $\mu_X, \mu_Y$  y varianzas  $\sigma_X^2, \sigma_Y^2$ , la *covarianza* y la *correlación de Pearson* se definen como:

$$\text{Cov}(X, Y) := \langle (X - \mu_X)(Y - \mu_Y) \rangle \quad (2.5)$$

$$\text{Corr}(X, Y) := \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \quad (2.6)$$

con la convención de  $\text{Cov}(X, X) = \sigma_X^2$ .

Dado un conjunto de  $N$  procesos estocásticos  $\{X_i = X_i(t) : i = 1, \dots, N\}$ , (por ejemplo, los rendimientos (logarítmicos) de  $N$  acciones al tiempo  $t$ :  $X_i(t) = \log\left(\frac{P_i(t)}{P_i(t-\Delta t)}\right)$ ), las covarianzas  $\text{Cov}(X_i(t), X_j(t)) := \langle (X_i(t) - \mu_i(t))(X_j(t) - \mu_j(t)) \rangle$  constituyen la base para el estudio de la intensidad de su interdependencia (*lineal*). En mercados financieros, donde se recopila una vasta cantidad de información (muchas muestras temporales  $\hat{X}_i(1), \dots, \hat{X}_i(T)$  de muchas variables  $X_i : i = 1, \dots, N$ ), las *matrices poblacionales*  $\mathbf{\Sigma} = (\sigma_{ij}) = (\text{Cov}(X_i, X_j))$  y

$\mathbf{C} = (c_{ij}) = (\text{Corr}(X_i, X_j))$ , deberían poder estimarse por las *matrices muestrales*  $\hat{\Sigma} = (\hat{\sigma}_{ij})$  y  $\hat{\mathbf{C}} = (\hat{c}_{ij})$ , con entradas:

$$\hat{\sigma}_{ij} = \frac{1}{T} \sum_{t=1}^T \left( \hat{X}_i(t) - \hat{\mu}_i \right) \left( \hat{X}_j(t) - \hat{\mu}_j \right) \quad (2.7)$$

$$\hat{c}_{ij} = \frac{1}{T} \sum_{t=1}^T \frac{\left( \hat{X}_i(t) - \hat{\mu}_i \right) \left( \hat{X}_j(t) - \hat{\mu}_j \right)}{\hat{\sigma}_i \hat{\sigma}_j} \quad (2.8)$$

(donde  $\hat{\mu}_i$  y  $\hat{\sigma}_j$  son estimadores de las medias  $\mu_i := \mu_{X_i}$  y varianzas  $\sigma_i^2 := \sigma_{X_i}^2$  poblacionales<sup>3</sup> de  $X_i$ , y se adoptan las convenciones  $\sigma_{ii} = \sigma_i^2$ ,  $\hat{\sigma}_{ii} = \hat{\sigma}_i^2$ ), *al menos en el caso de*  $1 \leq N \ll T$  [52].

Cabe mencionar que el énfasis sobre las correlaciones para estudiar los mercados financieros fue introducido hace más de 20 años por R. Mantegna, E. Stanley, J. P. Bouchaud *et al* [53-55]. Las matrices muestrales también son referidas como *empíricas*, y corresponden a lo que *miden* u *observan* los actores en los mercados financieros.

J. Wishart [56] fue el primero en analizar las covarianzas muestrales en el caso en que  $1 \ll N \leq T$  (con  $\lambda := N/T = \mathcal{O}(1)$ ) cuando  $\mathbf{X} := (X_1, \dots, X_N)^\top \in \mathbb{R}^N$  es un vector aleatorio normal y  $(\hat{\mathbf{X}}(1), \dots, \hat{\mathbf{X}}(T)) \in \mathbb{R}^{N \times T}$  son  $T$  vectores aleatorios independientes con la misma distribución que  $\mathbf{X}$ . La forma de la densidad de  $\hat{\Sigma}$  condicionada a  $\Sigma$ , así como un esbozo de sus propiedades, se da en el apéndice C.

Refiriéndonos a ese material, la densidad conjunta de valores propios  $\lambda_1, \dots, \lambda_N$  (que en lo sucesivo consideramos ordenados de forma ascendente:  $\lambda_1 \leq \dots \leq \lambda_N$ ) de  $\hat{\Sigma}$ ,  $\varphi(\lambda_1, \dots, \lambda_N | \Sigma)$ , dada por la fórmula (C.2), se anula cuando  $|\lambda_k - \lambda_j| \rightarrow 0$ . Ésto es llamado *repulsión de eigenvalores* o *niveles*, debido al contexto de matrices Hamiltonianas y los ensembles clásicos de matrices aleatorias surgidos a partir del trabajo de Eugene P. Wigner [57]). Experimentalmente, al considerar la distribución  $P = P(s)$  del espaciamiento  $s$  entre las energías de resonancia nucleares (eigenvalores) sucesivas<sup>4</sup> de átomos pesados, normalizada al promedio de todas ellas<sup>5</sup>, se cumple en buena aproximación la *conjetura de Wigner* (1957) [58]:

$$P(s) = \frac{\pi s}{2} \exp\left(-\frac{\pi}{4} s^2\right) \quad (2.9)$$

deducida por Wigner de forma exacta para matrices Hamiltonianas reales de  $2 \times 2$  invariantes bajo el grupo ortogonal, con elementos gaussianos independientes de media nula y varianza finita [59, cap. 2, sec. 1].

La Teoría de Matrices Aleatorias, o *RMT*<sup>6</sup>, tiene modelos relevantes para el estudio de matrices de covarianza y correlaciones, como los ensembles WOE<sup>7</sup> y CWOE<sup>8</sup>. En general, las

<sup>3</sup>Cuando  $T \gg 1$ , los estimadores de máxima verosimilitud y sin sesgo para  $\sigma_i^2$ , con factores  $T^{-1}$  y  $(T-1)^{-1}$ , respectivamente, son prácticamente iguales; además, éstos factores son irrelevantes para el cálculo de  $\hat{c}_{ij}$ . Para una discusión pertinente, véanse [50], [51, cap. 4, sec. 3] y [este artículo](#) (consultado el 15 de enero de 2023).

<sup>4</sup>*Nearest-neighbor eigenvalue spacings.*

<sup>5</sup>El procedimiento de transformar eigenvalores  $\lambda_i$  en  $\xi_i$ , con separación local media de 1, se conoce como desdoblamiento o *unfolding*, y facilita la comparación de datos experimentales con la teoría [55].

<sup>6</sup>*Random Matrix Theory.*

<sup>7</sup>*Wishart Orthogonal Ensemble.*

<sup>8</sup>*Correlated Wishart Orthogonal Ensemble.*



matrices de éstos ensembles poseen la misma estructura local que la de los miembros de los ensembles GOE, GUE y GSE<sup>9</sup> (es decir: separación de niveles, función de correlación de dos puntos, etc.), pero no tienen la misma distribución de niveles (ecuación (2.2)); véase [57, sec. 2.1]. En 1967, V. Marchenko y L. Pastur [60] demostraron que, para  $\Sigma = \sigma^2 \mathbf{I}$  y las mismas condiciones de independencia y distribución idéntica de  $\{\hat{\mathbf{X}}(t) : t = 1, \dots, T\} \subseteq \mathbb{R}^N$ , la *densidad espectral empírica* o *densidad de eigenvalores*:

$$\rho_{\Sigma}^N(x) := \frac{1}{N} \sum_{i=1}^N \delta(x - \lambda_i)$$

converge casi seguramente a la densidad (distribución) de niveles, o *densidad espectral límite*:

$$\rho_{\Sigma}(x) = \frac{1}{2\pi x \lambda \sigma^2} \sqrt{(\lambda_+ - x)(x - \lambda_-)} \mathbb{1}_{[\lambda_-, \lambda_+]} + \left(1 - \frac{1}{\lambda}\right) \delta(x) \mathbb{1}_{\lambda \geq 1} \quad (2.10)$$

(donde  $\lambda_{\pm} := \sigma^2 (1 \pm \sqrt{\lambda})^2$ ) cuando  $N, T \rightarrow \infty$  pero  $N/T \rightarrow \lambda > 0$  fija. Esta densidad incluye el caso en que  $N \geq T$  (correspondiente a  $\lambda > 1$ , excluido para la deducción de (C.1) y (C.2); véase el Apéndice C). Su importancia es que establece un modelo nulo (de correlaciones de ruido blanco) para inferir la existencia de correlaciones reales.

Hay que mencionar que, de forma análoga a la relación (2.2), se deduce que

$\langle \lambda_{\text{máx}} \rangle = \lambda_+ = \sigma^2 (1 + \sqrt{\lambda})^2 = \sigma^2 \left(1 + \sqrt{\frac{N}{T}}\right)^2$ , lo que contrasta con el mayor eigenvalor poblacional  $\sigma^2$  de  $\hat{\Sigma} = \sigma^2 \mathbf{I}$ . Por ejemplo, si  $\lambda = 1$  (es decir:  $N = T$ ), entonces  $\langle \lambda_{\text{máx}} \rangle = 4\sigma^2$ .

Otro caso para el que hay resultados asintóticos es el de correlación constante [61]: para  $0 < c < 1$  fijo, si  $\Sigma = \sigma^2 (c \mathbf{1} \otimes \mathbf{1} + (1 - c) \mathbf{I})$  (o equivalentemente, si  $\sigma^2 = \text{Var}(\hat{X}_i(t))$  para cada observación  $t$  de cada variable  $i$ , y  $\mathbf{C} = (c_{ij}) = (\delta_{ij} + c(1 - \delta_{ij}))$ ), la densidad espectral límite de  $\rho_{\Sigma}^N(x)$  es:

$$\rho_{\Sigma}(x) = \rho_{\Sigma}^{(0)}(x) + \frac{1}{N} \delta\left(x - \frac{\lambda \sigma^2 (Nc + 1 - c) \left(\frac{Nc}{\lambda} + 1 - c\right)}{Nc}\right) + \mathcal{O}(N^{-2}) \quad (2.11)$$

donde  $\rho_{\Sigma}^{(0)}(x)$  es la ley de Marchenko-Pastur (2.10) con la sustitución  $\sigma^2 \mapsto \sigma^2(1 - c)$  y la normalización  $\int \rho_{\Sigma}^{(0)}(x) dx = 1 - \frac{1}{N}$ . La suposición  $0 < c$ , además de evitar las complicaciones geométricas de la *anti-transitividad* de la correlación ( $c_{ij} = -1 \& c_{jk} = -1 \implies c_{ik} = 1$ ), es consistente con la escasez de correlaciones negativas observada en mercados financieros. Véase, por ejemplo, la figura 3.8(a).

En la literatura, a los eigenvalores en el soporte de  $\rho_{\Sigma}^{(0)}(x)$ :

$$[\lambda_-, \lambda_+] = \left[ \sigma^2 (1 - c) (1 - \sqrt{\lambda}), \sigma^2 (1 - c) (1 + \sqrt{\lambda}) \right]$$

se les conoce como del *bulk* [57]. Por otra parte, de (2.11) se deduce que la media sobre *ensemble* para el eigenvalor máximo muestral es:

$$\langle \lambda_{\text{máx}} \rangle = \frac{\lambda \sigma^2 (Nc + 1 - c) \left(\frac{Nc}{\lambda} + 1 - c\right)}{Nc} = \frac{\lambda \sigma^2 (Nc + 1 - c) \left(\frac{Nc}{\lambda} + 1 - c\right)}{Nc} = \frac{\sigma^2 (Nc + 1 - c) (Tc + 1 - c)}{Tc} \quad (2.12)$$

<sup>9</sup> Gaussian Orthogonal/Unitary/Symplectic Ensemble.

que comparada con el eigenvalor máximo poblacional  $\sigma^2(Nc + 1 - c)$  de  $\Sigma = \sigma^2(c\mathbf{1} \otimes \mathbf{1} + (1 - c)\mathbf{I})$ , difiere en un factor:

$$\frac{Tc + 1 - c}{Tc} = 1 + \frac{1 - c}{cT}$$

lo que indica un sesgo relativo de  $\frac{1-c}{cT}$ . Además, el eigenvalor aislado

$\langle \lambda_{\text{máx}} \rangle = \frac{\lambda\sigma^2(Nc + 1 - c)(\frac{Nc}{\lambda} + 1 - c)}{Nc}$  en (2.11) debe cumplir  $\langle \lambda_{\text{máx}} \rangle > \lambda_+$  (es decir, debe encontrarse afuera del *bulk*):

$$\begin{aligned} \frac{\lambda\sigma^2(Nc + 1 - c)(\frac{Nc}{\lambda} + 1 - c)}{Nc} &> \sigma^2(1 - c) \left(1 + \sqrt{\lambda}\right)^2 \\ \implies \sqrt{\lambda} &< \frac{Nc}{1 - c} \\ \therefore c > \frac{\sqrt{\lambda}}{N + \sqrt{\lambda}} &= \frac{1}{1 + T\sqrt{\lambda}} = \frac{1}{1 + \sqrt{NT}} \end{aligned} \quad (2.13)$$

De la penúltima relación se deduce lo siguiente:

1. El sesgo  $\frac{1-c}{cT}$  está acotado superiormente por  $\sqrt{\lambda}$ , así que  $1 \ll N \ll T$  mejora la estimación del máximo eigenvalor. En estas condiciones, las matrices muestrales  $\hat{\Sigma}$  son singulares. Además, si el sesgo  $\frac{1-c}{cT}$  no debe superar  $x < 1$ , entonces es necesario que  $\frac{1}{1+xT} \leq c$ .
2. Cuando  $cN = \mathcal{O}(1)$ , se obtiene la expresión relativamente simple  $\sqrt{\lambda} < Nc \iff c > \frac{\sqrt{\lambda}}{N}$  (ver la discusión en [61, sec. IV]). Esto se debe a que, para  $0 < c < 1$ ,  $\frac{c}{1-c} = c + c^2 + c^3 + \dots$ .

El eigenvector poblacional asociado con el máximo eigenvalor poblacional  $\sigma^2(Nc + 1 - c)$ ,  $\frac{1}{\sqrt{N}}(1, \dots, 1) \in \mathbb{R}^N$ , existe incluso para  $0 < c \ll 1$  pero tiene una extensión en sus componentes que contrasta con las de los eigenvectores del modelo sin correlación  $\Sigma = \sigma^2\mathbf{I}$ , que son gaussianas por la invariancia bajo el grupo ortogonal (fórmula (C.2)).

Una forma alternativa de plantear las correlaciones constantes, consiste en proceder de manera análoga al modelo de May (2.1) y la matriz  $\mathbf{J}$ : si la matriz aleatoria  $\hat{\mathbf{C}} = (\hat{c}_{ij}) \in \mathbb{R}^N$  es simétrica, con entradas independientes (sin necesariamente la misma distribución), acotadas y tales que:

$$\begin{aligned} \forall 1 \leq i < j \leq N : \quad \langle \hat{c}_{ij} \rangle &= c > 0, \quad \text{Var}(\hat{c}_{ij}) = \sigma^2 \\ \forall 1 \leq i \leq N : \quad \langle \hat{c}_{ii} \rangle &= \nu \end{aligned}$$

entonces  $\lambda_N$  es asintóticamente normal con media  $(N - 1)c + \nu + \sigma^2/c$  y varianza  $2\sigma^2$ , y los demás eigenvalores siguen la ley del semicírculo de Wigner (2.2) [62].

Un modelo más complejo es el de grupo [63]: los  $N$  renglones de  $(\hat{\mathbf{X}}(1), \dots, \hat{\mathbf{X}}(T))$ , previamente estandarizados, se agrupan en bloques  $\alpha_1, \dots, \alpha_M$  con tamaños  $N_1, \dots, N_M$  (con  $N_i > 0$  y  $N = N_1 + \dots + N_M$ ), y se asumen de la forma:

$$\hat{X}_i(t) = f_{\alpha_i}(t) + \varepsilon_i(t), \quad (t = 1, \dots, T) \quad (2.14)$$

donde  $\alpha_i$  representa el grupo al que pertenece  $(\hat{X}_i(1), \dots, \hat{X}_i(T))$ ,  $f_{\alpha_i}$  una variación común del grupo  $\alpha_i$ , y  $\varepsilon_i$  una variable aleatoria de ruido controlada por un parámetro de intensidad  $\epsilon_{\alpha_i}$  tal que  $\text{Var}(f_{\alpha_i}) = (1 + \epsilon_{\alpha_i}^2)^{-1}$  y  $\text{Var}(\varepsilon_i) = \epsilon_{\alpha_i}^2 (1 + \epsilon_{\alpha_i}^2)^{-1}$ . En estas condiciones, la matriz de correlación  $\hat{\mathbf{C}}$  toma la forma  $(\hat{c}_{ij}) = (\delta_{\alpha_i \alpha_j} + c_{ij}^1)$ . Para  $T \gg 1$  y  $\epsilon_{\alpha_i} = 0 \forall i$ ,  $\hat{\mathbf{C}}$  es una matriz de bloques diagonales con tamaños  $(N_1 \times N_1), \dots, (N_M \times N_M)$ , cada una de las cuales tiene eigenvalor máximo  $N_i$  y todos los demás eigenvalores nulos. En cambio, para  $\epsilon_{\alpha_i} \neq 0$ , se obtiene una expresión similar a (2.11) para cada bloque diagonal de  $\hat{\mathbf{C}}$ , con el eigenvalor máximo igual a  $N_i (1 + \epsilon_{\alpha_i}^2)^{-1}$  a orden  $\mathcal{O}(N^{-2})$ .

También hay resultados analíticos para la densidad de eigenvalores cuando  $N$  y  $T$  son finitos, para los que se utiliza el *método de supersimetría*; véanse [64, 65].

### 2.2.1. Matrices locales en el tiempo

Hay que resaltar que el planteamiento de las matrices muestrales es problemático por varias razones, como la no-estacionariedad de los datos [15, Caps. 6-7][53, 55] (los rasgos del mercado cambian con el tiempo, porque cambian las mismas expectativas de sus actores o las actividades de las compañías, que también nacen, quiebran o se fusionan), y la no-normalidad de los rendimientos (véase, por ejemplo, [5, sec. 2.1.1]).

Para ayudar a enfrentar estos problemas, en [66] se plantea la estimación de la versión muestral de la matriz de covarianza, o correlación, entre rendimientos  $X_1, \dots, X_N$  de precios de acciones para  $N$  compañías cotizando en un mercado de valores, sobre una época de tamaño  $q$  que finaliza en  $t$ :

$$\mathcal{E}(t; q) := \{\tau : t - q + 1 \leq \tau \leq t\},$$

contenida en un horizonte temporal discreto  $\{t : 1 \leq t \leq T\}$ , dando lugar a una *matriz de covarianza* o *correlación local*. Explícitamente: para  $T$  muestras consecutivas  $\hat{X}_i(1) = r_i(1), \dots, \hat{X}_i(T) = r_i(T)$  de cada uno de los rendimientos  $i = 1, \dots, N$ , una  $q \in \mathbb{N}$  fija tal que  $1 < q \leq T$ , y  $t \in \{q, \dots, T\}$ , las  $T - q + 1$  matrices locales  $\hat{\Sigma} = \hat{\Sigma}(t; q)$  y  $\hat{\mathbf{C}} = \hat{\mathbf{C}}(t; q)$  tienen elementos:

$$\begin{aligned} \hat{\sigma}_{ij}(t; q) &= \frac{1}{q} \sum_{\tau=t-q+1}^t (r_i(\tau) - \bar{r}_i(t; q)) (r_j(\tau) - \bar{r}_j(t; q)) \\ \hat{c}_{ij}(t; q) &= \frac{\hat{\sigma}_{ij}(t; q)}{\hat{\sigma}_i(t; q) \hat{\sigma}_j(t; q)} \end{aligned} \quad (2.15)$$

para  $1 \leq i, j \leq N$ , considerando las definiciones auxiliares:

$$\begin{aligned} \bar{r}_i(t; q) &:= \frac{1}{q} \sum_{\tau=t-q+1}^t r_i(\tau), \\ \hat{\sigma}_i(t; q) &:= \sqrt{\frac{1}{q} \sum_{\tau=t-q+1}^t (r_i(\tau) - \bar{r}_i(t; q))^2} \end{aligned}$$

Al tomar épocas cortas (es decir:  $q$  pequeña) las series de tiempo de rendimientos se vuelven débilmente estacionarias (lo cual cierto respaldo empírico [15, cap. 7, sec. 3]). Una serie de tiempo

$\{X(t)\}_{t \in \mathcal{T}}$  es débilmente estacionaria si  $\langle X(t) \rangle$  es independiente de  $t$  y  $\text{Cov}(X(t), X(t'))$  sólo depende de  $|t - t'|$ . Además, al tomar épocas cortas se obtienen perspectivas actualizadas (y por lo tanto valiosas) para los actores en un mercado financiero. Sin embargo, la cortedad de las épocas se traduce en una pérdida de información de la dinámica subyacente y a estadísticas pobres, pues las matrices de covarianza y correlación resultan singulares. En tal caso, la nulidad de varios eigenvalores evocaría el ruido característico del WOE [56, 60][67, Cap. 7][61, 68-70].

Además, en épocas de crisis (como la de las hipotecas *subprime* de 2008), la estacionariedad débil tampoco se cumple ni en períodos cortos [66, 71, 72]. En cualquier caso, las matrices de covarianza y correlación de las fórmulas (2.2.1) siguen estando formalmente definidas y se pueden estudiar, pero deben tenerse en cuenta las problemáticas descritas.

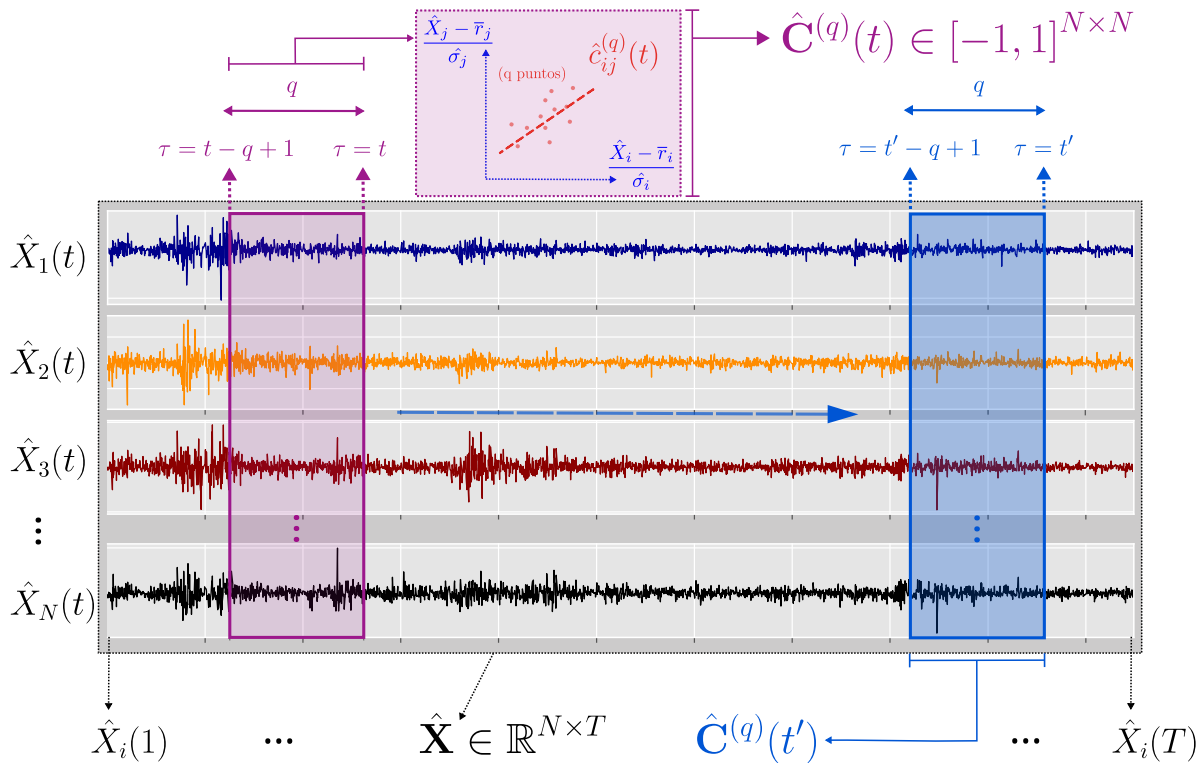


Figura 2.1: Esquema de cálculo de las correlaciones locales. Para la matriz de datos  $\hat{\mathbf{X}} \in \mathbb{R}^{N \times T}$  (relación (2.17)) con entradas  $(\hat{X}_i(t))$ , se fija un tamaño de época  $q$  ( $1 < q < T$ ) para delimitar submatrices contiguas o *ventanas* con  $N \times q$  números  $\{\hat{X}_i(\tau) : i = 1, \dots, N; \tau = t - q + 1, \dots, t\}$  (donde  $t = q, \dots, T$ ). Para cada una de estas  $T - q + 1$  ventanas (que en la figura son ejemplificadas por los bloques rectangulares traslúcidos morado y azul), se obtienen las correlaciones  $\hat{\mathbf{C}} = \hat{\mathbf{C}}(t; q) \in [-1, 1]^{N \times N}$  (relación (2.2.1)) para  $t = q, \dots, T$ .

Una posible solución al problema de la singularidad es usar la técnica del *power map* o *función de Guhr-Kälber*, introducida originalmente en [66, 73] para identificar y cuantificar el ruido en matrices de correlación debido a épocas cortas, pero aprovechada en [74, 75] para detectar correlaciones mediante la obtención de un espectro emergente de matrices de correlación singulares. De hecho, en [74] se utiliza el *power map* para mejorar la estimación de un portafolio óptimo de inversión, en el marco de la teoría de Markowitz. (Más adelante, mencionaremos otras aplicaciones [41, 42] de (2.16)). La transformación está dada para los elementos de la matriz de

correlación:

$$c_{ij} \mapsto c_{ij}^{(\epsilon)} := \text{sgn}(c_{ij}) |c_{ij}|^{1+\epsilon} \quad (2.16)$$

donde el parámetro  $\epsilon \in (0, 1)$  remueve la singularidad de las matrices cuando  $q \ll N$ . De este modo, dado un horizonte temporal de longitud  $T$  (típicamente del orden de 10 años de actividad bursátil, o unos 3000 puntos) de  $N$  series de tiempo, y un tamaño de época  $q$  ( $1 < q \ll T$ , que en [66]  $q = 13$  se muestra como óptimo), se puede calcular una secuencia de  $T - q + 1$  matrices locales de correlación  $\{\hat{\mathbf{C}}^{(\epsilon)}(t; q) : t = q, \dots, T\}$ , transformadas con el *power map*, representativas de las *instantáneas* (en la época de tamaño  $q$  terminando en  $t$ ) del mercado representado por la matriz de datos  $\hat{\mathbf{X}}$ :

$$\hat{\mathbf{X}} := \begin{pmatrix} \hat{X}_1(1) & \hat{X}_1(2) & \cdots & \hat{X}_1(T) \\ \hat{X}_2(1) & \hat{X}_2(2) & \cdots & \hat{X}_2(T) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{X}_N(1) & \hat{X}_N(2) & \cdots & \hat{X}_N(T) \end{pmatrix} \quad (2.17)$$

### 2.2.2. Análisis en componentes principales

A través de la descomposición diádica (o espectral), podemos pensar en que la matriz de correlación empírica  $\hat{\mathbf{C}} = \hat{\mathbf{C}}(t)$  entre las variaciones de los precios de  $N$  acciones en una época  $\mathcal{E} = \mathcal{E}(t; q)$ <sup>10</sup>, tiene una expresión en una base diagonal  $|n\rangle$ :

$$\begin{aligned} \hat{\mathbf{C}} &= \sum_{n=1}^N \lambda_n |n\rangle\langle n| \\ |n\rangle &= \sum_{j=1}^N a_j(n) |s_j\rangle \end{aligned} \quad (2.18)$$

donde  $\{|s_j\rangle\}_{j=1}^N$  son los estados que representan a las acciones de las compañías con la propiedad  $\hat{c}_{ij} = \langle s_i | \hat{\mathbf{C}} | s_j \rangle$ , y  $\{|n\rangle\}_{n=1}^N$  es la base de eigenvectores, que en lo sucesivo supondremos ordenados de acuerdo con los valores propios de manera ascendente  $\{\lambda_{\min} := \lambda_1 \leq \dots \leq \lambda_N =: \lambda_{\max}\}$ .

En la mayor parte de las aplicaciones de RMT en datos experimentales u observacionales, se ha visto que las funciones de correlación espectral dentro del *bulk* son universales en el límite de dimensiones grandes o al normalizar la separación entre niveles a 1 (es decir, hacer *unfolding*) [55, 72, 76, 77]. En general, la conjetura de universalidad de Wigner-Dyson-Mehta afirma que el comportamiento local alrededor del máximo de la distribución de los eigenvalores de matrices aleatorias, en el límite de dimensiones grandes, es independiente de la distribución particular de los elementos de matriz; véase [77].

Para mercados de valores, eso implica que la mayor parte del contenido espectral de la matriz de correlación es aleatoria en épocas suficientemente grandes [78]. Esto se confirma especialmente para el espaciamiento entre primeros vecinos (o separación entre niveles, que es común al GOE (2.2)) y la varianza de número.

El valor propio máximo (o aquellos eigenvalores que se encuentren significativamente afuera del *bulk*), por otra parte, corresponde al comportamiento colectivo o al efecto de una *influencia*

<sup>10</sup>Tomaremos el tamaño de época  $q$  como un parámetro fijo en el contexto correspondiente.

*común* (tal como un momento de pánico en una bolsa, el efecto de un tsunami o una pandemia) sobre las variaciones de los precios de acciones bajo investigación [5, 15, 70, 78], lo que es reminiscente del concepto de auto-organización microscópica y observables macroscópicos emergentes. Al mismo tiempo, la ortogonalidad de los eigenvectores del *bulk* se debe posiblemente al ruido que influye en el sistema, o al que ocurre durante el registro de los datos. En consecuencia, puede decirse que la mayor parte de la información está en las componentes asociadas con los valores propios más grandes, que a su vez generan un subespacio de menor dimensión.

El *análisis en componentes principales* o *PCA*<sup>11</sup> corresponde al análisis de los eigenvectores asociados con los eigenvalores más grandes, y permite, por ejemplo, identificar las variables relevantes de una matriz de datos de  $N$  variables con  $T$  observaciones [79, cap. 7, sec. 2]. De hecho, en el caso más general de la matriz de covarianza  $\Sigma$ , las direcciones propias son análogas a los *ejes principales* de un cuerpo rígido en Mecánica Clásica [80, págs. 406-409], y corresponden a las direcciones de mayor varianza.

Para estudiar las direcciones de mayor varianza es pertinente la razón de participación, que en general sirve para medir el grado de mezcla en un sistema discreto:

$$\text{PR} := 1 / \sum_{j=1}^N |a_j(N)|^4 \quad (2.19)$$

PR es máxima e igual a  $N$  si  $a_j(N) = 1/\sqrt{N}$  para toda  $j$ , y mínima e igual a 1 si  $a_j(N) = \delta_{ij}$  para alguna  $i$  ( $1 \leq i \leq N$ ). (Cabe recordar que los coeficientes  $a_1(N), \dots, a_N(N)$  conforman la representación del eigenvector principal  $|N\rangle$  en la base  $\{|s_j\rangle\}_{j=1}^N$ , introducida tras la ecuación (2.18)). Además, si el eigenvector principal  $|N\rangle$  está normalizado, sus componentes al cuadrado  $p_j := |a_j(N)|^2$  conforman una masa de probabilidad, por lo que tiene sentido pensar en la entropía de la información (o binaria) de Shannon:

$$\text{H} := - \sum_{j=1}^N p_j \log_2 p_j \quad (2.20)$$

Ésta función alcanza el máximo  $\log_2(N)$  cuando  $a_j(N) = 1/\sqrt{N}$  para toda  $j$ , y tiene mínimo 0 cuando  $a_j(N) = \delta_{ij}$  para alguna  $i$  ( $1 \leq i \leq N$ ), lo que es análogo al caso de la razón de participación. Nada impide que las definiciones (2.19) y (2.20) (y más adelante, (2.22)) también se apliquen a  $|n\rangle$  y sus coeficientes  $a_1(n), \dots, a_N(n)$  para  $1 \leq n < N$ . Usaremos el subíndice  $n$  para distinguir tales casos:  $\text{PR}_n, \text{H}_n, \text{IPR}_n$ .

En [81] se habla sobre cómo las subidas y bajadas en los mercados financieros de Alemania y los EE.UU., representados en ese estudio por los índices y las componentes del DAX<sup>12</sup> y del DIJA<sup>13</sup>, están relacionadas con cambios en la matriz de correlación de rendimientos de precios y la entropía de la información de sus eigenvectores: las disminuciones del índice tienden a reducir la varianza promedio y a separar significativamente el eigenvector principal de la matriz de correlación (donde las caídas de mayor intensidad se traducen en una separación más pronunciada del modo colectivo), mientras que las subidas tienen el efecto opuesto, haciendo que los eigenvalores estén más uniformemente distribuidos. Así, los decrementos simultáneos de los precios

<sup>11</sup>Principal component analysis.

<sup>12</sup>Deutsche Aktienindex.

<sup>13</sup>Dow Jones Industrial Average.

de las acciones de la mayoría de las compañías participantes ocurren en las caídas de los mercados, pero en las subidas de éstos no suele haber incrementos simultáneos de los precios.

Tomando como referencia el límite de la fórmula (2.20) en el ensemble GOE para  $N \gg 1$  [82]:

$$H_{\text{GOE}} = \frac{1}{\ln 2} \left( \psi \left( \frac{N}{2} + 1 \right) - \psi \left( \frac{3}{2} \right) \right) \approx \frac{\ln \left( \frac{1}{2} N \alpha \right)}{\ln 2} + \mathcal{O}(1/N) \quad (2.21)$$

(donde  $\psi(x) := \frac{d}{dx} \ln \Gamma(x)$  es la función digamma y  $\alpha \approx 0.96$ ), así como su valor máximo (alcanzado para una distribución uniforme  $p_j = 1/N$ ; es decir,  $a_j(N) = 1/\sqrt{N}$ ):  $H_{\text{Unif}} = \log_2 N$ , Drozd *et al* reportaron en [81], utilizando datos reales, que la función  $H = H(t)$ , donde  $t$  es la variable que localiza ventanas de ancho  $q = 30$ , alcanza el valor  $H_{\text{Unif}}$  en el eigenvector principal sólo cuando el índice correspondiente baja, y que en promedio se tiene  $H = H_{\text{GOE}}$  para los demás vectores propios de la matriz de correlación, con desviaciones anticorrelacionadas con el índice.

De hecho, en el ensemble GOE se tiene  $\text{PR}_{\text{GOE}} = N/3$  en el límite  $N \gg 1$ , por la normalidad e independencia asintóticas de las componentes [57, sec. 3.5.1][76, sec. VII.B]:  $a_j(n) \sim \mathcal{N}(0, \frac{1}{N})$ <sup>14</sup>. En el modelo de Wishart de correlación constante, el eigenvector principal  $\frac{1}{\sqrt{N}}(1, \dots, 1) \in \mathbb{R}^N$  tiene una razón de participación  $\text{PR}_{\text{Unif}} = N$ .

También existe la razón de participación inversa:

$$\text{IPR} := \sum_{j=1}^N |a_j(N)|^4 \quad (2.22)$$

cuyo negativo tiene propiedades matemáticas similares a las de la entropía de Shannon [25, sec. 2], y resulta una función analítica de las variables  $a_j(N)$ . Aunque las fórmulas (2.22) y (2.19) son recíprocas, sus medias sobre ensemble no lo son, pues en general  $\langle f^{-1} \rangle \neq \langle f \rangle^{-1}$ .

Plerou *et al* [55] compararon la razón de participación inversa del eigenvector principal de matrices de correlación, tanto para series de tiempo sintéticas con la distribución de rendimientos sabida empíricamente, como para series de tiempo de rendimientos de acciones con resolución de días o de media hora. En las series de tiempo sintéticas reportaron un valor promedio para la razón de participación inversa de  $1/N$  (tanto para el eigenvector principal como para los demás), mientras que en los datos reales encontraron valores cercanos a  $3/5$  (datos con resolución de media hora) y  $3/4$  (datos con resolución diaria), en promedio, en el caso de la razón de participación inversa del eigenvector principal.

Para el modelo de grupo (2.14) introducido en [63], se reportan resultados numéricos para la razón de participación inversa  $\text{IPR} \propto N_i^{-1}$  del eigenvector principal de cada bloque, cuando se asume una intensidad de ruido de la forma  $\epsilon_{\alpha_i} = 1 - \frac{1}{N_i}$ , y que el eigenvalor correspondiente está por encima del límite superior  $\lambda_+$  de la ley de Marchenko-Pastur.

---

<sup>14</sup>Para  $X \sim \mathcal{N}(\mu, \sigma^2)$ , se sabe que  $\langle (X - \mu)^p \rangle = \begin{cases} 0 & \text{si } p \text{ es impar} \\ \sigma^p (p-1)!! & \text{si } p \text{ es par} \end{cases}$ . Cuando  $\sigma^2 = N^{-1}$  y  $p = 4$ , se tiene  $\langle (X - \mu)^p \rangle = 3N^{-2}$ , de donde se deduce  $\text{PR}_{\text{GOE}} = N/3$ .

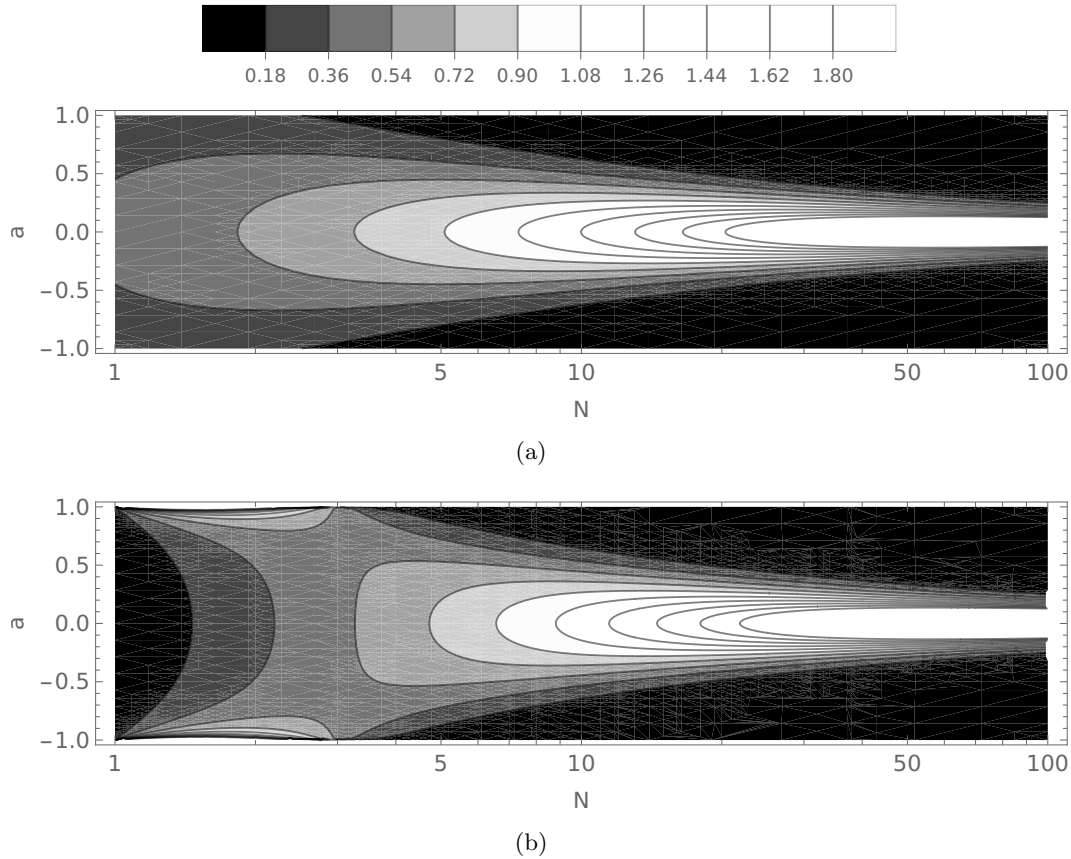


Figura 2.2: Para una componente  $a_j(n)$  de cualquiera de los eigenvectores de una matriz de  $N \times N$  del GOE (particularmente del WOE), la densidad de probabilidad es  $P_N(a) = \Gamma(N/2) [\sqrt{\pi} \Gamma((N-1)/2)]^{-1} (1-a^2)^{(N-3)/2}$  (figura 2.2(a)). Cuando  $N \gg 1$ ,  $P_N(a) = [N/(2\pi)]^{1/2} \exp(-Na^2/2)$  (figura 2.2(b)).

### 2.2.3. Matrices de correlación relativa y proyecciones de Guhr

Para medir el grado de dependencia entre parejas de variables aleatorias  $X, Y$ , la covarianza y la correlación son quizá la forma más simple. No obstante, hay que mencionar las siguientes desventajas:

1. Covarianza o correlación nula no implican independencia. Como ejemplo, puede tomarse el caso de  $X \sim \text{Unif}(-1, 1)$  y  $Y = X^2$ , o en general de cualquier variable aleatoria continua  $X$  con densidad simétrica, y  $Y = f(X)$  con  $f$  par<sup>15</sup>.
2.  $|\text{Corr}(X, Y)| = 1$  si y sólo si hay una relación lineal entre  $X$  y  $Y$  [83, sec. 4.5, teo. 7]. Esto excluye la correlación como una medida apta de la dependencia, cuando  $X$  y  $Y$  se relacionan de manera no lineal.

Pese a esto, la correlación y la covarianza son susceptibles de refinarse. Una posibilidad es

<sup>15</sup>Véase por ejemplo [esta discusión](#) (consultada el 3 de noviembre de 2023).



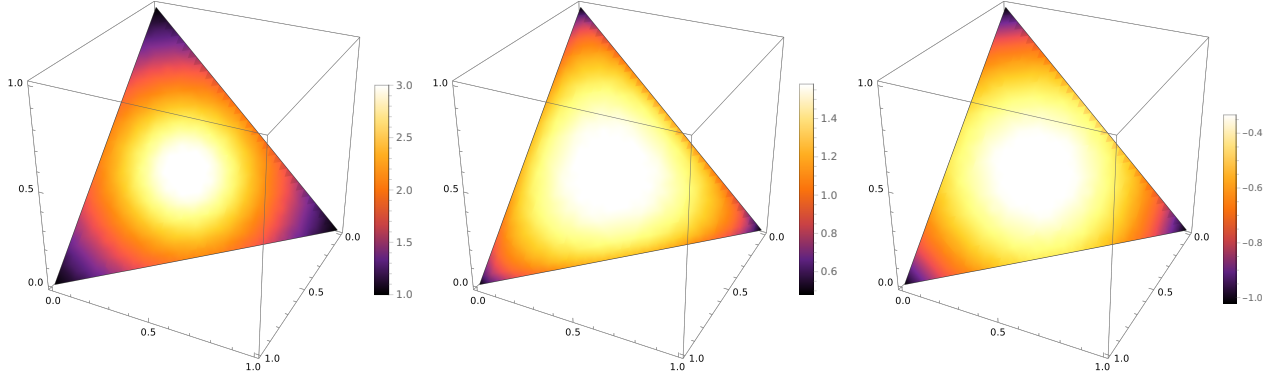


Figura 2.3: Gráficas de calor de la razón de participación (izquierda; fórmula (2.19)), entropía de Shannon (centro; fórmula (2.20)) y (opuesto de la) razón de participación inversa (derecha; fórmula (2.22)), definidas en el *simplex*  $\{(p_1, p_2, p_3) \in \mathbb{R}^+ : p_1 + p_2 + p_3 = 1\}$  para ejemplificar el caso  $N = 3$ . Los valores máximos son, respectivamente,  $N$ ,  $\log_2 N$  y  $-1/N$ ; mientras que los mínimos son 1, 0 y  $-1$ , respectivamente. Nótese que los códigos de color están normalizados al rango de la gráfica correspondiente.

haciendo una estimación de la dependencia lineal entre un conjunto de variables aleatorias, cuando se remueve la influencia de un subconjunto propio de ellas:

Si un vector aleatorio normal  $\mathbf{X} := (X_1, \dots, X_{N+1})^\top \in \mathbb{R}^{N+1}$  con media  $\mu$  y varianza  $\Sigma$  se parte en bloques como:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

donde  $\mathbf{X}_1, \mu_1 \in \mathbb{R}^k$ ,  $\mathbf{X}_2, \mu_2 \in \mathbb{R}^{N+1-k}$ ,  $\Sigma_{11} \in \mathbb{R}^{k \times k}$ ,  $\Sigma_{22} \in \mathbb{R}^{(N+1-k) \times (N+1-k)}$  y  $k \in \{1, \dots, N\}$  es fijo, entonces la distribución condicional de  $\mathbf{X}_1$  dada  $\mathbf{X}_2$  es normal con media  $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{X}_2 - \mu_2)$  y covarianza [51, cap. 5, sec. 3]:

$$\Sigma_{11}^{(2)} := \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = (\sigma_{ij:k+1,\dots,N+1}) \quad (2.23)$$

Los  $k \times k$  elementos de matriz  $(\sigma_{ij:k+1,\dots,N+1})$  son los coeficientes de la *covarianza parcial de  $\mathbf{X}_1$  relativa a  $\mathbf{X}_2$* , o *con respecto a  $\mathbf{X}_2$* . Estos coeficientes, que representan las covarianzas entre las variables aleatorias  $X_i$  y  $X_j$  (con  $1 \leq i \leq j \leq k$ ) de acuerdo con la distribución de  $\mathbf{X}_1$  condicionada a  $\mathbf{X}_2$ , conducen también a la definición de una *matriz de correlación parcial  $\mathbf{C}_{11}^{(2)}$  de  $\mathbf{X}_1$  relativa a  $\mathbf{X}_2$*  (o con respecto a  $\mathbf{X}_2$ ), con coeficientes:

$$c_{ij:k+1,\dots,N+1} = \frac{\sigma_{ij:k+1,\dots,N+1}}{\sqrt{\sigma_{ii:k+1,\dots,N+1} \cdot \sigma_{jj:k+1,\dots,N+1}}} \quad (2.24)$$

En el caso de la correlación parcial, cuando  $k = N$ , se obtiene la fórmula:

$$c_{ij:N+1} = \frac{c_{ij} - c_{i(N+1)}c_{j(N+1)}}{\sqrt{(1 - c_{i(N+1)}^2)(1 - c_{j(N+1)}^2)}} \quad (2.25)$$

donde  $c_{mn}$  son los coeficientes de correlación de Pearson de la fórmula (2.5). En este caso, y por simplicidad, denominamos a la matriz con coeficientes  $c_{ij:N+1}$  la *correlación relativa a  $X_{N+1}$* . En el contexto de series de tiempo financieras, algunos ejemplos de las series de tiempo de referencia  $X_{N+1}$  son (véanse las figuras 2.6 y 2.5):

- Rendimientos de un índice.
- Promedio, al tiempo  $t$ , sobre las  $N$  series componentes:

$$X_{N+1}(t) = \frac{1}{N} \sum_{i=1}^N X_i(t)$$

- Media geométrica, al tiempo  $t$ , sobre las  $N$  series componentes:

$$X_{N+1}(t) = \sqrt[N]{\prod_{i=1}^N X_i(t)}$$

- El máximo eigenvalor de  $\hat{\mathbf{C}}^{(\epsilon)}(t; q)$ :  $\lambda_{\text{máx}}(t)$ .
- El promedio de  $\hat{\mathbf{C}}^{(\epsilon)}(t; q)$  sobre sus entradas encima de la diagonal principal<sup>16</sup>:

$$\bar{c}(t) = \frac{2}{N(N-1)} \sum_{1 < i < j < N} \hat{c}_{ij}.$$

- Cualquier estadística de  $\hat{\mathbf{C}}^{(\epsilon)}(t; q)$ , como:

- La desviación estándar:  $\sigma_c := \sqrt{c^2 - \bar{c}^2}$ .
- La asimetría:  $\mathcal{S}_c = \frac{(c - \bar{c})^3}{\sigma_c^3}$ .
- La kurtosis:  $\kappa_c = \frac{(c - \bar{c})^4}{\sigma_c^4} - 3$ .

Los coeficientes  $c_{ij:N+1}$  tienen una interpretación geométrica provechosa: Si  $a, b, c \in \mathbb{R}^3$  son vectores linealmente independientes y  $\langle a|b \rangle = a_1b_1 + a_2b_2 + a_3b_3$  denota el producto interior de  $a$  y  $b$ , entonces el ángulo  $\alpha$  entre  $a$  y  $b$  ( $\alpha \in [0, \pi]$ ) se determina por la relación  $\cos \alpha = \langle a|b \rangle / \sqrt{\langle a|a \rangle \langle b|b \rangle} \in [-1, 1]$ . Al proyectar los vectores  $a, b$  al subespacio ortogonal a  $c$ , se obtienen vectores  $a', b'$  cuyo ángulo mutuo (digamos  $\alpha'$ ) será distinto a  $\alpha$ . En términos de  $a, b$  y  $c$ , los vectores  $a'$  y  $b'$  son:

$$a' = a - \frac{\langle c|a \rangle}{\langle c|c \rangle} c, \quad b' = b - \frac{\langle c|b \rangle}{\langle c|c \rangle} c$$

Cuando  $a, b$  y  $c$  están normalizados:

$$\langle a|a \rangle = \langle b|b \rangle = \langle c|c \rangle = 1$$

la norma (cuadrada) de  $a'$  y  $b'$  resulta:

$$\langle a'|a' \rangle = 1 - \langle c|a \rangle^2 \quad \langle b'|b' \rangle = 1 - \langle c|b \rangle^2$$

Así, el ángulo  $\alpha'$  se relaciona con los productos interiores de  $a, b$  y  $c$  de la siguiente manera:

<sup>16</sup>En este trabajo, usaremos la notación  $\overline{g(c)} = \overline{g(c)}(t)$  para representar promedios sobre las entradas encima de la diagonal de una matriz simétrica  $\mathbf{C} = \mathbf{C}(t) = (c_{ij})$  de  $N \times N$ :  $\overline{g(c)} = \frac{2}{N(N-1)} \sum_{1 < i < j < N} g(c_{ij})$ , extendiendo consistentemente la nomenclatura utilizada en fórmulas como la (2.2.1).

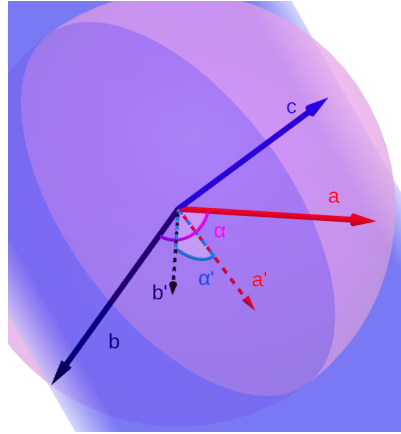


Figura 2.4: Interpretación geométrica de la correlación relativa: el coseno del ángulo  $\alpha$  entre dos vectores  $a$  y  $b$  corresponde a su correlación mutua, y si esos vectores se proyectan sobre el espacio ortogonal a un tercer vector de referencia  $b$ , sus proyecciones  $a', b'$  formarán otro ángulo  $\alpha'$  cuyo coseno corresponde a la correlación relativa de  $a$  y  $b$  con respecto a  $c$ .

$$\cos \alpha' = \frac{\langle a'|b' \rangle}{\sqrt{\langle a'|a' \rangle \langle b'|b' \rangle}} = \frac{\langle a|b \rangle - \langle a|c \rangle \langle c|b \rangle}{\sqrt{(1 - \langle a|c \rangle^2) (1 - \langle c|b \rangle^2)}} \quad (2.26)$$

que tiene la misma forma que la ecuación (2.25).

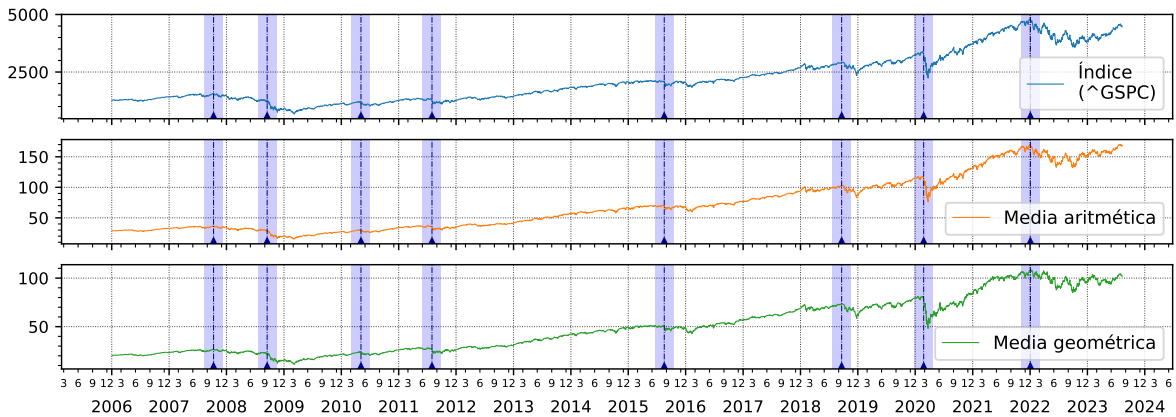


Figura 2.5: De arriba abajo: serie de tiempo de precios de cierre ajustados del índice S&P 500 (con índice bursátil o *ticker*  $\hat{GSPC}$ ), de la media aritmética, y de la media geométrica de las acciones de las compañías componentes del S&P 500 listadas en la tabla A.1, cuya inclusión en el índice fue continua de enero de 2006 a octubre de 2023. Las franjas verticales traslúcidas de color azul marcan períodos de crisis; véase la tabla A.2.

La fórmula (2.26) es válida en cualquier  $\mathbb{R}$ -espacio vectorial  $V$  con producto interior

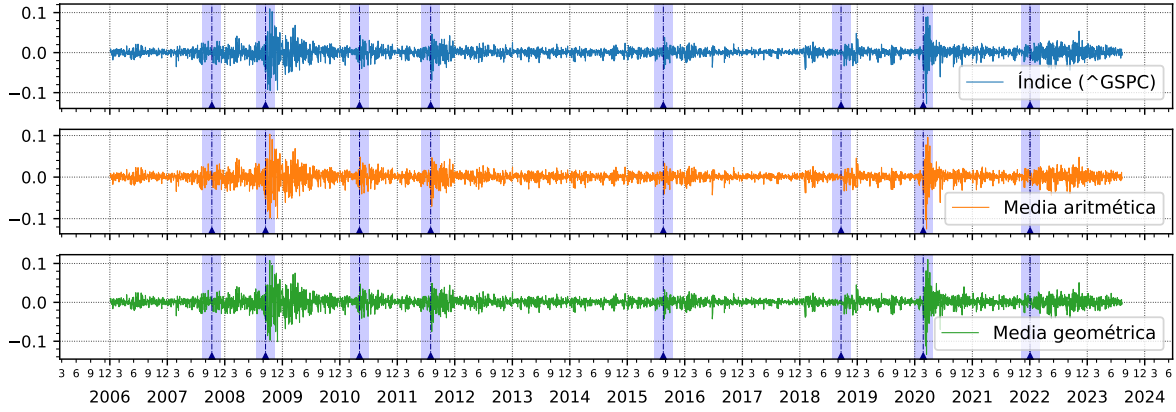


Figura 2.6: De arriba abajo: serie de tiempo de los rendimientos logarítmicos de los precios de cierre ajustados del índice S&P 500 (con índice bursátil o *ticker*  $\hat{GSPC}$ ), de la media aritmética, y de la media geométrica de las acciones de las compañías componentes del S&P 500 listadas en la tabla A.1.

$\langle \cdot | \cdot \rangle : V \times V \longrightarrow \mathbb{R}$ , con las propiedades:

$$\begin{aligned} \forall u, u', v \in V, \forall \lambda \in \mathbb{R} : \quad & \langle \lambda u + u' | v \rangle = \lambda \langle u | v \rangle + \langle u' | v \rangle \\ \forall u, v \in V : \quad & \langle u | v \rangle = \langle v | u \rangle \\ \forall u \in V : \quad & (\langle u | u \rangle \geq 0) \ \& \ (\langle u | u \rangle = 0 \implies u = \vec{0}) \end{aligned} \quad (2.27)$$

pues para éstas, es válida la desigualdad de Cauchy-Schwarz:

$$\forall u, v \in V : |\langle u | v \rangle| \leq \|u\| \|v\| \quad (2.28)$$

y, por lo tanto, la noción de ángulo:

$$\angle(u, v) := \arccos \left( \frac{\langle u | v \rangle}{\|u\| \|v\|} \right)$$

Si  $X, Y$  son variables aleatorias reales tales que  $\langle X^2 \rangle, \langle Y^2 \rangle < \infty$ , entonces  $\langle X | Y \rangle := \langle XY \rangle$  es un producto interior. Además, la correlación entre  $X$  y  $Y$  es precisamente el coseno del ángulo entre las versiones estandarizadas de  $X$  y  $Y$ :

$$\text{corr}(X, Y) = \frac{\langle X - \mu_X | Y - \mu_Y \rangle}{\|X - \mu_X\| \|Y - \mu_Y\|} \quad (2.29)$$

(donde  $\mu_X := \langle X \rangle$ ,  $\mu_Y := \langle Y \rangle$  y  $\|Z\| := \sqrt{\langle Z^2 \rangle} = \sqrt{\langle Z | Z \rangle}$ ; véase la figura 2.4).

Un procedimiento alternativo de filtración consiste en partir de la expresión diádica (2.18) de la matriz de correlación  $\hat{\mathbf{C}}$ , y truncar la suma hasta el penúltimo término:

$$\Sigma_L := \sum_{n=1}^{N-1} \lambda_n |n\rangle \langle n| \quad (2.30)$$

Esta matriz  $\Sigma_L$  tiene el mismo rango que  $\hat{\mathbf{C}}$  reducido en 1, y aunque no es una matriz de correlación, puede transformarse en una mediante la relación (2.5). En forma matricial, esto equivale a:

$$\hat{\mathbf{C}}_L = \Xi_L^{-1/2} \Sigma_L \Xi_L^{-1/2} \quad (2.31)$$

donde  $\Xi_L = \text{diag}(\Sigma_L)$  es la matriz con la parte diagonal de  $\Sigma_L$ . Esta matriz fue introducida en [84], y fue denominada *matriz de correlación de rango reducido con el método de correlación*. En el mismo artículo, también se introduce la siguiente variante: con la descomposición diádica de  $\hat{\Sigma}$ , análoga a (2.18), truncada hasta el penúltimo término (digamos  $\Sigma_B$ ) y estandarizada:

$$\hat{\mathbf{C}}_B = \Xi_B^{-1/2} \Sigma_B \Xi_B^{-1/2} \quad (2.32)$$

se obtiene la *matriz de correlación de rango reducido con el método de covarianza*, donde, como antes,  $\Xi_B := \text{diag}(\Sigma_B)$ . En este trabajo denominaremos a las matrices  $\hat{\mathbf{C}}_L$  y  $\hat{\mathbf{C}}_B$  *las proyecciones de Guhr* o *las matrices proyectadas de Guhr*, puesto que las correlaciones de series de tiempo cortas tienen rango incompleto, y es ambiguo hablar de un rango reducido ulterior. Para nuestros propósitos, los resultados con la variante de covarianza son menos atractivos.

### 2.3. Agrupamiento y reducción dimensional

En Estadística Multivariada, el análisis exploratorio de datos es el punto de partida intuitivo para entender la relación entre las variables de un sistema: valores atípicos, agrupaciones, tendencias, etc. Cuando las variables de interés son continuas, es natural pensar en su representación en coordenadas cartesianas. Fijando dos observaciones  $\mathbf{x} = (x_1, \dots, x_n)^\top, \mathbf{y} = (y_1, \dots, y_n)^\top \in \mathbb{R}^n$ , su disimilitud puede pensarse, en principio, como una relación simétrica y positiva definida, como la dada por la distancia euclidiana:

$$d_2(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} = \sqrt{(\mathbf{x} - \mathbf{y})^\top (\mathbf{x} - \mathbf{y})} \quad (2.33)$$

o más generalmente, para  $p \in [1, \infty)$ , por la distancia de Minkowski:

$$d_p(\mathbf{x}, \mathbf{y}) = \sqrt[p]{\sum_{i=1}^n |x_i - y_i|^p} \quad (2.34)$$

con el límite:

$$d_\infty(\mathbf{x}, \mathbf{y}) = \lim_{p \rightarrow \infty} d_p(\mathbf{x}, \mathbf{y}) = \max_{i=1, \dots, n} \{|x_i - y_i|\} \quad (2.35)$$

La desigualdad del triángulo que satisfacen las funciones  $d_p : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, \infty)$ :

$$d_p(\mathbf{x}, \mathbf{z}) \leq d_p(\mathbf{x}, \mathbf{y}) + d_p(\mathbf{y}, \mathbf{z})$$

hace posible la clasificación de  $m$  observaciones  $X = \{\mathbf{x}_j\}_{j=1}^m \subseteq \mathbb{R}^n$  en grupos de elementos similares entre sí (es decir: con baja disimilitud). Un algoritmo ampliamente utilizado para este propósito, tanto por su simplicidad como por su rapidez y desempeño, es el *k-means* [85, 86]. Este algoritmo, de la categoría de aprendizaje estadístico o *machine learning* no supervisado, tiene como objetivo definir una partición. Es decir: para un conjunto  $X$  con  $m \geq 1$  elementos, una partición  $\pi$  es una familia de subconjuntos no vacíos de  $X$ , disjuntos dos a dos, cuya unión es  $X$ :

$$\begin{aligned} \pi &\subseteq 2^X : \emptyset \notin \pi; \\ \bigcup \pi &= X; \\ p, q \in \pi &\Rightarrow (p \cap q = \emptyset \Leftrightarrow p \neq q). \end{aligned}$$

Los elementos de  $\pi$  se llaman *partes*. de  $X$  en un número predefinido  $k \in \mathbb{N}$  ( $k > 1$ ) de partes (grupos, cúmulos o *clusters*), de tal modo que las observaciones en cada cúmulo sean lo más similares entre sí, a la vez que observaciones en grupos distintos sean tan disímiles como sea posible.

A grandes rasgos, el algoritmo k-means inicializa  $k$  *centroides*  $\{\tilde{\mathbf{x}}_j\}_{j=1}^k$ , usualmente escogidos aleatoriamente de entre los miembros de  $X$ ; después, se llevan a cabo los siguientes pasos de forma recursiva:

1. Cada  $\mathbf{x}_i \in X$  se asocia con el centroide más cercano en términos de una distancia (normalmente la euclidiana), resultando  $k$  partes  $\{C_j\}_{j=1}^k$ .
2. Los centroides se actualizan a la media de los miembros de su grupo correspondiente:  

$$\tilde{\mathbf{x}}_j \mapsto \frac{1}{|C_j|} \sum_{\mathbf{x} \in C_j} \mathbf{x} \quad (j = 1, \dots, k).$$

El procedimiento se detiene ya sea cuando los centroides no cambian sus posiciones más allá de una tolerancia dada, o bien cuando se alcanza un número máximo de iteraciones. El criterio de optimización es minimizar la *suma de los cuadrados de las desviaciones* o *ESS*<sup>17</sup>, también llamada *inerencia*:

$$\text{ESS} = \sum_{j=1}^k \sum_{\mathbf{x}_i \in C_j} (d_2(\mathbf{x}_i, \tilde{\mathbf{x}}_j))^2 \quad (2.36)$$

Algunas de las posibles desventajas del K-means son la sensibilidad a las condiciones iniciales<sup>18</sup>; su tendencia a producir *clusters* de tamaños similares o a alcanzar mínimos locales [87, págs. 563-564][79, págs. 423-424]. Por otra parte, el número de cúmulos  $k$  se debe establecer *a priori*, y no hay un criterio único para elegirlo [41, 42, 88].

El *escalamiento multidimensional* o *MDS*<sup>19</sup> es una técnica de reducción dimensional, cuyo objetivo es asociarle a un conjunto de  $m$  observaciones  $X = \{\mathbf{x}_j\}_{j=1}^m \subseteq \mathbb{R}^n$  otro conjunto del mismo tamaño pero en dimensión menor:  $X' = \{\mathbf{x}'_j\}_{j=1}^m \subseteq \mathbb{R}^{n'}$  ( $n' < n$ ), de tal manera que la matriz de similaridad:

$$\eta_p(i, j) := d_p(\mathbf{x}_i, \mathbf{x}_j), \quad 1 \leq i \leq j \leq m \quad (2.37)$$

con las  $m^2$  distancias entre observaciones del espacio original, se reproduzca lo mejor posible en el espacio reducido:

$$\eta_p'(i, j) = d_p(\mathbf{x}'_i, \mathbf{x}'_j), \quad 1 \leq i \leq j \leq m$$

Normalmente, el MDS se emplea con  $n' \in \{2, 3\}$  para posibilitar la visibilidad de la estructura subyacente en las observaciones  $X \subseteq \mathbb{R}^n$  con  $n \geq 4$  [89, Cap. 12.6].

Hay principalmente dos tipos de MDS [90]:

- **Métrico:** Las distancias del espacio reducido son parecidas a las del espacio completo, módulo una constante multiplicativa  $b > 0$ :  $\eta_p' \approx b\eta_p$ . El criterio de optimización es

<sup>17</sup>Error sum of squares.

<sup>18</sup>Aunque esto también puede ser una ventaja, pues se puede crear un *ensemble* de agrupaciones variando las condiciones iniciales, añadiendo robustez estadística.

<sup>19</sup>Multidimensional Scaling.

minimizar la *función de estrés*:

$$\text{Stress} = \sum_{i < j} \eta_p(i, j) - \eta_p'(i, j)$$

- **No métrico:** Las distancias ordenadas del espacio reducido son una función monótona creciente de las distancias ordenadas del espacio completo:

$\eta_p(i, j) \leq \eta_p(k, l) \implies \eta_p'(i, j) \leq \eta_p'(k, l)$ . El criterio de optimización es minimizar la *función de estrés normalizada*:

$$\text{Stress}_1 = \sqrt{\frac{\sum_{i < j} (\eta_p(i, j) - \eta_p'(i, j))^2}{\eta_p^2(i, j)}}$$

En este trabajo consideraremos la variante métrica, con  $p = 2$ .

### 2.3.1. Estados de mercado

Habiendo fijado un tamaño de época  $q$ , un valor para el parámetro  $\epsilon$  de la función (2.16), y aprovechando la simetría de las correlaciones, la secuencia de  $T - q + 1$  matrices

$\{\hat{\mathbf{C}}^{(\epsilon)}(t; q) : t = q, \dots, T\}$  de la matriz de datos  $\hat{\mathbf{X}} \in \mathbb{R}^{N \times T}$  puede pensarse como una muestra de una trayectoria en  $\mathbb{R}^{N(N-1)/2}$  a través del *aplanamiento*:

$$\begin{aligned} \hat{\mathbf{C}} = (\hat{c}_{ij}) &\mapsto \vec{c} = (c^{(1)}, \dots, c^{(N \times (N-1)/2)}) \in \mathbb{R}^{N \times (N-1)/2} \\ \forall i < j : \quad \hat{c}_{ij} &= c^{\binom{Ni+j-(i+2)(i+1)}{2}} \end{aligned} \quad (2.38)$$

donde las  $N \times (N - 1) / 2$  entradas  $c_{ij}$  de la parte triangular superior (sin la diagonal) de cada  $\hat{\mathbf{C}}$  corresponden a las componentes de un vector  $\vec{c}$ .

Las distintas regiones visitadas por  $\hat{\mathbf{C}}(t; q)$  en este espacio de correlaciones no-redundantes, resultan estar organizadas en *clusters*, conocidos en Econofísica como *estados de mercado* [40]. (Hay que distinguir esta noción de estado de mercado (*market state*) y *estado del mercado* (*state of the market*), dado por el mayor eigenvalor de la matriz de correlación, tal como se utiliza, por ejemplo, en [15]).

Parte de esa estructura geométrica es reflejada por la matriz de similaridad (2.37) cuando  $(i, j) = (t, t')$  para tiempos  $t, t' \in \{q, \dots, T\}$  (véase la Figura 2.7), que en el contexto de Sistemas Dinámicos es conocida como *gráfica de recurrencia* [91]. En el artículo de [40] se usa una variante de *clustering* disociativo o *top-down* combinada con K-means, en lugar de fijar de antemano el número de cúmulos: las  $m$  matrices locales de correlación, tomadas en épocas de dos meses sin traslape, conforman un *cluster* inicial que, tras haber escogido una de las  $\binom{m}{2}$  posibles parejas de entre aquéllas, es dividido en dos grupos con el K-means descrito en 2.3 y una tolerancia de 0.1465 veces la distancia promedio entre clusters. Aquella división que minimice una distancia intra-cluster promedio (proporcional a una modificación de la fórmula (2.36), con  $p = 1$ ) se fija como óptima y el procedimiento se repite recursivamente para cada subcluster resultante, hasta que se obtengan tantos cúmulos como  $m$ .

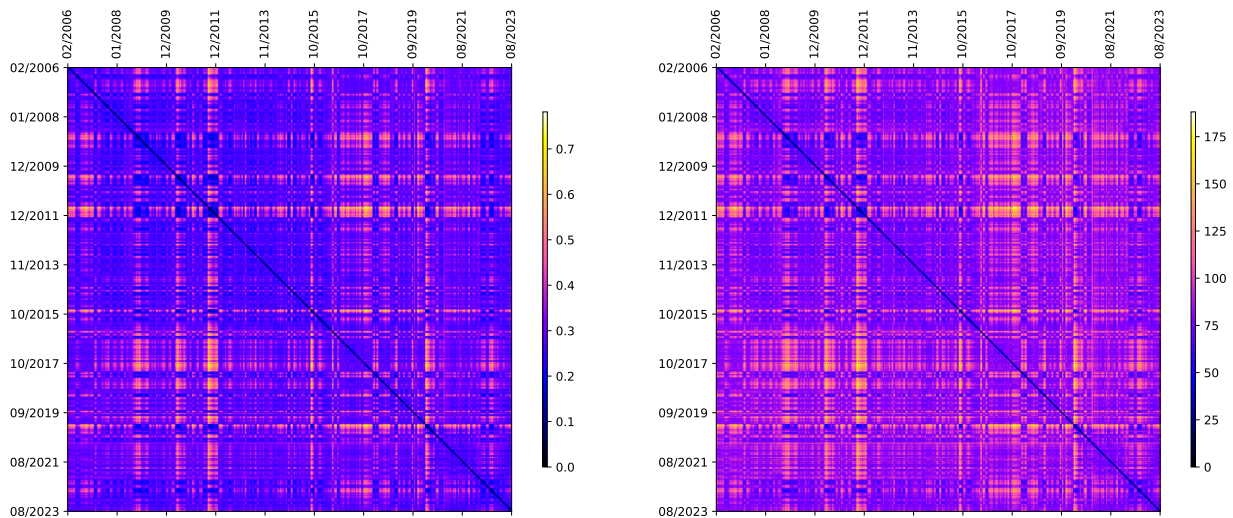


Figura 2.7: Matrices de similaridad  $\eta_p(t, t') = d_p(\vec{c}(t; q), \vec{c}(t'; q))$  de la secuencia de matrices locales de correlación de Pearson  $\{\hat{C}^{(\epsilon)}(t; q) : t = q, \dots, T\}$ , calculadas a partir de los  $T = 4420$  rendimientos de precios de cierre ajustado de  $N = 322$  acciones de compañías estadounidenses listadas en el Apéndice A, cuando se consideran épocas sucesivas de  $q = 20$  días bursátiles, con 19 días de traslape, y  $\epsilon = 0$ . Las métricas utilizadas son  $p = 1$  (lado izquierdo) y  $p = 2$  (lado derecho). En cada subgráfica el código de color se normaliza de acuerdo con el rango de datos correspondiente.

Distintos estudios [41, 42, 72, 92-94] han profundizado al respecto, enfatizando los estados de mercado como atractores en los que transita la matriz de correlación, pasando de uno a otro en una dinámica cuasi-estacionaria reflejando no sólo una colectividad de las compañías, sino también su organización en sectores industriales interna y externamente relacionados, de acuerdo con los altibajos de los mercados de valores internacionales. Lo mismo parece ocurrir en otros sistemas organizados en redes, como el cerebro [95], en cuya actividad (medida a partir de señales EEG<sup>20</sup> o fMRI<sup>21</sup> tomadas de pacientes en distintas situaciones: jugando, durmiendo, antes o durante un ataque epiléptico, etc.) se pueden distinguir estados o regímenes de colectividad [96, 97].

Otro estudio de relevancia especial es el de Pharasi *et al* [41], donde los parámetros  $(k, \epsilon)$  del K-means y el *power map* (2.16) se varían para minimizar la varianza de (2.36) para distintas condiciones iniciales, en la representación tridimensional de las matrices de correlación locales (calculadas sobre épocas con  $q = 20$  y 50% de traslape), dada por el MDS métrico con  $p = 1$ . Esto se hace para datos de los mercados estadounidense y japonés, de rendimientos de precios de cierre ajustado de acciones de compañías representativas, en un lapso de 1985 a 2016; los estados de mercado obtenidos son  $k = 4$  para Estados Unidos y  $k = 5$  para Japón. Además, en el trabajo mencionado se analizan las transiciones entre distintos estados a partir de la serie de tiempo:

$$S(t) = j \text{ si la matriz dimensionalmente escalada } \hat{C}_{3D}^{(\epsilon)}(t; q) \text{ está en el } j\text{-ésimo cúmulo} \quad (2.39)$$

donde  $j \in \{1, \dots, k\}$  y los estados están ordenados de forma ascendente con la correlación promedio de los miembros del cúmulo correspondiente.

<sup>20</sup>Electroencefalografía.

<sup>21</sup>Functional magnetic resonance imaging.



Las transiciones:

$$M(i, j) := |\{t : S(t) = i \text{ y } S(t+1) = j\}| \quad (2.40)$$

resultan aproximadamente simétricas ( $M(i, j) \approx M(j, i)$ ) y son más frecuentes entre estados adyacentes  $i, j$ . Además, la forma de la matriz  $\mathbf{W}$  con elementos  $w_{ij}$  dados por:

$$w_{ij} := \frac{M(j, i)}{|\{t : S(t) = i\}|} \quad (2.41)$$

(es decir: cada renglón de la matriz  $M(i, j)$  se divide entre la suma de sus elementos) es consistente con la suposición de que  $S = S(t)$  es una cadena de Markov estacionaria [12, cap. 1] en un espacio de estados finito  $\mathcal{S} = \{1, \dots, k\}$ :

$$\mathbb{P}(S(t+1) = s_{t+1} | S(0) = s_0, \dots, S(t) = s_t) = \mathbb{P}(S(t+1) = s_{t+1} | S(t) = s_t)$$

(donde  $t \in \mathbb{N}$ ,  $\{s_j\}_{j=0}^{t+1} \subseteq \mathcal{S}$  y  $\mathbb{P}(S(t+1) = y | S(t) = x)$  es independiente de  $t$ ), pues se tiene la relación:

$$\forall i \in \{1, \dots, k\} : \sum_{j=1}^k w_{ij} = 1$$

para los renglones de la matriz  $\mathbf{W} = (w_{ij})$ . Así, es posible determinar un vector de probabilidades estacionarias  $\boldsymbol{\pi}^* = (\pi_1^*, \dots, \pi_k^*)$  si identificamos  $w_{ij}$  con la probabilidad de ocupar el estado  $S(t+1) = j$  dado que el estado previo es  $S(t) = i$  a partir de la condición:

$$\boldsymbol{\pi}^* \mathbf{W} = \boldsymbol{\pi}^* \quad (2.42)$$

que es, en nuestro caso, equivalente a:

1.  $\boldsymbol{\pi}^* = \lim_{n \rightarrow \infty} \boldsymbol{\pi}^{(0)}$  para cualquier distribución inicial<sup>22</sup>  $\boldsymbol{\pi}^{(0)} = (\pi_1^{(0)}, \dots, \pi_k^{(0)})$ .
2.  $\boldsymbol{\pi}^{*\top}$  es el eigenvector (apropiadamente normalizado) de la matriz  $\mathbf{W}^\top$ , correspondiente al eigenvalor 1.

Nótese que la reducción dimensional indicada en (2.39) no es necesaria para definir la matriz  $\mathbf{W} = (w_{ij})$ , pues el K-means puede aplicarse en el espacio completo de dimensión  $N(N-1)/2$  de las matrices de correlación locales en el tiempo (véase la relación (2.3.1)). Este será nuestro abordaje en los siguientes capítulos.

---

<sup>22</sup>Es decir:  $0 \leq \pi_j^{(0)} \leq 1$  para cada  $1 \leq j \leq k$ , y  $\sum_{j=1}^k \pi_j^{(0)} = 1$ .

## Capítulo 3

# Correlaciones relativas y proyecciones de Guhr

En este capítulo utilizaremos matrices de correlación relativa y proyecciones de Guhr, descritas en el Capítulo 2, para estudiar los estados de mercado. Compararemos la dinámica de las correlaciones locales en el tiempo en las formulaciones de Pearson, relativa y proyectadas de Guhr, a través del análisis de sus trayectorias en  $\mathbb{R}^{N(N-1)/2}$  (véase la relación (2.3.1)), de sus estados recurrentes (entendidos como los cúmulos visitados en  $\mathbb{R}^{N(N-1)/2}$ ) y sus transiciones. Una parte de los resultados será publicada en [98], y otra ya se encuentra disponible en [99].

### 3.1. Preliminares

La figura 3.1 muestra las matrices de correlación de horizonte completo, en los casos de Pearson (fórmula (2.7); subfigura 3.1(a)), proyecciones de Guhr (fórmulas (2.31), (2.32); subfiguras 3.1(b), 3.1(c), respectivamente) y relativas (fórmula (2.25), subfiguras 3.1(d), 3.1(e), 3.1(f)) con respecto a distintas series de tiempo representativas. Es notable la similitud entre las correlaciones relativas y las proyecciones de Guhr en el caso de horizonte completo (es decir, tomando los datos sobre todo el tiempo), y en cómo los sectores UT y EG (véase la tabla A.1) mantiene su coloración azul intensa en todos los casos. Dicha similitud se verifica cuantitativamente por la tabla 3.1 y el histograma de la figura 3.2. Sin embargo, veremos más adelante que el análisis de correlaciones locales en el tiempo muestra diferencias importantes.

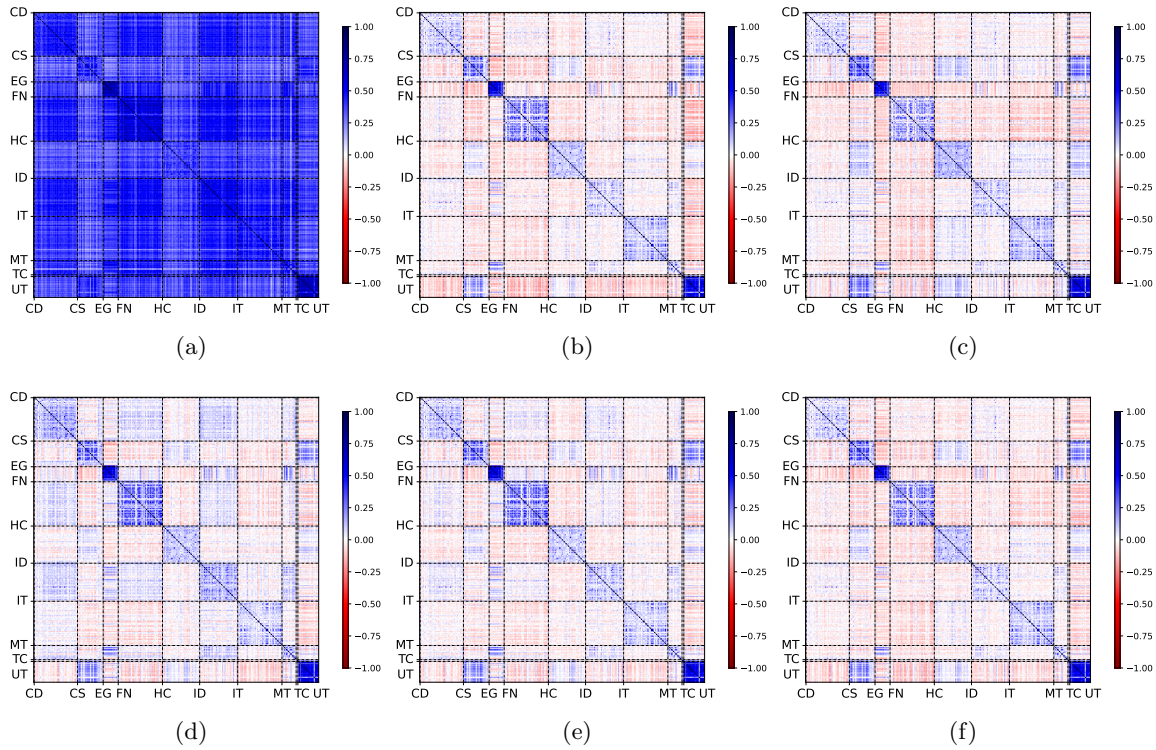


Figura 3.1: De izquierda a derecha y de arriba abajo: Matrices de correlación de Pearson [3.1\(a\)](#), proyecciones de Guhr ([3.1\(b\)](#)): método de correlación; [3.1\(c\)](#): método de covarianza) y relativas (a la serie de tiempo del índice: [3.1\(d\)](#); de la media aritmética: [3.1\(e\)](#); de la media geométrica: [3.1\(f\)](#)), calculadas para las 322 compañías del S&P 500 listadas en el apéndice (ver tabla [A.1](#)), sobre el horizonte temporal completo: del 3 de enero de 2006 al 10 de octubre de 2023.

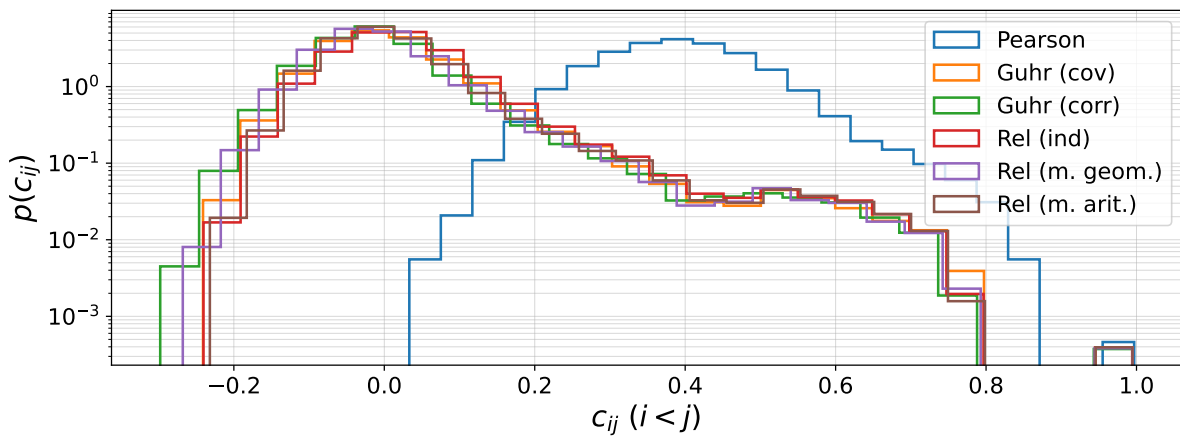


Figura 3.2: Histogramas de los elementos de matriz no diagonales de las matrices de correlación de horizonte temporal completo de la figura [3.1](#). Los *bins* de estos histogramas se calcularon con el [método de Doane](#). Las estadísticas básicas de estos datos se encuentran en la tabla [3.1](#).

	Pearson	Guhr (cov.)	Guhr (corr.)	Relativa (índice)	Relativa (media geom.)	Relativa (media arit.)
$\bar{c}$	0.394	0.010	-0.003	0.026	-0.000	0.012
$\sigma_c$	0.102	0.101	0.099	0.101	0.098	0.099
$\mathcal{S}_c$	0.432	2.144	2.281	2.065	2.378	2.443
$\kappa_c$	0.804	9.332	10.419	8.751	10.804	10.864
$\min\{c_{ij}\}_{1 \leq i < j \leq N}$	0.033	-0.241	-0.298	-0.240	-0.268	-0.232
$Q_1$	0.325	-0.051	-0.058	-0.033	-0.057	-0.045
$Q_2$	0.389	-0.004	-0.015	0.014	-0.014	-0.003
$Q_3$	0.456	0.049	0.031	0.065	0.033	0.046
$\max\{c_{ij}\}_{1 \leq i < j \leq N}$	0.997	0.995	0.995	0.994	0.995	0.995

Cuadro 3.1: Estadísticas básicas de las matrices de correlación de horizonte temporal completo de la figura 3.1: correlación promedio  $\bar{c}$ , desviación estándar  $\sigma_c := \sqrt{c^2 - \bar{c}^2}$ , asimetría (*skewness*)  $\mathcal{S}_c$ , kurtosis  $\kappa_c$ , mínimo; primer, segundo y tercer cuartiles  $Q_1$ ,  $Q_2$  y  $Q_3$ ; y máximo.

### 3.2. Correlación local en el tiempo

Para la secuencia de  $T - q + 1 = 4411$  matrices de correlación (ya sean de Pearson, relativas o proyecciones de Guhr)  $\mathcal{C} := \{\hat{\mathbf{C}}^{(\epsilon)}(t; q) : t = q, \dots, T\}$  (cuyo miembro  $\hat{\mathbf{C}}^{(\epsilon)}(t = 10/11/2016; q)$  se muestra en la figura 3.3(a) para las formulaciones de Pearson y relativa a la serie de tiempo del índice) calculadas a partir de los rendimientos logarítmicos de los precios de las  $N = 322$  compañías del S&P 500 listadas en el apéndice (ver tabla A.1), sin *power map* (es decir:  $\epsilon = 0$ ) y en épocas sucesivas de tamaño  $q = 20$ , hay distintas estadísticas en forma de series de tiempo, que dan cuenta del *estado del mercado*. En las siguientes subsecciones, daremos cuenta de algunas de ellas.

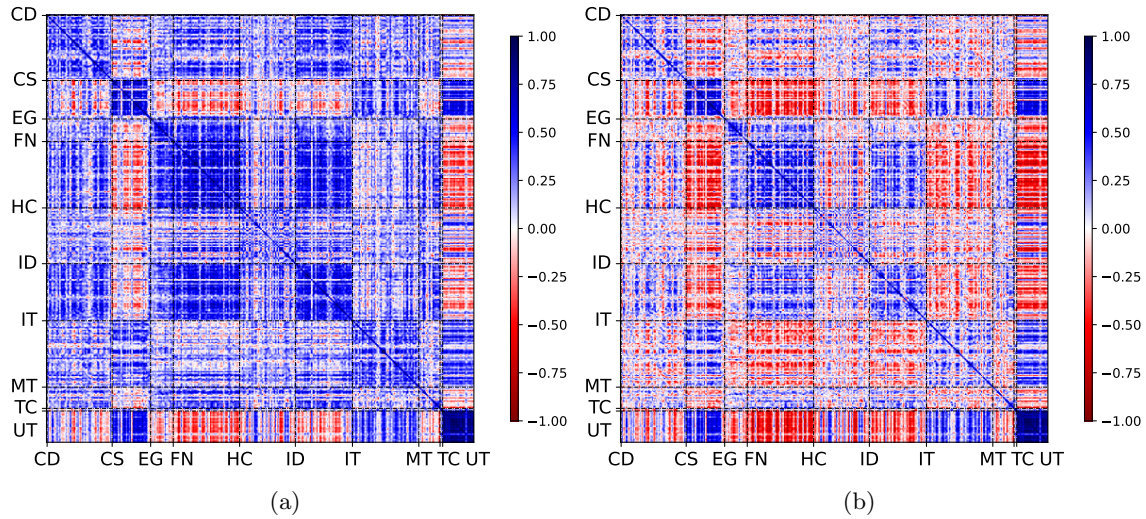


Figura 3.3: Matrices de correlación de Pearson (subfigura 3.3(a)) y relativa a la serie de tiempo del índice (subfigura 3.3(b); véase la fórmula (2.25)) en una época con longitud de  $q = 20$  días bursátiles que finaliza en el tiempo  $t = 10/11/2016$ . Nótese la diferencia en el balance de correlaciones (color azul) y anticorrelaciones (color rojo). La distribución de los elementos no diagonales de esta matriz se muestra en la figura 3.4.

Una forma conveniente de visualizar la evolución global de las correlaciones en su conjunto, es mediante un mapa de calor. En la figura 3.5 se muestra la distribución de los elementos de matriz, calculada con el método *Kernel Density Estimation*, KDE (véase el Apéndice B). El eje horizontal corresponde al tiempo; el eje vertical corresponde a distintos valores de la correlación de Pearson (subfigura 3.5(a)) y relativa al índice (subfigura 3.5); el código de color representa la intensidad con la que se distribuyen las correlaciones en el intervalo  $[-1, 1]$ .

#### Momentos de las matrices locales de correlación

Otra forma de visualizar el estado del mercado, es mediante el máximo eigenvalor de las matrices de correlación, y los momentos centrales (como el promedio y la desviación estándar) de los elementos de matriz sin considerar los 1's de la diagonal principal (véase la lista 2.2.3). Tales métricas se muestran en la figura 3.6, donde no solamente se observa que los valles y crestas de las series de tiempo son particularmente pronunciados en épocas críticas, sino además que  $\lambda_{\text{máx}}$  y  $\bar{c}$

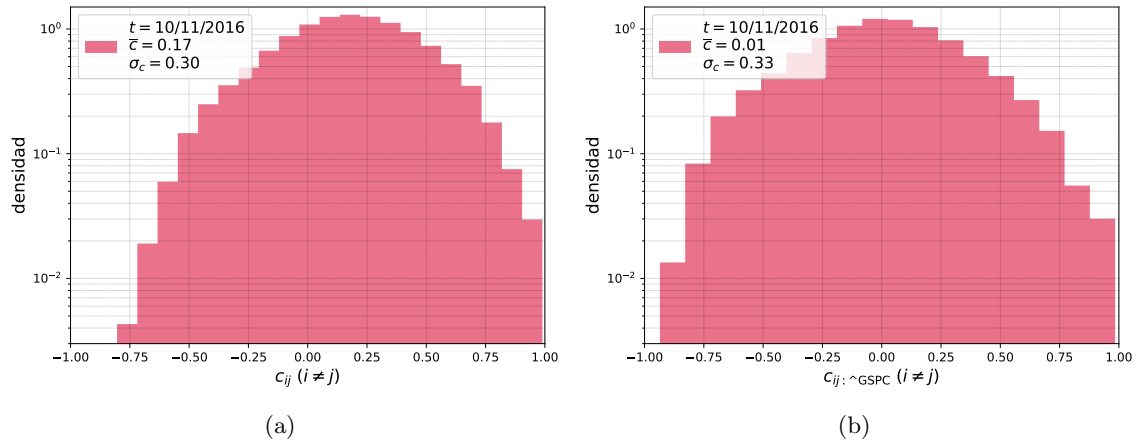


Figura 3.4: Distribución de los elementos no diagonales de las matrices de correlación de la figura 3.3. Dentro de cada marco se despliega el promedio y desviación estándar de los datos correspondientes, que a su vez pueden graficarse como función de  $t$ ; véanse las figuras 3.6 y 3.7. Nótese la asimetría en la distribución de la subfigura 3.4(a) (lo cual es típico en todas las épocas; véase la figura 3.8(a)) en contraste con la de la subfigura 3.4(b), así como la elección de la escala vertical. Los *bins* de este histograma se calcularon con el método de Doane.

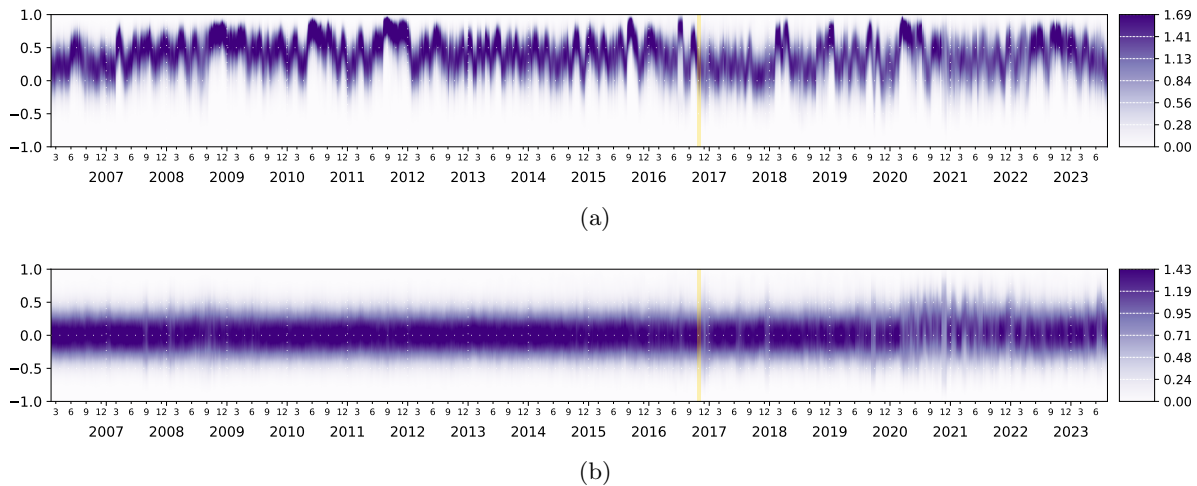


Figura 3.5: Representación de la dinámica de las correlaciones mediante mapas de calor. Para cada  $t$ , las entradas no diagonales de la matriz de correlación tienen una cierta distribución, cuya densidad de probabilidad puede obtenerse con el KDE. La densidad correspondiente, soportada en el intervalo  $[-1, 1]$ , se marca con púrpura en distintas intensidades que varían tanto a lo largo del eje vertical, como del eje horizontal. La banda vertical delgada de color amarillo, ubicada alrededor de noviembre de 2016, marca la época de la figura 3.3, donde la media de la distribución es  $\bar{c} \approx 0.2$  para la subfigura 3.5(a), y  $\bar{c} \approx 0.0$  para la subfigura 3.5(b). Cabe resaltar que la escasez de correlaciones negativas en la subfigura 3.5(a) queda marcada por la asimetría de las regiones claras encima y debajo de la línea horizontal  $c = 0$ .

están fuertemente correlacionadas (con un coeficiente de correlación global aproximadamente igual a 0.991; véase el primer renglón, segunda columna de la figura 3.10). También es notable una fuerte anticorrelación entre  $\bar{c}(t)$  y  $\sigma_c = \sigma_c(t)$ , aproximadamente igual a  $-0.68$  (véase el segundo

renglón, tercera columna de la figura 3.10).

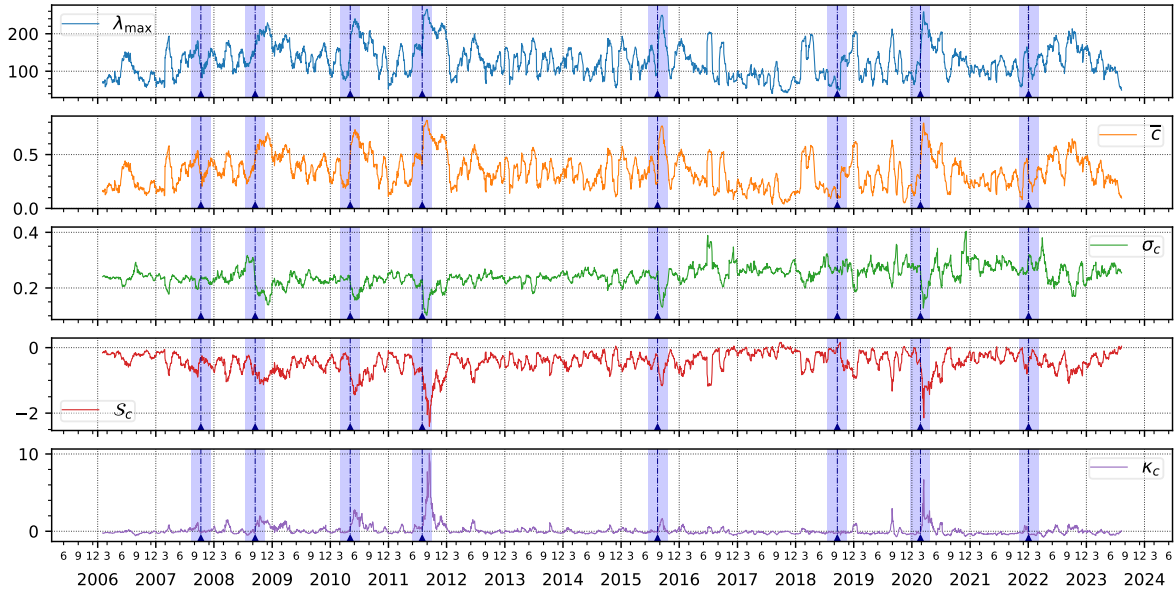


Figura 3.6: Series de tiempo del mayor eigenvalor  $\lambda_{\text{máx}}$ , la correlación promedio  $\bar{c}$ , la desviación estándar  $\sigma_c := \sqrt{c^2 - \bar{c}^2}$ , así como su asimetría (*skewness*)  $\mathcal{S}_c$  y kurtosis  $\kappa_c$ , de las matrices de correlación de Pearson locales en el tiempo. Las franjas verticales traslúcidas de color azul marcan períodos de crisis; véase la tabla A.2.

Un caso de interés particular es el de la región a finales de febrero de 2020 (penúltima franja azul traslúcida de cada subgráfica en la figura 3.6), donde la correlación promedio  $\bar{c} = \bar{c}(t)$  llega a un valor de aproximadamente 0.8, y la desviación estándar  $\sigma_c = \sigma_c(t)$  baja hasta cerca de 0.15; ésto indica un fuerte acoplamiento en el comportamiento de las acciones de las compañías del S&P 500, y es consistente tanto con una caída fuerte del índice correspondiente, como con su alta volatilidad (traducida en un frenesí de la serie de tiempo de los rendimientos del índice).

También en la figura 3.6 es notable el comportamiento de la kurtosis  $\kappa_c = \kappa_c(t)$  (último renglón en la gráfica referida). Se sabe que esta métrica es nula para datos con distribución normal, y aquél es el caso para esta serie de tiempo. Aquellas épocas donde la kurtosis es grande, las colas de la distribución de los datos son pesadas, y para las matrices de Pearson eso se cumple para  $\bar{c}$  cercana a 1.

De manera análoga a la figura 3.6, la figura 3.7 muestra algunas estadísticas de la correlación relativa promedio. Los valles y crestas de las series de tiempo en los renglones primero y segundo de esa figura, no tienen la misma pronunciación que en el caso de las correlaciones de Pearson, salvo en el lapso de 2020 a 2022, donde hay fluctuaciones altas. Por otra parte, hay en todo el horizonte de tiempo una correlación menor (en comparación con las correlaciones de Pearson), pero todavía positiva, entre  $\bar{c}(t)$  y  $\lambda_{\text{máx}}(t)$  de aproximadamente 0.709, mientras que la correlación global entre  $\bar{c}(t)$  y  $\sigma_c(t)$  se vuelve positiva e igual a 0.677.

Para medir la presencia de correlaciones negativas, calculamos  $|\bar{c}| = |\bar{c}(t)|$ ,  $\bar{c} = \bar{c}(t)$  (primer renglón de las figuras 3.8(a) y 3.8(b)), así como su diferencia (segundo renglón de las mismas subfiguras). La mayor simetría entre correlación y anticorrelación en las matrices relativas al

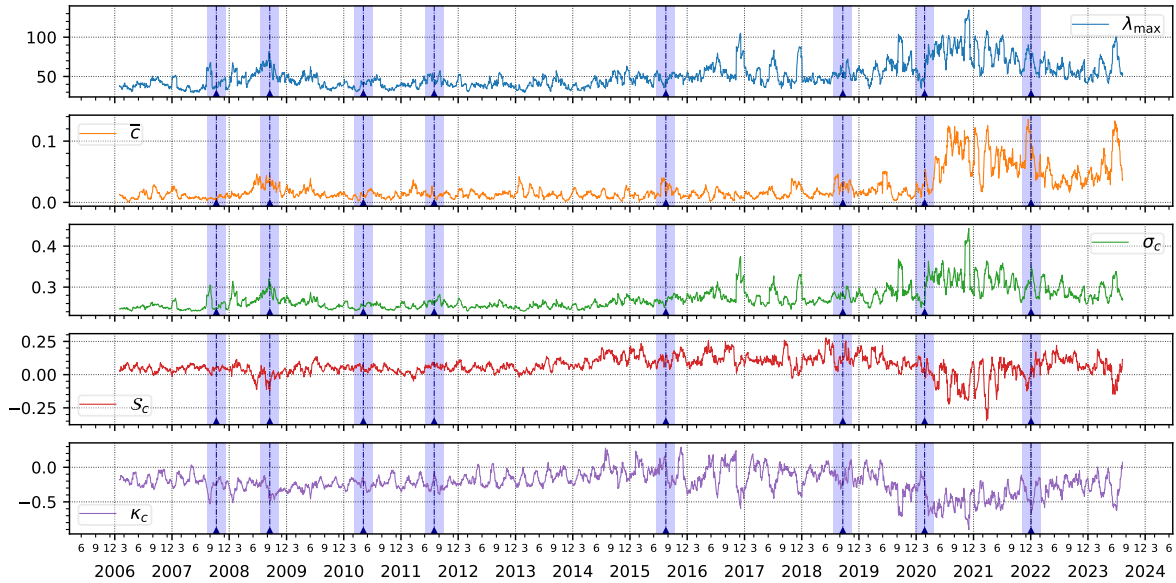


Figura 3.7: Series de tiempo del mayor eigenvalor  $\lambda_{\text{máx}}$ , la correlación promedio  $\bar{c}$ , la desviación estándar  $\sigma_c := \sqrt{c^2 - \bar{c}^2}$ , así como su asimetría (*skewness*)  $\mathcal{S}_c$  y kurtosis  $\kappa_c$ , de las matrices de correlación relativas al índice, locales en el tiempo. Las franjas verticales traslúcidas de color azul marcan períodos de crisis; véase la tabla A.2.

índice se refleja en los mínimos de la serie de tiempo  $|\bar{c}| - \bar{c}(t)$  de la figura 3.8(b), y en la centralidad de la figura 3.5(b).

### Componentes principales

También son de interés las estadísticas inducidas por los eigenvectores principales de la secuencia  $\mathcal{C}^{\text{Pearson}} := \left\{ \hat{\mathbf{C}}^{(\epsilon)}(t; q) : t = q, \dots, T \right\}$ . En la figura 3.9 se muestran la razón de participación  $\text{PR}_N = \text{PR}_N(t)$  (ecuación (2.19)) y la entropía de las componentes al cuadrado del eigenvector principal de cada matriz  $\hat{\mathbf{C}}^{(\epsilon)}(t; q)$ ,  $\text{H}_N = \text{H}_N(t)$  (ecuación (2.20)). La alta correlación entre  $\text{PR}_N(t)$  y  $\bar{c}(t)$ , así como entre  $\text{H}_N(t)$  y  $\bar{c}(t)$ , (ver renglón 2, columnas 6 y 7 de la gráfica 3.10), pone en evidencia que una mayor inestabilidad o criticidad del mercado, ya reconocida en la literatura como función de un mayor acoplamiento en la dinámica de los rendimientos de las acciones (alta correlación), también está ligada a valores altos de  $\text{PR}_N(t)$  y  $\text{H}_N(t)$ , que en las ecuaciones (2.19) y (2.20) de la sección 2.2.2 fueron reconocidos como estados colectivos (véase también la figura 3.9). Por otro lado, la heterogeneidad de las correlaciones corresponde a mayores valores para  $\sigma_c(t)$ , lo que parece ser una condición necesaria (mas no suficiente) para la estabilidad del mercado (ver renglón 2, columna 3 de la figura 3.10).

En el caso de las correlaciones de Pearson, las líneas horizontales en los renglones primero y segundo marcan las cotas superior (color negro) e inferior (color rojo) correspondientes a los valores de correlación constante y WOE, respectivamente. En cambio, para las correlaciones relativas al índice, solamente se muestran las líneas horizontales en los renglones primero y segundo para las cotas inferiores (color rojo) correspondientes a los valores de WOE, pues las



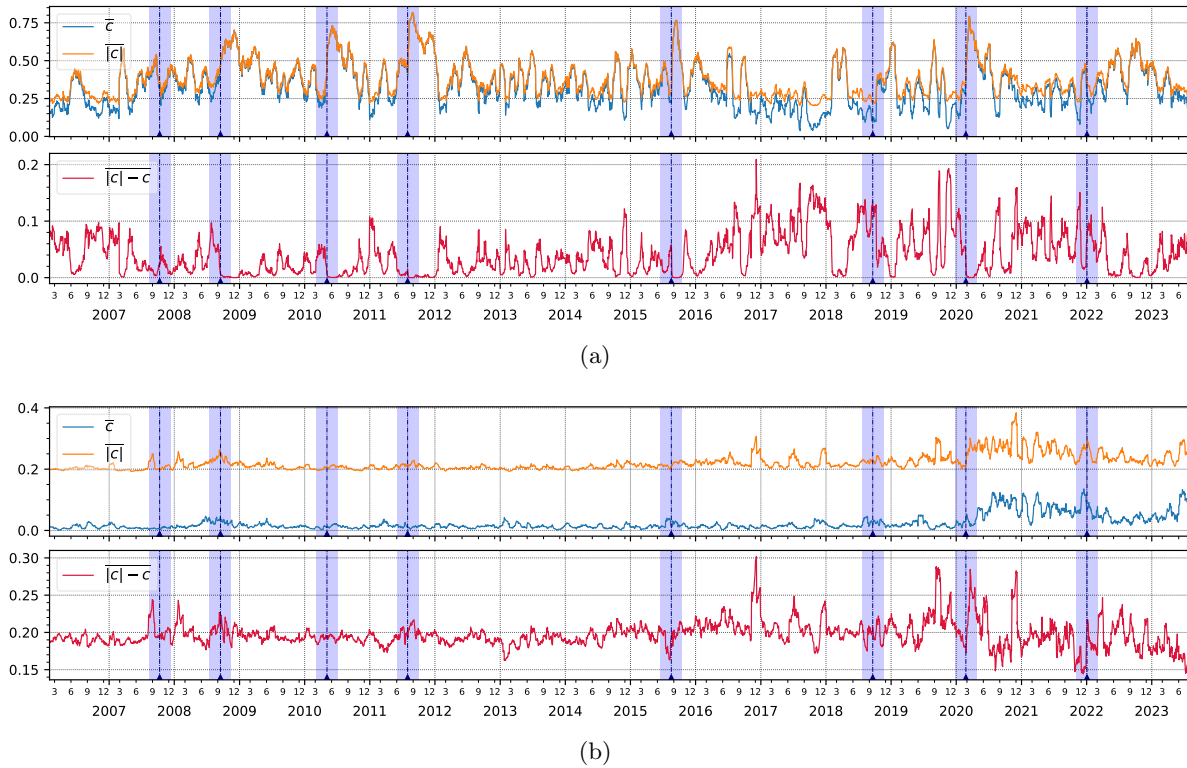


Figura 3.8: Tomando en cuenta las  $T - q + 1 = 4411$  matrices de correlación, los valores promedio de  $|c| - \bar{c}$  sobre  $t$  son de aproximadamente 0.041 (Pearson, 3.8(a)) y 0.197, (relativa al índice, 3.8(b)), con desviaciones estándar de 0.037 (Pearson) y 0.017 (relativa), y cuartiles de 0.011, 0.032 (mediana) y 0.063 (Pearson), y 0.188, 0.195 (mediana) y 0.204 (relativa).

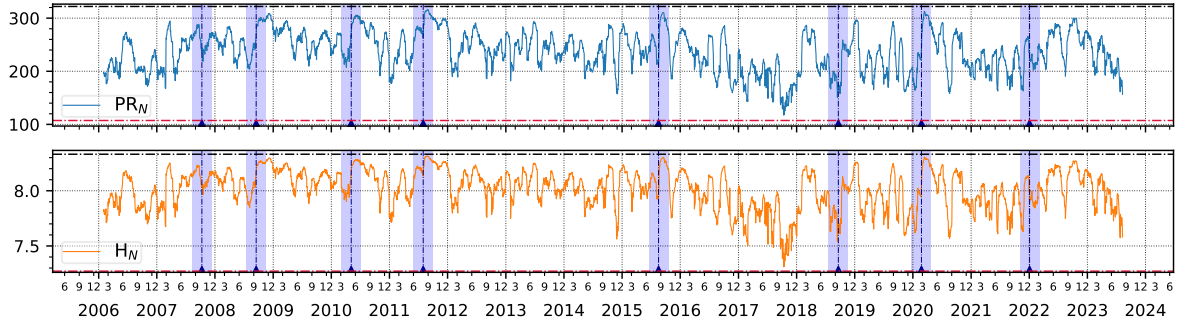
cotas superiores correspondientes a correlación constante están muy por encima de los correspondientes rangos de variación.

Los gráficos de dispersión de  $\lambda_{\text{máx}}$ ,  $\bar{c}$ ,  $\sigma_c$ ,  $\mathcal{S}_c$ ,  $\kappa_c$ ,  $\text{PR}_N$  y  $\text{H}_N$  se muestran en las figuras 3.10 y 3.11. En éstas figuras, se incluye un código de color para representar el tiempo, y una diagonal con la densidad empírica de cada serie de tiempo, calculada con el método *kernel density estimation* (KDE; véase el Apéndice B).

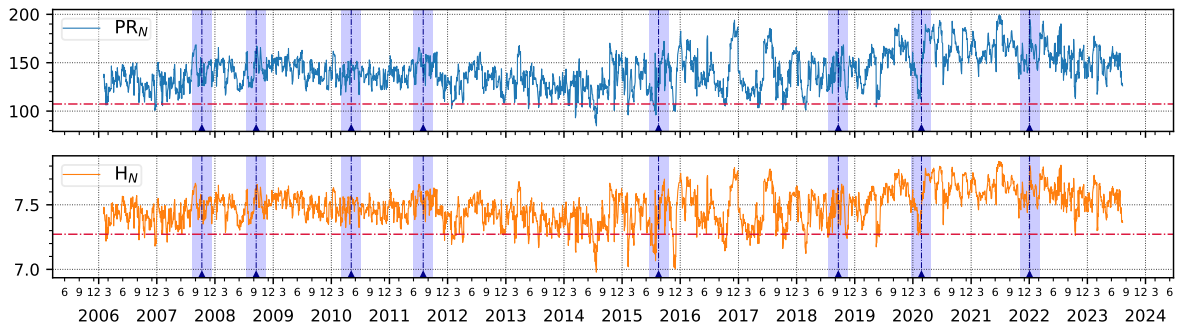
Estos gráficos permiten diferenciar con mayor detalle la geometría de las matrices de correlación relativa, y la de las matrices de correlación de Pearson. Un ejemplo importante es el de la dispersión del mayor eigenvalor con la desviación estándar (véase el renglón 3, columna 2 de dichas figuras): en Pearson, una subida de  $\lambda_{\text{máx}}$  (o equivalentemente, de  $\bar{c}$ ) es acompañada por una bajada de  $\sigma_c$ ; en cambio, para las correlaciones relativas, una subida de  $\lambda_{\text{máx}}$  es acompañada también por una subida de  $\sigma_c$ . Esto tiene implicaciones en la clasificación de los estados del mercado, o de los cúmulos de matrices de correlación locales, que veremos más adelante.

### Eigenvalores dominantes

En las matrices de correlación de Pearson, el eigenvalor máximo  $\lambda_{\text{máx}}$  es lo que define el estado del mercado, y los siguientes tres eigenvalores dominantes parecen ser una copia especular



(a)



(b)

Figura 3.9: Series de tiempo de la razón de participación del eigenvector principal  $PR_N$  y entropía de Shannon del (módulo al cuadrado de las componentes del) eigenvector principal  $H_N$ , para las correlaciones locales de Pearson (subfigura 3.9(a)) y relativas al índice (subfigura 3.9(b)). Las franjas verticales traslúcidas de color azul marcan períodos de crisis; véase la tabla A.2.

de aquél (véase la figura 3.12). Esto no ocurre con la misma intensidad con las matrices de correlación relativa (véase la figura 3.12).

<b>Pearson</b>	$\lambda_N$	$\lambda_{N-1}$	$\lambda_{N-2}$	$\lambda_{N-3}$	<b>Relativa</b>	$\lambda_N$	$\lambda_{N-1}$	$\lambda_{N-2}$	$\lambda_{N-3}$
$\lambda_N$	1	-0.574	-0.773	-0.924	$\lambda_N$	1	0.394	-0.058	-0.420
$\lambda_{N-1}$	-0.574	1	0.687	0.576	$\lambda_{N-1}$	0.394	1	0.328	-0.108
$\lambda_{N-2}$	-0.773	0.687	1	0.794	$\lambda_{N-2}$	-0.058	0.328	1	0.456
$\lambda_{N-3}$	-0.924	0.576	0.794	1	$\lambda_{N-3}$	-0.420	-0.108	0.456	1

Cuadro 3.2: Correlación entre las series de tiempo de los cuatro mayores eigenvalores de las matrices locales de correlación de Pearson (izquierda) y relativas al índice (derecha).

En el apéndice C se abunda sobre las características de los eigenvalores de cada tipo de correlación. En esta sección, mencionamos los puntos que consideramos más importantes:

- $\lambda_{N-1}$  de las matrices de correlación de Pearson, tiene una correlación alta con  $\lambda_{\text{máx}}$  de las matrices de correlación relativa.
- Como las matrices de correlación tienen traza invariante igual a 1, la fracción  $\frac{1}{N} \sum_{i=0}^n \lambda_{N-i}$  es creciente con  $n = 0, \dots, N-1$ , asumiendo que los eigenvalores están ordenados de forma

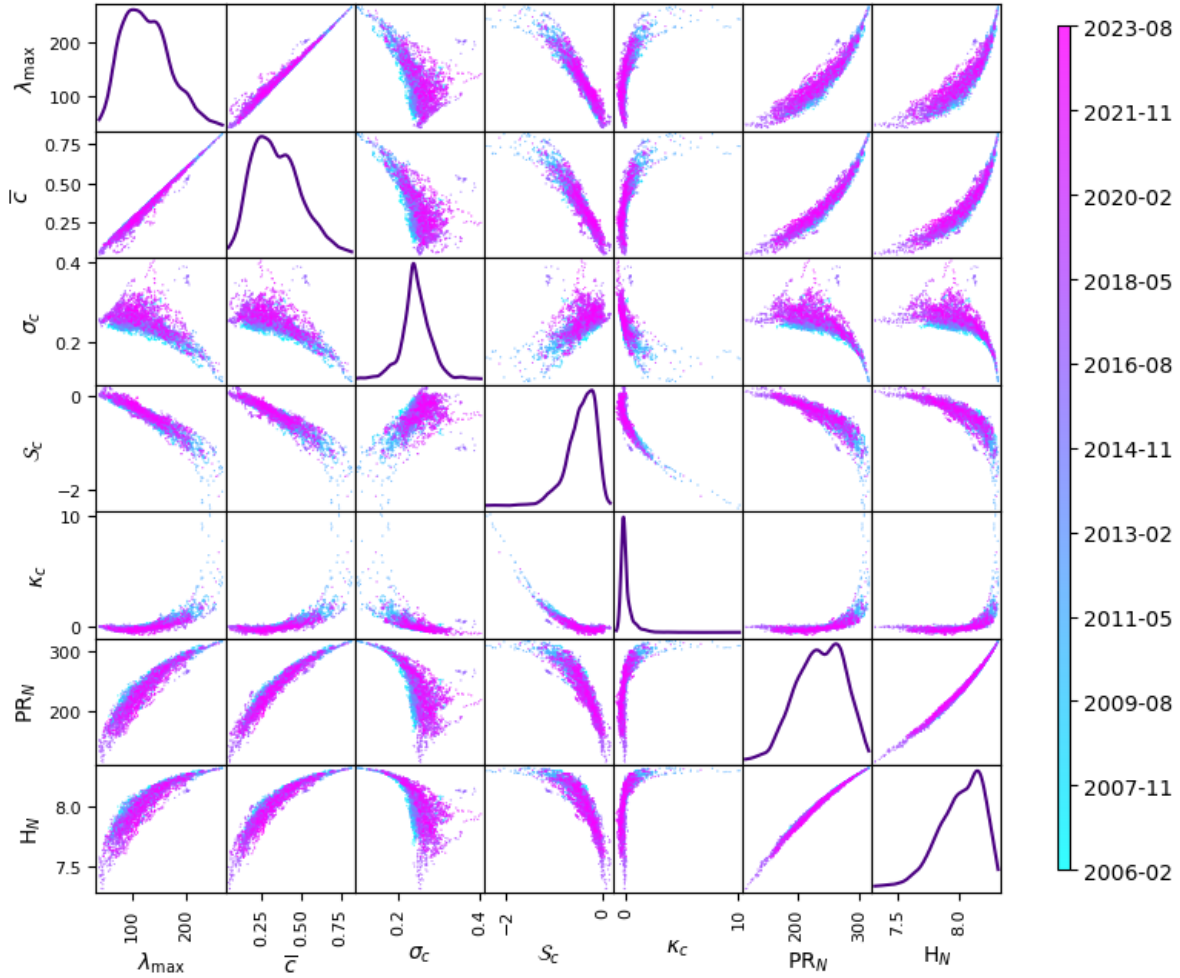


Figura 3.10: Gráficos de dispersión entre las series de tiempo del eigenvalor máximo  $\lambda_{\max}$ , correlación promedio  $\bar{c}$ , desviación estándar de las correlaciones  $\sigma_c$ , asimetría  $S_c$ , kurtosis  $\kappa_c$ , razón de participación del eigenvalor principal  $PR_N$  y entropía de Shannon del (módulo cuadrado de las componentes del) eigenvalor principal  $H_N$ , calculadas a partir de las correlaciones locales de Pearson. En la diagonal de esta gráfica se representa la distribución empírica de cada serie de tiempo sobre el horizonte completo, calculada con el método KDE.

creciente. Resulta que, para las matrices de correlación de Pearson, la mayor parte de la traza se concentra en los primeros eigenvalores dominantes, mientras que en las correlaciones relativas esa traza se distribuye en más eigenvalores.

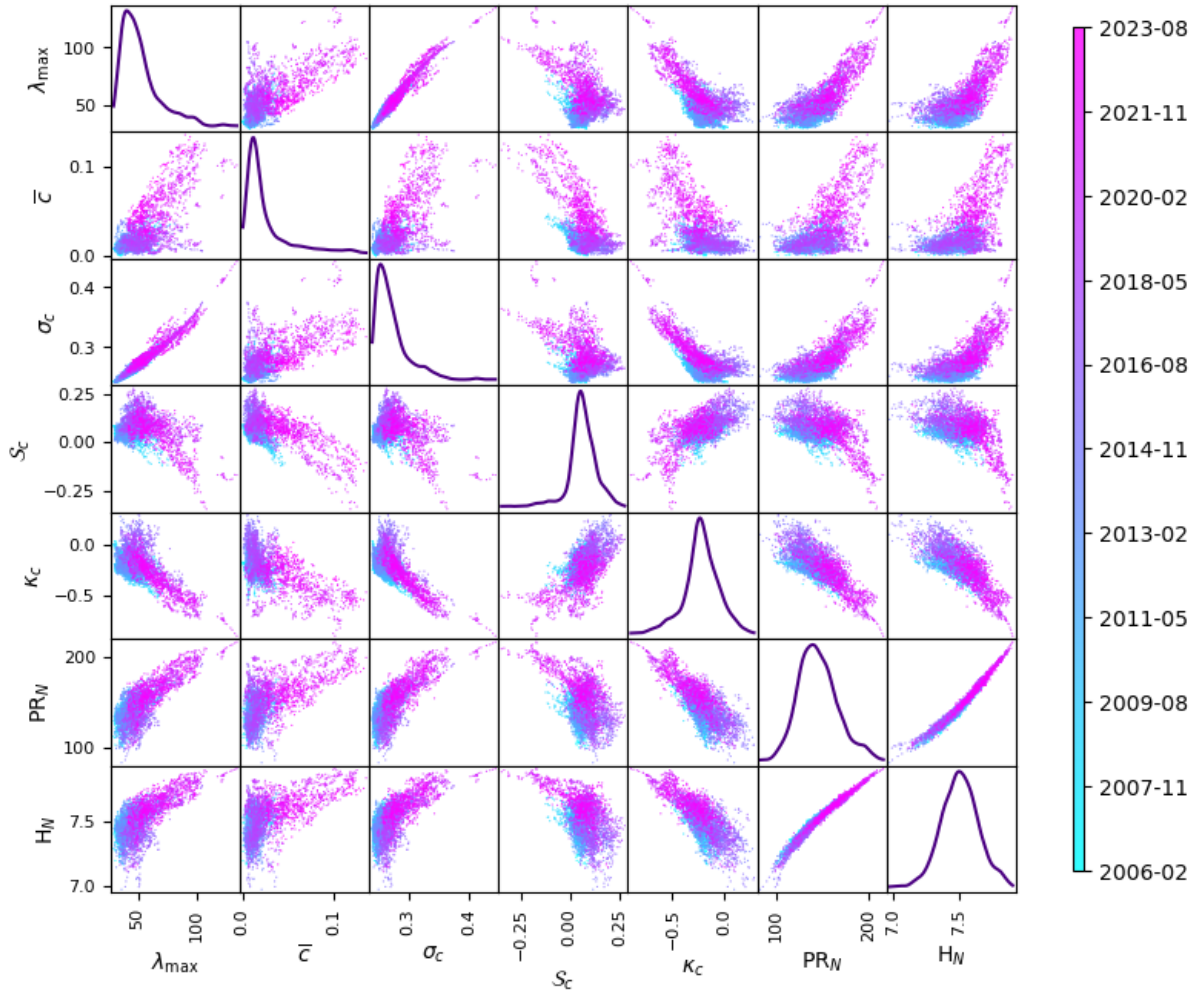
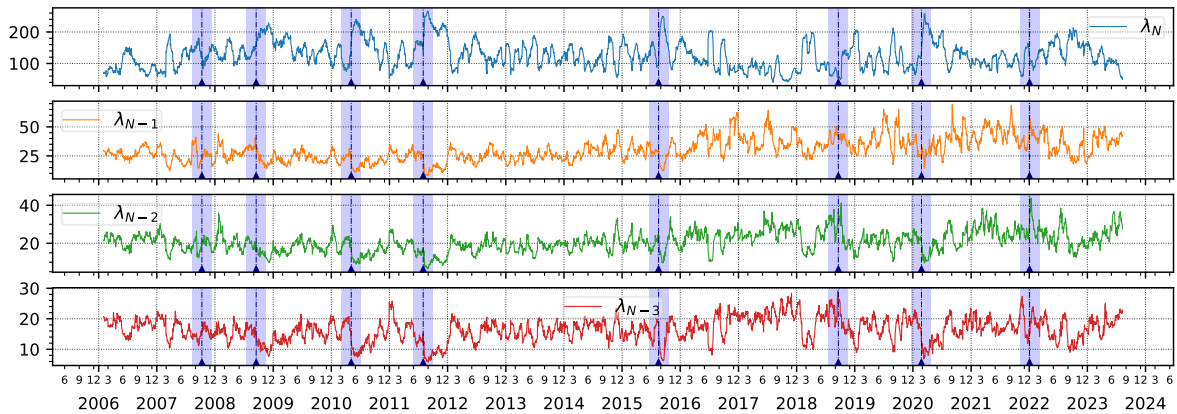
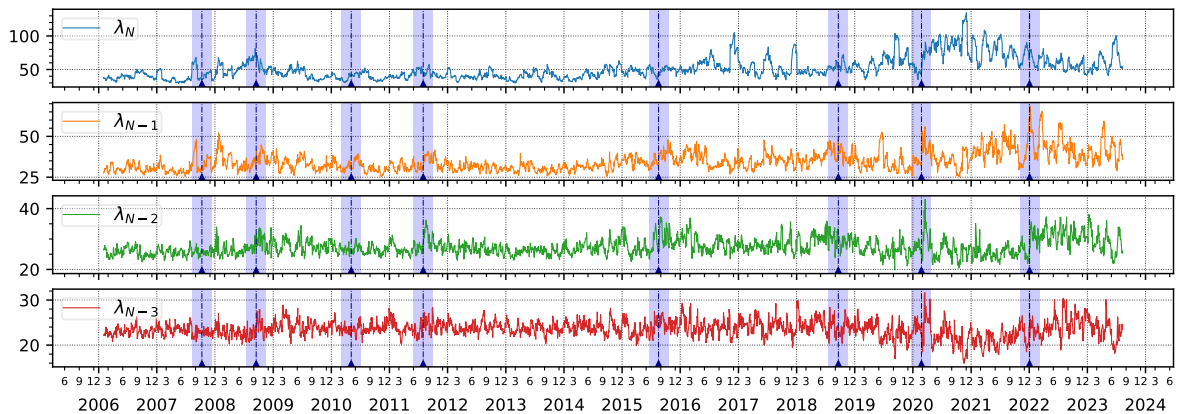


Figura 3.11: Gráficos de dispersión entre las series de tiempo del eigenvalor máximo  $\lambda_{\max}$ , correlación promedio  $\bar{c}$ , desviación estándar de las correlaciones  $\sigma_c$ , asimetría  $\mathcal{S}_c$ , kurtosis  $\kappa_c$ , razón de participación del eigenvector principal  $PR_N$  y entropía de Shannon del (módulo cuadrado de las componentes del) eigenvector principal  $H_N$ , calculadas a partir de las correlaciones locales relativas al índice. En la diagonal de esta gráfica se representa la distribución empírica de cada serie de tiempo sobre el horizonte completo, calculada con el método KDE.



(a)



(b)

Figura 3.12: Series de tiempo de los cuatro mayores eigenvalores  $\lambda_N = \lambda_{\text{máx}}$ ,  $\lambda_{N-1}$ ,  $\lambda_{N-2}$  y  $\lambda_{N-3}$  de las matrices de correlación locales en el tiempo de Pearson (subfigura 3.12(a)) y relativas al índice (subfigura 3.12(b)). Las franjas verticales traslúcidas de color azul marcan períodos de crisis; véase la tabla A.2. En el caso de las matrices de Pearson, las series de tiempo  $\lambda_{N-1}$ ,  $\lambda_{N-2}$  y  $\lambda_{N-3}$  se anticorrelacionan con  $\lambda_N$ , lo que no pasa con las matrices de correlación relativas al índice. Véanse la tabla 3.2 y la figura 3.13.

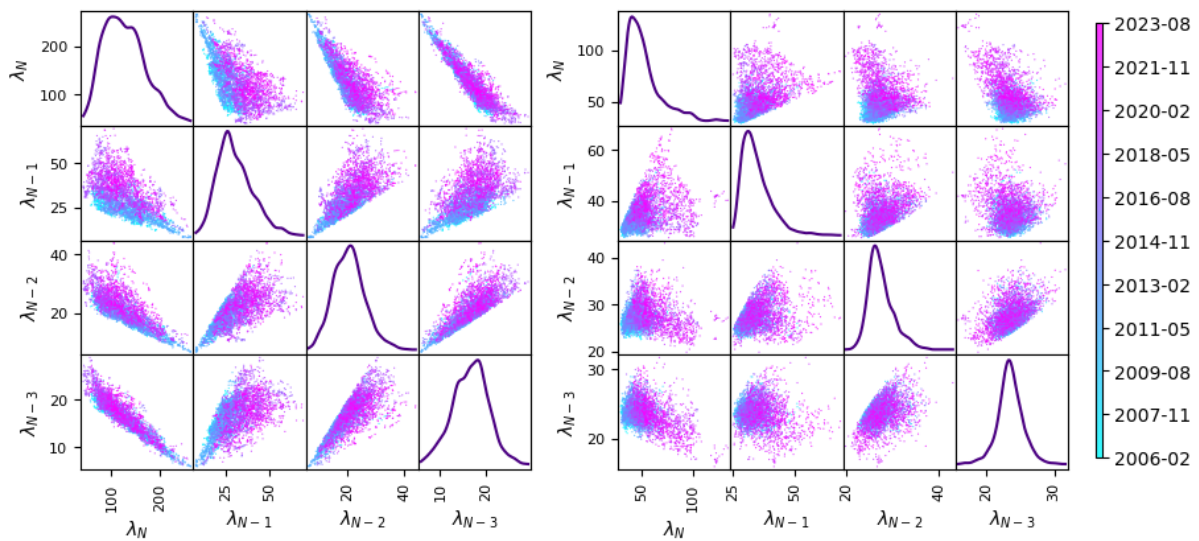


Figura 3.13: Gráficos de dispersión entre las series de tiempo de los mayores eigenvalores  $\lambda_N = \lambda_{\text{máx}}, \lambda_{N-1}, \lambda_{N-2}$  y  $\lambda_{N-3}$  de las matrices de correlación locales en el tiempo de Pearson (izquierda) y relativas al índice (derecha). Las anticorrelaciones mostradas por la dispersión se reflejan en la tabla izquierda del 3.2. En la diagonal de esta gráfica se representa la distribución empírica de cada serie de tiempo sobre el horizonte completo, calculada con el método KDE.

### 3.3. Estados de mercado en el S&P 500

Para estudiar los estados de mercado, consideramos el *aplanamiento* de las matrices de correlación, de las secuencias  $\mathcal{C}^{\text{Pearson}}, \mathcal{C}^{\text{AGSPC}} \subseteq \mathbb{R}^{N \times N}$ , dado por la función (2.3.1). A los elementos de estas secuencias les aplicamos el *power map* (fórmula (2.16)) con distintos exponentes. Luego, calculamos sus matrices de similaridad (ecuación (2.37) con  $p = 2$ ); véanse las figuras 3.14 y 3.15. Las matrices de similaridad son la base para delimitar las matrices locales de

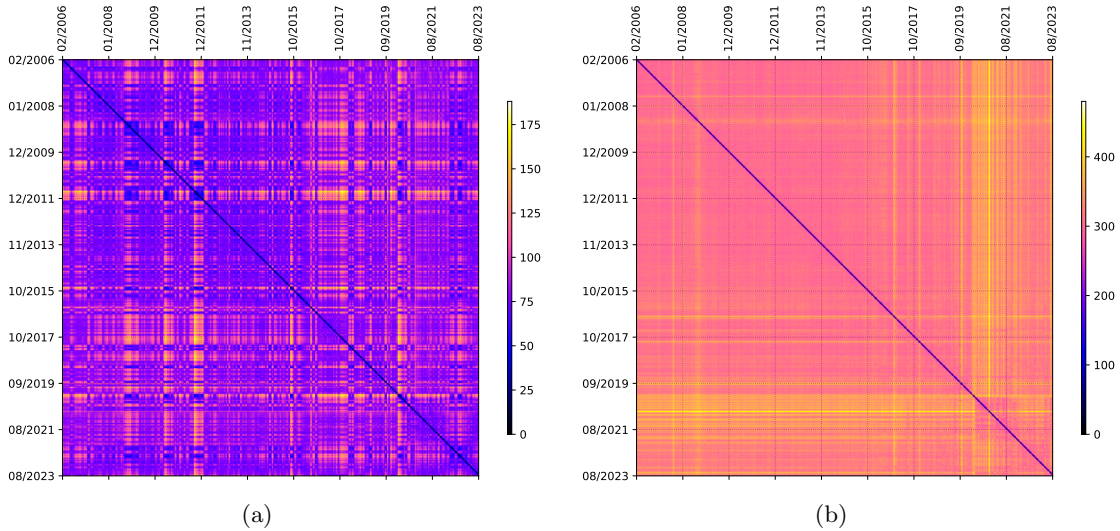


Figura 3.14: Matrices de similaridad de las secuencias de matrices de correlación de Pearson (subfigura 3.14(a)) y relativas al índice (subfigura 3.14(b)) locales en el tiempo, cuando se consideran épocas sucesivas de  $q = 20$  días bursátiles, con 19 días de traslape, y sin haberse aplicado el *power map* ( $\epsilon = 0$ ). La métrica considerada es  $p = 2$ , y en cada subgráfica el código de color se normaliza de acuerdo con el rango de datos correspondiente.

correlación en cúmulos. En principio, porque sus gráficas permiten identificar las épocas donde hay una menor similitud en términos de la distancia, a través de franjas distintivas horizontales verticales que se relacionan entre sí por una simetría con respecto a la diagonal principal, cuyo valor numérico es nulo. Luego, porque los bloques de estas matrices con colores similares indican épocas donde los patrones de correlación son parecidos. Además, las regiones donde el color tiene un aspecto constante, indican un período con cierta estabilidad.

Mediante la implementación en **Scikit-Learn** del algoritmo de K-means (que fue descrito en la sección 2.3), con el ajuste de parámetros que se detalla más adelante en el Apéndice B, y una variación de cúmulos desde  $k = 5$  hasta  $k = 12$ , obtuvimos una secuencia

$\{S(t) : t = q, \dots, T\} \subseteq \{1, \dots, k\}$  (ver relación (2.39)) para indicar el cúmulo

$\mathcal{C}_1^{\text{Pearson}}, \dots, \mathcal{C}_k^{\text{Pearson}}$  al que pertenece la matriz  $\hat{\mathbf{C}}^{(\epsilon)}(t; q) \in \mathcal{C}^{\text{Pearson}}$  al tiempo  $t$ . La asignación de las etiquetas  $\{1, \dots, k\}$  a los  $k$  estados de mercado, a diferencia de la elegida en [41], es de acuerdo con la regla:

$$\forall m, n \in \{1, \dots, k\} : m < n \implies \overline{\lambda_{\text{máx}}^{(m)}} \leq \overline{\lambda_{\text{máx}}^{(n)}} \quad (3.1)$$

donde  $\overline{\lambda_{\text{máx}}^{(m)}}$  es el promedio de los eigenvalores máximos de las matrices de correlación de la parte  $\mathcal{C}_m^{\text{Pearson}}$ . No obstante, la alta correlación lineal entre  $\lambda_{\text{máx}}(t)$  y  $\bar{c}(t)$  en las correlaciones de

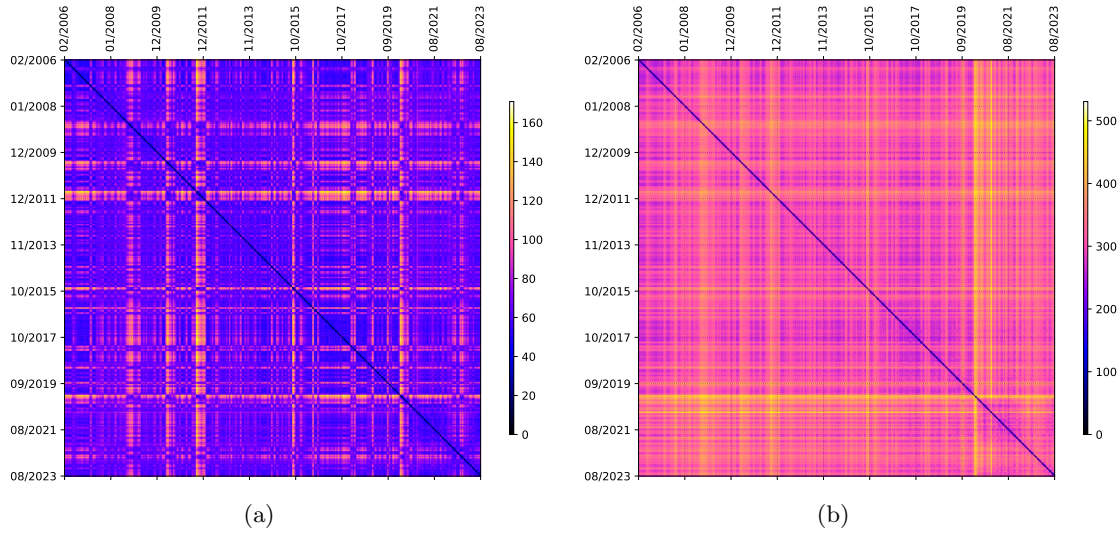


Figura 3.15: Matrices de similitud de las secuencias de matrices de correlación de Pearson (subfigura 3.15(a)) y relativas al índice (subfigura 3.15(b)) locales en el tiempo, cuando se consideran épocas sucesivas de  $q = 20$  días bursátiles, con 19 días de traslape, y habiéndose aplicado el *power map* con parámetro  $\epsilon = 0.5$ . La métrica considerada es  $p = 2$ , y en cada subgráfica el código de color se normaliza de acuerdo con el rango de datos correspondiente.

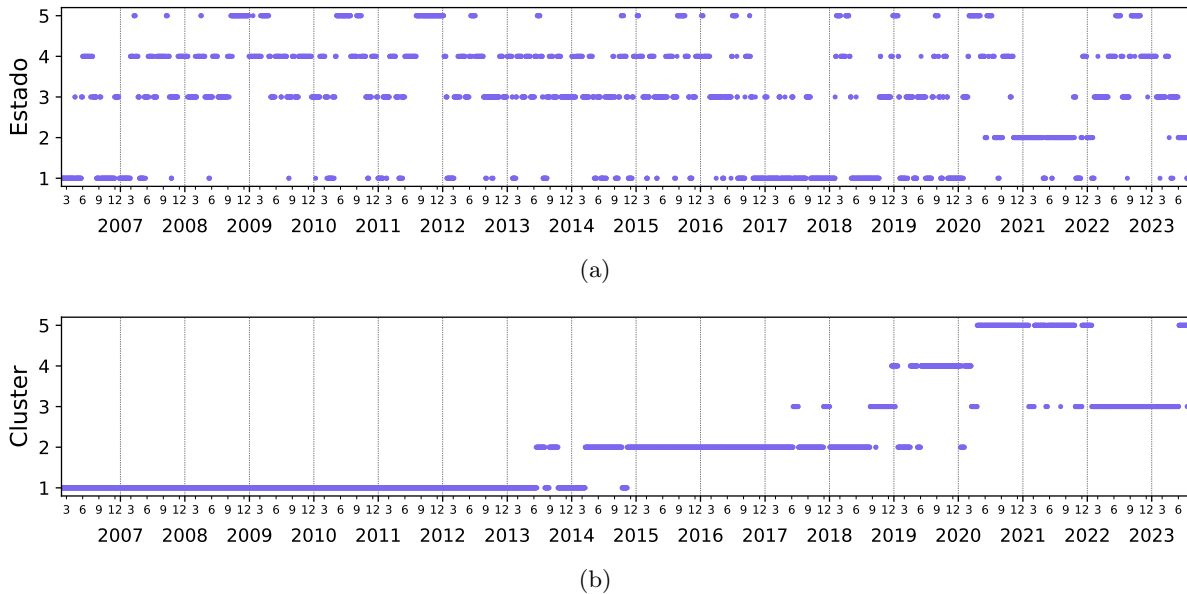


Figura 3.16: Evolución de los estados de mercado para las correlaciones de Pearson  $\mathcal{C}^{\text{Pearson}}$  y relativas al índice  $\mathcal{C}^{\text{GSPC}}$ , cuando  $k = 5$  y no se aplica el *power map*. El ordenamiento de los estados es creciente con el promedio del eigenvalor máximo de los miembros de cada estado o cúmulo (relación (3.1)). Nótese que la etiqueta del eje vertical en la subfigura 3.16(b) es *Cluster*, para diferenciar los cúmulos de las matrices de correlación relativa de los cúmulos, o estados de mercado, de las matrices de correlación de Pearson.



Pearson (véase la figura 3.10), hace que este ordenamiento sea prácticamente equivalente<sup>1</sup>. El

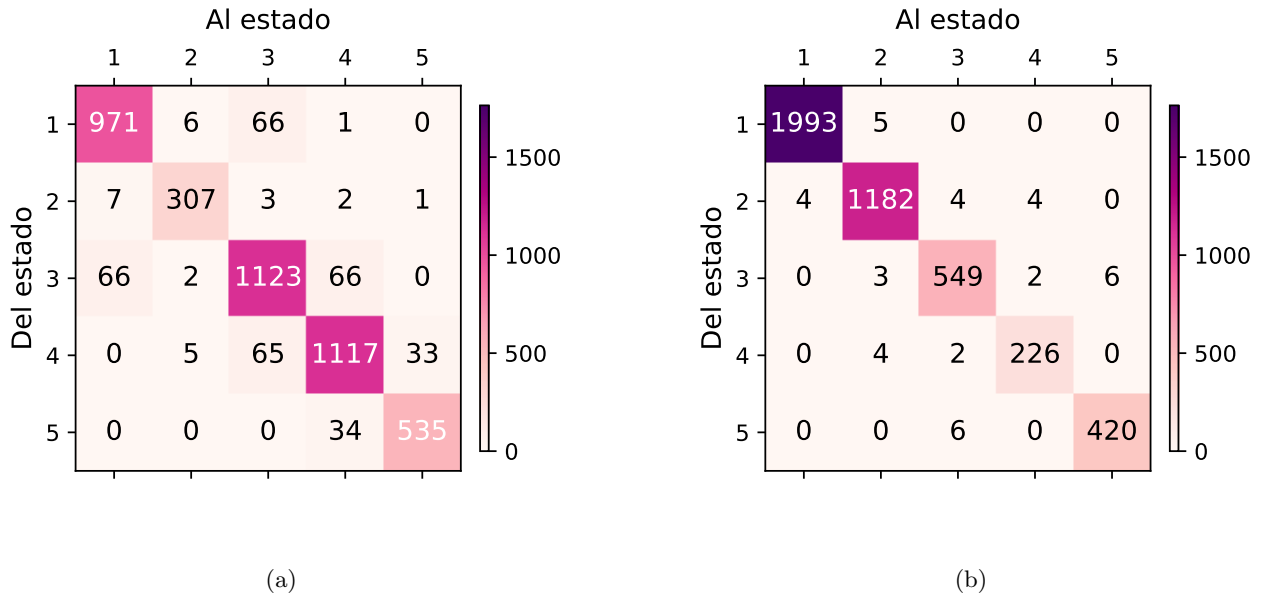


Figura 3.17: Matrices de transición para las correlaciones de Pearson  $\mathcal{C}^{\text{Pearson}}$  y relativas al índice  $\mathcal{C}^{\text{GSPC}}$ , cuando  $k = 5$  y no se aplica el *power map*. La dinámica de los estados de la figura 3.16 se traduce en matrices de transición cuasi-simétricas y tridiagonales. La subfigura 3.17(a) corresponde a los estados de mercado de las correlaciones de Pearson, mientras que la subfigura 3.17(b) corresponde a los *clusters* de las correlaciones relativas al índice.

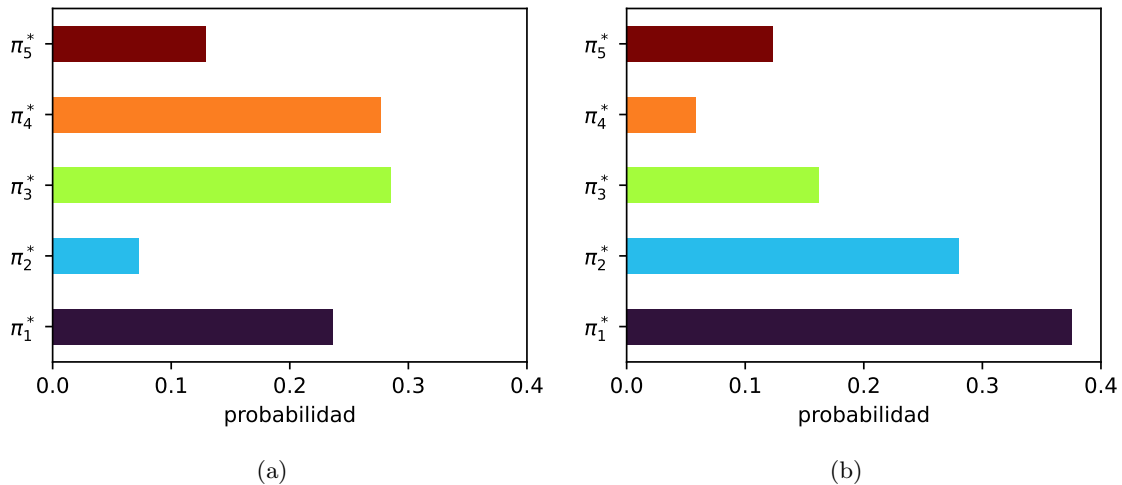


Figura 3.18: Distribuciones de probabilidad estacionarias, calculadas mediante la condición (2.42), a partir de las matrices de transición mostradas en la figura 3.17, cuando aquellas se normalizan de acuerdo con la fórmula (2.41). Los valores son: Pearson:  $\pi^* = (0.2367, 0.0726, 0.285, 0.2766, 0.129)$ ; correlaciones relativas:  $\pi^* = (0.3756, 0.2806, 0.162, 0.0587, 0.1232)$ .

<sup>1</sup>Ordenar con respecto al eigenvalor máximo tiene sentido, justo por la alineación preponderante de las matrices de correlación con los ejes principales.

cúmulo ocupado por las matrices de correlación de Pearson y relativas al índice, en cada época  $t$ , cuando  $k = 5$ , se muestra en la figura 3.16. Nótese que el estado 2 de la subfigura 3.16(a) no existía antes de mediados del 2020, y que las épocas correspondientes a ese estado casi coinciden en las épocas del *cluster* 5 de la subfigura 3.16(b). En [99], estudiamos a detalle estas épocas, a las que nombramos *(d)el estado de COVID*.

Inicialmente, el estado COVID (numerado como 2 para las correlaciones de Pearson) se *esconde* en el pánico de compra-venta debido a la caída del índice (que en las correlaciones de Pearson se refleja por la dominancia del eigenvalor máximo; véase la figura 2.5), pues su inicio solo se da hasta junio de 2020 (véase la figura 3.16(a)). La correlación relativa, en cambio, permite quitar esa predominancia del estado del mercado (es decir: *the state of the market* de la literatura), lo que se refleja en el acoplamiento del correspondiente eigenvalor máximo con el nivel de fluctuaciones de la correlación promedio en las matrices de correlación relativa (o sea: la desviación estándar; véase el tercer renglón de la figura 3.7, y el primer renglón, segunda columna de cada matriz de dispersión en la figura 3.13). Por eso, el estado COVID en las correlaciones relativas al índice, numerado como 5, hace su aparición por primera vez en marzo de 2020 [99]; véase la figura 3.16(b).

La dinámica de las transiciones  $M(i, j)$  (relación (2.40)) se muestra en la figura 3.17. El

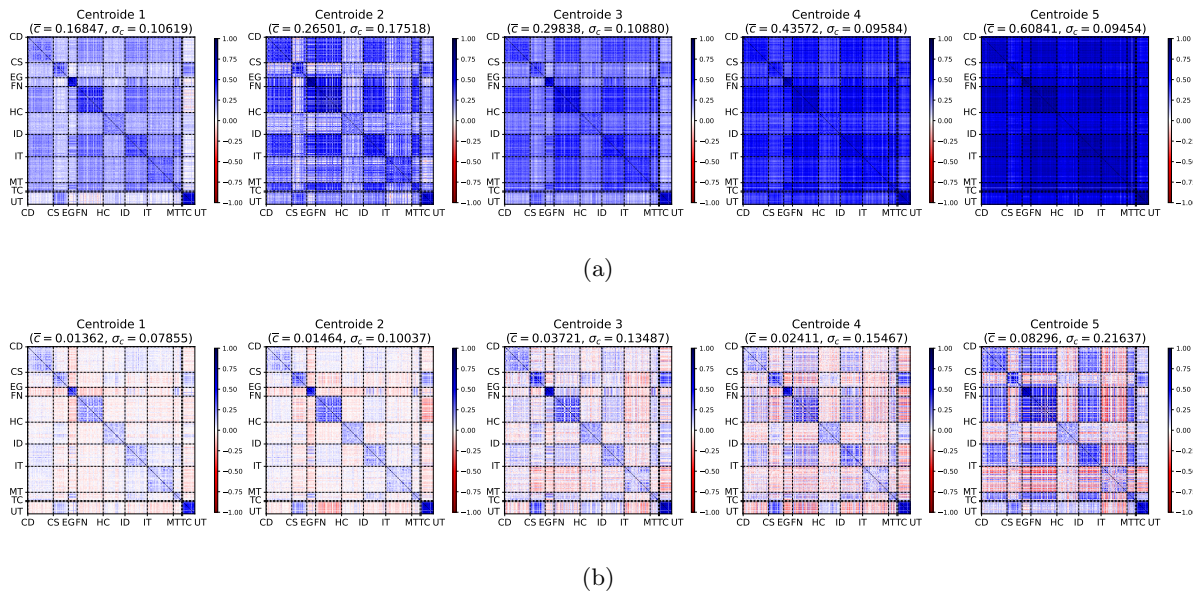


Figura 3.19: Centroides de los estados de mercado para las correlaciones de Pearson  $\mathcal{C}^{\text{Pearson}}$  y relativas al índice  $\mathcal{C}^{\wedge\text{GSPC}}$ , cuando  $k = 5$  y no se aplica el *power map*. Como el ordenamiento de los estados (relación (3.1)) es creciente con el promedio del eigenvalor máximo de los miembros de cada estado o cúmulo, entonces  $\bar{c}$  de las entradas de los centroides es creciente con el índice del estado en las correlaciones de Pearson (pues en ellas el mayor eigenvalor de cada matriz se correlaciona fuertemente con el promedio de sus entradas), mientras que  $\sigma_c$  es creciente con el índice del estado (*cluster*) en las correlaciones relativas (pues en ellas el mayor eigenvalor de cada matriz está fuertemente correlacionado con la desviación estándar de sus entradas).

centroide de cada cúmulo,  $\tilde{\mathbf{C}}_m$  ( $m \in \{1, \dots, k\}$ ) puede graficarse mediante la inversa del *aplanamiento* (2.3.1), y también en el caso  $k = 5$ , se obtiene la figura 3.19. Las densidades y

algunas estadísticas de las entradas de dichos centroides, se pueden ver en la figura 3.20 y las tablas 3.3.

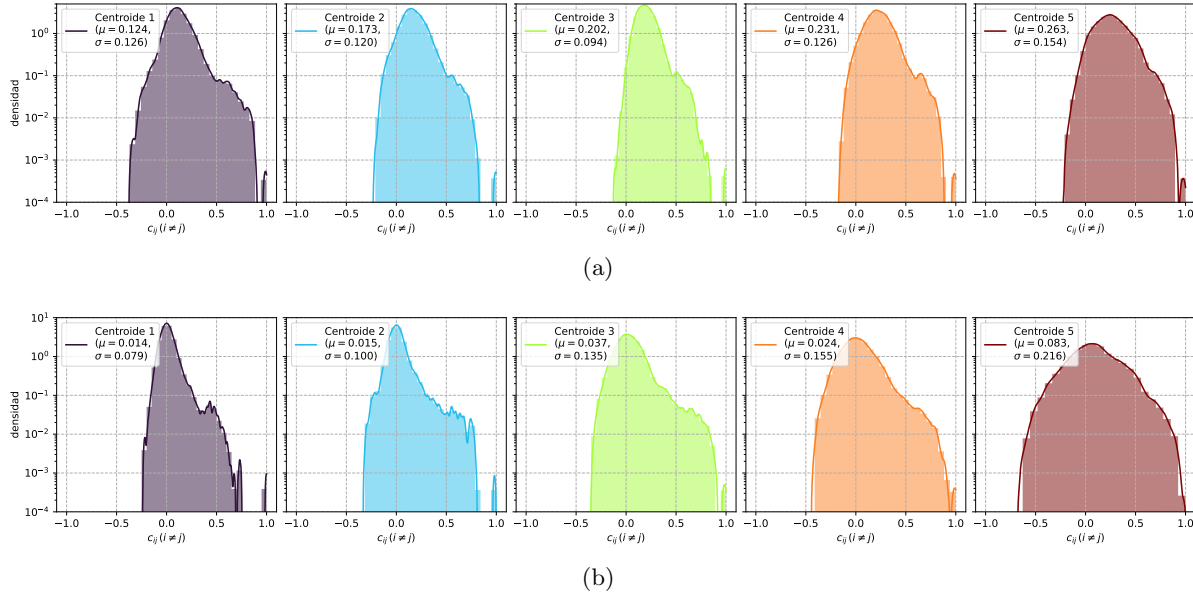


Figura 3.20: KDE de las entradas no diagonales de los centroides de los estados de mercado para las correlaciones de Pearson  $\mathcal{C}^{\text{Pearson}}$  (subfigura 3.20(a)) y relativas al índice  $\mathcal{C}^{\text{GSPC}}$  (subfigura 3.20(b)), cuando  $k = 5$  y no se aplica el *power map*.

Nótese que en la tabla superior 3.3,  $\bar{c}$  es creciente con el índice de cada cúmulo, pues la manera en que se definió el ordenamiento (relación (3.1)) es a través de  $\lambda_{\text{máx}}$ , y ésta se correlaciona fuerte y positivamente con  $\bar{c}$  (véase el renglón 1, columna 2 de la figura 3.10). En contraste, para la tabla inferior 3.3 es  $\sigma_c$  quien crece con el índice de cada cúmulo, pues  $\sigma_c$  se correlaciona fuerte y positivamente con  $\lambda_{\text{máx}}$  en las correlaciones relativas (véase el renglón 1, columna 3 de la figura 3.11).

Por otro lado, las mismas estadísticas de las entradas no diagonales pero sobre todos los elementos de cada cúmulo, se muestran en la figura 3.21 y las tablas 3.4.

El eigenvector principal  $|N^{(j)}\rangle$  de cada centroide  $j = 1, \dots, k$  ( $k = 5$ ) se representa en la figura 3.22, donde se grafican las frecuencias de las componentes normalizadas con la orientación que hace su primera componente positiva. Análogamente, se muestra el histograma de las componentes del eigenvector principal de cada una de los miembros de  $\mathcal{C}^{\text{Pearson}}$  y relativas al índice  $\mathcal{C}^{\text{GSPC}}$ .

Son notables las diferencias en las distribuciones de las componentes de  $|N\rangle$  en la figura 3.23, para la que se elige una escala semilogarítmica. Cabe resaltar que, *a ojo*, la distribución de  $a_j(N)$  para el *cluster* 1 de la subfigura 3.23(b) es normal, como en el resultado de RMT graficado en la figura 2.2 del Capítulo 2.

El escalamiento multidimensional, MDS, es útil para tener una idea de la distribución espacial de los miembros de cada cúmulo. En la figura 3.24 se muestran las proyecciones tridimensionales, en planos principales, correspondientes a las secuencias de matrices (aplanadas)  $\mathcal{C}^{\text{Pearson}}$  y  $\mathcal{C}^{\text{GSPC}}$ .

<b>Pearson</b>	Centroide 1	Centroide 2	Centroide 3	Centroide 4	Centroide 5
$\bar{c}$	0.124	0.173	0.202	0.231	0.263
$\sigma_c$	0.126	0.120	0.094	0.126	0.154
$\mathcal{S}_c$	1.127	0.739	1.054	0.685	0.459
$\kappa_c$	4.031	1.987	2.876	1.670	0.343

<b>Relativa</b>	Centroide 1	Centroide 2	Centroide 3	Centroide 4	Centroide 5
$\bar{c}$	0.014	0.015	0.037	0.024	0.083
$\sigma_c$	0.079	0.100	0.135	0.155	0.216
$\mathcal{S}_c$	2.143	2.439	1.419	0.905	0.270
$\kappa_c$	10.177	12.062	4.888	2.301	0.382

Cuadro 3.3: Primeros momentos centrales de las entradas no diagonales de las matrices de correlación que representan los centroides de los  $k = 5$  clusters de las matrices de correlación de Pearson (tabla en la parte superior) y relativas al índice (tabla inferior), representados en la figura 3.19.

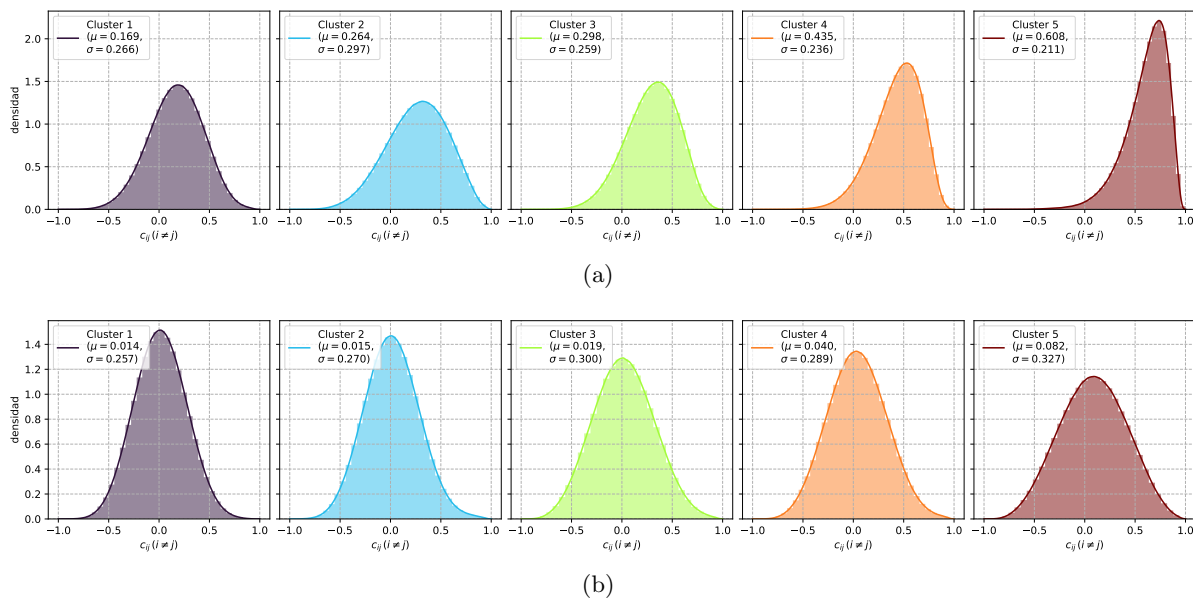


Figura 3.21: KDE de las entradas no diagonales de las matrices que conforman cada uno de los  $k = 5$  cúmulos de las correlaciones de Pearson  $\mathcal{C}^{\text{Pearson}}$  (subfigura 3.21(a)) y relativas al índice  $\mathcal{C}^{\wedge\text{GSPC}}$  (subfigura 3.21(b)). No se aplica el *power map*.

Las características que consideramos más importantes son:

1. El eje principal de la proyección de las correlaciones de Pearson, está alineado con el índice del estado.
2. El eje principal de la proyección de las correlaciones relativas, está alineado con el índice del estado, pero su secuencia es difusa.

<b>Pearson</b>	Cúmulo 1	Cúmulo 2	Cúmulo 3	Cúmulo 4	Cúmulo 5
$\bar{c}$	0.169	0.264	0.298	0.435	0.608
$\sigma_c$	0.266	0.297	0.259	0.236	0.211
$\mathcal{S}_c$	-0.144	-0.279	-0.339	-0.602	-1.093
$\kappa_c$	-0.183	-0.340	-0.167	0.220	1.608

<b>Relativa</b>	Cúmulo 1	Cúmulo 2	Cúmulo 3	Cúmulo 4	Cúmulo 5
$\bar{c}$	0.014	0.015	0.019	0.040	0.082
$\sigma_c$	0.257	0.270	0.300	0.289	0.327
$\mathcal{S}_c$	0.048	0.128	0.080	0.081	-0.035
$\kappa_c$	-0.202	-0.060	-0.279	-0.237	-0.472

Cuadro 3.4: Primeros momentos centrales de las entradas no diagonales de las matrices que conforman cada uno de los  $k = 5$  cúmulos de las matrices de correlación de Pearson (tabla en la parte superior) y relativas al índice (tabla inferior).

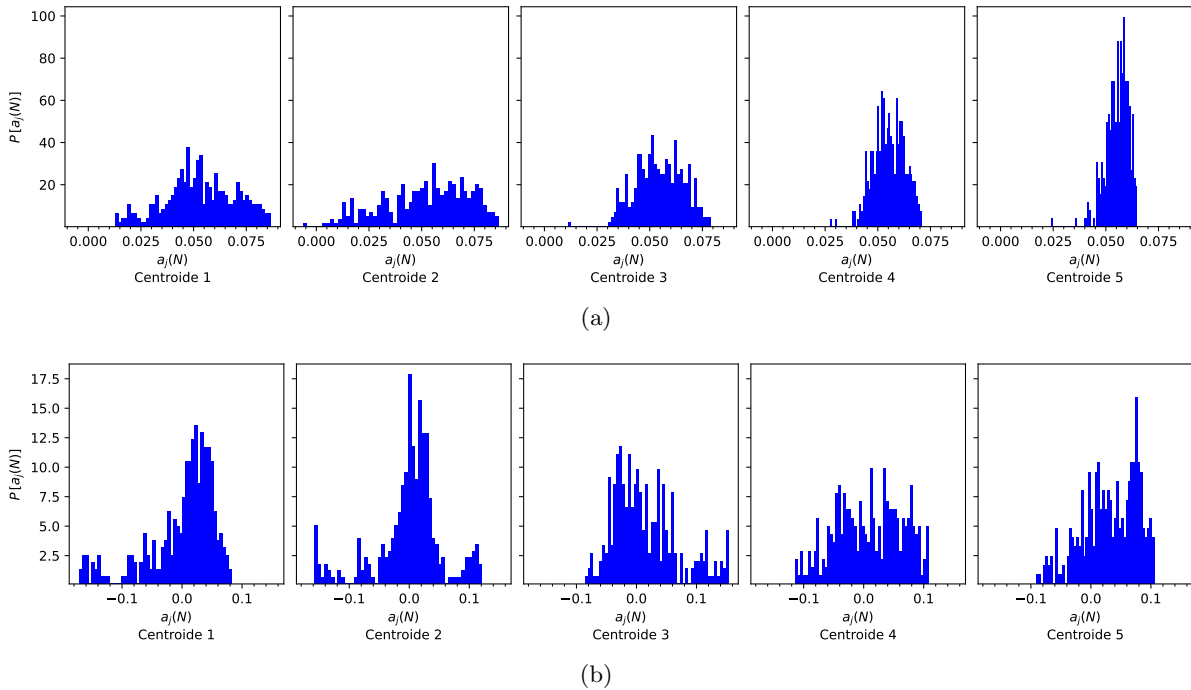


Figura 3.22: Histograma de las componentes del eigenvector principal  $|N^{(j)}\rangle$  de cada centroide  $j = 1, \dots, k$  de los estados de mercado para las correlaciones de Pearson  $\mathcal{C}^{\text{Pearson}}$  (subfigura 3.22(a)) y relativas al índice  $\mathcal{C}^{\text{GSPC}}$  (subfigura 3.22(b)), cuando  $k = 5$  y no se aplica el *power map*. Los eigenvectores están normalizados, y se orientan de tal manera que  $a_1(N) \geq 0$  (véase la relación (2.18)).

3. Los puntos representantes del estado 2 (coloreadas en azul) de la subfigura 3.24(a) son atípicas.
4. La distribución espacial de los puntos en 3.24(b) está menos distinguida por los colores

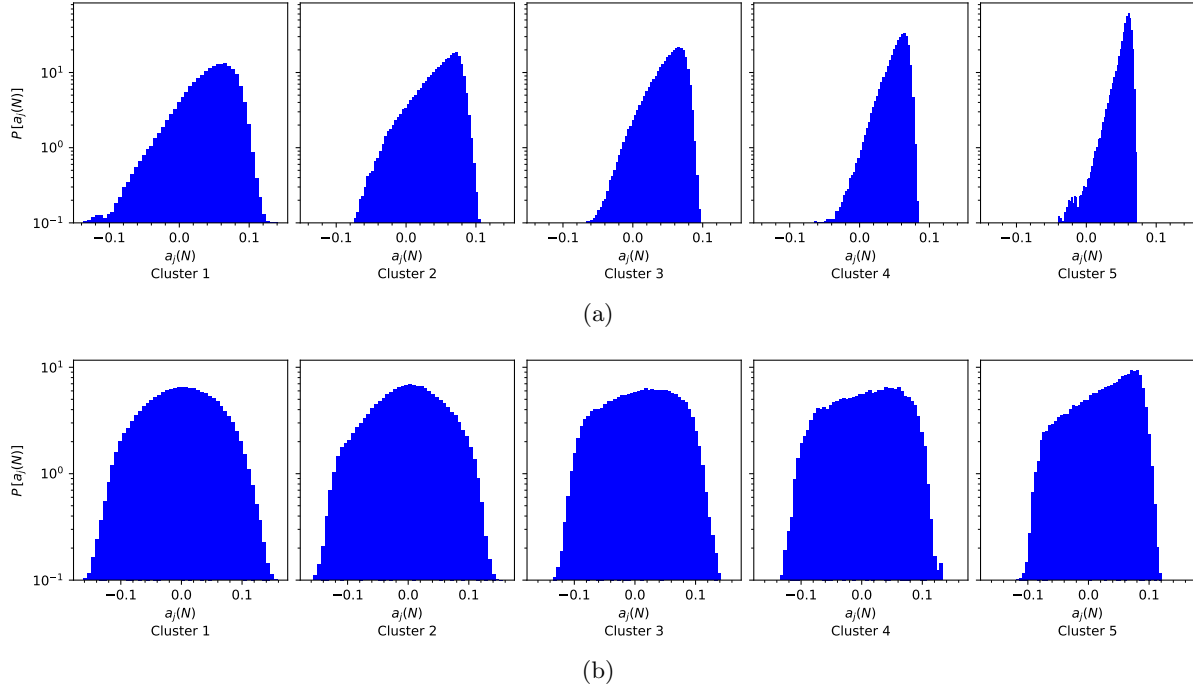


Figura 3.23: Histograma de las componentes del eigenvector principal de los elementos de  $\mathcal{C}^{\text{Pearson}}$  (subfigura 3.23(a)) y  $\mathcal{C}^{\text{GSPC}}$  (subfigura 3.23(a)), cuando  $k = 5$  y no se aplica el *power map*. Los eigenvectores están normalizados, y se orientan de tal manera que  $a_1(N) \geq 0$  para el eigenvector principal del centroide del cúmulo correspondiente ( $|N^{(j)}\rangle$  para el cúmulo  $j \in \{1, \dots, k = 5\}$ ; véase la relación (2.18)), y  $\langle N|N^{(j)}\rangle > 0$ .

<b>Pearson</b>	Centroide 1	Centroide 2	Centroide 3	Centroide 4	Centroide 5
$\overline{a(N)}$	-0.001	0.001	-0.015	-0.009	-0.032
$\sigma_{a(N)}$	0.056	0.056	0.054	0.055	0.046
$\mathcal{S}_{a(N)}$	1.395	-0.784	-0.795	0.156	0.497
$\kappa_{a(N)}$	1.503	1.498	0.057	-0.930	-0.534

<b>Relativa</b>	Centroide 1	Centroide 2	Centroide 3	Centroide 4	Centroide 5
$\overline{a(N)}$	0.053	0.052	0.055	0.055	0.055
$\sigma_{a(N)}$	0.017	0.020	0.011	0.007	0.005
$\mathcal{S}_{a(N)}$	-0.161	-0.481	-0.164	-0.267	-1.110
$\kappa_{a(N)}$	-0.454	-0.558	-0.266	-0.002	3.045

Cuadro 3.5: Primeros momentos centrales de las componentes del eigenvector principal  $|N\rangle$  de cada centroide de los estados de mercado para las correlaciones de Pearson  $\mathcal{C}^{\text{Pearson}}$  (subfigura 3.22(a)) y relativas al índice  $\mathcal{C}^{\text{GSPC}}$  (subfigura 3.22(b)), cuando  $k = 5$  y no se aplica el *power map*.

asignados.

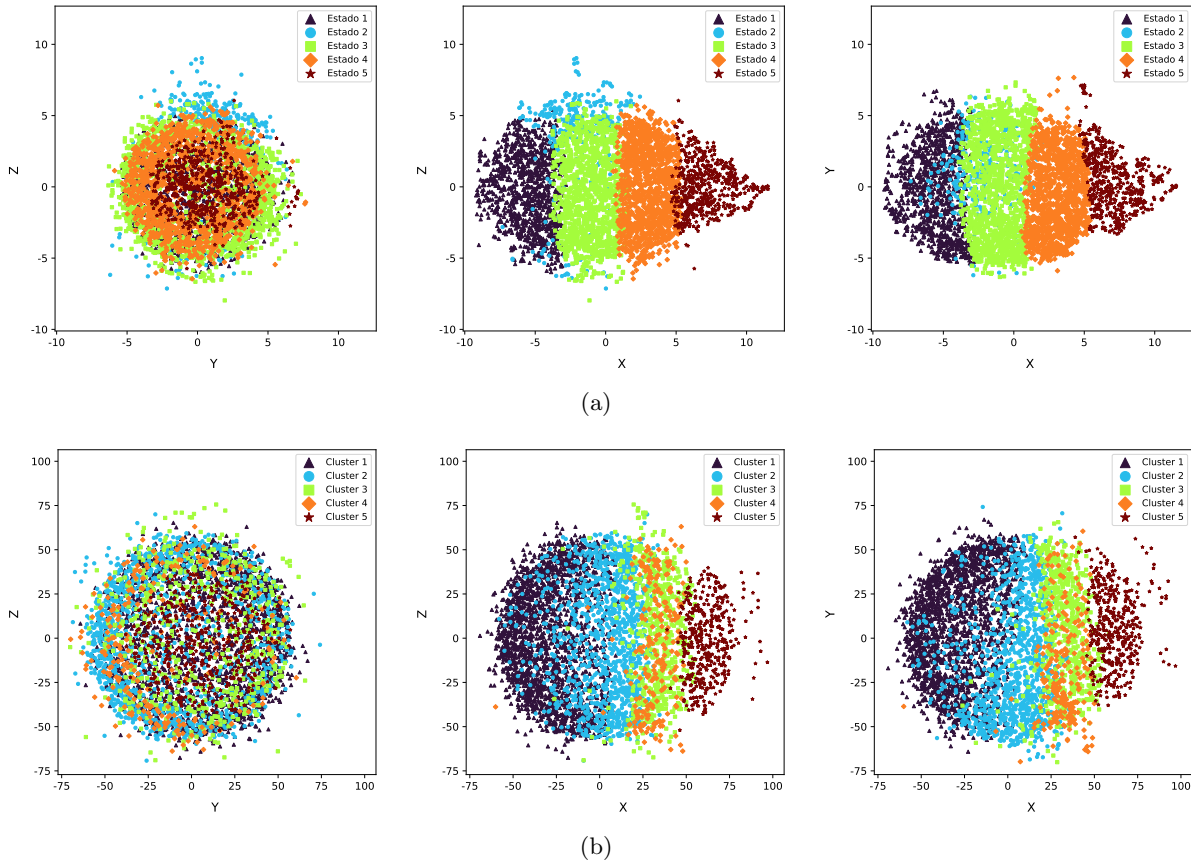


Figura 3.24: MDS de los estados de mercado para las correlaciones de Pearson  $\mathcal{C}^{\text{Pearson}}$  (subfigura 3.24(a)), y de los *clusters* de las matrices de correlación relativas al índice  $\mathcal{C}^{\text{AGSPC}}$  (subfigura 3.24(b)), cuando  $k = 5$  y no se aplica el *power map*. El k-means se hace en el espacio  $N(N - 1)/2$ -dimensional, para luego obtenerse las coordenadas 3D con el MDS. Además, el sistema de coordenadas  $x, y, z$  del espacio reducido se alinea con los ejes de mayor varianza (es decir, los ejes principales), con el origen en el centro geométrico de la proyección.

Al combinar otras métricas con las coordenadas MDS, y su respectivo estado ocupado, pudimos también observar que:

- En las correlaciones de Pearson: la coordenada  $x$  del MDS describe esencialmente el estado del mercado, excepto para el estado COVID, que parece un chichón (ver las Figuras 3.24(a) y 3.25(a)).
- En las correlaciones relativas al índice: la coordenada  $x$  del MDS también se alinea con el índice del cúmulo, pero la transición entre estados adyacentes es mucho más difusa que en los estados de mercado de Pearson (ver la figura 3.24(b) y 3.25(b)).

Es importante mencionar que las figuras 3.24(b), 3.25(b) y 3.26(b) deben interpretarse con cuidado, ya que el MDS distorsiona fuertemente las distancias del espacio  $N(N - 1)/2$ -dimensional al tridimensional. Muestra de esto son las figuras 5.2 y 4.9<sup>2</sup>.

<sup>2</sup>Por otro lado, las proyecciones MDS de las matrices de  $\mathcal{C}^{\text{Pearson}}$  no tienen una distorsión tan dramática; véase la

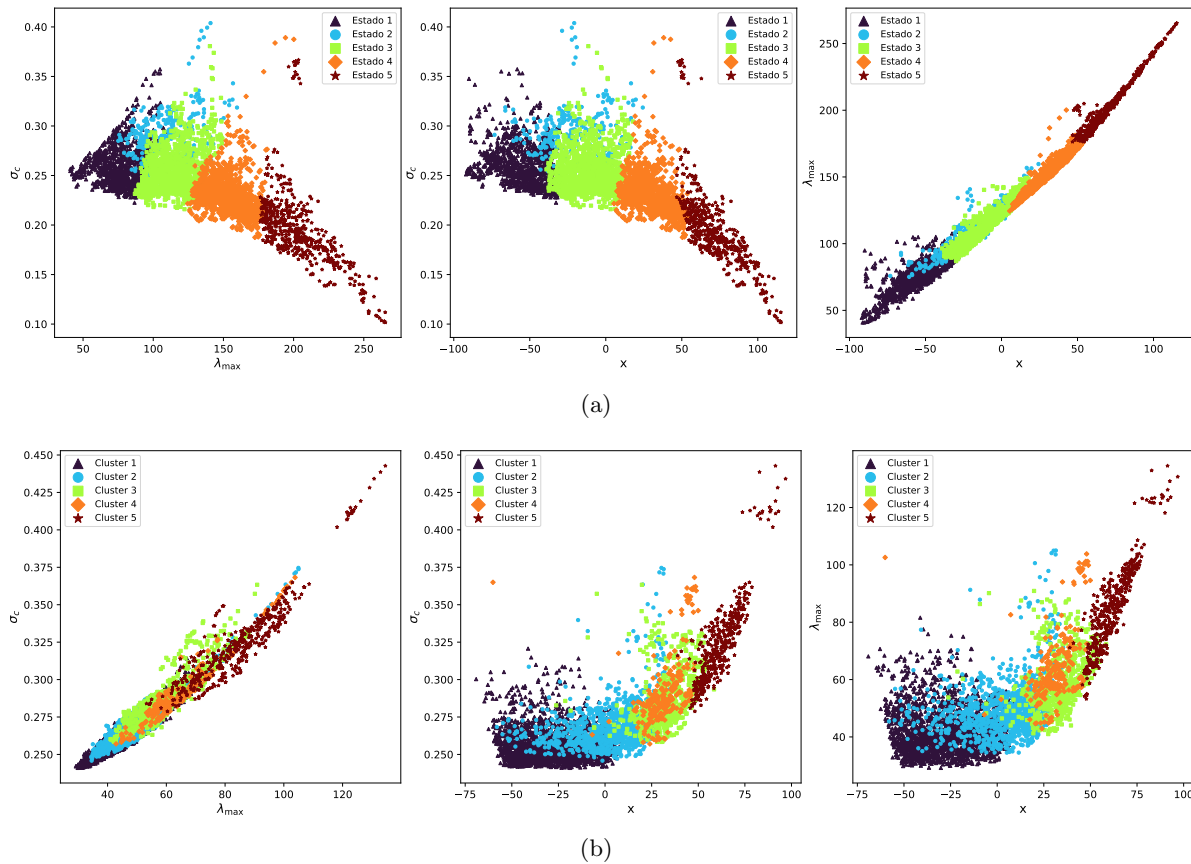


Figura 3.25: Diagramas de dispersión de la coordenada  $x$  del MDS, el mayor eigenvalor  $\lambda_{\max}$  y la desviación estándar  $\sigma_c$ . El código de color corresponde a los estados de mercado (Pearson) o *clusters* (correlaciones relativas al índice).

### 3.3.1. Estado de COVID

La aparición del estado 2 en la figura 3.16(a) sólo hasta el primer semestre de 2020; su configuración anormal en la proyección MDS de la figura 3.24(a); así como su coincidencia con el *cluster* 5 de la figura 3.16(b), hace pensar que se trata de un estado especial del mercado. En [99] lo llamamos *estado COVID*, precisamente por la coincidencia temporal con el momento álgido de la pandemia, cuyas restricciones socioeconómicas tuvieron consecuencias en los mercados financieros.

Llevamos a cabo cálculos numéricos que sustentan las siguientes afirmaciones:

- El estado COVID es robusto ante la variación de  $k$  desde 5 hasta 12.
- En las correlaciones de Pearson aparece el estado COVID tres meses después de que apareciera el estado 5 de las correlaciones relativas. Esto podría deberse a que en ese lapso son las altas correlaciones quienes dominan el estado, provocadas a su vez por las ventas de pánico y su consecuente *crash* financiero [100]. De hecho, éste *crash* casi coincide con la

figura 2.7.



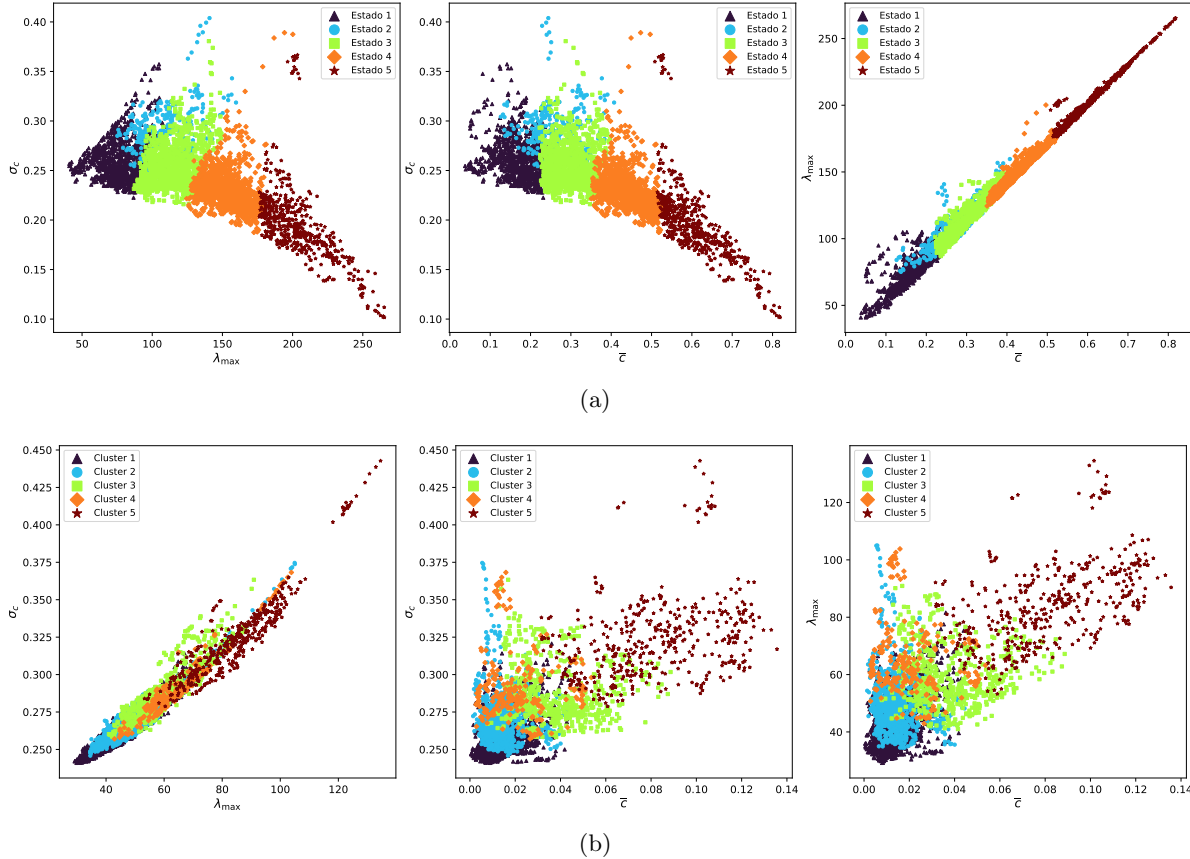


Figura 3.26: Diagramas de dispersión de la correlación promedio  $\bar{c}$ , el mayor eigenvalor  $\lambda_{\max}$  y la desviación estándar  $\sigma_c$  de las correlaciones de Pearson (subfigura 3.26(a)) y relativas al índice (subfigura 3.26(b)). El código de color corresponde a los estados de mercado (Pearson) o *clusters* (correlaciones relativas al índice).

aparición del estado 5 en las correlaciones relativas, así como en el incremento sustancial de  $\lambda_{\max}$ ,  $\bar{c}$  y  $\sigma_c$  de la figura 3.7; las franjas claras diagonalmente simétricas en la matriz de similaridad 3.14(b); y la volatilidad de la mediana en la dinámica de la distribución de los elementos de matriz de las correlaciones relativas sobre todo el horizonte temporal 3.5(b). De este modo, al remover la componente global del mercado con la correlación relativa al índice, se pueden apreciar otras componentes.

- La continuidad del estado COVID en las correlaciones de Pearson termina en febrero de 2022, pero reaparece esporádicamente por el resto del lapso analizado.
- Las métricas de la correlación relativa amplían los parámetros relevantes para describir la dinámica de los mercados financieros en términos de sus correlaciones.
- El *power map* reduce las correlaciones sutiles que devienen en la observación del estado COVID.

### 3.3.2. Correlaciones proyectadas de Guhr

El método de Guhr *et al* [84, 101, 102] permite remover la componente del mercado, entendida como el eigenvector principal de la correlación de Pearson. Los resultados de las secciones anteriores dan un resultado que recuerda a aquellos de las proyecciones de Guhr (y que veremos más adelante, en este Capítulo y el Apéndice C), pero también tienen un atractivo potencial para la comunidad economista y financiera: *remover la componente del mercado corresponde a filtrar el índice*.

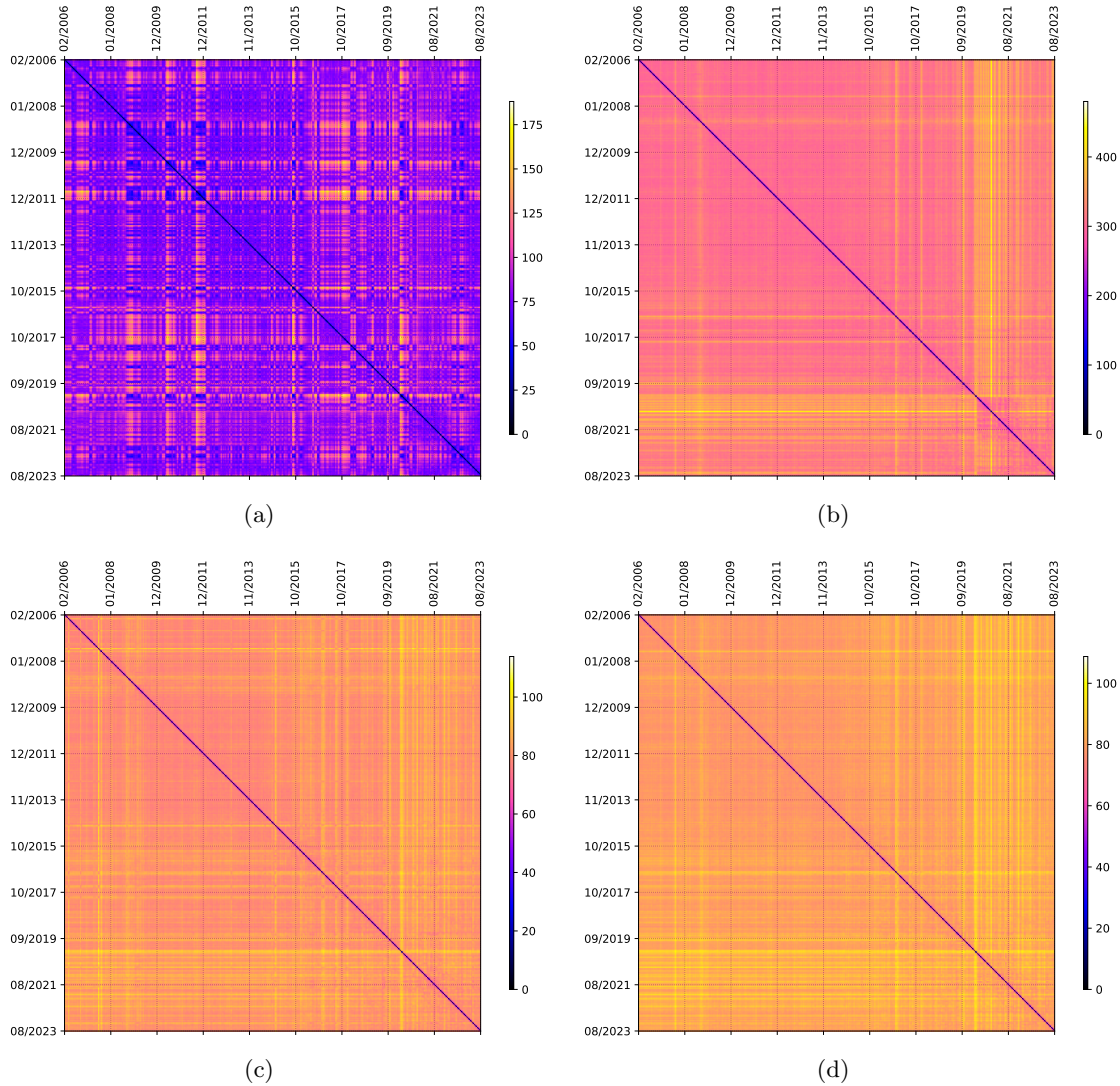


Figura 3.27: Matrices de similitud de las secuencias de matrices de correlación de Pearson (subfigura 3.27(a)), relativas al índice (subfigura 3.27(b)), y proyectadas de Guhr con los métodos de covarianza (subfigura 3.27(c)) y de correlación (subfigura 3.27(d)) locales en el tiempo, cuando se consideran épocas sucesivas de  $q = 20$  días bursátiles, con 19 días de traslape, y sin haberse aplicado el *power map* ( $\epsilon = 0$ ). La métrica considerada es  $p = 2$ , y en cada subgráfica el código de color se normaliza de acuerdo con el rango de datos correspondiente.

Podemos comenzar por comparar las matrices de similitud de las secuencias  $\mathcal{C}^{\text{Pearson}}$ ,

$\mathcal{C}^{\wedge\text{GSPC}}$ ,  $\mathcal{C}^{\text{B}}$  y  $\mathcal{C}^{\text{L}}$  mediante la figura 3.27. El parecido se hace más evidente al aplicar el *power map*, lo que se ejemplifica para  $\epsilon = 0.5$  en la figura 3.28. Ésto también se refleja en el mapa de calor de la figura 3.29, análoga a 3.5. Las matrices de correlación proyectadas de época corta, ubicadas en las regiones resaltadas en amarillo de las figuras mencionadas en la sentencia previa, las pospusimos al Apéndice C junto con las distribuciones de sus elementos no diagonales.

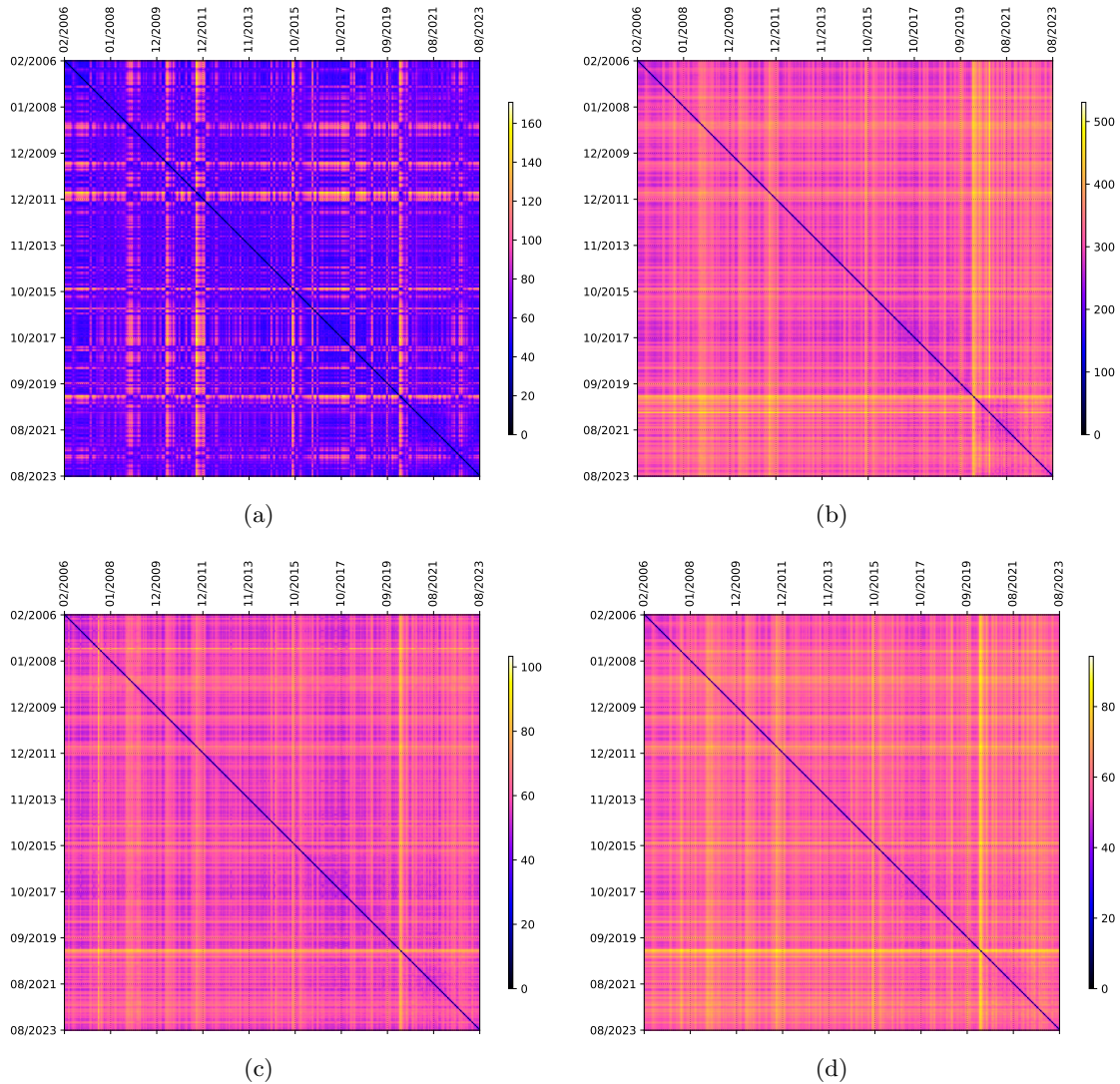


Figura 3.28: Matrices de similitud de las secuencias de matrices de correlación de Pearson (subfigura 3.28(a)), relativas al índice (subfigura 3.28(b)), y proyectadas de Guhr con los métodos de covarianza (subfigura 3.28(c)) y de correlación (subfigura 3.28(d)) locales en el tiempo, cuando se consideran épocas sucesivas de  $q = 20$  días bursátiles, con 19 días de traslape, y habiéndose aplicado el *power map* con parámetro  $\epsilon = 0.5$ . La métrica considerada es  $p = 2$ , y en cada subgráfica el código de color se normaliza de acuerdo con el rango de sus datos.

Las gráficas en 3.29 deben compararse con aquellas de la figura 3.5. Las épocas de alta volatilidad, particularmente las que parten de 2020, pueden observarse en todos los casos de las correlaciones reducidas, al igual que la centralidad predominante de las distribuciones.

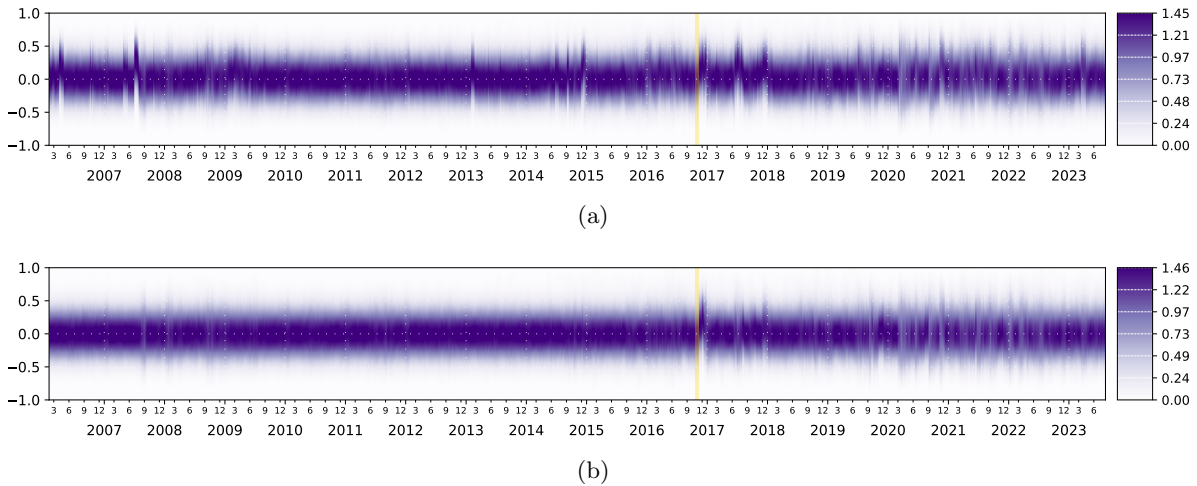


Figura 3.29: Representación de la dinámica de las correlaciones proyectadas locales de Guhr mediante mapas de calor. Para cada  $t$ , las entradas no diagonales de la matriz proyectada tienen una cierta distribución, cuya densidad de probabilidad puede obtenerse con el método *Kernel Density Estimation*. La densidad correspondiente, soportada en el intervalo  $[-1, 1]$ , se marca con púrpura en distintas intensidades que varían tanto a lo largo del eje vertical, como del eje horizontal. La banda vertical delgada de color amarillo, ubicada alrededor de noviembre de 2016, marca la época de la figura C.1, donde la media de la distribución es  $\bar{c} \approx 0.17$  para la subfigura 3.29(a), y  $\bar{c} \approx 0.0$  para la subfigura 3.29(b).

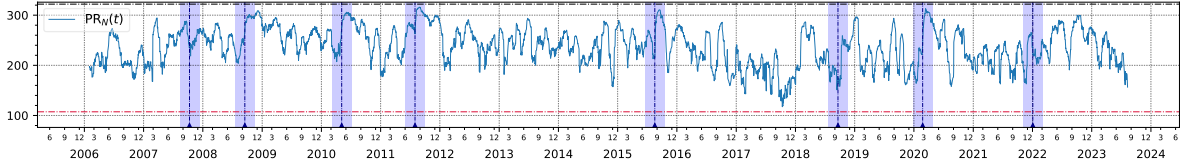
La figura 3.30 muestra las series de tiempo de la razón de participación del eigenvector principal  $PR_N$  de las matrices de correlación locales de Pearson  $\mathcal{C}^{\text{Pearson}}$ , relativas al índice  $\mathcal{C}^{\text{GSPC}}$ , y proyectadas de Guhr  $\mathcal{C}^{\text{B}}$  (método de covarianza) y  $\mathcal{C}^{\text{L}}$  (método de correlación). Puede notarse que, a partir del segundo renglón, las series de tiempo tienen correlación; véase la figura 3.31.

La matriz de dispersión de la figura 3.31 da evidencia de que la razón de participación  $PR_{N-1}$  de las matrices de correlación locales de Pearson, guarda cierta similitud con la razón de participación  $PR_N$  de las demás matrices. Sin embargo, éstas últimas tienen estadísticas más parecidas entre sí, y muestran una correlación comparativamente mayor.

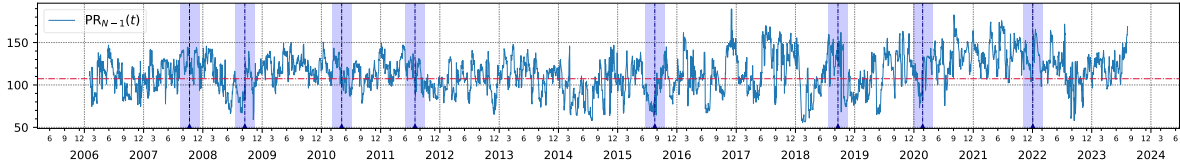
El efecto del estado de COVID se sigue observando en las proyecciones de Guhr, aunque su primera aparición y la numeración correspondiente son distintas; véase la figura 3.32. Esto podría deberse a que el número apropiado de estados es distinto de  $k = 5$  (véanse las figuras C.18 y C.20 del Apéndice C).

Los centroides de los  $k = 5$  cúmulos se muestran en la gráfica 3.33. Es notable que las correlaciones positivas predominan en la proyección de Guhr con el método de covarianza. Las distribuciones de las entradas correspondientes se pueden observar y comparar en la figura C.17 del Apéndice C, complementada a su vez con la tabla C.3 (mismo apéndice), donde se reportan los primeros cuatro momentos centrales de las entradas no diagonales de los centroides de los *clusters* en  $\mathcal{C}^{\text{B}}$  y  $\mathcal{C}^{\text{L}}$ . En el caso de  $\mathcal{C}^{\text{L}}$  sólo es monótono el incremento de la media con el índice del centroide, y en el caso de  $\mathcal{C}^{\text{B}}$ , es notable la disminución monótona de la kurtosis.

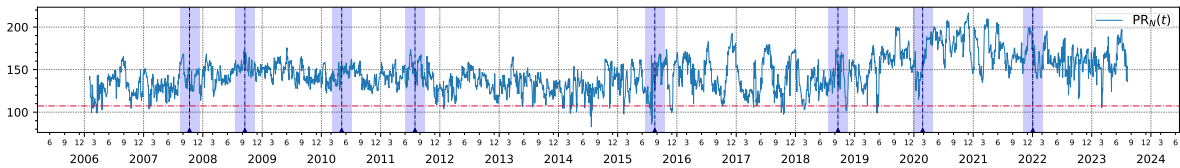
En cuanto a las matrices de transición, puede verse también una simetría aproximada, y que las transiciones entre estados no adyacentes son raras (aunque menos numerosas en el caso de las



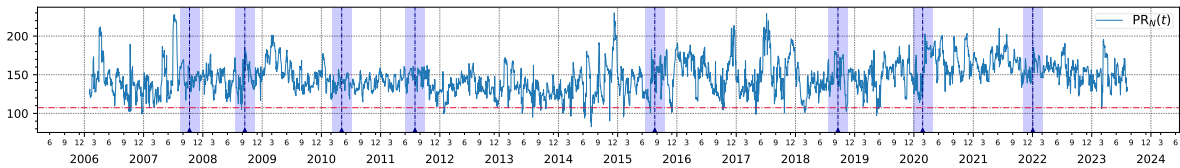
(a)



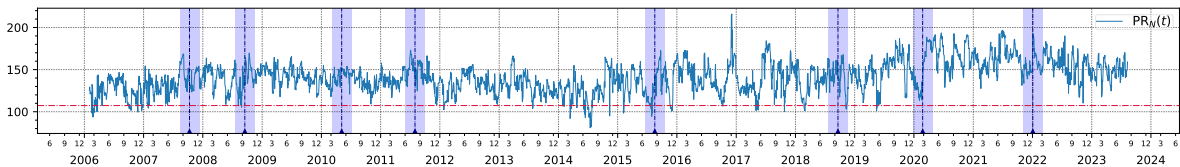
(b)



(c)



(d)



(e)

Figura 3.30: Series de tiempo de la razón de participación de los eigenvectores principales  $PR_N$  y  $PR_{N-1}$  de las matrices de correlación locales de Pearson (subfiguras 3.30(a) y 3.30(b)), y de la razón de participación del eigenvector principal  $PR_N$  para las correlaciones locales relativas al índice (figura 3.30(c)), y proyectadas de Guhr con los métodos de covarianza (figura 3.30(d)) y correlación (figura 3.30(e)). Las franjas verticales translúcidas de color azul marcan períodos de crisis; véase la tabla A.2. Las líneas horizontales discontinuas corresponden a los valores del WOE (color rojo) y del CWOE de correlación constante (color negro).

matrices  $\mathcal{C}^L$ ); véase la figura 3.34. Para las transiciones en  $\mathcal{C}^L$ , es bastante menor la ocupación del cúmulo 2 (subfigura 3.34(b)).

Las distribuciones de probabilidad estacionarias, asociadas a las matrices de transición graficadas en la figura 3.34, se muestran en la figura 3.35. El valor de  $\pi_2^*$  en la gráfica 3.35(b),

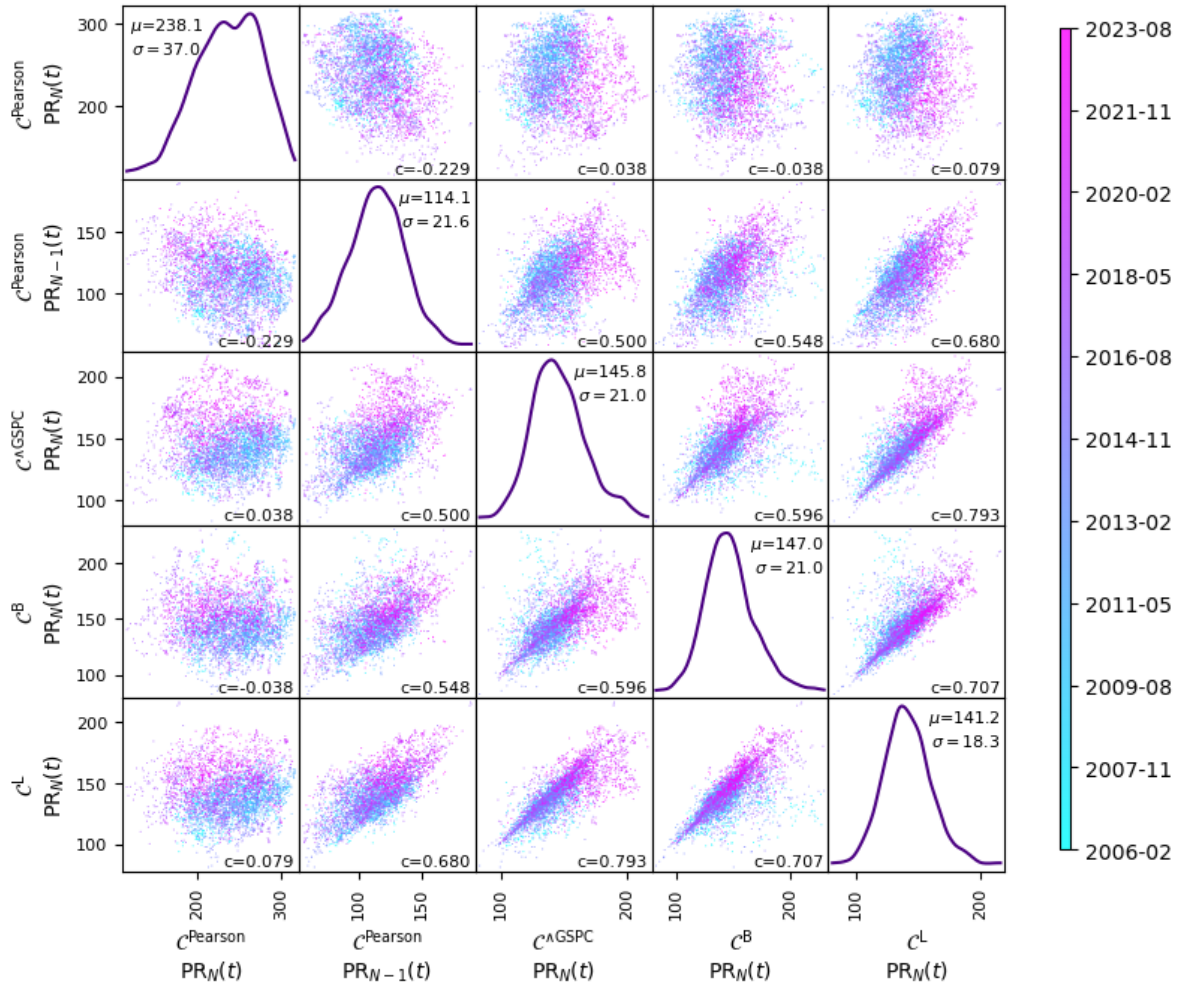
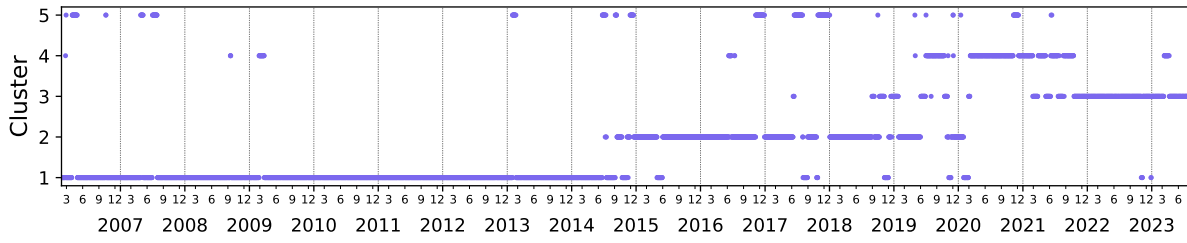


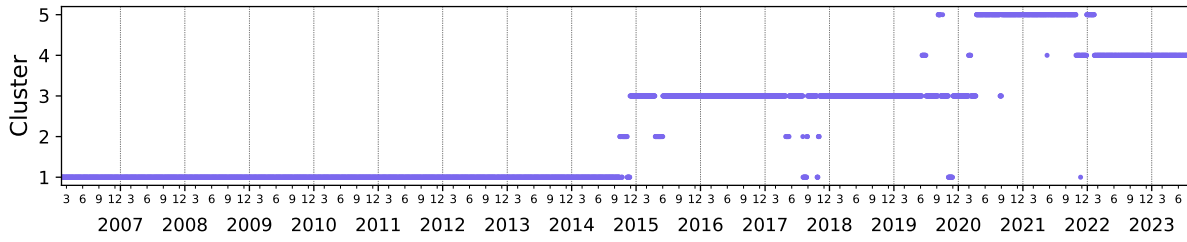
Figura 3.31: Matriz de dispersión de las series de tiempo de la razón de participación de los eigenvectores principales  $PR_N$  y  $PR_{N-1}$  de las matrices de correlación locales de Pearson  $C^{\text{Pearson}}$ , y de la razón de participación del eigenvector principal  $PR_N$  para las correlaciones locales relativas al índice  $C^{\wedge\text{GSPC}}$ , y proyectadas de Guhr con los métodos de covarianza  $C^{\text{B}}$  y correlación  $C^{\text{L}}$ ; véase la figura 3.30. En la diagonal está graficada la distribución KDE de cada serie de tiempo, y se muestran media y desviación estándar de los datos. En los bloques no-diagonales, se muestra el coeficiente de correlación de Pearson de los datos correspondientes.

sugiere que el cúmulo 2 en no está bien diferenciado en  $C^{\text{L}}$ ; ésto parece reafirmarse por el parecido entre los centroides 1 y 2 de la gráfica 3.33(b), y posiblemente los dos cúmulos correspondientes son realmente uno solo.

Las distribuciones de las componentes de los eigenvectores principales, tanto de cada centroide como de los miembros del cúmulo, se pospone al Apéndice C, haciendo el ajuste de cúmulos a  $k = 4$ .

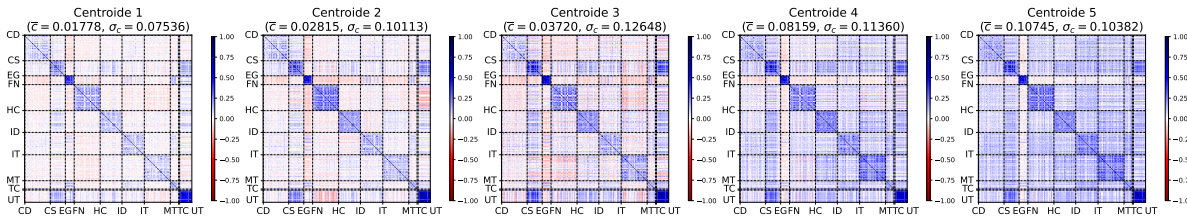


(a)

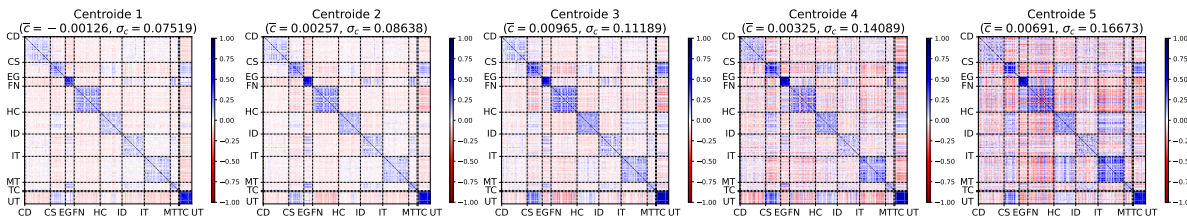


(b)

Figura 3.32: Evolución de los estados de mercado para las correlaciones proyectadas de Guhr  $\mathcal{C}^B$  (método de covarianza) y  $\mathcal{C}^L$  (método de correlación), cuando  $k = 5$  y no se aplica el *power map*. El ordenamiento de los estados es creciente con el promedio del eigenvalor máximo de los miembros de cada estado o cúmulo (relación (3.1)). Nótese que la etiqueta del eje vertical en la subfigura 3.16(b) es *Cluster*, para diferenciar los cúmulos de las matrices de correlación proyectadas de Guhr, de los estados de mercado de las matrices de correlación de Pearson.



(a)



(b)

Figura 3.33: Centroides de los estados de mercado para las correlaciones proyectadas de Guhr, cuando  $k = 5$  y no se aplica el *power map*.

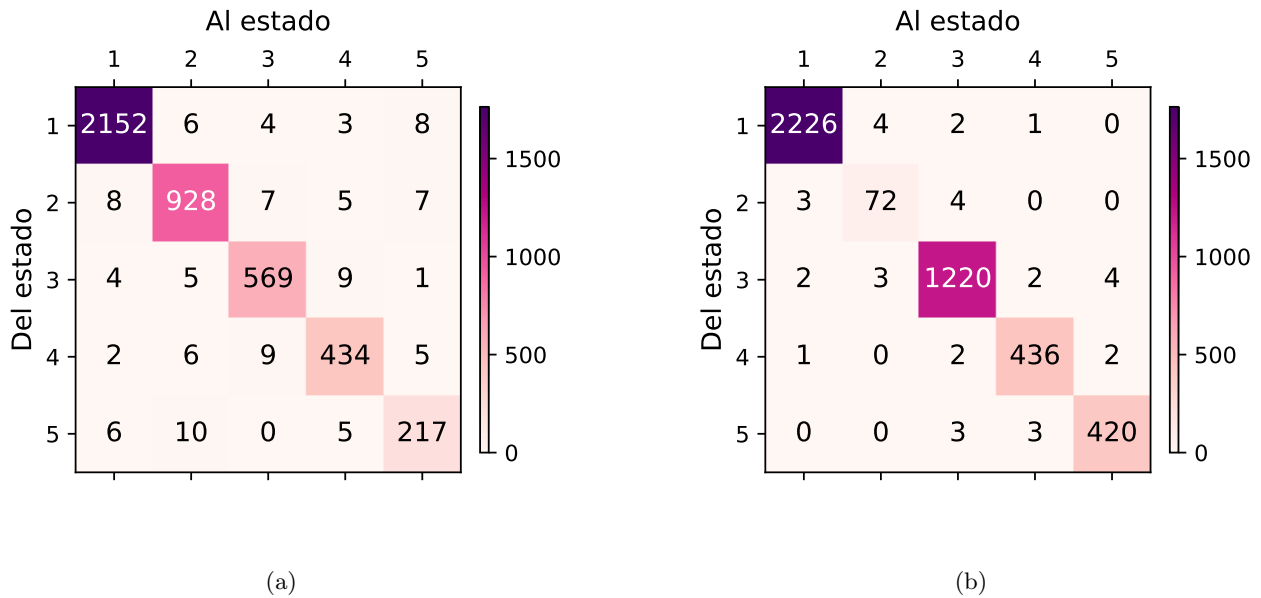


Figura 3.34: Matrices de transición para las correlaciones proyectadas de Guhr  $\mathcal{C}^B$  (método de covarianza) y  $\mathcal{C}^L$  (método de correlación), cuando  $k = 5$  y no se aplica el *power map*. La dinámica de los estados de la figura 3.32 se traduce en matrices de transición cuasi-simétricas, pero no tridiagonales como en 3.17. La subfigura 3.34(a) corresponde a los cúmulos visitados por las correlaciones  $\mathcal{C}^B$ , mientras que la subfigura 3.34(b) corresponde a los *clusters* ocupados por las correlaciones  $\mathcal{C}^L$ .

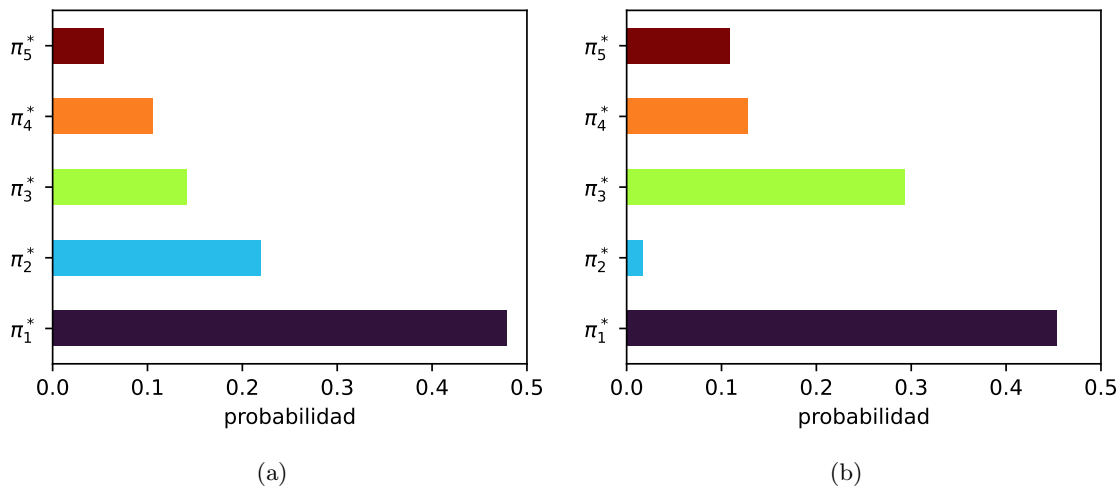


Figura 3.35: Distribuciones de probabilidad estacionarias, calculadas mediante la condición (2.42), a partir de las matrices de transición mostradas en la figura 3.34, cuando aquellas se normalizan de acuerdo con la fórmula (2.41). Los valores son: método de covarianza:  $\pi^* = (0.479, 0.2192, 0.1418, 0.106, 0.0541)$ ; método de correlación:  $\pi^* = (0.453, 0.0172, 0.2935, 0.1275, 0.1088)$ .





## Capítulo 4

# Correlaciones de grano grueso

Los centros geométricos de los cúmulos que conforman los estados de mercado, como los mostrados en la Figura 3.19(a), muestran a simple vista la persistencia de altas correlaciones intrasectoriales. Las correlaciones intersectoriales, por otra parte, parecen evolucionar en patrones más complejos. En este capítulo estudiaremos la medida en que, considerando solamente las correlaciones a escala sectorial, los estados de mercado resultantes preservan la información de las correlaciones por pares. Utilizamos datos de acciones de compañías representativas de los EE.UU. y Japón, y los resultados fueron publicados en [103].

El método empleado fue introducido por Rinn *et al* en [93]. La idea es que las  $N \times N$  componentes de una matriz de correlación (en nuestro caso, de rendimientos logarítmicos de precios de acciones), previamente organizada en  $M$  bloques (en nuestro caso, industriales), se promedian en bloques internos y externos (es decir: intra e intersectoriales) para reducir de manera importante la dimensión de los datos y disminuir el ruido de estimación de épocas cortas. Cabe resaltar que:

- La diagonal principal de esta nueva matriz (que en lo sucesivo denominaremos *de grano grueso* o *matriz sectorial de Guhr*), no es trivial como en las matrices de correlación de Pearson, relativas o proyectadas. De esta manera, la matriz de grano grueso se asemeja más a una matriz de covarianza.
- Los bloques pueden ser (y normalmente son) de distintos tamaños.
- Los  $N(N-1)/2$  números independientes de una matriz de correlación, se relacionan con  $M(M+1)/2$  números de la matriz de grano grueso. Por ejemplo, para los datos descritos en el Apéndice A, una matriz de correlación de  $N \times N = 350 \times 350 = 122,500$  se convierte en una de  $M \times M = 10 \times 10 = 100$  entradas, reduciendo efectivamente  $N(N-1)/2 = 61,075$  coeficientes independientes en  $M(M+1)/2 = 55$  números (véase la fórmula (2.3.1)). Esto es muy ventajoso en términos computacionales.
- En [93] no se especifica si los coeficientes de la diagonal principal deben ser excluidos. En nuestro caso, optaremos por hacerlo.

Esta idea de dividir a una matriz simétrica de  $N \times N$  en  $M \times M$  bloques ( $1 \leq M < N$ ), podría aplicarse también cuando los datos tienen otra estructura *natural*, como en el caso de las

correlaciones entre rendimientos de acciones en mercados de países distintos. Igualmente, podría prescindirse de esa *naturalidad* y más bien considerarse todas las posibles agrupaciones. Sin embargo, si consideramos el ejemplo más simple de  $M = 2$  y  $N$  par, entonces hay un total de  $\binom{N}{N/2}$  posibles matrices de grano grueso de  $2 \times 2$ . Con la aproximación de Stirling, resulta  $\binom{N}{N/2} \approx 2^N$ , lo que está fuera de nuestro alcance en términos computacionales si pretendemos ser exhaustivos. Podría ser interesante si se considera un *ensemble* de agrupaciones con un muestreo adecuado [103].

Por otra parte, en lugar de promediar los bloques, pueden sumarse sus elementos para evitar la heterogeneidad en el número de componentes por cada sector. Si todos los bloques son del mismo tamaño, la única diferencia entre la matriz sectorial de promedios y la de sumas es de una constante multiplicativa. En este trabajo nos vamos a concentrar en matrices sectoriales de promedios, para los datos del S&P 500 y Nikkei 225, en el lapso de enero de 2006 a diciembre de 2019. Nótese que los datos de acciones estadounidenses difieren de los analizados en el Capítulo 3, dada la aparición del estado de mercado anómalo.

## 4.1. Generalidades de las matrices sectoriales de Guhr

Las figuras 4.1 y 4.1 muestran las matrices de correlación de horizonte completo, para los datos del S&P 500 y Nikkei 225 (véase el Apéndice A), respectivamente.

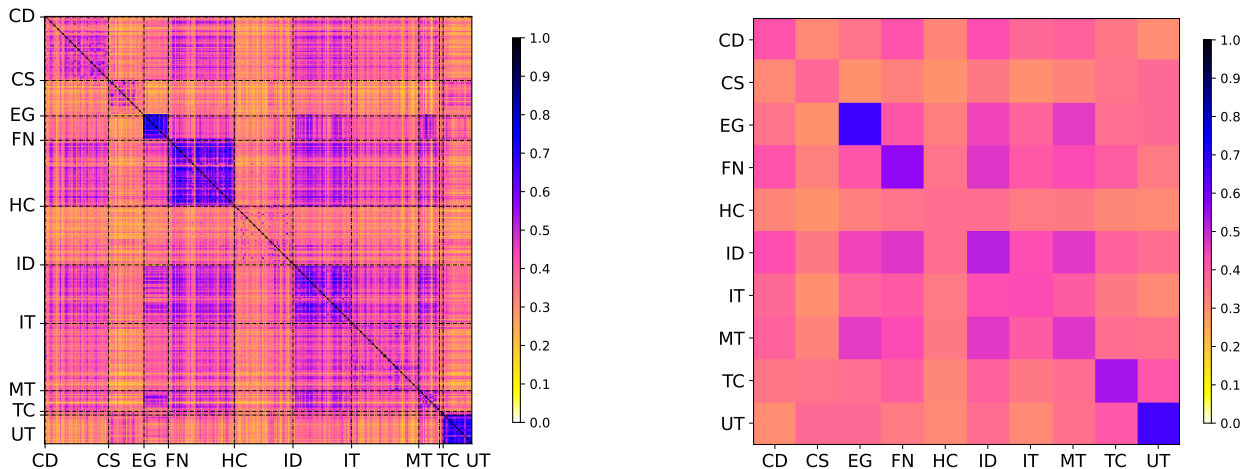


Figura 4.1: Matrices de correlación de Pearson (izquierda) y grano grueso (derecha) para datos del S&P 500, sobre el lapso de enero de 2006 a diciembre de 2019. Los sectores industriales están indicados en la tabla A.1. El haber tomado el promedio sobre las correlaciones inter e intrasectoriales (excluyendo, en éste caso, a las autocorrelaciones), asigna de forma indirecta un peso mayor a aquellos sectores industriales con un menor número de componentes.

### 4.1.1. Estado del mercado

Para empezar a estudiar si las matrices sectoriales de Guhr locales en el tiempo retienen *la información* de las matrices de correlación de Pearson (que en este capítulo también

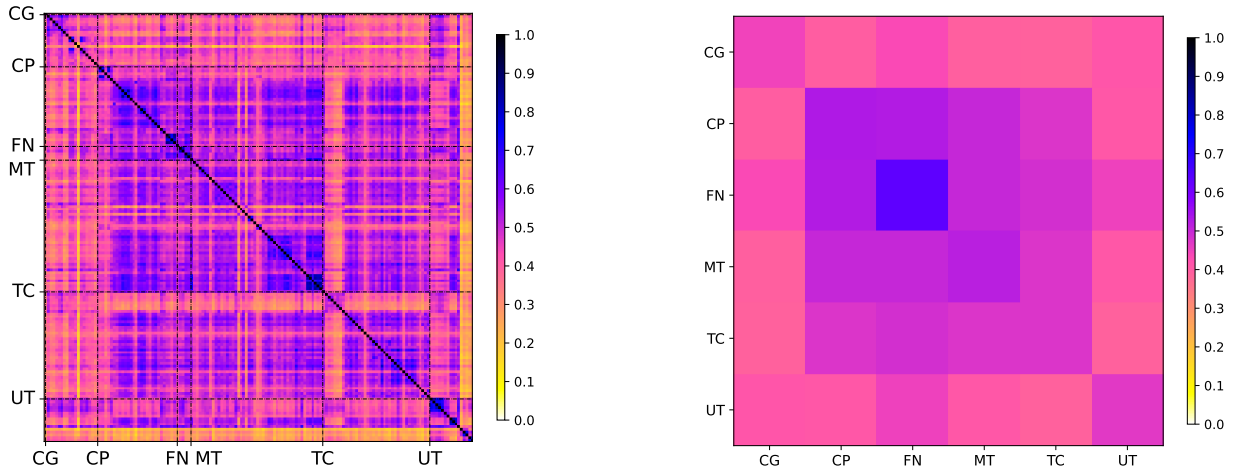


Figura 4.2: Matrices de correlación de Pearson (izquierda) y grano grueso (derecha) para datos del Nikkei 225, sobre el lapso de enero de 2006 a diciembre de 2019. Los sectores industriales están indicados en la tabla A.4. El haber tomado el promedio sobre las correlaciones inter e intrasectoriales (excluyendo, en éste caso, a las autocorrelaciones), asigna de forma indirecta un peso mayor a aquellos sectores industriales con un menor número de componentes.

denominaremos *matrices (de correlación) completas* locales, obtuvimos las series de tiempo  $\lambda_{\max}$  para épocas de longitud  $q = 20$  sin *power map* ( $\epsilon = 0$ ), reconocidas como reflejantes del *estado del mercado*, para ambos conjuntos de datos: estadounidenses y japoneses; véanse las figuras 4.3 y 4.4. Puede observarse que, en efecto, el máximo eigenvalor en ambos tipos de matrices está fuertemente correlacionado: en una cierta manera, es *suficiente* conocer la correlación promedio intra e intersectorial para definir el estado del mercado.

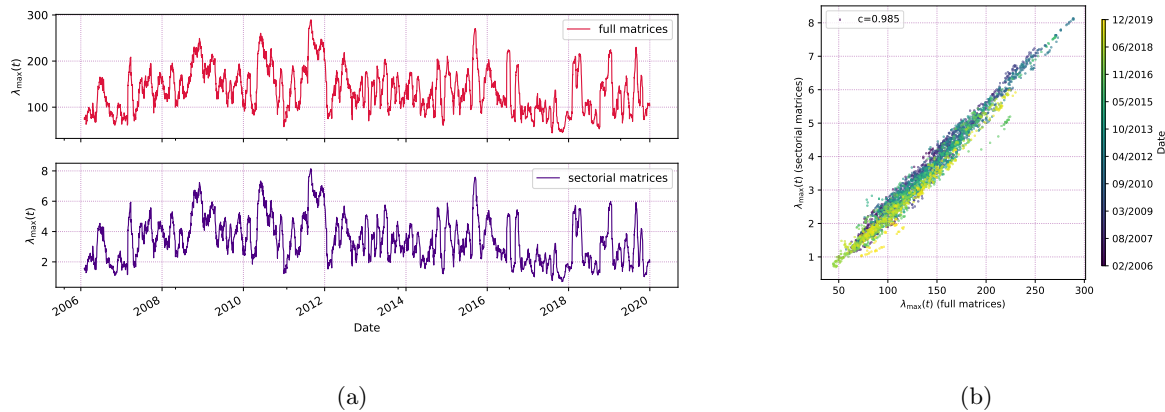


Figura 4.3: Comparación entre las series de tiempo de eigenvalores máximos, para las matrices de correlación locales de Pearson (figura 4.3(a), parte superior) y sectoriales de Guhr (figura 4.3(a), parte inferior) en los datos del S&P 500. La figura 4.3(b) muestra el diagrama de dispersión de las series de tiempo en 4.3(a), con el coeficiente de correlación de Pearson enmarcado en la esquina superior izquierda. Las épocas son de 20 días bursátiles, y el traslape entre matrices consecutivas es de 19 días.

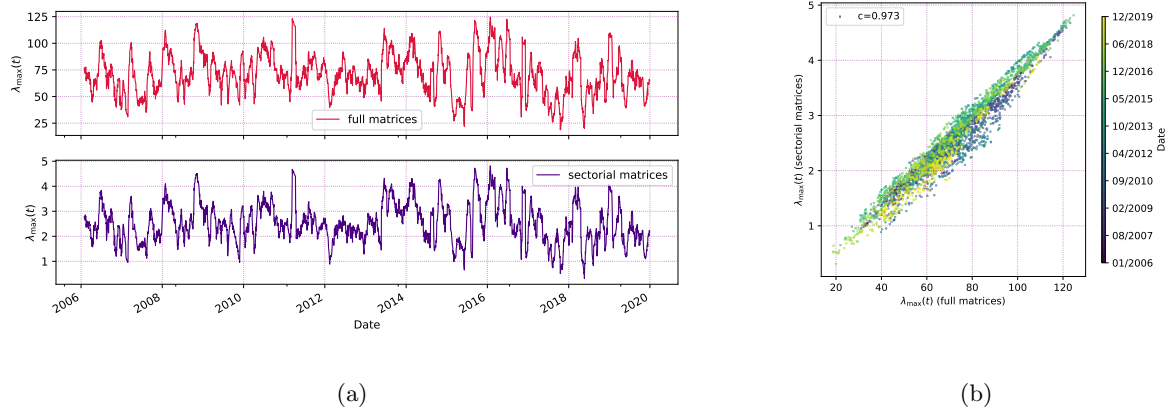


Figura 4.4: Comparación entre las series de tiempo de eigenvalores máximos, para las matrices de correlación locales de Pearson (figura 4.4(a), parte superior) y sectoriales de Guhr (figura 4.4(a), parte inferior) en los datos del Nikkei 225. La figura 4.4(b) muestra el diagrama de dispersión de las series de tiempo en 4.4(a), con el coeficiente de correlación de Pearson enmarcado en la esquina superior izquierda. Las épocas son de 20 días bursátiles, y el traslape entre matrices consecutivas es de 19 días.

Más aún: se preserva la correlación hasta el segundo eigenvalor máximo,  $\lambda_{N-1}$ , aunque a un grado menor; véanse las figuras 4.5 y 4.6.

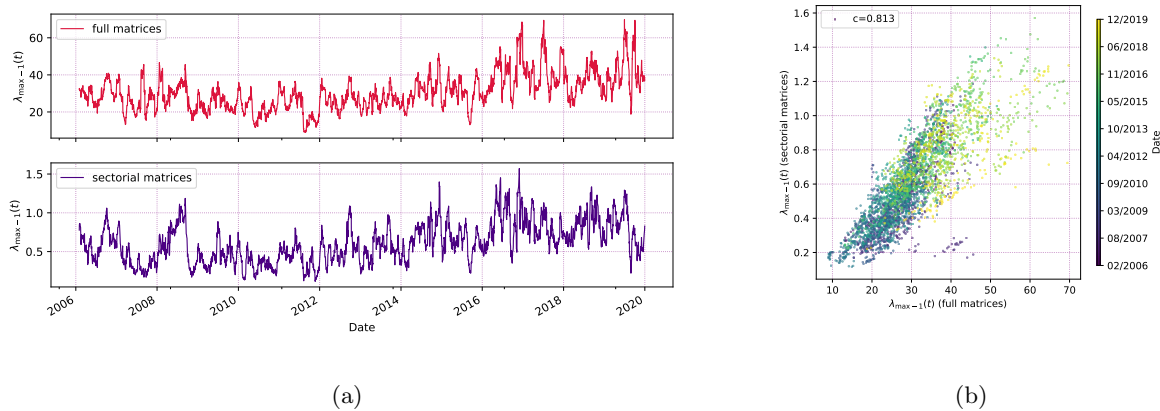


Figura 4.5: Comparación entre las series de tiempo del segundo eigenvalor más grande, representado por  $\lambda_{\max-1}$ , para las matrices de correlación locales de Pearson (figura 4.5(a), parte superior) y sectoriales de Guhr (figura 4.5(a), parte inferior) en los datos del S&P 500. La figura 4.5(b) muestra el diagrama de dispersión de las series de tiempo en 4.5(a), con el coeficiente de correlación de Pearson enmarcado en la esquina superior izquierda. Las épocas son de 20 días bursátiles, y el traslape entre matrices consecutivas es de 19 días.

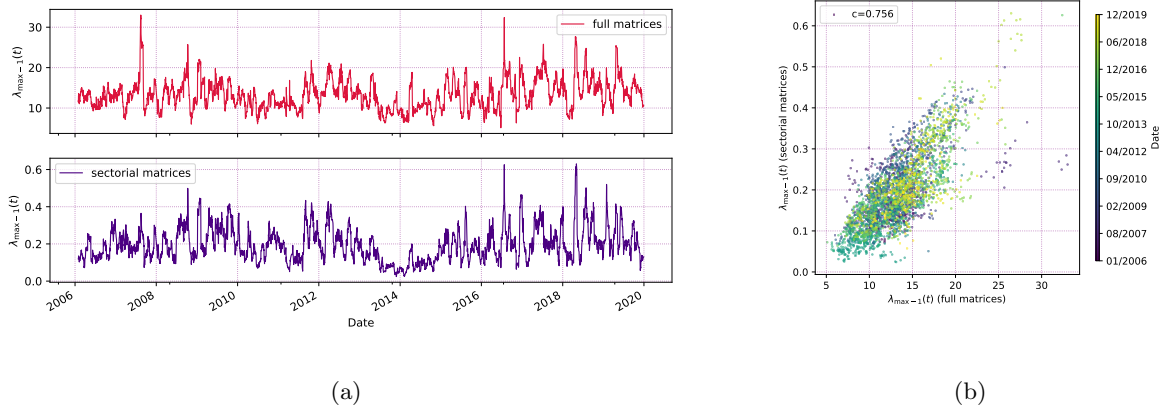


Figura 4.6: Comparación entre las series de tiempo del segundo eigenvalor más grande, representado por  $\lambda_{\text{máx}-1}$ , para las matrices de correlación locales de Pearson (figura 4.6(a), parte superior) y sectoriales de Guhr (figura 4.6(a), parte inferior) en los datos del Nikkei 225. La figura 4.6(b) muestra el diagrama de dispersión de las series de tiempo en 4.6(a), con el coeficiente de correlación de Pearson enmarcado en la esquina superior izquierda. Las épocas son de 20 días bursátiles, y el traslape entre matrices consecutivas es de 19 días.

#### 4.1.2. Matrices de similitud

En las figuras 4.7 y 4.8 se muestran las matrices de similitud (con la métrica euclidiana (2.37)) de dimensión  $(T - q + 1) \times (T - q + 1)$ , donde  $T$  es el número de puntos temporales de cada serie de tiempo de rendimientos, y  $q = 20$  es el tamaño de época para las matrices locales de correlación sucesivas, con un traslape de 19 días. Además del parecido visual, se puede considerar como medida de similitud la correlación entre las matrices de similitud de las correlaciones de Pearson y de las correlaciones sectoriales de Guhr, tomando en cuenta solamente las entradas de sus partes triangulares superiores (véase la ecuación (2.3.1)); en este caso, resulta un valor de 0.932 (ver la figura 4.9(a)) para el S&P 500, y 0.883 para Nikkei 225 (ver la figura 4.10(a)).

Para ahondar en la comparación de las matrices de similitud, estudiamos la distribución de sus elementos no diagonales y graficamos su dispersión; véanse las figuras 4.9 y 4.10. En éstas, se puede apreciar una preservación de la información de las distancias, tanto por la linealidad en la dispersión como por la forma de la distribución. Dicho de otro modo: el grano grueso sobre los sectores preserva bien las distancias, y ésto apoya la hipótesis de que los estados de mercado se preservan.

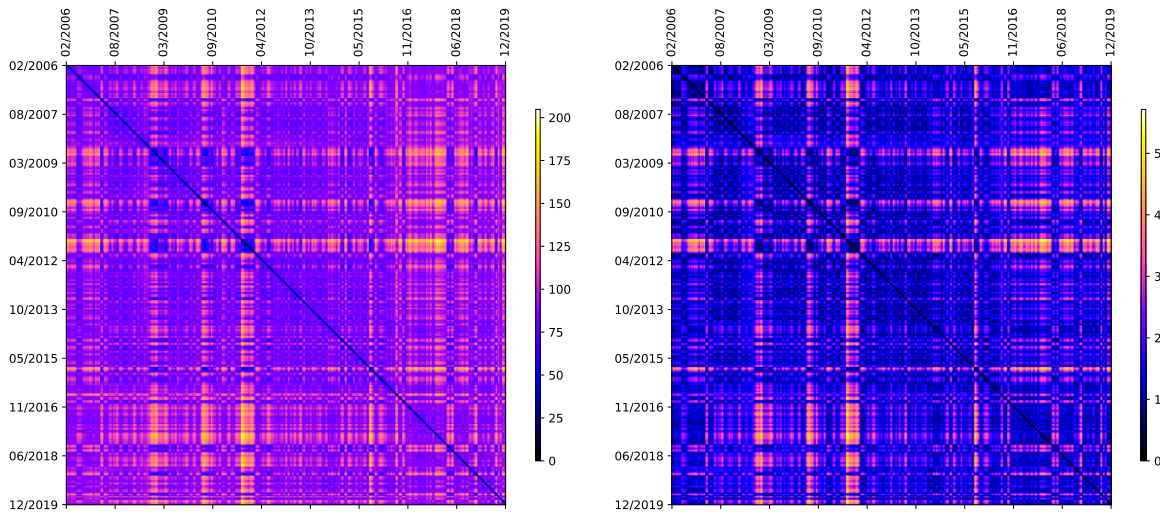


Figura 4.7: Matrices de similaridad de las secuencias de matrices de correlación locales de Pearson (izquierda) y de grano grueso (derecha), cuando se consideran épocas sucesivas de  $q = 20$  días bursátiles, con 19 días de traslape. La métrica considerada es  $p = 2$ , y en cada subgráfica el código de color se normaliza de acuerdo con el rango de datos correspondiente.

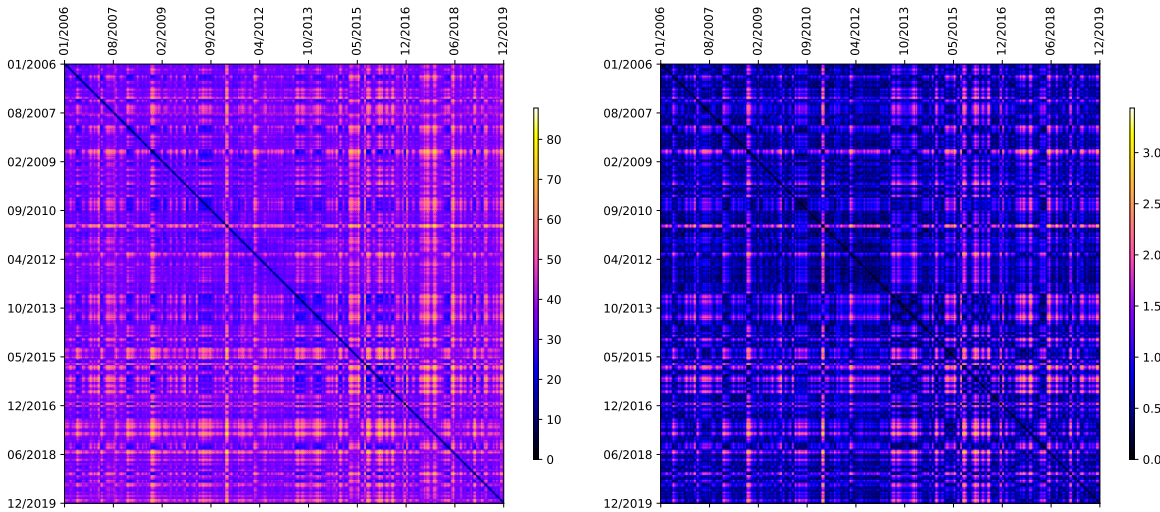


Figura 4.8: Matrices de similaridad de las secuencias de matrices de correlación locales de Pearson (izquierda) y de grano grueso (derecha), cuando se consideran épocas sucesivas de  $q = 20$  días bursátiles, con 19 días de traslape. La métrica considerada es  $p = 2$ , y en cada subgráfica el código de color se normaliza de acuerdo con el rango de datos correspondiente.

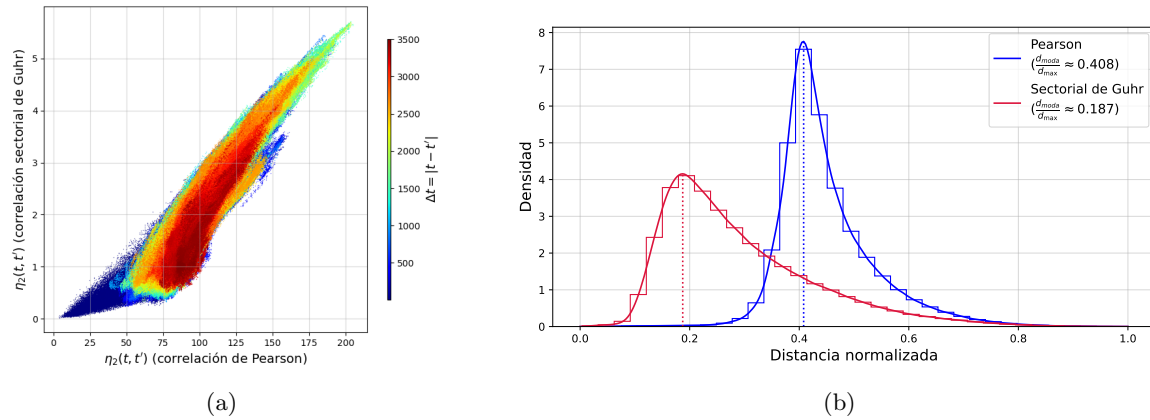


Figura 4.9: Comparación de las distancias euclidianas entre matrices de correlación de Pearson y sectoriales de Guhr. En las distribuciones de la figura 4.9(b), las distancias máximas entre las matrices de Pearson, o entre las matrices de correlación sectorial de Guhr, son 204.793 y 5.714, respectivamente, y éstas distancias sirven para normalizar las densidades. Las líneas suaves representan una *Kernel Density Estimation* con núcleos gaussianos y el método de Silverman, y las líneas escalonadas representan un histograma cuyos bins se calcularon con el método de Doane.

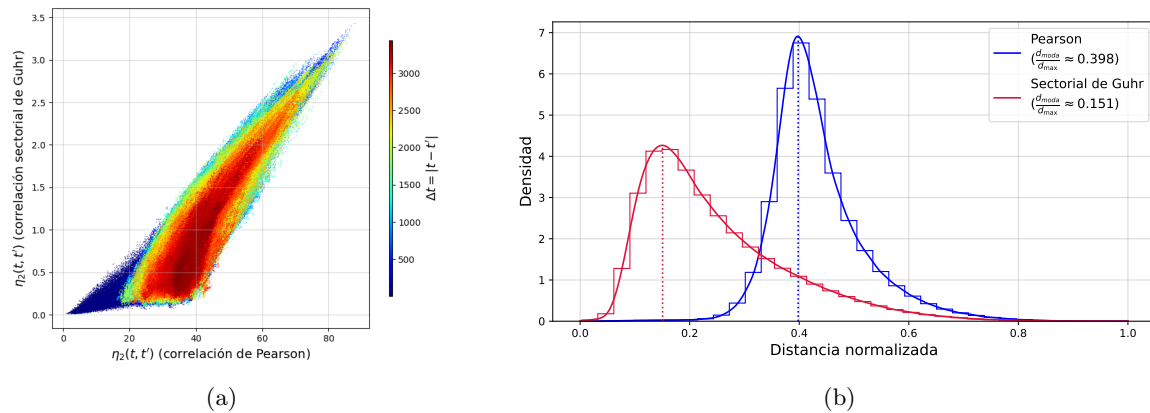


Figura 4.10: Comparación de las distancias euclidianas entre matrices de correlación de Pearson y sectoriales de Guhr. En las distribuciones de la figura 4.10(b), las distancias máximas entre las matrices de Pearson, o entre las matrices de correlación sectorial de Guhr, son 87.844 y 3.437, respectivamente, y éstas distancias sirven para normalizar las densidades. Las líneas suaves representan una *Kernel Density Estimation* con núcleos gaussianos y el método de Silverman, y las líneas escalonadas representan un histograma cuyos bins se calcularon con el método de Doane.



## 4.2. Estados de mercado

A continuación, describiremos los resultados de la acumulación de las matrices de correlación locales en el tiempo, tanto completas como sectoriales. Dividiremos los resultados de acuerdo con el índice de donde se obtuvieron las acciones estadounidenses y japonesas analizadas.

### 4.2.1. S&P 500

En la figura 4.11 se muestra la evolución temporal de los estados de mercado, utilizando el K-means (variante *k-means ++*, implementada en `scikit-learn`) con  $k = 5$  cúmulos. Específicamente: por 15 corridas, cada una con 10 condiciones iniciales, aplicamos el k-means descrito en la sección 2.3 sobre la secuencia de  $T - q + 1 = 3503$  matrices sectoriales de Guhr, en su versión *aplanada*.

Para cada corrida, tras un cierto número de iteraciones se obtiene convergencia para cada condición inicial, y se admite como válido el resultado con el menor ESS (véase la ecuación (2.36)). Así, resultan 15 secuencias de  $T - q + 1$  números con rango  $\{0, \dots, k - 1\}$ , representantes de los cúmulos hallados por el algoritmo pero sin un orden especial. Una forma canónica de reetiquetar los cúmulos de las 15 corridas, es a través del promedio de las componentes del centroide de cada cúmulo<sup>1</sup>; véase la figura B.1 del Apéndice B.

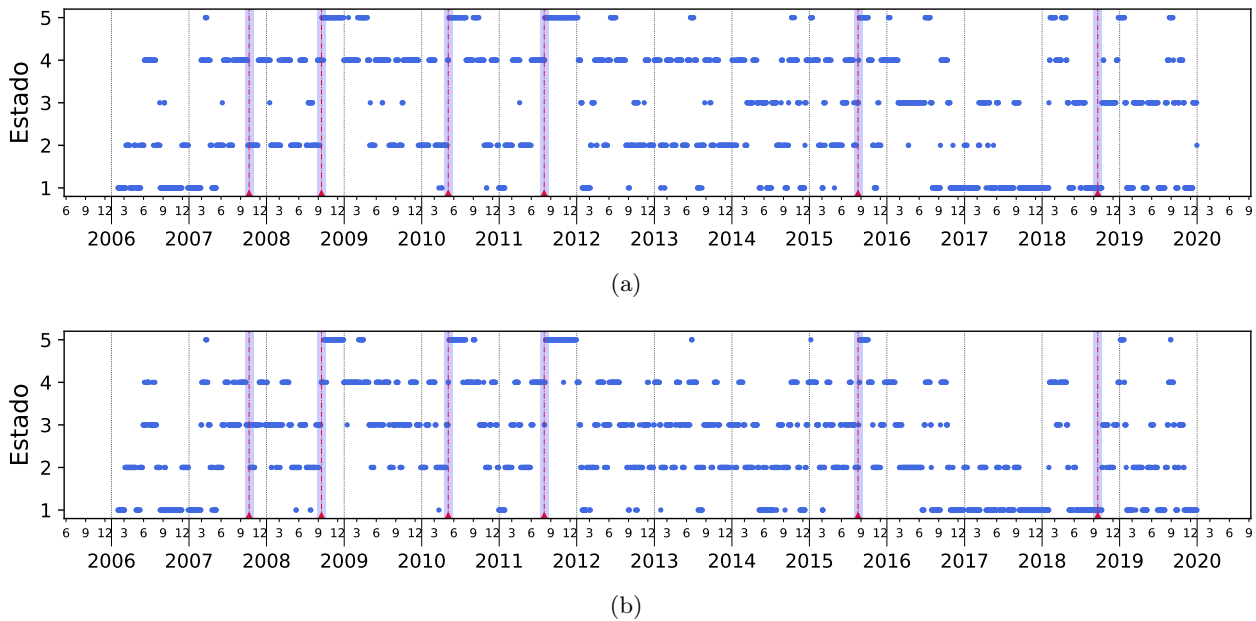


Figura 4.11: Figura 4.11(a): evolución de los estados de mercado en el S&P 500, cuando se utilizan las matrices de correlación de Pearson en épocas sucesivas de  $q = 20$  días. Figura 4.12(b): evolución de los estados de mercado en el S&P 500, cuando se utilizan las matrices sectoriales de Guhr en las mismas condiciones. Las bandas verticales de color azul, cuyo ancho abarca 40 días naturales, marcan épocas críticas del mercado estadounidense (ver la tabla A.2).

<sup>1</sup>Los centroides mismos son matrices de correlación, por lo que el promedio de sus componentes equivale a la correlación promedio de los miembros del cúmulo.

Estado	Pearson	Guhr	Coincidencias
1	749	773	668
2	774	931	512
3	504	798	78
4	1033	716	559
5	443	285	285
Total	3503	3503	2102

Cuadro 4.1: Ocupación total de cada estado en las  $T - q + 1 = 3503$  épocas consideradas, tanto para las matrices de Pearson como para las de Guhr del S&P 500. El total de coincidencias es del  $2102/3503 \approx 60\%$ .

Las figuras 4.12 y 4.13 muestran la comparación de las matrices que conforman los centroides de los estados de mercado en las correlaciones de Pearson, de grano grueso, y el grano grueso de los centroides de las correlaciones de Pearson. Puede observarse que el proceso de granular y clasificar (figura 4.12(b)), da un resultado parecido al de clasificar y luego granular (figura 4.12(c)).

Las matrices de transición correspondientes a la dinámica mostrada en la Figura 4.11, se encuentran en la Figura 4.13. Puede notarse una mayor simetría y tridiagonalidad en las transiciones de los estados de mercado sectoriales, lo que se ajusta más a la markovianidad de las transiciones. Los estados 3 y 5 son los que tienen una probabilidad de ocupación más disímil, aunque los centroides correspondientes tienen una alta similitud cuando se granulan las matrices de Pearson; véase la Figura 4.12. También las proyecciones MDS de las matrices de correlación local, tienen una configuración similar; véase la Figura 4.14. Quizá la diferencia más evidente, sea la alineación de los estados en las matrices sectoriales con el eje  $x$  (elegido como el de mayor varianza), lo que explica la mayor tridiagonalidad de la matriz 4.13(b): las transiciones son más comunes entre estados geoméricamente adyacentes. Sin embargo, cuando utilizamos el *power map* en las matrices de correlación completas, la configuración de sus proyecciones MDS se asemeja más a aquella de las matrices sectoriales proyectadas; véanse [103] (particularmente la sección 3.1 y el Apéndice) y la subsección siguiente. Por último, notamos que el eje de mayor varianza de la proyección MDS está fuertemente correlacionado con el mayor eigenvalor, por lo que describe bien el estado del mercado (es decir:  $\lambda_{\text{máx}}$ ). Véase la Figura 4.15.

#### 4.2.2. Nikkei 225

Siguiendo la línea de [41], repetimos el procedimiento descrito en la sección anterior para varios valores de  $(k, \epsilon)$ , buscando optimizar (minimizar) la siguiente cantidad:

$$\sigma_{d_{intra}} = \sqrt{\left\langle \left( d_{intra}^{(m)} \right)^2 \right\rangle_m - \left( \left\langle d_{intra}^{(m)} \right\rangle_m \right)^2} \quad (4.1)$$

donde:

$$d_{intra}^{(m)} = \frac{1}{k} \sum_{j=1}^k \left( \frac{1}{n_j} \sum_{l=1}^{n_j} \left\| C_j - p_l^{(j)} \right\| \right), \quad m = 1, \dots, 1000 \quad (4.2)$$

y  $m$  indica una condición inicial del  $k$ -means. El resultado de la optimización puede verse en la Figura 4.2.2. En la figura 4.17 se muestra la evolución temporal de los estados de mercado en el

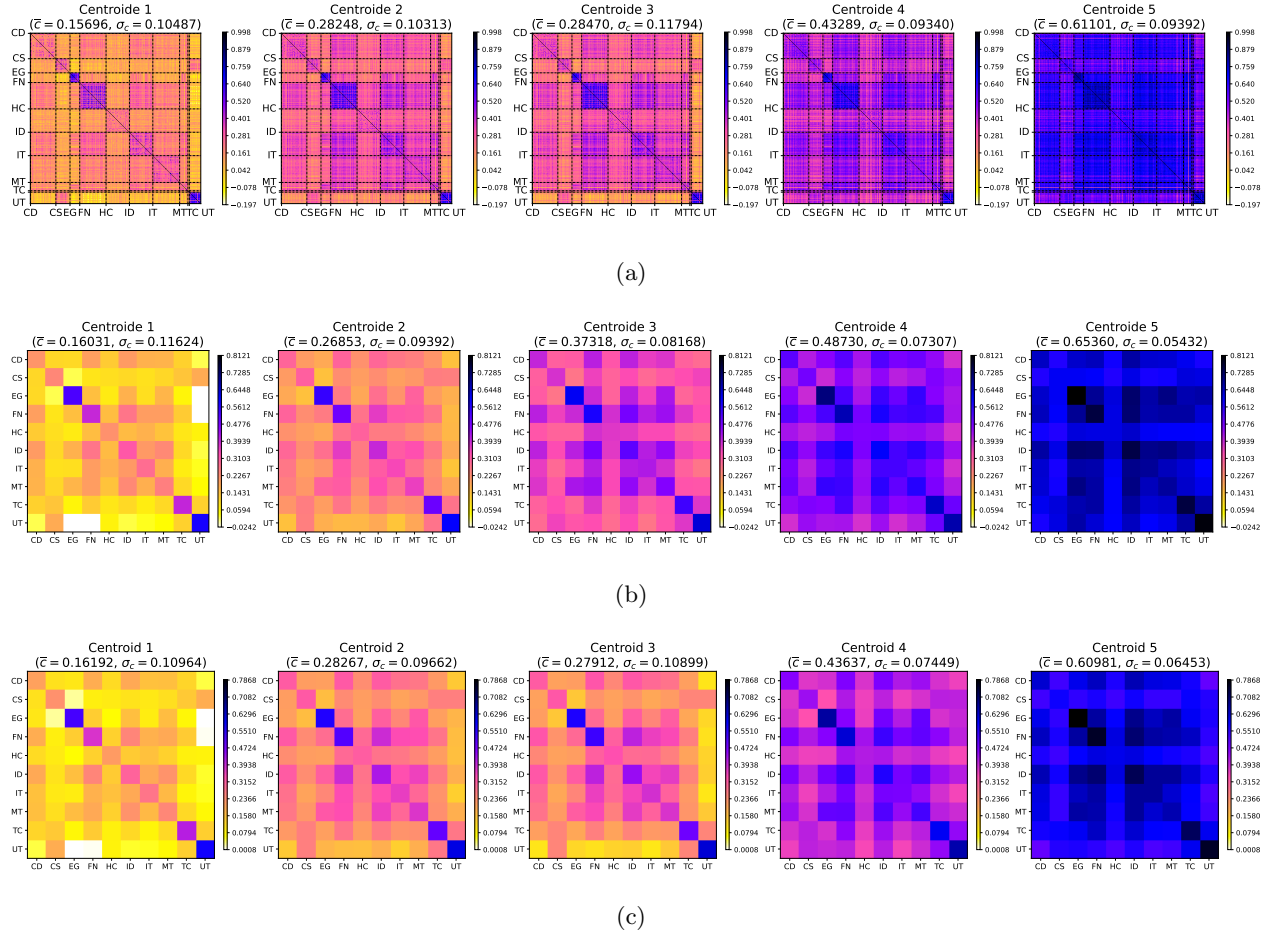


Figura 4.12: (S&P 500) Figura 4.12(a): Centroides (o centros geométricos) en el espacio  $N(N-1)/2$ -dimensional de correlaciones de Pearson, correspondientes a los estados (cúmulos) ocupados en la figura 4.11(a). Figura 4.12(b): Centroides en el espacio  $M(M+1)/2$ -dimensional de correlaciones sectoriales de Guhr, correspondientes a los estados ocupados en la figura 4.11(b). Figura 4.12(c): Matrices sectoriales de Guhr calculadas a partir de los centroides de la figura 4.12(a). Encima de cada matriz se encuentran el promedio y desviación estándar de cada centroide, cuando se utilizan solamente las componentes de la matriz triangular superior (excluyendo la diagonal principal en el caso de las matrices de Pearson, e incluyéndola en el caso de las matrices sectoriales de Guhr). Pensando en las matrices de las figuras 4.12(b) y 4.12(c) como vectores, tomando solamente las partes triangulares superiores, el parecido de los centroides de cada estado puede medirse a través de la correlación; resulta 0.9950, 0.9960, 0.9458, 0.9822, 0.9343, respectivamente.

Nikkei 225, utilizando  $k = 6$  cúmulos y  $\epsilon = 0.3$ . La proporción de épocas donde la clasificación del estado según el análisis de las matrices completas, y el de las matrices sectoriales, es de aproximadamente 53%; los detalles de ocupación pueden verse en la Tabla 4.2.2.

Las figuras 4.18 y 4.19 muestran la comparación de las matrices que conforman los centroides de los estados de mercado en las correlaciones de Pearson, de grano grueso, y el grano grueso de los centroides de las correlaciones de Pearson. Puede observarse, como antes, que el proceso de granular y clasificar (figura 4.18(b)), da un resultado parecido al de clasificar y luego granular (figura 4.18(c)).

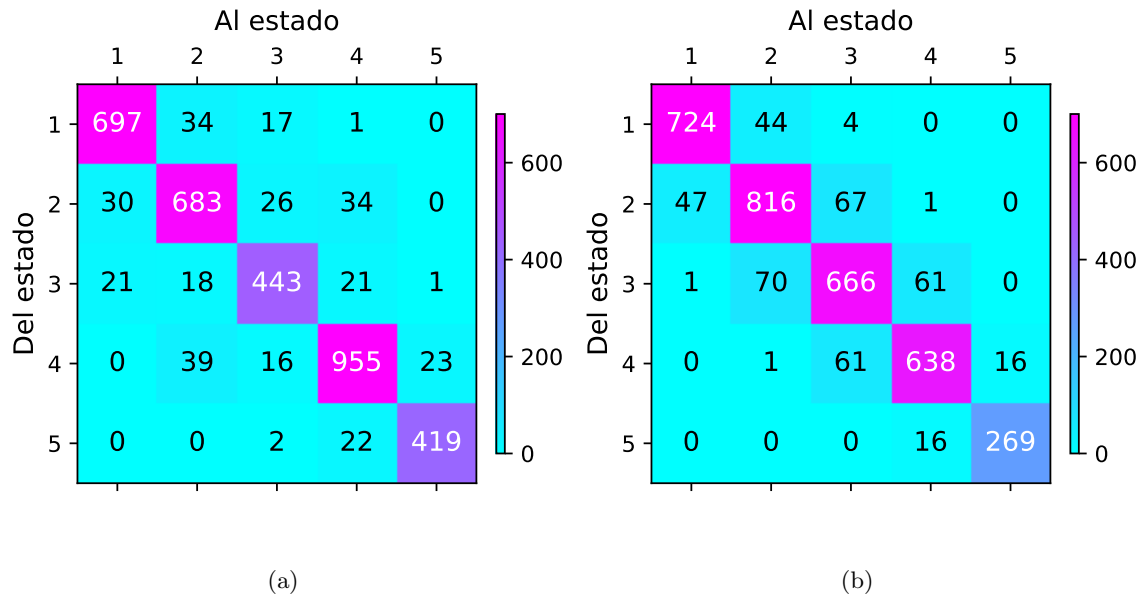


Figura 4.13: (S&P 500) Matrices de transición (o transferencia) de los estados mostrados en la figura 4.11. A la izquierda (4.13(a)): transiciones de los estados de mercado para las matrices de Pearson; a la derecha (4.13(b)): transiciones de los estados de mercado para las matrices sectoriales de Guhr. Las distribuciones estacionarias de probabilidad son: 4.13(a) (0.211, 0.222, 0.144, 0.296, 0.127) y 4.13(b) (0.220, 0.266, 0.228, 0.204, 0.082).

Estado	Pearson	Guhr	Coincidencias
1	679	528	501
2	569	924	411
3	893	386	62
4	299	714	27
5	654	564	504
6	345	323	311
<i>Total</i>	3439	3439	1816

Cuadro 4.2: (Nikkei 225) Ocupación total de cada estado en las  $T - q + 1 = 3439$  épocas consideradas, tanto para las matrices de Pearson como para las de Guhr. El total de coincidencias es del  $1816/3439 \approx 53\%$ .

Las matrices de transición correspondientes a la dinámica mostrada en la Figura 4.17, se encuentran en la Figura 4.19. Puede notarse, como en el caso del S&P 500, hay una mayor simetría y tridiagonalidad en las transiciones de los estados de mercado sectoriales, y que los estados 3 y 4 son los que tienen una probabilidad de ocupación más disímil. Por otra parte, también se observa que los centroides correspondientes tienen una alta similitud cuando se granulan las matrices de Pearson, a excepción de los centroides 3 y 4; véase la Figura 4.18. Éstos centroides parecen estar intercambiados, cuando se pasa de la granulación de los centroides de las correlaciones de Pearson (Figura 4.18(c)), a los centroides de las matrices granuladas (Figura 4.18(b)). El aparente intercambio de los centroides 3 y 4 de las matrices completas, y de las matrices sectoriales, también se manifiesta en las proyecciones MDS de las matrices locales en

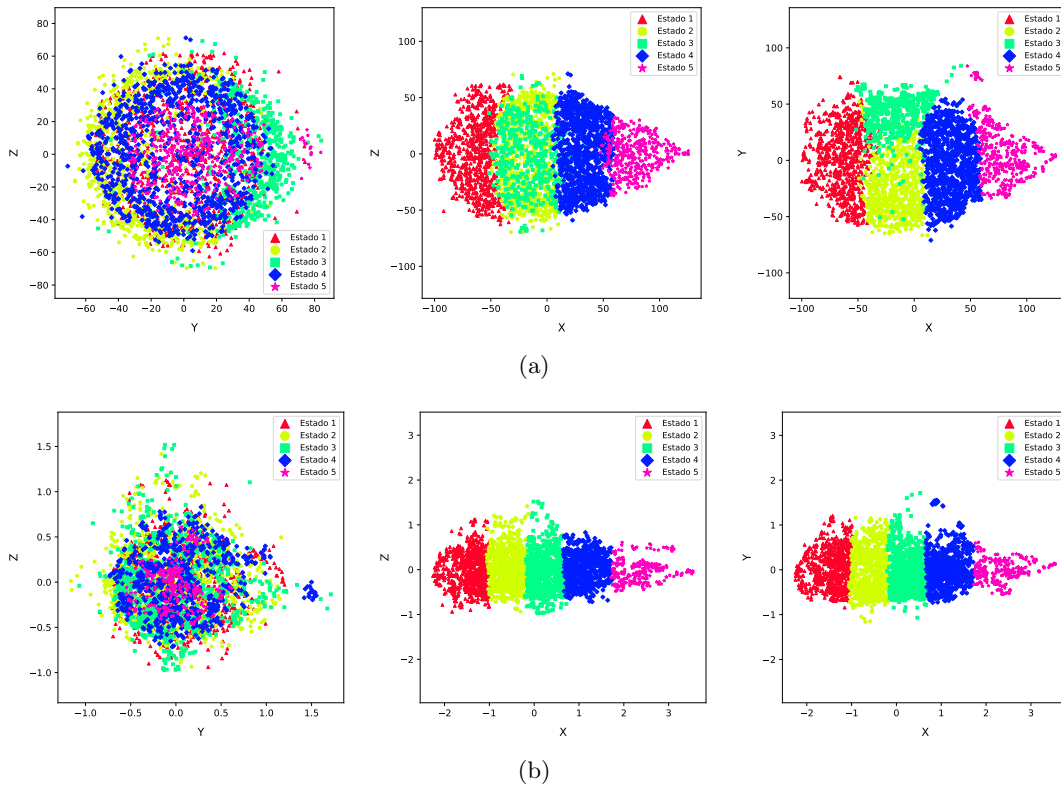


Figura 4.14: Representación tridimensional de las matrices de correlación de Pearson (figura 4.14(a)) y sectoriales de Guhr (figura 4.14(b)) del S&P 500, tras aplicar el escalamiento multidimensional.

el tiempo. En la Figura 4.20 se puede observar, a diferencia del análisis en el S&P 500, que la configuración geométrica es parecida, salvo por una reflexión en el plano YZ. Igualmente al caso del S&P 500, notamos que el eje de mayor varianza de la proyección MDS está fuertemente correlacionado con el mayor eigenvalor, lo que refuerza la evidencia de que dicho eje describe bien el estado del mercado (es decir:  $\lambda_{\text{máx}}$ ). Véase la Figura 4.21.

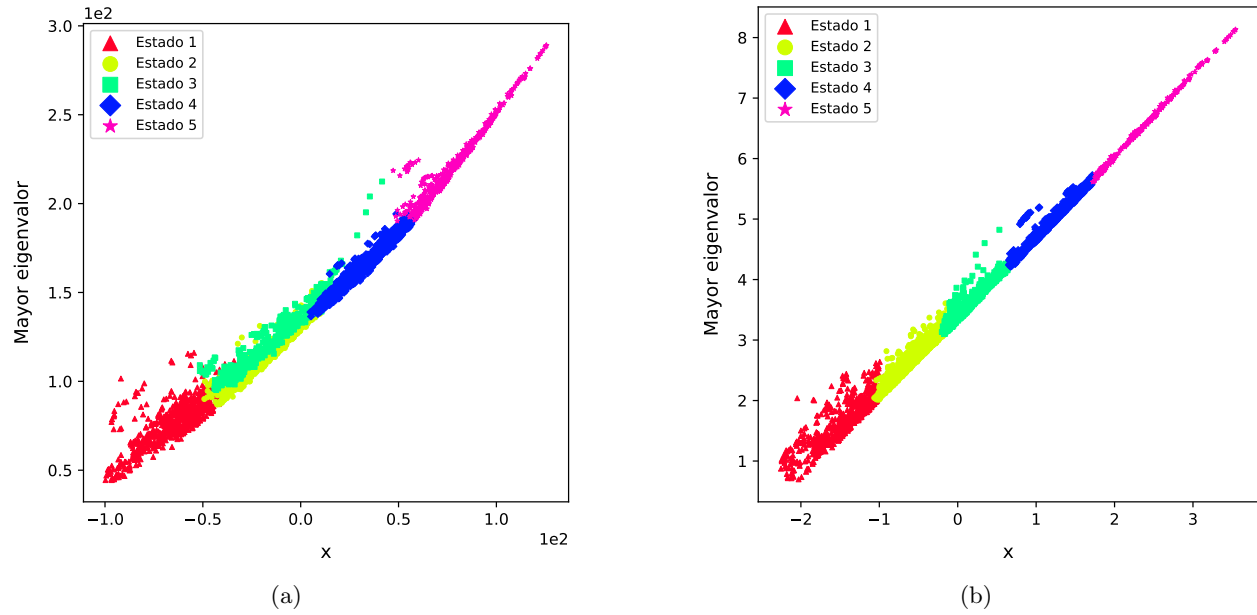


Figura 4.15: Diagramas de dispersión de las componentes  $x$  en las proyecciones tridimensionales, contra el mayor eigenvalor, para las matrices de correlación de Pearson (4.15(a)) y sectoriales de Guhr (4.15(b)) en el S&P 500. En cada caso, el color y forma de los puntos representa el estado ocupado.

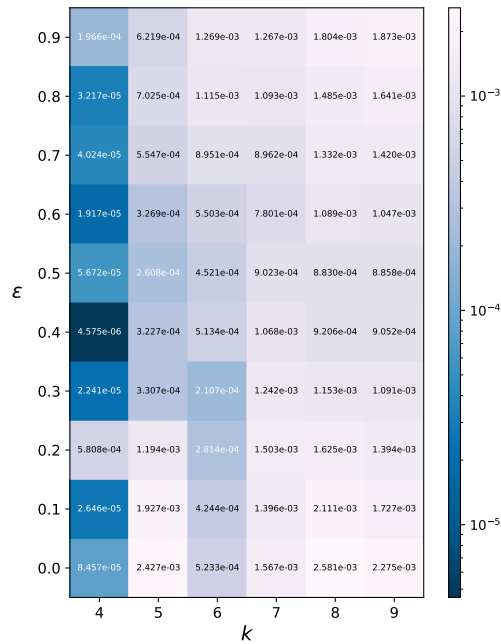
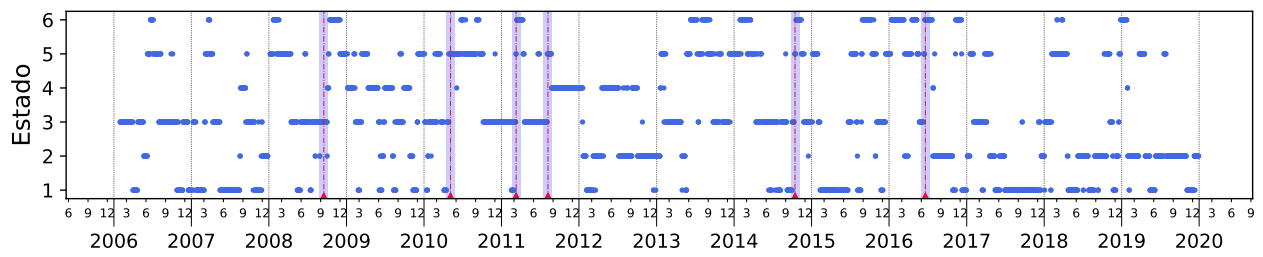
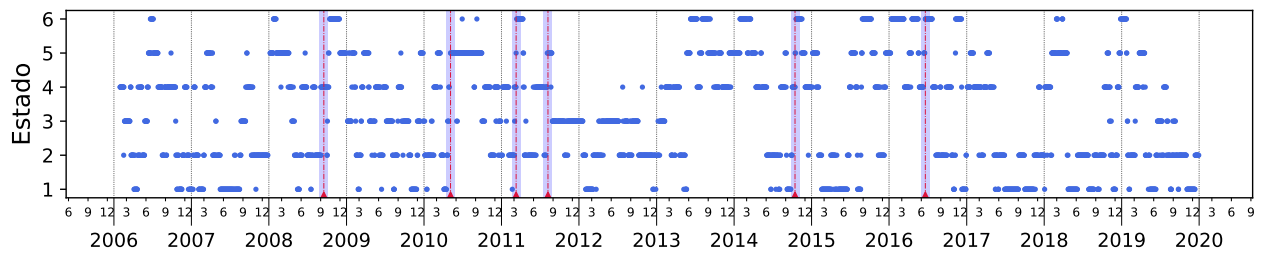


Figura 4.16: Gráfica de la desviación estándar de la distancia intra-cluster  $\sigma_{d_{intra}}$  (fórmula (4.2)) como función del número de clusters  $k$  y el exponente del *power map* (relación (2.16)), a partir de los datos del Nikkei 225. Considerando  $k > 4$ , como en [41], el mínimo de  $\sigma_{d_{intra}}$  está en  $(k, \epsilon) = (6, 0.3)$ .



(a)



(b)

Figura 4.17: Figura 4.17(a): evolución de los estados de mercado en el Nikkei 225, cuando se utilizan las matrices de correlación de Pearson en épocas sucesivas de  $q = 20$  días. Figura 4.18(b): evolución de los estados de mercado en el Nikkei 225, cuando se utilizan las matrices sectoriales de Guhr en las mismas condiciones.

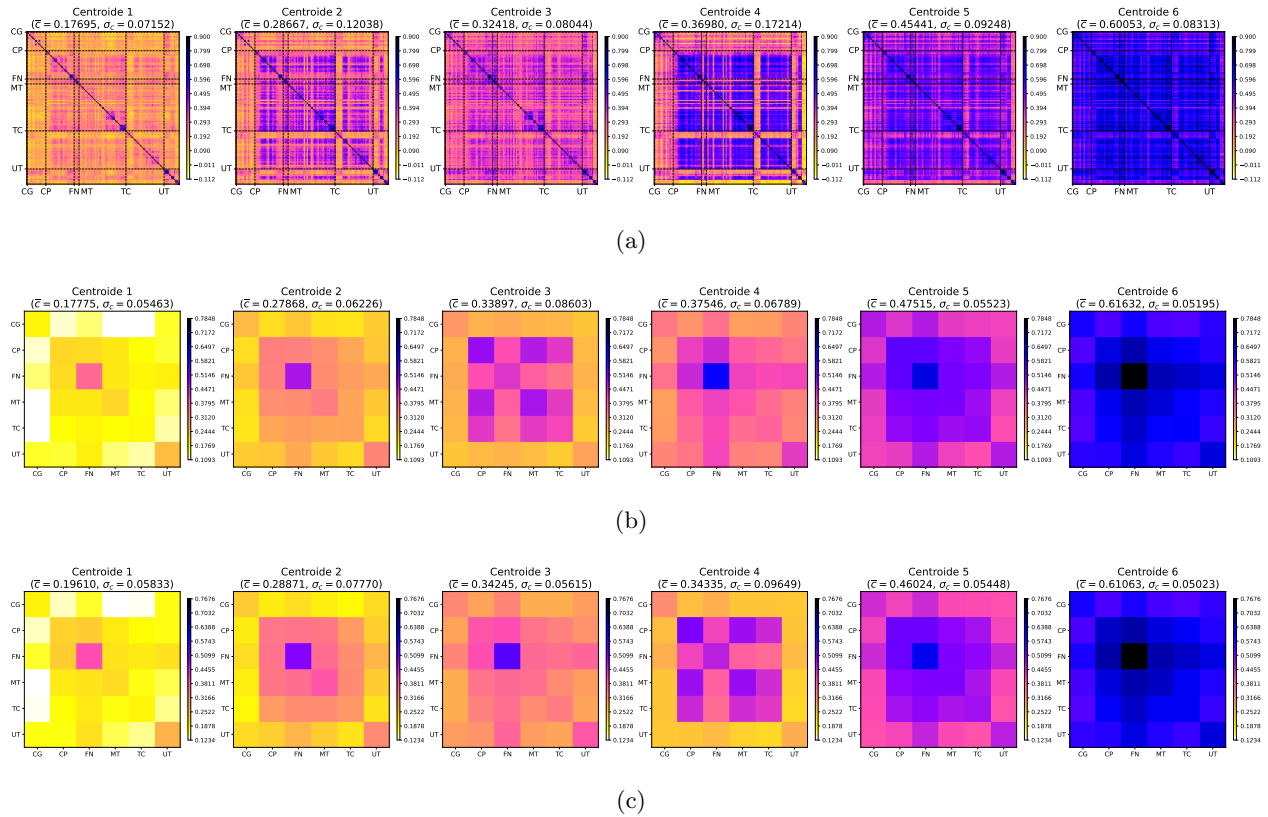


Figura 4.18: (Nikkei 225) Figura 4.18(a): Centroides (o centros geométricos) en el espacio  $N(N - 1)/2$ -dimensional de correlaciones de Pearson, correspondientes a los estados (cúmulos) ocupados en la figura 4.17(a). Figura 4.18(b): Centroides en el espacio  $M(M + 1)/2$ -dimensional de correlaciones sectoriales de Guhr, correspondientes a los estados ocupados en la figura 4.17(b). Figura 4.18(c): Matrices sectoriales de Guhr calculadas a partir de los centroides de la figura 4.18(a). Encima de cada matriz se encuentran el promedio y desviación estándar de cada centroide, cuando se utilizan solamente las componentes de la matriz triangular superior (excluyendo la diagonal principal en el caso de las matrices de Pearson, e incluyéndola en el caso de las matrices sectoriales de Guhr). Pensando en las matrices de las figuras 4.18(b) y 4.18(c) como vectores, tomando solamente las partes triangulares superiores, el parecido de los centroides de cada estado puede medirse a través de la correlación; resulta 0.9920, 0.9777, 0.5558, 0.5649, 0.9969, 0.9945, respectivamente.



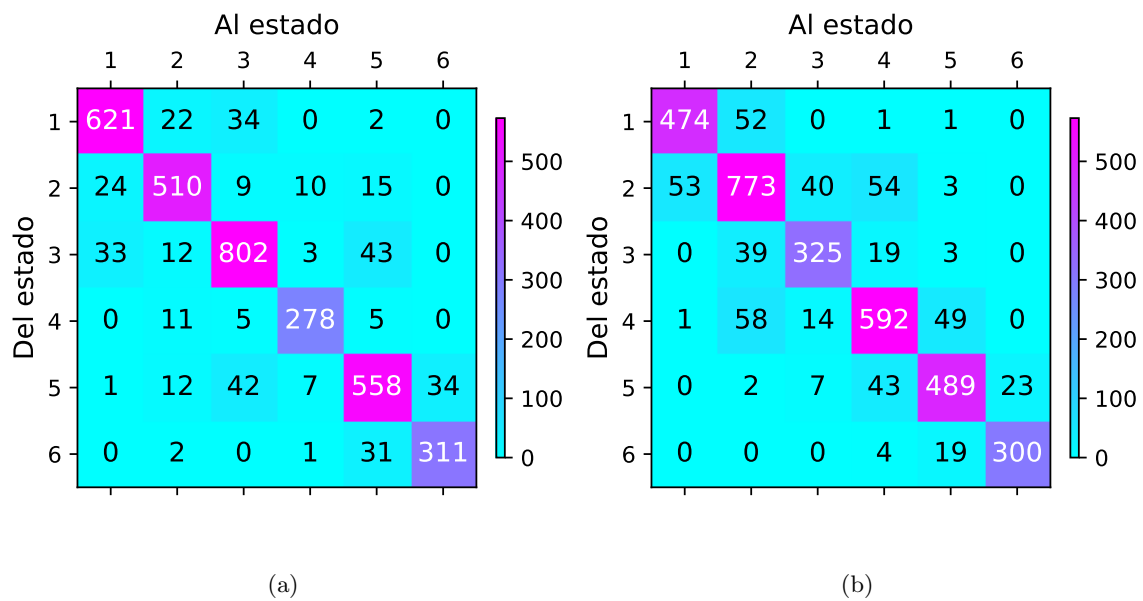


Figura 4.19: (Nikkei 225) Matrices de transición (o transferencia) de los estados mostrados en la figura 4.17. A la izquierda (4.19(a)): transiciones de los estados de mercado para las matrices de Pearson; a la derecha (4.19(b)): transiciones de los estados de mercado para las matrices sectoriales de Guhr. Las distribuciones estacionarias de probabilidad son: 4.19(a) (0.198, 0.168, 0.257, 0.087, 0.19, 0.1) y 4.19(b) (0.155, 0.27, 0.112, 0.206, 0.163, 0.093).

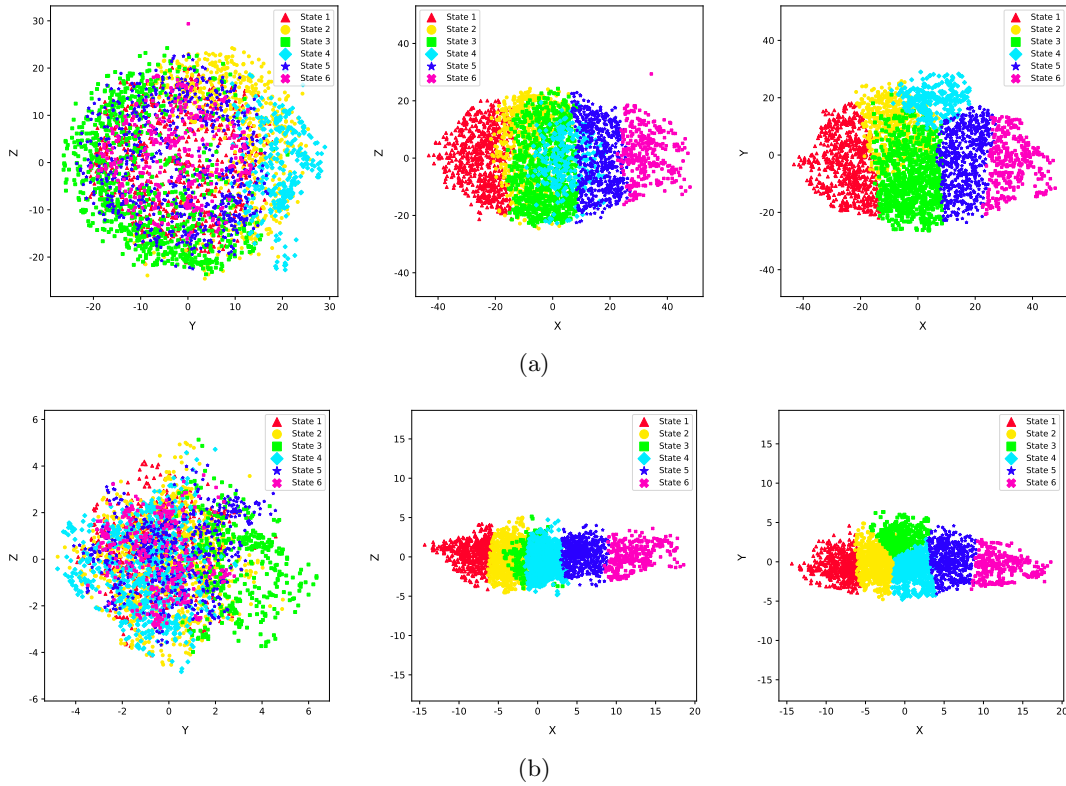


Figura 4.20: Representación tridimensional de las matrices de correlación de Pearson (figura 4.20(a)) y sectoriales de Guhr (figura 4.20(b)) en el Nikkei 225, tras aplicar el escalamiento multidimensional.

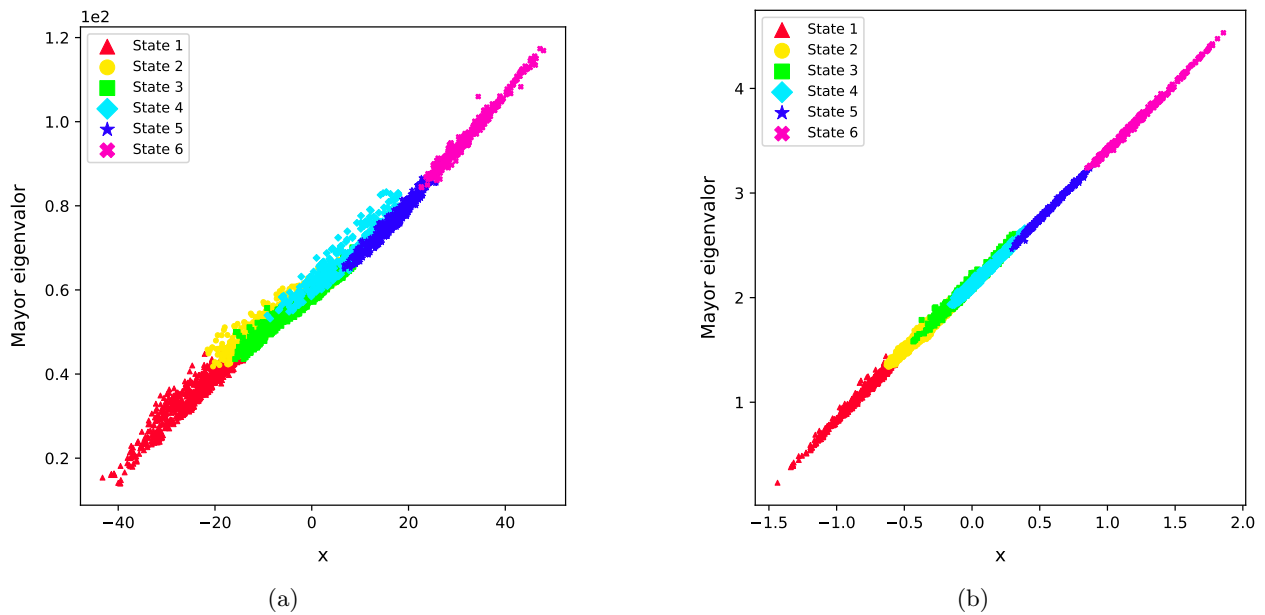


Figura 4.21: Diagramas de dispersión de las componentes  $x$  en las proyecciones tridimensionales, contra el mayor eigenvalor, para las matrices de correlación de Pearson (4.21(a)) y sectoriales de Guhr (4.21(b)) en el Nikkei 225. En cada caso, el color y forma de los puntos representa el estado ocupado.



## Capítulo 5

# Conclusión

En este trabajo estudiamos mercados financieros mediante las series de tiempo de rendimientos logarítmicos diarios de precios al cierre ajustados. Configuramos *ensembles* de matrices de correlación locales en el tiempo y estudiamos sus propiedades: visitación recurrente de regiones o *clusters*, directamente asociados a patrones de correlación típicos, así como la frecuencia de sus transiciones. Observamos que, en el caso de las proyecciones de Guhr y las correlaciones relativas, las matrices de similitud de las matrices de correlación locales en el tiempo, son parecidas; lo mismo ocurre con las matrices de correlación de horizonte completo, y en los centroides de la acumulación de las matrices locales, donde se pueden observar de manera diferenciada los patrones intra e intersectoriales.

En el caso de los datos del S&P 500 (para  $N = 322$  compañías, que abarcan el lapso de 2006 a 2023), encontramos un estado de mercado atípico, que asociamos con la época álgida del COVID (a inicios de 2020). Éste estado es más claramente visible cuando se filtra la componente del mercado con las correlaciones relativas, o con las proyecciones de Guhr. Además, para este mismo tipo de correlaciones *filtradas*, las estadísticas  $\lambda_{\text{máx}}(t)$  y  $\bar{c}(t)$  dejan de estar fuertemente correlacionadas, posiblemente porque las anticorrelaciones son mucho más frecuentes. A su vez, observamos una fuerte correlación entre las estadísticas  $\lambda_{\text{máx}}(t)$  y  $\sigma_c(t)$ , lo que sugiere una noción de *estado sectorial* como extensión del *estado del mercado* (dado por  $\lambda_{\text{máx}}$ , o  $\bar{c}(t)$ , en las correlaciones de Pearson). Esto puede observarse en la volatilidad de los elementos de matriz de las correlaciones filtradas, en las figuras 3.5(b) y 3.29 (especialmente durante el período COVID).

También estudiamos la medida en que las regiones recurrentes, en el espacio  $N(N-1)/2$ -dimensional que contiene las trayectorias temporales de las matrices de correlación locales, son todavía observables tras la granulación de los bloques sectoriales (es decir: las matrices sectoriales de Guhr). Los resultados son satisfactorios, en cuanto a que la dinámica de los estados (véase, por ejemplo, la figura 4.11), y de sus transiciones mutuas (v.gr. figura 4.13), es similar.

A través de la variación de los parámetros  $(k, \epsilon)$  (número de cúmulos y exponente del *power map* (2.16)), pudimos determinar valores óptimos en los mercados japonés y estadounidense, para los que se logra una aglomeración de las matrices de correlación locales en el tiempo con una configuración similar para las matrices de correlación de Pearson y sectoriales de Guhr. (En este escrito solo mostramos las figuras del mercado japonés, pero los resultados de ambos mercados se reportan en [103]). Así, demostramos una fuerte ventaja del análisis sectorial: que el número de

variables se reduce en órdenes de magnitud en comparación al análisis de las correlaciones completas.

## 5.1. Perspectivas

A continuación, listamos una serie de posibles líneas de investigación ulterior:

- Todavía hay que encontrar o aclarar el papel de la correlación relativa. El efecto de reducir las correlaciones entre fluctuaciones de precios con respecto a las fluctuaciones de los precios del índice, es similar al método de proyecciones de Guhr en cuanto a la conglomeración de estados de mercado. Sin embargo, hay otras maneras de aproximar una matriz (en este caso, de correlación) *lo mejor posible* con otra de **de menor rango**. Además, la simetría alrededor del 0 de las correlaciones *filtradas* (es decir, proyecciones de Guhr y correlaciones relativas) tiene implicaciones sobre el uso del *power map* (ecuación (2.16)), pues en el caso de las correlaciones de Pearson son los valores negativos los que sufren un mayor cambio bajo dicha función.
- El método de proyección de Guhr se puede iterar, y el espectro resultante en cada paso puede contener o no información relevante.
- La distribución de las correlaciones por pares podría estudiarse **con un modelo de mezcla**.
- Los distintos estados de mercado se pueden relacionar con un régimen cuasi-estacionario, y los patrones de las correlaciones resultantes reflejan una estructura macro y mesoscópica que recuerda a las particiones de un conjunto. Éstas particiones conforman un orden parcial con respecto a la contención, y dependiendo de una granularidad dada (entendida como el número de partes y la homogeneidad de sus tamaños), habrá más o menos microestados. Los patrones de correlación podrían asociarse con una o varias particiones de las  $N$  series de tiempo en una época  $t$ , indicando una región particular dentro del diagrama de Hasse del orden parcial. Quizás esto dé luz sobre riesgo sistémico y dinámica de redes [104] en mercados financieros.
- Tomando en cuenta que la matriz de correlación local en el tiempo, habiendo fijado un tamaño de época, representa el estado del mercado con una mayor resolución que su eigenvalor máximo, tiene sentido pensar en un modelo dinámico lineal como el filtro de Kalman, para predecir y actualizar [78, sec. 1] el estado del mercado en términos de la covarianza o la correlación.

### 5.1.1. Modelos de correlación con grafos

En un sistema complejo, la interacción de sus  $N$  partes constituyentes puede modelarse mediante una red (o grafo) del tipo  $G := (V, E)$ , con un conjunto de vértices (o nodos)  $V$  (tal que  $|V| = N$ ) representante de las partes en interacción, y otro conjunto de aristas  $E \subseteq V \times V$  representante de dichas interacciones<sup>1</sup>. Los mercados financieros son sistemas complejos que

<sup>1</sup>Que, por simplicidad, asumiremos como *interacciones binarias (o por pares) simétricas*:  $(i, j) \equiv (j, i)$ .

pueden ser modelados por un grafo cuyos nodos son compañías o sectores industriales. Sin embargo, hemos de diferenciar esta consideración de *los modelos de agentes* o ABM<sup>2</sup>, donde el modelado del comportamiento individual de las partes (como inversionistas, sus empleados o sus computadoras), denominadas *agentes*, permite obtener patrones emergentes o colectivos [105].

Si un sistema complejo está sujeto a influencias aleatorias, o tiene una dinámica determinista pero extremadamente complicada, el grafo que modela sus interacciones puede dotarse de una estructura estocástica. Específicamente, esto se logra mediante un espacio de probabilidad:

$$\mathcal{G}_N := (\Omega, \mathcal{F}, \mathbb{P}) \quad (5.1)$$

(donde  $\Omega$  es el espacio muestral que modela las propiedades de  $V$ , y  $\mathcal{F} \subseteq \mathcal{P}(\Omega)$  es un álgebra representativa de los estados o eventos del sistema), o a través de la especificación de una *densidad*  $\phi : \Omega \rightarrow \mathbb{R}^+$ :

$$\forall E \in \mathcal{F} : \quad \mathbb{P}(E) = \sum_{e \in E} \phi(e)$$

Las matrices de correlación se pueden pensar como una representación, quizás incompleta, de las interacciones instantáneas o *locales* entre las compañías que cotizan en una bolsa de valores, quienes tienen entre sí conexiones comerciales. Podemos formular esto con más rigor: las compañías cuyas series de tiempo financieras estudiamos, conforman los nodos de un grafo, y la correlación mutua se representa por el peso de una arista dada.

Como modelo inicial, se tiene la gráfica de Erdős-Rényi con  $N$  nodos. En ésta gráfica, las  $\binom{N}{2} = \frac{N(N-1)}{2}$  posibles aristas entre los  $N$  nodos se escogen de manera aleatoria. Específicamente: si la matriz de adyacencia<sup>3</sup> es  $\mathbf{A} = (a_{ij})$ , entonces la conexión entre dos nodos cualesquiera es una variable aleatoria  $a_{ij} = X \sim \text{Bernoulli}(\alpha)$ , con soporte  $\{0, 1\}$  y probabilidad de éxito  $\alpha \in [0, 1]$ . La conectividad  $C$  de una gráfica, definida en general como la cantidad promedio de aristas incidentes en un vértice, resulta en este modelo igual a  $C = \langle Y \rangle = (N-1)\alpha$ , donde  $Y \sim \text{Bin}(N-1, \alpha)$  suma las posibles aristas incidentes en un nodo arbitrario, provenientes de los  $N-1$  vértices restantes<sup>4</sup>. A su vez, esta conectividad  $C$  se puede asociar con la intensidad de acoplamiento o interacción de un modelo tipo Ising en la gráfica de Erdős-Rényi:

$$C = (N-1)\alpha = \left\langle \sum_{i \neq j} a_{ij} \right\rangle = \sum_{i \neq j} \langle a_{ij} \rangle = (N-1)\mathbb{P}(a_{ij} = 1), \quad \forall j \in V \quad (5.2)$$

Entonces, para  $N$  espines  $V = \{s_i = \pm 1\}_{i=1}^N$  colocados en esta gráfica, se obtiene un Hamiltoniano de interacción:

$$\mathcal{H} := - \sum_{(i,j) \in E} J_{ij} s_i s_j - \sum_{i \in V} h_i s_i. \quad (5.3)$$

donde  $\mathbf{J} = (J_{ij})$  es proporcional a  $\mathbf{A}$  (cuyas entradas cumplen  $\mathbb{P}(a_{ij} = 1) = C/(N-1)$ ) y  $\{h_i : i \in V\}$  son números reales fijos (representables en un vector  $\vec{h} = (h_1, \dots, h_N)$  si  $V$  está ordenado). En este caso, la densidad que modela un estado  $\vec{\sigma} := (\sigma_1, \dots, \sigma_N) \in \{-1, 1\}^N =: \Omega$  en equilibrio del sistema está dada por la distribución de Boltzmann-Gibbs:

$$\phi(\sigma_1, \dots, \sigma_N) := \mathbb{P}[s_1 = \sigma_1, \dots, s_N = \sigma_N] = \frac{1}{Z(\beta, \mathbf{J})} \exp(-\beta \mathcal{H}) \quad (5.4)$$

<sup>2</sup>Agent Based Models.

<sup>3</sup>Definida como  $a_{ij} = 1$  si  $\{i, j\} \in E$  y  $a_{ij} = 0$  de otro modo.

<sup>4</sup>En este trabajo, consideraremos que no hay aristas conectando vértices consigo mismos.

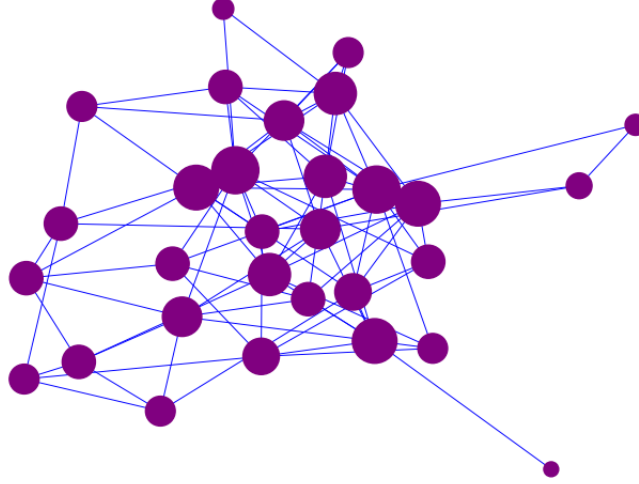


Figura 5.1: Realización de una gráfica de Erdős-Rényi, con parámetros  $N = 30$ ,  $\alpha = 1/5$ . En este caso, resulta  $C = \frac{1}{N(N-1)} \sum_{i \neq j} a_{ij} = 5.7\bar{3} \approx (N-1)\alpha$ .

donde  $\mathbf{J} = (J_{ij}) \propto \mathbf{A}$ ;  $\beta := 1/k_B T$  es el factor de Boltzmann; y  $Z = Z(\beta, \mathbf{J}, \vec{h})$  es la función de partición que garantiza normalidad.

(En este ejemplo, cabe aclarar que  $\Omega$  es el conjunto de las  $\binom{N}{2}$  posibles aristas  $(i, j)$ ; que el álgebra de estados  $\mathcal{F}$  del sistema corresponde a todos los subconjuntos de  $\Omega$ :  $\mathcal{F} := \mathcal{P}(\{1, -1\}^N)$ , y que su densidad  $\phi$  queda definida tanto por el Hamiltoniano de Ising como por las interacciones mutuas (aristas existentes), descritas a su vez por una realización particular de un subconjunto  $E \subseteq \Omega$  de aristas (o de la matriz de adyacencia  $\mathbf{A}$ ) con probabilidad  $\mathbb{P}(E) = \alpha^{|E|} (1 - \alpha)^{\binom{N}{2} - |E|}$  [106, cap. 2]).

Por otro lado, la maximización de la entropía:

$$S(\vec{\sigma}) = -k_B \sum_{\vec{\sigma} \in \Omega} \phi(\vec{\sigma}) \log(\phi(\vec{\sigma})) \quad (5.5)$$

sujeta a las condiciones:

$$\sum_{\vec{\sigma} \in \Omega} \phi(\vec{\sigma}) = 1 \quad (5.6)$$

$$q_i := \langle s_i \rangle = \sum_{\vec{\sigma} \in \Omega} s_i \phi(\vec{\sigma}) \quad (5.7)$$

$$q_{ij} := \langle s_{ij} \rangle := \sum_{\vec{\sigma} \in \Omega} s_i s_j \phi(\vec{\sigma}) \quad (5.8)$$

da precisamente como resultado la densidad de Boltzmann-Gibbs (5.4) con el Hamiltoniano (5.3), donde los multiplicadores de Lagrange  $J_{ij}$  y  $h_i$  representan las magnitudes de interacción y *preferencias*, variables conjugadas a su vez a las orientaciones y correlaciones promedio [25, secs. 2 y 3]:

$$\frac{\partial S(\vec{\sigma})}{\partial \beta q_i} = -h_i, \quad \frac{\partial S(\vec{\sigma})}{\partial \beta q_{ij}} = -J_{ij} \quad (5.9)$$

Este tipo de modelos es relevante si, por ejemplo, las variables binarias  $\{s_i\}_{i=1}^N$  representan una *instantánea* de una métrica financiera a la alza o baja (v.gr. precio de cierre de una acción mayor o menor al de apertura, respectivamente). El cálculo de la función de partición  $Z = Z(\beta, \mathbf{J}, \vec{h})$  permitiría deducir la entropía en (5.5) (y por lo tanto, deducir  $q_i$  y  $q_{ij}$ ) mediante  $S = \frac{\partial}{\partial T} (k_B T \log Z) = -\frac{\partial F}{\partial T}$ , donde  $F := -k_B T \log Z$  es la energía libre de Helmholtz.

Este cálculo conlleva varias dificultades. Primero, el hecho de que hay una  $F$  para cada realización de  $\mathbf{J}$  (o equivalentemente, de  $E$ ): tomando un *promedio configuracional*  $[\dots]$ [107, cap. 2] sobre las  $2^{\binom{N}{2}}$  posibles gráficas  $\mathcal{G}_N$  de Erdős-Rényi:

$$[F] = -k_B T \sum_{E \subseteq V \times V} \log Z(\beta, \mathbf{J}(E), \vec{h}) \mathbb{P}(E) = -k_B T \sum_{E \subseteq V \times V} \log Z(\beta, \mathbf{J}(E), \vec{h}) \alpha^{|E|} (1 - \alpha)^{\binom{N}{2} - |E|} \quad (5.10)$$

todavía faltaría saber la energía libre  $F$  para una configuración particular. El *método de réplicas* [108, cap. 1, sec. 3] asume que para un *ensemble* de  $n \in \mathbb{N}$  realizaciones o *réplicas* del sistema, la función de partición correspondiente  $\mathcal{Z} \equiv Z^n$  puede continuarse analíticamente a  $n \in (0, 1)$  y, en consecuencia, dar lugar a:

$$[\log Z] = \lim_{n \rightarrow 0^+} \frac{[Z^n] - 1}{n} = \lim_{n \rightarrow 0^+} \frac{\log [Z^n]}{n} \quad (5.11)$$

Los promedios configuracionales  $[Z^n]$  son más tratables[107, cap. 2, sec. 1.2].

Como modelo más refinado, podemos pensar en una densidad condicionada  $P(W|C) \sim \exp(-\frac{T}{2} \text{tr} WW^T C^{-1})$  [65, 109], tal que:

$$\langle WW^T \rangle = C \quad (5.12)$$

### 5.1.2. Técnica de ensembles

Para una matriz de  $N$  series de tiempo sobre una época (horizonte temporal) de tamaño  $T$ :  $X = (x_i(j)) \in \mathbb{R}^{N \times T}$ , la matriz de covarianza de las  $N$  series en esa época:

$$C = \frac{1}{T} X X^t = (c_{\iota\kappa})$$

es singular si  $T < N$ , en cuyo caso pueden elegirse  $m$  ( $1 < m < N$ ) de ellas, de tal modo que  $m \leq T$  induzca matrices de covarianza no singulares.

Para un conjunto con  $N$  elementos hay  $\binom{N}{m} = \frac{N!}{m!(N-m)!} \approx 2^{NH_2(m/N)}$  posibles agrupaciones (no ordenadas, sin reemplazo) de tamaño  $m$ , donde  $H_2(x) := -x \log_2(x) - (1-x) \log_2(1-x)$  es la función de entropía binaria. De todas estas posibles agrupaciones (*selecciones*), se toman  $k$  de ellas. Por ejemplo, si  $N = 100$ ,  $m = 10$ , entonces hay  $\binom{100}{10} = 17310309456440$  posibles agrupaciones de tamaño  $m$ . Un *ensemble* de selección aleatoria



(NSRSE <sup>5</sup> [109]) de 5 miembros sería:

$$\begin{aligned} & \{ \{3, 31, 41, 49, 58, 71, 81, 84, 92, 99\}, \\ & \{4, 15, 18, 29, 31, 45, 65, 79, 85, 91\}, \\ & \{2, 10, 26, 32, 39, 55, 57, 77, 80, 94\}, \\ & \{0, 1, 8, 55, 62, 77, 78, 86, 94, 95\}, \\ & \{0, 18, 29, 30, 36, 44, 54, 80, 85, 97\} \end{aligned}$$

Así, fijando una selección (digamos,  $S_i$ ) de tamaño  $m$ , tomando de la matriz de datos  $m$  renglones al azar:  $\{S_i(1), \dots, S_i(m)\}$ :

$$X(t; m, i) = \begin{pmatrix} x_{S_i(1)}(1) & \cdots & x_{S_i(1)}(T) \\ \vdots & \ddots & \vdots \\ x_{S_i(m)}(1) & \cdots & x_{S_i(m)}(T) \end{pmatrix} \quad (i = 1, \dots, k; 1 \ll k \ll \binom{N}{m})$$

Para obtener un análisis como función del tiempo, procedemos como con la relación (2.2.1) fijando de antemano:

- Una matriz de datos:  $X = (x_{ij}) \in \mathbb{R}^{N \times T}$ .
- Los parámetros:
  - Tamaño de época (ventana)  $q$ .
  - Tamaño de selección  $m$  (tal que  $m + 1 \leq q$  para obtener un *ensemble* de matrices de correlación no singulares).
  - Miembros del *ensemble*  $k$  (típicamente del orden de  $k = 10^3$ , de acuerdo con [109]).

La *técnica de ensembles* permite hacer estadística sobre el espectro de las matrices de correlación, que ya no presentarían el problema de la singularidad por la elección de épocas cortas. En ese sentido, tendría sentido examinar, por ejemplo, la *varianza de número*:

$$\Sigma^2(l) := \left\langle [n(\xi, l) - l]^2 \right\rangle_{\xi} \quad (5.13)$$

Además, el examen minucioso de dichas estadísticas podría dar más luz sobre el problema de la identificación de los parámetros del estado del mercado, más allá del máximo eigenvalor.

## Métodos de reducción dimensional

Los efectos de las proyecciones de Guhr, así como de las correlaciones relativas, sobre el espaciado entre las matrices de correlación locales, parece tener consecuencias sobre la calidad del MDS. Las figuras 5.2 y 5.3 muestran cómo las distancias de la proyección a tres dimensiones de las matrices locales de correlación relativa al índice, sufren una distorsión importante. Esto implica que, entre otras, la figura 3.24(b) no es fiel a la configuración espacial relativa de los cúmulos en el espacio  $N(N-1)/2$ -dimensional. Probablemente la geometría de las correlaciones proyectadas de Guhr, y de las correlaciones relativas, sea más intrincada.

<sup>5</sup>Non-Singular Random Selection Ensemble.

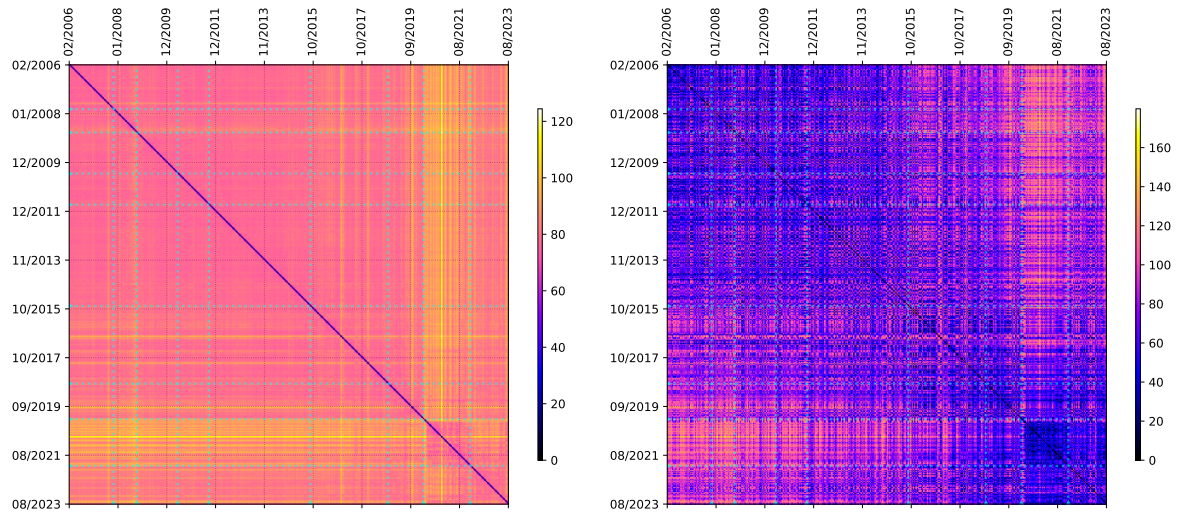


Figura 5.2: Matrices de similitud de las secuencias de matrices de correlación locales relativas al índice, antes (izquierda) y después de la reducción dimensional (derecha). Se consideran épocas sucesivas de  $q = 20$  días bursátiles, con 19 días de traslape. La métrica considerada es  $p = 2$ , y en cada subgráfica el código de color se normaliza de acuerdo con el rango de datos correspondiente.

De esta manera, las regiones visitadas por las correlaciones proyectadas o relativas podrían tener una distribución fuertemente no-lineal, que no pueda ser capturada por el k-means. En este sentido, algunas alternativas para la agrupación o *clustering* que podrían funcionar mejor son el *tSNE*<sup>6</sup> o el *HDBSCAN*<sup>7</sup>, por mencionar algunas.

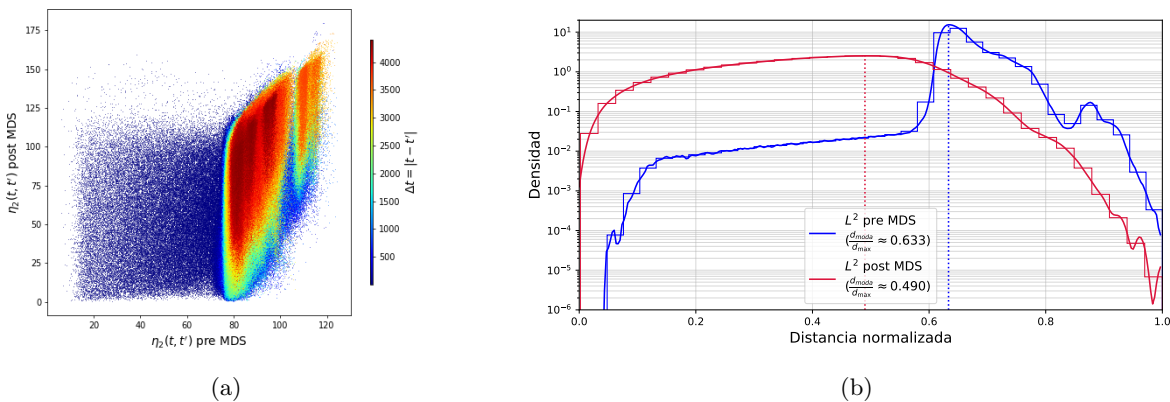


Figura 5.3: Comparación de las distancias euclidianas entre las matrices locales de correlación relativas al índice, y sus proyecciones MDS. En las distribuciones de la figura 5.3(a), las distancias máximas entre las matrices del espacio completo, o entre sus proyecciones MDS, son 124.436 y 179.867, respectivamente, y éstas distancias sirven para normalizar las densidades. Las líneas suaves representan una *Kernel Density Estimation* con núcleos gaussianos y el método de Silverman, y las líneas escalonadas representan un histograma cuyos bins se calcularon con el método de Doane.

<sup>6</sup> *T-distributed Stochastic Neighbor Embedding*.

<sup>7</sup> *Hierarchical Density-Based Spatial Clustering of Applications with Noise*.

### 5.1.3. Comparación de mercados

Decidimos posponer los resultados sobre mercados europeos, delegando parte de su análisis a estudiantes de licenciatura y maestría de nuestro grupo de trabajo en el ICF. La extensión del trabajo tiene, en particular, la posibilidad de realizar el análisis de grano grueso en dos variantes: por país/región, o por sector. También son de mucho interés las particularidades de la bolsa de Shanghai.

# Apéndices



## Apéndice A

# Sobre los datos utilizados

La razones para escoger precios de cierre ajustado son:

1. Los datos intradía no son accesibles de forma gratuita, y suelen ser costosos. Además, las estadísticas de los precios de transacción se asemejan a las del movimiento browniano.
2. Los datos de apertura tienen mayor volatilidad que los precios de cierre ajustado, y no es claro el momento en que los precios máximo y mínimo se alcanzan para cada acción.
3. Los datos de cierre ajustado reflejan mejor las decisiones de los actores (inversionistas o grupos de inversión) de la bolsa.

### A.1. S&P 500

Para este trabajo, utilizamos dos conjuntos de datos de bolsas estadounidenses, correspondientes a un subconjunto de compañías listadas en el índice S&P 500:

- $N = 350, T = 3523$ , comprendiendo precios ajustados de cierre del 3 enero del 2006 al 31 de diciembre de 2019.
- $N = 322, T = 4431$ , comprendiendo precios ajustados de cierre del 3 de enero del 2006 al 10 de agosto de 2023.

Las compañías componentes del segundo grupo de datos son un subconjunto del primero, y corresponden a las acciones que no salieron del índice en el período extendido. La tabla [A.1](#) muestra las  $N = 350$  componentes en el intervalo de tiempo menor, con las 28 compañías excluidas del grupo de datos en el intervalo de tiempo mayor marcadas en negritas y precedidas por un trébol (♣).

Cabe mencionar que, entre finales de 2019 y mediados de 2023, la compañía *Everest Re Group Ltd.* (sector FN) cambió su nombre a *Realty Income Corp.*, de modo que su *ticker* pasó de **RE** a **O**.

<b>Etiqueta GICS</b>	<b>Sector</b>
CD	Consumer Discretionary
CS	Consumer Staples
EG	Energy
FN	Financials
HC	Health Care
ID	Industrials
IT	Information Technology
MT	Materials
TC	Telecommunication Services
UT	Utilities

Cuadro A.1: Sectores de las compañías que componen el índice S&amp;P 500.

<b>Evento</b>	<b>Fecha</b>
United States bear market of 2007–2009	11/Oct/2007
Financial crisis of 2007–2008	16/Sep/2008
European sovereign debt crisis	6/May/2010
August 2011 stock markets fall	1/Ago/2011
2015–16 Chinese stock market crash	18/Ago/2015
2018 cryptocurrency crash	20/Sep/2018
2020 stock market crash	24/Feb/2020
2022 stock market decline	3/Ene/2022

Cuadro A.2: Lista de eventos críticos en los mercados de valores estadounidenses, según [Wikipedia](#).

<b>Etiqueta GICS</b>	<b>Ticker</b>	<b>Compañía</b>
CD	AAP	Advance Auto Parts
CD	AMZN	Amazon.com Inc.
CD	AZO	AutoZone Inc
CD	BBY	Best Buy Co. Inc.
CD	BKNG	Booking Holdings Inc
CD	CCL	Carnival Corp.
CD	CMCSA	Comcast Corp.
CD	DHI	D. R. Horton
CD	♣ DISCA	Discovery Inc. Class A
CD	DISH	Dish Network
CD	DLTR	Dollar Tree
CD	EXPE	Expedia Inc.
CD	F	Ford Motor
CD	GPC	Genuine Parts

Cuadro A.3: Lista de las compañías estadounidenses, incluidas en el índice S&amp;P 500, continuamente cotizando en las bolsas Nasdaq y NYSE.

Etiqueta GICS	Ticker	Compañía
CD	GPS	Gap Inc.
CD	GRMN	Garmin Ltd.
CD	HAS	Hasbro Inc.
CD	HD	Home Depot
CD	HOG	Harley-Davidson
CD	IPG	Interpublic Group
CD	JWN	Nordstrom
CD	KMX	Carmax Inc
CD	KSS	Kohl's Corp.
CD	♣ <b>LB</b>	L Brands Inc.
CD	LEG	Leggett & Platt
CD	LEN	Lennar Corp.
CD	LKQ	LKQ Corporation
CD	LOW	Lowe's Cos.
CD	M	Macy's Inc.
CD	MAR	Marriott Int'l.
CD	MCD	McDonald's Corp.
CD	MGM	MGM Resorts International
CD	MHK	Mohawk Industries
CD	NKE	Nike
CD	NWL	Newell Brands
CD	OMC	Omnicom Group
CD	ORLY	O'Reilly Automotive
CD	PHM	Pulte Homes Inc.
CD	PVH	PVH Corp.
CD	RL	Polo Ralph Lauren Corp.
CD	ROST	Ross Stores
CD	SBUX	Starbucks Corp.
CD	SWK	Stanley Black & Decker
CD	TGT	Target Corp.
CD	♣ <b>TIF</b>	Tiffany & Co.
CD	TJX	TJX Companies Inc.
CD	TPR	Tapestry, Inc.
CD	UAA	Under Armour Class A
CD	VFC	V.F. Corp.
CD	WHR	Whirlpool Corp.
CD	WYNN	Wynn Resorts Ltd
CD	YUM	Yum! Brands Inc
CS	ADM	Archer-Daniels-Midland Co
CS	CAG	Conagra Brands
CS	CHD	Church & Dwight
CS	CL	Colgate-Palmolive
CS	CLX	The Clorox Company
CS	COST	Costco Wholesale Corp.
CS	CPB	Campbell Soup

Cuadro A.3: Lista de las compañías estadounidenses, incluidas en el índice S&P 500, continuamente cotizando en las bolsas Nasdaq y NYSE.



<b>Etiqueta GICS</b>	<b>Ticker</b>	<b>Compañía</b>
CS	CVS	CVS Health
CS	EL	Estee Lauder Cos.
CS	GIS	General Mills
CS	HRL	Hormel Foods Corp.
CS	HSY	The Hershey Company
CS	K	Kellogg Co.
CS	KMB	Kimberly-Clark
CS	KO	Coca-Cola Company (The)
CS	KR	Kroger Co.
CS	MDLZ	Mondelez International
CS	MKC	McCormick & Co.
CS	MNST	Monster Beverage
CS	MO	Altria Group Inc
CS	PEP	PepsiCo Inc.
CS	PG	Procter & Gamble
CS	SJM	JM Smucker
CS	STZ	Constellation Brands
CS	SYY	Sysco Corp.
CS	TAP	Molson Coors Brewing Company
CS	TSN	Tyson Foods
CS	WBA	Walgreens Boots Alliance
CS	WMT	Wal-Mart Stores
EG	APA	Apache Corporation
EG	♣ <b>COG</b>	Cabot Oil & Gas
EG	COP	ConocoPhillips
EG	CVX	Chevron Corp.
EG	DVN	Devon Energy Corp.
EG	EOG	EOG Resources
EG	FTI	TechnipFMC
EG	HAL	Halliburton Co.
EG	HES	Hess Corporation
EG	HP	Helmerich & Payne
EG	MRO	Marathon Oil Corp.
EG	♣ <b>NBL</b>	Noble Energy Inc
EG	NOV	National Oilwell Varco Inc.
EG	OKE	ONEOK
EG	PXD	Pioneer Natural Resources
EG	SLB	Schlumberger Ltd.
EG	VLO	Valero Energy
EG	WMB	Williams Cos.
EG	♣ <b>XEC</b>	Cimarex Energy
EG	XOM	Exxon Mobil Corp.
FN	AFL	AFLAC Inc
FN	AIG	American International Group, Inc.
FN	AIZ	Assurant Inc.

Cuadro A.3: Lista de las compañías estadounidenses, incluidas en el índice S&P 500, continuamente cotizando en las bolsas Nasdaq y NYSE.

<b>Etiqueta GICS</b>	<b>Ticker</b>	<b>Compañía</b>
FN	AJG	Arthur J. Gallagher & Co.
FN	AMG	Affiliated Managers Group Inc
FN	AMP	Ameriprise Financial
FN	AON	Aon plc
FN	AXP	American Express Co
FN	BAC	Bank of America Corp
FN	BEN	Franklin Resources
FN	BK	The Bank of New York Mellon Corp.
FN	BLK	BlackRock
FN	C	Citigroup Inc.
FN	CINF	Cincinnati Financial
FN	CMA	Comerica Inc.
FN	CME	CME Group Inc.
FN	♣ <b>ETFC</b>	E*Trade
FN	FITB	Fifth Third Bancorp
FN	GS	Goldman Sachs Group
FN	HBAN	Huntington Bancshares
FN	HIG	Hartford Financial Svc.Gp.
FN	HRB	Block H&R
FN	ICE	Intercontinental Exchange
FN	IVZ	Invesco Ltd.
FN	JPM	JPMorgan Chase & Co.
FN	KEY	KeyCorp
FN	L	Loews Corp.
FN	LNC	Lincoln National
FN	MCO	Moody's Corp
FN	MET	MetLife Inc.
FN	MMC	Marsh & McLennan
FN	MS	Morgan Stanley
FN	MTB	M&T Bank Corp.
FN	NDAQ	Nasdaq, Inc.
FN	NTRS	Northern Trust Corp.
FN	♣ <b>PBCT</b>	People's United Financial
FN	PFG	Principal Financial Group
FN	PGR	Progressive Corp.
FN	PNC	PNC Financial Services
FN	PRU	Prudential Financial
FN	RE	Everest Re Group Ltd.
FN	RF	Regions Financial Corp.
FN	RJF	Raymond James Financial Inc.
FN	SCHW	Charles Schwab Corporation
FN	♣ <b>SIVB</b>	SVB Financial
FN	SPGI	S&P Global, Inc.
FN	STT	State Street Corp.
FN	TROW	T. Rowe Price Group

Cuadro A.3: Lista de las compañías estadounidenses, incluidas en el índice S&P 500, continuamente cotizando en las bolsas Nasdaq y NYSE.

<b>Etiqueta GICS</b>	<b>Ticker</b>	<b>Compañía</b>
FN	TRV	The Travelers Companies Inc.
FN	UNM	Unum Group
FN	USB	U.S. Bancorp
FN	WFC	Wells Fargo
FN	♣ <b>WLTW</b>	Willis Towers Watson
FN	ZION	Zions Bancorp
HC	A	Agilent Technologies Inc
HC	ABC	AmerisourceBergen Corp
HC	ABT	Abbott Laboratories
HC	♣ <b>AGN</b>	Allergan, Plc
HC	ALGN	Align Technology
HC	♣ <b>ALXN</b>	Alexion Pharmaceuticals
HC	AMGN	Amgen Inc.
HC	♣ <b>ANTM</b>	Anthem Inc.
HC	BAX	Baxter International Inc.
HC	BDX	Becton Dickinson
HC	BIIB	Biogen Inc.
HC	BMJ	Bristol-Myers Squibb
HC	BSX	Boston Scientific
HC	♣ <b>CERN</b>	Cerner
HC	CI	CIGNA Corp.
HC	CNC	Centene Corporation
HC	COO	The Cooper Companies
HC	DGX	Quest Diagnostics
HC	DVA	DaVita Inc.
HC	EW	Edwards Lifesciences
HC	GILD	Gilead Sciences
HC	HOLX	Hologic
HC	HSIC	Henry Schein
HC	HUM	Humana Inc.
HC	IDXX	IDEXX Laboratories
HC	ILMN	Illumina Inc
HC	INCY	Incyte
HC	ISRG	Intuitive Surgical Inc.
HC	JNJ	Johnson & Johnson
HC	LH	Laboratory Corp. of America Holding
HC	LLY	Lilly (Eli) & Co.
HC	MDT	Medtronic plc
HC	MRK	Merck & Co.
HC	MTD	Mettler Toledo
HC	♣ <b>MYL</b>	Mylan N.V.
HC	PFE	Pfizer Inc.
HC	♣ <b>PKI</b>	PerkinElmer
HC	PRGO	Perrigo
HC	REGN	Regeneron

Cuadro A.3: Lista de las compañías estadounidenses, incluidas en el índice S&P 500, continuamente cotizando en las bolsas Nasdaq y NYSE.

<b>Etiqueta GICS</b>	<b>Ticker</b>	<b>Compañía</b>
HC	RMD	ResMed
HC	SYK	Stryker Corp.
HC	TMO	Thermo Fisher Scientific
HC	UHS	Universal Health Services, Inc.
HC	UNH	United Health Group Inc.
HC	VRTX	Vertex Pharmaceuticals Inc
HC	WAT	Waters Corporation
HC	XRAY	Dentsply Sirona
HC	ZBH	Zimmer Biomet Holdings
ID	AAL	American Airlines Group
ID	ALK	Alaska Air Group Inc
ID	AME	AMETEK Inc.
ID	AOS	A.O. Smith Corp
ID	♣ ARNC	Arconic Inc.
ID	BA	Boeing Company
ID	CAT	Caterpillar Inc.
ID	CHRW	C. H. Robinson Worldwide
ID	CMI	Cummins Inc.
ID	CSX	CSX Corp.
ID	CTAS	Cintas Corporation
ID	DE	Deere & Co.
ID	DOV	Dover Corp.
ID	EFX	Equifax Inc.
ID	EMR	Emerson Electric Company
ID	ETN	Eaton Corporation
ID	EXPD	Expeditors International
ID	FAST	Fastenal Co
ID	FDX	FedEx Corporation
ID	FLS	Flowserve Corporation
ID	GD	General Dynamics
ID	GE	General Electric
ID	GWV	Grainger (W.W.) Inc.
ID	♣ IR	Ingersoll-Rand PLC
ID	ITW	Illinois Tool Works
ID	JBHT	J. B. Hunt Transport Services
ID	JCI	Johnson Controls International
ID	♣ KSU	Kansas City Southern
ID	LMT	Lockheed Martin Corp.
ID	LUV	Southwest Airlines
ID	MAS	Masco Corp.
ID	NOC	Northrop Grumman Corp.
ID	NSC	Norfolk Southern Corp.
ID	PCAR	PACCAR Inc.
ID	PH	Parker-Hannifin
ID	PNR	Pentair Ltd.

Cuadro A.3: Lista de las compañías estadounidenses, incluidas en el índice S&P 500, continuamente cotizando en las bolsas Nasdaq y NYSE.

Etiqueta GICS	Ticker	Compañía
ID	PWR	Quanta Services Inc.
ID	RHI	Robert Half International
ID	ROK	Rockwell Automation Inc.
ID	ROP	Roper Technologies
ID	RSG	Republic Services Inc
ID	♣ RTN	Raytheon Co.
ID	TXT	Textron Inc.
ID	UNP	Union Pacific
ID	UPS	United Parcel Service
ID	URI	United Rentals, Inc.
ID	♣ UTX	United Technologies
ID	WM	Waste Management Inc.
IT	AAPL	Apple Inc.
IT	ACN	Accenture plc
IT	ADBE	Adobe Systems Inc
IT	ADI	Analog Devices, Inc.
IT	ADP	Automatic Data Processing
IT	♣ ADS	Alliance Data Systems
IT	ADSK	Autodesk Inc.
IT	AKAM	Akamai Technologies Inc
IT	AMAT	Applied Materials Inc.
IT	AMD	Advanced Micro Devices Inc
IT	ANSS	ANSYS
IT	APH	Amphenol Corp
IT	ATVI	Activision Blizzard
IT	CDNS	Cadence Design Systems
IT	CRM	Salesforce.com
IT	CSCO	Cisco Systems
IT	CTSH	Cognizant Technology Solutions
IT	♣ CTXS	Citrix Systems
IT	DXC	DXC Technology
IT	EA	Electronic Arts
IT	EBAY	eBay Inc.
IT	FFIV	F5 Networks
IT	FIS	Fidelity National Information Services
IT	♣ FISV	Fiserv Inc
IT	♣ FLIR	FLIR Systems
IT	GLW	Corning Inc.
IT	GOOG	Alphabet Inc Class C
IT	GOOGL	Alphabet Inc Class A
IT	GPN	Global Payments Inc.
IT	HPQ	HP Inc.
IT	IBM	International Business Machines
IT	INTC	Intel Corp.
IT	INTU	Intuit Inc.

Cuadro A.3: Lista de las compañías estadounidenses, incluidas en el índice S&P 500, continuamente cotizando en las bolsas Nasdaq y NYSE.

<b>Etiqueta GICS</b>	<b>Ticker</b>	<b>Compañía</b>
IT	IT	Gartner Inc
IT	JNPR	Juniper Networks
IT	KLAC	KLA-Tencor Corp.
IT	LRCX	Lam Research
IT	MCHP	Microchip Technology
IT	MSFT	Microsoft Corp.
IT	MSI	Motorola Solutions Inc.
IT	MU	Micron Technology
IT	NFLX	Netflix Inc.
IT	NTAP	NetApp
IT	NVDA	Nvidia Corporation
IT	ORCL	Oracle Corp.
IT	PAYX	Paychex Inc.
IT	QCOM	QUALCOMM Inc.
IT	SNPS	Synopsys Inc.
IT	STX	Seagate Technology
IT	SWKS	Skyworks Solutions
IT	TTWO	Take-Two Interactive
IT	TXN	Texas Instruments
IT	VRSN	Verisign Inc.
IT	WDC	Western Digital
IT	♣ XLNX	Xilinx Inc
MT	APD	Air Products & Chemicals Inc
MT	AVY	Avery Dennison Corp
MT	♣ BLL	Ball Corp
MT	CF	CF Industries Holdings Inc
MT	ECL	Ecolab Inc.
MT	FCX	Freeport-McMoRan Inc.
MT	FMC	FMC Corporation
MT	IFF	Intl Flavors & Fragrances
MT	IP	International Paper
MT	MOS	The Mosaic Company
MT	NEM	Newmont Mining Corporation
MT	NUE	Nucor Corp.
MT	PKG	Packaging Corporation of America
MT	PPG	PPG Industries
MT	SEE	Sealed Air
MT	SHW	Sherwin-Williams
MT	VMC	Vulcan Materials
TC	♣ CTL	CenturyLink Inc
TC	T	AT&T Inc.
TC	VZ	Verizon Communications
UT	AEE	Ameren Corp
UT	AEP	American Electric Power
UT	AES	AES Corp

Cuadro A.3: Lista de las compañías estadounidenses, incluidas en el índice S&P 500, continuamente cotizando en las bolsas Nasdaq y NYSE.

<b>Etiqueta GICS</b>	<b>Ticker</b>	<b>Compañía</b>
UT	CMS	CMS Energy
UT	CNP	CenterPoint Energy
UT	D	Dominion Energy
UT	DTE	DTE Energy Co.
UT	DUK	Duke Energy
UT	ED	Consolidated Edison
UT	EIX	Edison Int'l
UT	ES	Eversource Energy
UT	ETR	Entergy Corp.
UT	EXC	Exelon Corp.
UT	FE	FirstEnergy Corp
UT	LNT	Alliant Energy Corp
UT	NEE	NextEra Energy
UT	NI	NiSource Inc.
UT	NRG	NRG Energy
UT	PEG	Public Serv. Enterprise Inc.
UT	PNW	Pinnacle West Capital
UT	SO	Southern Co.
UT	SRE	Sempra Energy
UT	WEC	Wec Energy Group Inc
UT	XEL	Xcel Energy Inc

Cuadro A.3: Lista de las compañías estadounidenses, incluidas en el índice S&P 500, continuamente cotizando en las bolsas Nasdaq y NYSE.

## A.2. Nikkei 225

Siguiendo el mismo criterio de exclusión de datos que en la sección anterior, el grupo de datos del mercado japonés consiste en los precios de cierre diario de  $N = 156$  acciones en  $T = 3459$  días bursátiles, comprendiendo el intervalo del 4 de enero de 2006 al 30 de diciembre de 2019.

<b>Etiqueta</b>	<b>Ticker</b>	<b>Compañía</b>
CG	1332.T	NIPPON SUISAN KAISHA, LTD.
CG	2002.T	NISSHIN SEIFUN GROUP INC.
CG	2282.T	NH FOODS LTD.
CG	2502.T	ASAHI GROUP HOLDINGS, LTD.
CG	2503.T	KIRIN HOLDINGS CO., LTD.

Cuadro A.5: Lista de las compañías japonesas, incluidas en el índice Nikkei 225, continuamente cotizando en la bolsa de Tokio en el período considerado.

<b>Etiqueta</b>	<b>Ticker</b>	<b>Compañía</b>
CG	2531.T	TAKARA HOLDINGS INC.
CG	2801.T	KIKKOMAN CORP.
CG	2802.T	AJINOMOTO CO., INC.
CG	2871.T	NICHIREI CORP.
CG	3086.T	J.FRONT RETAILING CO., LTD.
CG	3099.T	ISETAN MITSUKOSHI HOLDINGS LTD.
CG	4704.T	TREND MICRO INC.
CG	8028.T	FAMILYMART UNY HOLDINGS CO., LTD.
CG	8233.T	TAKASHIMAYA CO., LTD.
CG	8267.T	AEON CO., LTD.
CG	9602.T	TOHO CO., LTD
CG	9681.T	TOKYO DOME CORP.
CG	9766.T	KONAMI HOLDINGS CORP.
CG	9983.T	FAST RETAILING CO., LTD.
CP	1721.T	COMSYS HOLDINGS CORP.
CP	1801.T	TAISEI CORP.
CP	1803.T	SHIMIZU CORP.
CP	1808.T	HASEKO CORP.
CP	1812.T	KAJIMA CORP.
CP	1928.T	SEKISUI HOUSE, LTD.
CP	1963.T	JGC CORP.
CP	5631.T	THE JAPAN STEEL WORKS, LTD.
CP	6113.T	AMADA HOLDINGS CO., LTD.
CP	6301.T	KOMATSU LTD.
CP	6302.T	SUMITOMO HEAVY IND., LTD.
CP	6305.T	HITACHI CONST. MACH. CO., LTD.
CP	6326.T	KUBOTA CORP.
CP	6361.T	EBARA CORP.
CP	6366.T	CHIYODA CORP.
CP	6367.T	DAIKIN INDUSTRIES, LTD.
CP	6471.T	NSK LTD.
CP	6472.T	NTN CORP.
CP	7003.T	MITSUI E&S HOLDINGS CO., LTD.
CP	7004.T	HITACHI ZOSEN CORP.
CP	7011.T	MITSUBISHI HEAVY IND., LTD.
CP	7012.T	KAWASAKI HEAVY IND., LTD.
CP	7013.T	IHI CORP.
CP	7912.T	DAI NIPPON PRINTING CO., LTD.
CP	7951.T	YAMAHA CORP.
CP	8801.T	MITSUI FUDOSAN CO., LTD.
CP	8802.T	MITSUBISHI ESTATE CO., LTD.
CP	8804.T	TOKYO TATEMONO CO., LTD.
CP	8830.T	SUMITOMO REALTY & DEVELOPMENT CO., LTD.
FN	8253.T	CREDIT SAISON CO., LTD.
FN	8331.T	THE CHIBA BANK, LTD.

Cuadro A.5: Lista de las compañías japonesas, incluidas en el índice Nikkei 225, continuamente cotizando en la bolsa de Tokio en el período considerado.



<b>Etiqueta</b>	<b>Ticker</b>	<b>Compañía</b>
FN	8355.T	THE SHIZUOKA BANK, LTD.
FN	8601.T	DAIWA SECURITIES GROUP INC.
FN	8604.T	NOMURA HOLDINGS, INC.
MT	2768.T	SOJITZ CORP.
MT	3101.T	TOYOBO CO., LTD.
MT	3103.T	UNITIKA, LTD.
MT	3401.T	TEIJIN LTD.
MT	3402.T	TORAY INDUSTRIES, INC.
MT	3405.T	KURARAY CO., LTD.
MT	3407.T	ASAHI KASEI CORP.
MT	3861.T	OJI HOLDINGS CORP.
MT	4004.T	SHOWA DENKO K.K.
MT	4005.T	SUMITOMO CHEMICAL CO., LTD.
MT	4021.T	NISSAN CHEMICAL CORP.
MT	4042.T	TOSOH CORP.
MT	4061.T	DENKA CO., LTD.
MT	4063.T	SHIN-ETSU CHEMICAL CO., LTD.
MT	4183.T	MITSUI CHEMICALS, INC.
MT	4208.T	UBE INDUSTRIES, LTD.
MT	4272.T	NIPPON KAYAKU CO., LTD.
MT	4452.T	KAO CORP.
MT	4631.T	DIC CORP.
MT	4901.T	FUJIFILM HOLDINGS CORP.
MT	4911.T	SHISEIDO CO., LTD.
MT	5101.T	THE YOKOHAMA RUBBER CO., LTD.
MT	5108.T	BRIDGESTONE CORP.
MT	5202.T	NIPPON SHEET GLASS CO., LTD.
MT	5232.T	SUMITOMO OSAKA CEMENT CO., LTD.
MT	5233.T	TAIHEIYO CEMENT CORP.
MT	5301.T	TOKAI CARBON CO., LTD.
MT	5333.T	NGK INSULATORS, LTD.
MT	5401.T	NIPPON STEEL CORP.
MT	5406.T	KOBE STEEL, LTD.
MT	5541.T	PACIFIC METALS CO., LTD.
MT	5703.T	NIPPON LIGHT METAL HOLDINGS CO., LTD.
MT	5706.T	MITSUI MINING & SMELTING CO.
MT	5707.T	TOHO ZINC CO., LTD.
MT	5711.T	MITSUBISHI MATERIALS CORP.
MT	5713.T	SUMITOMO METAL MINING CO., LTD.
MT	5714.T	DOWA HOLDINGS CO., LTD.
MT	5801.T	FURUKAWA ELECTRIC CO., LTD.
MT	5802.T	SUMITOMO ELECTRIC IND., LTD.
MT	5803.T	FUJIKURA LTD.
MT	5901.T	TOYO SEIKAN GROUP HOLDINGS, LTD.
MT	6988.T	NITTO DENKO CORP.

Cuadro A.5: Lista de las compañías japonesas, incluidas en el índice Nikkei 225, continuamente cotizando en la bolsa de Tokio en el período considerado.

<b>Etiqueta</b>	<b>Ticker</b>	<b>Compañía</b>
MT	8001.T	ITOCHU CORP.
MT	8002.T	MARUBENI CORP.
MT	8015.T	TOYOTA TSUSHO CORP.
MT	8031.T	MITSUI & CO., LTD.
MT	8053.T	SUMITOMO CORP.
MT	8058.T	MITSUBISHI CORP.
TC	3105.T	NISSHINBO HOLDINGS INC.
TC	4151.T	KYOWA KIRIN CO., LTD.
TC	4502.T	TAKEDA PHARMACEUTICAL CO., LTD.
TC	4503.T	ASTELLAS PHARMA INC.
TC	4506.T	SUMITOMO DAINIPPON PHARMA CO., LTD.
TC	4507.T	SHIONOGI & CO., LTD.
TC	4519.T	CHUGAI PHARMACEUTICAL CO., LTD.
TC	4543.T	TERUMO CORP.
TC	4902.T	KONICA MINOLTA, INC.
TC	6479.T	MINEBEA MITSUMI INC.
TC	6503.T	MITSUBISHI ELECTRIC CORP.
TC	6504.T	FUJI ELECTRIC CO., LTD.
TC	6506.T	YASKAWA ELECTRIC CORP.
TC	6701.T	NEC CORP.
TC	6702.T	FUJITSU LTD.
TC	6703.T	OKI ELECTRIC IND. CO., LTD.
TC	6724.T	SEIKO EPSON CORP.
TC	6758.T	SONY CORP.
TC	6762.T	TDK CORP.
TC	6770.T	ALPS ALPINE CO., LTD.
TC	6841.T	YOKOGAWA ELECTRIC CORP.
TC	6902.T	DENSO CORP.
TC	6952.T	CASIO COMPUTER CO., LTD.
TC	6954.T	FANUC CORP.
TC	6976.T	TAIYO YUDEN CO., LTD.
TC	7201.T	NISSAN MOTOR CO., LTD.
TC	7202.T	ISUZU MOTORS LTD.
TC	7203.T	TOYOTA MOTOR CORP.
TC	7205.T	HINO MOTORS, LTD.
TC	7211.T	MITSUBISHI MOTORS CORP.
TC	7261.T	MAZDA MOTOR CORP.
TC	7269.T	SUZUKI MOTOR CORP.
TC	7270.T	SUBARU CORP.
TC	7272.T	YAMAHA MOTOR CO., LTD.
TC	7731.T	NIKON CORP.
TC	7733.T	OLYMPUS CORP.
TC	7735.T	SCREEN HOLDINGS CO., LTD.
TC	7752.T	RICOH CO., LTD.
TC	8035.T	TOKYO ELECTRON LTD.

Cuadro A.5: Lista de las compañías japonesas, incluidas en el índice Nikkei 225, continuamente cotizando en la bolsa de Tokio en el período considerado.

<b>Etiqueta</b>	<b>Ticker</b>	<b>Compañía</b>
UT	9001.T	TOBU RAILWAY CO., LTD.
UT	9005.T	TOKYU CORP.
UT	9007.T	ODAKYU ELECTRIC RAILWAY CO., LTD.
UT	9008.T	KEIO CORP.
UT	9009.T	KEISEI ELECTRIC RAILWAY CO., LTD.
UT	9062.T	NIPPON EXPRESS CO., LTD.
UT	9064.T	YAMATO HOLDINGS CO., LTD.
UT	9101.T	NIPPON YUSEN K.K.
UT	9104.T	MITSUI O.S.K.LINES, LTD.
UT	9107.T	KAWASAKI KISEN KAISHA, LTD.
UT	9301.T	MITSUBISHI LOGISTICS CORP.
UT	9501.T	TOKYO ELECTRIC POWER COMPANY HOLDINGS,
UT	9502.T	CHUBU ELECTRIC POWER CO., INC.
UT	9503.T	THE KANSAI ELECTRIC POWER CO., INC.
UT	9531.T	TOKYO GAS CO., LTD.
UT	9532.T	OSAKA GAS CO., LTD.

Cuadro A.5: Lista de las compañías japonesas, incluidas en el índice Nikkei 225, continuamente cotizando en la bolsa de Tokio en el período considerado.

<b>Etiqueta GICS</b>	<b>Sector</b>
CG	Consumer Goods
CP	Capital Goods / Others
FN	Financials
MT	Materials
TC	Technology
UT	Transportation & Utilities

Cuadro A.4: Sectores de las compañías que componen el índice Nikkei 225.



## Apéndice B

# Detalles metodológicos

### B.1. Lenguajes de programación y bibliotecas

El lenguaje de programación utilizado para los cálculos de este trabajo es *Python* [110], versión 3.11.6. Nos valimos principalmente de las bibliotecas:

1. *Numpy* [111] (versión 1.24.4), para almacenar arreglos numéricos, operarlos y diagonalizar matrices.
2. *pandas* [112, 113], para crear y analizar estructuras de datos (*dataframes*), especialmente series de tiempo.
3. *Matplotlib* [114, 115], para graficar.
4. *SciPy* [116] (versión 1.11.3), para rutinas de cómputo especiales, como el cálculo de matrices de similitud, de estadísticas (véase la sección siguiente) o de momentos de orden superior a 2 en datos numéricos (v.gr. asimetría y kurtosis).

Otras más se especificarán en las siguientes subsecciones.

### B.2. K-means y MDS

Para efectuar el K-means y el MDS, utilizamos la biblioteca *Scikit-learn* [117] (versión 1.3.1) particularmente las clases `sklearn.cluster` (K-means) y `sklearn.cluster` (MDS).

#### B.2.1. K-means

En el caso del K-means, utilizamos el algoritmo de Lloyd [85] (también llamado *naïve* K-means) un número variable de *clusters*  $k$  ( $k = 4, \dots, 12$ ), la inicialización *greedy k-means++* con 10 condiciones iniciales, un tope de 300 iteraciones (que de hecho nunca fueron alcanzadas) de no

haber convergencia, cuya tolerancia se como  $10^{-4}$ ; véase [esta documentación](#). Este procedimiento se repitió 15 veces, dando lugar a las secuencias  $\{S^l(1), \dots, S^l(T')\} \subseteq \{1, \dots, k\}$  para  $l = 1, \dots, 15$  (con  $T' = T - q + 1$ ; véanse la figura 2.1 y la relación 2.39).

Para cada  $l$ , la enumeración de los  $k$  cúmulos no tiene una ordenación canónica: las etiquetas  $S^l(t)$  sólo distinguen la pertenencia a uno u otro grupo. El criterio 3.1, definido en el Capítulo 3, permite comparar para cada  $t$  las ocupaciones  $S^l(t)$  y  $S^{l'}(t)$  para  $l \neq l'$ , y definir  $\{S(t) = \text{moda}(S^1(t), \dots, S^{15}(t)) : t = 1, \dots, T'\}$  como *el estado ocupado*. Así, la gráfica 4.11(b) (y las otras análogas) sintetiza de manera consistente los resultados de varias corridas del K-means; véase la figura B.1.

	iter 1	iter 2	iter 3	iter 4	iter 5	iter 6	iter 7	iter 8	iter 9	iter 10	iter 11	iter 12	iter 13	iter 14	iter 15
Date															
2006-02-01	1	1	1	2	0	4	3	3	4	2	0	0	3	4	2
2006-02-02	1	1	1	2	0	4	3	3	4	2	0	0	3	4	2
2006-02-03	1	1	1	2	0	4	3	3	4	2	0	0	3	4	2
2006-02-06	1	1	1	2	0	4	3	3	4	2	0	0	3	4	2
2006-02-07	1	1	1	2	0	4	3	3	4	2	0	0	3	4	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

(a)

	iter 1	iter 2	iter 3	iter 4	iter 5	iter 6	iter 7	iter 8	iter 9	iter 10	iter 11	iter 12	iter 13	iter 14	iter 15
Date															
2006-02-01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2006-02-02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2006-02-03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2006-02-06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2006-02-07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

(b)

Figura B.1: Ordenamiento canónico de los cúmulos obtenidos en distintas iteraciones del k-means, de acuerdo con el criterio 3.1. El orden resultante coincide entre las 15 iteraciones en el 99.32 % de las épocas, y en este caso determina la gráfica 4.11(b).

Para evaluar el desempeño de las soluciones se calculó, para cada  $l = 1, \dots, 15$ , una serie de métricas ofrecida por la biblioteca `Scikit-learn`: *Silhouette Coefficient* (mayor: mejor), *Calinski-Harabasz index* (mayor: mejor) y *Davies-Bouldin index* (menor: mejor). En el caso del k-means efectuado sobre las matrices de correlación sectoriales de Guhr en el S&P 500 (Capítulo 4), estas métricas se muestran en la figura B.2, junto con el número de iteraciones hasta llegar a la convergencia, y la inercia resultante. Nótese que el número de iteraciones no supera 40, y que los valores de las métricas y de la inercia son similares, pese a que se utilizaron 10 condiciones iniciales en cada una de las  $l = 1, \dots, 15$  corridas.

El estudio de estas mismas métricas para evaluar el desempeño de la convergencia, da resultados similares (convergencia rápida, uniformidad de los valores de los parámetros) en los otros casos en que aplicamos el k-means.

	n_iter	inertia	silhouette	calinski-harabasz	davies_bouldin
1	35	1716.827848	0.191238	2145.637862	1.534313
2	22	1716.827848	0.191238	2145.637862	1.534313
3	33	1716.827848	0.191238	2145.637862	1.534313
4	17	1716.778900	0.191344	2145.723971	1.531730
5	17	1716.827848	0.191238	2145.637862	1.534313
6	15	1716.772913	0.191379	2145.734502	1.530302
7	40	1716.827848	0.191238	2145.637862	1.534313
8	16	1716.777574	0.191336	2145.728817	1.532021
9	31	1716.827848	0.191238	2145.637862	1.534313
10	18	1716.779072	0.191321	2145.725909	1.532171
11	32	1716.827848	0.191238	2145.637862	1.534313
12	36	1716.827848	0.191238	2145.637862	1.534313
13	24	1716.791131	0.191310	2145.723217	1.532303
14	25	1716.827848	0.191238	2145.637862	1.534313
15	32	1716.827848	0.191238	2145.637862	1.534313

Figura B.2: Comparación de algunas métricas del desempeño del k-means, en el caso de las correlaciones de grano grueso en el S&P 500.

### B.2.2. MDS

En el caso del MDS, calculamos la inmersión a 3 dimensiones de las secuencias de correlaciones aplanadas  $\{\vec{c}(t) : t = 1, \dots, T'\} \subseteq \mathbb{R}^{N(N-1)/2}$  mediante la variante métrica del algoritmo (véase la sección 2.3) con la distancia euclidiana 2.33, con 10 inicializaciones, un tope de 300 iteraciones (que de hecho nunca fueron alcanzadas) de no haber convergencia, cuya tolerancia se definió de  $10^{-3}$ ; véase esta documentación. Este procedimiento resultó en secuencias  $\{\vec{c}^*(t) : t = 1, \dots, T'\} \subseteq \mathbb{R}^3$  de puntos con centro geométrico en el origen  $(0, 0, 0)$ , pero con una orientación arbitraria (véase la figura B.3).

Con ayuda de la clase `sklearn.decomposition.PCA` se obtuvo la base de coordenadas de los ejes principales  $(x, y, z)$  en orden decreciente de la inercia de los puntos  $\{\vec{c}'(t) : t = 1, \dots, T'\} \subseteq \mathbb{R}^3$ , pensados como masas puntuales iguales.

El procedimiento descrito es el mismo para todos los tipos de correlación estudiados en este trabajo.



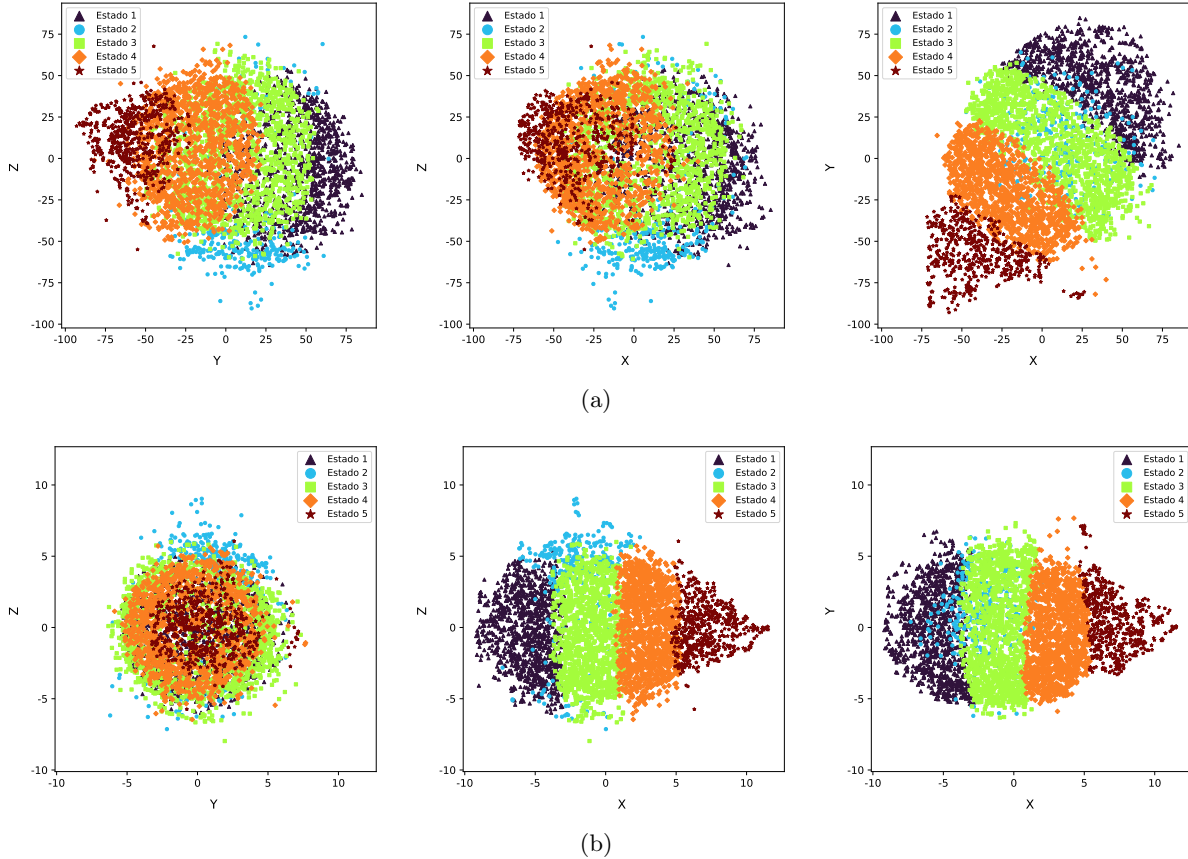


Figura B.3: MDS de los estados de mercado para las correlaciones de Pearson  $\mathcal{C}^{\text{Pearson}}$  antes y después de cambiar las coordenadas a las de los ejes de mayor varianza (es decir, los ejes principales). En este caso,  $k = 5$  y no se aplica el *power map*. El K-means se hace en el espacio completo (en este caso,  $N(N - 1)/2$ -dimensional), para luego obtenerse las coordenadas 3D con el MDS coloreadas de acuerdo con el estado ocupado en el espacio completo.

### B.3. Histogramas y estimación de densidades de probabilidad

Para estimar la densidad de probabilidad de una variable aleatoria real  $X$  a partir de una muestra de tamaño  $n$ ,  $\{x_1, \dots, x_n\}$ , se pueden utilizar dos enfoques [118, 119]:

1. Paramétrico: asumiendo que  $X$  tiene una distribución  $f(x; \mu_i)$  con parámetros  $\mu_i$ , éstos se determinan como función de  $\{x_1, \dots, x_n\}$ .
2. No paramétrico: sin asumir una distribución particular, se determina una función adaptada a los datos  $\{x_1, \dots, x_n\}$  de manera heurística.

La estimación de densidad de *kernel* o KDE<sup>1</sup> cae dentro del enfoque no paramétrico, y su heurística se basa en los histogramas: a cada  $x_i$  se le asocia un *bin*, no necesariamente rectangular y sin la restricción de evitar los traslapes. Específicamente, habiendo fijado un *híperparámetro*

<sup>1</sup>Kernel Density Estimation.

$h > 0$ , el *ancho de banda*; una función *kernel*  $K = K(u)$  no negativa, con dominio simétrico con respecto al origen, tal que  $\int K(u) du = 1$ ,  $K(u) = K(-u)$ , y  $\int u^2 K(u) du > 0$ ; y una métrica  $d = d(x, y)$  en  $\mathbb{R}$ , el estimador de densidad de *kernel* se define como [119, cap. 6, sec. 1]:

$$\hat{f}_n(x) := \frac{1}{nh} \sum_{i=1}^n K\left(\frac{d(x, x_i)}{h}\right) \equiv \frac{1}{n} \sum_{i=1}^n K_h(d(x, x_i)) \quad (\text{B.1})$$

donde  $K_h(u) := K(u/h)/h$  es el *kernel* reescalado.

Intuitivamente, el ancho de banda  $h$  controla la suavidad de B.1: cuando  $h \rightarrow 0$ ,  $\hat{f}$  se vuelve una superposición de deltas de Dirac; cuando  $h \rightarrow \infty$ ,  $\hat{f}$  se vuelve uniforme. Hay diversos criterios para elegir el ancho de banda  $h$  y el *kernel*  $K$ , que en general buscan minimizar una *función de costo* o *función de error*, como el error cuadrático integrado medio o MISE<sup>2</sup>:

$$R(f, \hat{f}) := \mathbb{E} \left[ \int [f(x) - \hat{f}(x)]^2 dx \right] \quad (\text{B.2})$$

entre la *densidad real*  $f$  y el estimador  $\hat{f}$ , que en sí mismo puede pensarse como una variable aleatoria en función de  $h$ . En [120, cap. 20, secs. 2 y 3] se demuestra que para los *estimadores de histograma*  $\hat{g}_n = \hat{g}_n(x)$ :

$$\hat{g}_n(x) := \sum_{i=1}^m \frac{\hat{p}_i}{h} \mathbb{1}_{B_i} \quad (\text{B.3})$$

(donde  $B_1, \dots, B_m$  son intervalos disjuntos de la misma longitud que conforman el soporte de  $x$ ,  $h = 1/m$  y  $\hat{p}_i = \nu_i/n$  para  $\nu_i$  el número de observaciones  $\{x_1, \dots, x_n\}$  en  $B_i$ ), el MISE alcanza el valor mínimo  $R(f, \hat{g}_n) \approx \frac{C}{n^{2/3}}$  (con  $C$  una constante) para la  $h$  óptima, mientras que el MISE del estimador KDE B.1 alcanza el mínimo  $R(f, \hat{f}_n) \approx \frac{C'}{n^{4/5}}$  bajo condiciones generales. En otras palabras: los histogramas tienen una tasa de convergencia de  $\mathcal{O}(n^{-2/3})$ , que es superada por la tasa de convergencia  $\mathcal{O}(n^{-4/5})$  del estimador KDE. Además, ningún estimador converge más rápidamente que  $\mathcal{O}(n^{-4/5})$ .

Los criterios anteriores son la base para obtener una  $h$  óptima en la práctica, de acuerdo con los datos a la mano. Son comunes los métodos de Silverman [121] y Scott [122, 123]; véanse [124] y [125].

Algunos tipos comunes de *kernels* son:

1. Gaussiano:  $K(u) = \frac{e^{-u^2/2}}{\sqrt{2\pi}}$
2. Rectángulo (caja, sombrero de copa):  $K(u) = \begin{cases} 1/2 & \text{si } |u| \leq 1 \\ 0 & \text{en otro caso} \end{cases}$
3. Exponencial:  $K(u) = \frac{e^{-|u|}}{2}$
4. Coseno (truncado):  $K(u) = \begin{cases} \frac{\pi}{4} \cos\left(\frac{\pi u}{2}\right) & \text{si } |u| \leq 1 \\ 0 & \text{en otro caso} \end{cases}$

---

<sup>2</sup>Mean Integrated Squared error.

En este trabajo, utilizamos las bibliotecas `KDEpy` [126] y `SciPy` [116] (particularmente la clase `scipy.stats.gaussian_kde`). Las matrices de dispersión, como en la figura 3.10, tienen en su diagonal estimaciones KDE hechas con `SciPy`, con *kernel* gaussiano y ancho de banda determinado por la regla de Scott; véase esta documentación. Las densidades de probabilidad de  $c_{ij}$  ( $i \neq j$ ), como las graficadas en 3.20, 3.21 y 3.5, fueron calculadas con la biblioteca `KDEpy`, *kernel* coseno y ancho de banda determinado por la regla de Silverman; véase esta documentación.

Cuando es posible, se complementan las estimaciones KDE con histogramas (por ejemplo, en la Figura 4.9), cuyos *bins* se determinaron con el método de Doane [127], que es apropiado para datos no gaussianos; véase esta documentación.

# Apéndice C

## Material suplementario

### C.1. Esbozo de las características de las matrices de Wishart

La densidad de  $\hat{\Sigma}$  condicionada a  $\Sigma$  [51, cap. 3, cor. 3.2.2] es<sup>1</sup>:

$$f(\hat{\Sigma}|\Sigma) = \left(\frac{T}{2}\right)^{NT/2} \frac{(\det \hat{\Sigma})^{(T-N-1)/2}}{\Gamma_N(T/2) (\det \Sigma)^{T/2}} \exp\left(-\frac{T}{2} \text{tr} \Sigma^{-1} \hat{\Sigma}\right) \quad (\text{C.1})$$

donde  $\Gamma_N(t)$  es la *función gamma multivariada*<sup>2</sup>. De esta expresión, se puede deducir [128, teo. 7.1.2][51, cap. 3, teo. 3.2.18 y ej. 3.13] la densidad conjunta de valores propios  $\lambda_1, \dots, \lambda_N$ <sup>3</sup> de  $\hat{\Sigma}$ :

$$\begin{aligned} \varphi(\lambda_1, \dots, \lambda_N|\Sigma) &= \left(\frac{T}{2}\right)^{NT/2} \frac{\pi^{N^2/2}}{\Gamma_N(N/2) \Gamma_N(T/2) (\det \Sigma)^{T/2}} \prod_{i=1}^N \lambda_i^{(T-N-1)/2} \prod_{j < k} (\lambda_k - \lambda_j) \times \\ &\times \int_{\mathbb{O}(N)} \exp\left(-\frac{T}{2} \text{tr} \left(\Sigma^{-1} \mathbf{H} \text{diag}(\lambda_1, \dots, \lambda_N) \mathbf{H}^\top\right)\right) d\mu(\mathbf{H}) \quad (\text{C.2}) \end{aligned}$$

donde  $\mathbb{O}(N)$  es el grupo ortogonal de orden  $N$  con volumen  $\frac{2^N \pi^{N^2/2}}{\Gamma_N(N/2)}$ , y  $d\mu(\mathbf{H})$  es la medida de Haar (o *invariante*) normalizada que corresponde<sup>4</sup>:  $\int_{\mathbb{O}(N)} d\mu(\mathbf{H}) = 1$ .

<sup>1</sup>Cabe resaltar que: (1) La matriz  $\hat{\Sigma}$  es positiva definida con probabilidad 1 si, y sólo si,  $T < N$  [51, cap. 3, teo. 3.1.4]; (2) en el caso  $N = 1$ , la fórmula C.1 para la densidad de  $\hat{\Sigma} \equiv \hat{\sigma}^2$  condicionada a  $\Sigma \equiv \sigma^2$  resulta  $f(\hat{\sigma}^2|\sigma^2) = \left(\frac{T}{2\sigma^2}\right)^{T/2} \frac{(\hat{\sigma}^2)^{(T/1)-1}}{\Gamma\left(\frac{T}{2}\right)} \exp\left(-\frac{T\hat{\sigma}^2}{2\sigma^2}\right)$ , así que la variable aleatoria  $\frac{T\hat{\sigma}^2}{\sigma^2}$  tiene distribución  $\chi_T^2$  [79, págs. 62-64].

<sup>2</sup>Definida como una integral sobre el espacio  $S_N^+$  de matrices  $A$  de  $N \times N$ , simétricas y positivas definidas:  $\Gamma_N(t) := \int_{S_N^+} (\det A)^{t-(N+1)/2} e^{-\text{tr} A} dA$ . A su vez, esta integral se puede expresar como  $\Gamma_N(t) = \pi^{N(N-1)/4} \prod_{k=1}^N \Gamma\left(t - \frac{k-1}{2}\right)$ , donde  $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$  es la función gamma.

<sup>3</sup>Que en lo sucesivo consideramos ordenados de forma ascendente:  $\lambda_1 \leq \dots \leq \lambda_N$ .

<sup>4</sup>Explícitamente:  $d\mu(\mathbf{H}) = \frac{\Gamma_N(N/2)}{2^N \pi^{N^2/2}} \mathbf{H}^\top d\mathbf{H}$ , y  $\mathbf{H}^\top d\mathbf{H} = \prod_{1 \leq i < j \leq N} \mathbf{h}_j^\top d\mathbf{h}_i$  para las columnas  $\mathbf{h}_k$  de  $\mathbf{H} \in \mathbb{O}(N)$ :

$$\mathbf{H} = (\mathbf{b}_1 \quad \dots \quad \mathbf{b}_N)$$

Al considerar el caso general del espaciamiento entre niveles (véase la fórmula 2.9) en el *Gaussian Orthogonal Ensemble* o GOE (donde sólo se suponen matrices reales de  $N \times N$ , invariantes bajo el grupo  $\mathbb{O}(N)$  y con elementos cuyos momentos son finitos<sup>5</sup>), la fórmula 2.9 es una excelente aproximación [47, Cap. 7].

Cuando  $\Sigma = \sigma^2 \mathbf{I}$ , la densidad C.1 se reduce a:

$$\begin{aligned} f(\hat{\Sigma} | \sigma^2 \mathbf{I}) &= \left(\frac{T}{2}\right)^{NT/2} \frac{(\det \hat{\Sigma})^{(T-N-1)/2}}{\Gamma_N(N/2) \Gamma_N(T/2) (\sigma^2)^{NT/2}} \exp\left(-\frac{T}{2\sigma^2} \text{tr} \hat{\Sigma}\right) \\ &= \left(\frac{T}{2\sigma^2}\right)^{NT/2} \frac{1}{\Gamma_N(N/2) \Gamma_N(T/2)} \exp\left(-\frac{T}{2\sigma^2} \text{tr}(\hat{\Sigma}) + \frac{T-N-1}{2} \text{tr}(\log \hat{\Sigma})\right) \end{aligned} \quad (\text{C.3})$$

donde la última igualdad se debe a la fórmula de Jacobi:  $\det(\exp(\mathbf{B})) = \exp(\text{tr}(\mathbf{B}))$ , y a que  $\log(\mathbf{A})$  tiene sentido para  $\mathbf{A}$  positiva definida. Las matrices  $\hat{\Sigma}$  que siguen la distribución C.3 forman el *Wishart Orthogonal Ensemble* o WOE, y aunque tiene ciertas similitudes con el GOE (como la distribución 2.9), se distingue de éste porque la densidad C.2 toma la forma:

$$\begin{aligned} \varphi(\lambda_1, \dots, \lambda_N | \sigma^2 \mathbf{I}) &= \left(\frac{T}{2\sigma^2}\right)^{NT/2} \frac{\pi^{N^2/2}}{\Gamma_N(N/2) \Gamma_N(T/2)} \times \\ &\quad \times \prod_{i=1}^N \lambda_i^{(T-N-1)/2} \prod_{j < k} (\lambda_k - \lambda_j) \exp\left(-\frac{T}{2\sigma^2} \sum_{i=1}^N \lambda_i\right) \end{aligned} \quad (\text{C.4})$$

La densidad 2.10, mencionada en el Capítulo 2, se sigue alcanzando bajo condiciones menos restrictivas. Por ejemplo, en [130, sec. 2.2.2] se muestra que basta independencia en las columnas de la matriz  $(\hat{\mathbf{X}}(1) \cdots \hat{\mathbf{X}}(T)) \in \mathbb{R}^{N \times T}$ , con media  $\mu = \langle \hat{X}_i(t) \rangle$  y varianza  $\sigma^2 = \text{Var}(\hat{X}_i(t))$  comunes, más la condición:

$$\forall \eta > 0 : \quad \frac{1}{\eta NT} \sum_{i=1}^N \sum_{t=1}^T \left\langle \left| \hat{X}_i(t) \right|^2 \mathbb{1}_{\{|\hat{X}_i(t)| \geq \eta \sqrt{T}\}} \right\rangle \rightarrow 0$$

para que la densidad espectral de  $\hat{\Sigma}$  converja a 2.10, con  $N, T \rightarrow \infty$  pero  $N/T \rightarrow \lambda > 0$  fija. Cabe mencionar que  $\lambda_{\text{máx}}$  y  $\lambda_{\text{mín}} := \min\{\lambda_i : \lambda_i \neq 0\}$  convergen con probabilidad 1 a  $\lambda_+$  y  $\lambda_-$  si, y sólo si, las entradas de la matriz  $(\hat{\mathbf{X}}(1) \cdots \hat{\mathbf{X}}(T)) \in \mathbb{R}^{N \times T}$  son independientes, idénticamente distribuidas y tienen momento de orden 4 finito [131-133][129, sec. 5].

## C.2. Otras estadísticas de las matrices locales de correlación

### C.2.1. Proyecciones de Guhr

<sup>5</sup>Frecuentemente se caracteriza a las matrices del GOE como aquellas de la forma  $\frac{1}{\sqrt{2}}(\mathbf{A} + \mathbf{A}^T)$ , donde  $\mathbf{A}$  es una matriz de  $N \times N$  con entradas independientes e idénticamente distribuidas acorde a  $\mathcal{N}(0, 1)$  [129, sec. 5].

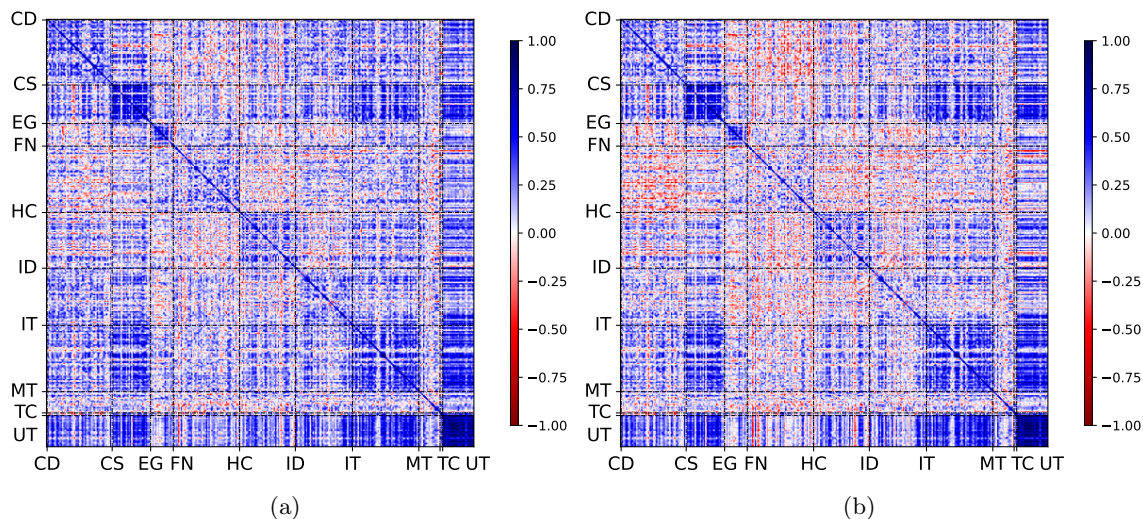


Figura C.1: Matrices de correlación proyectadas de Guhr, con los métodos de covarianza (subfigura C.1(a)) y correlación (subfigura C.1(b)) en una época con longitud de  $q = 20$  días bursátiles que finaliza en el tiempo  $t = 10/11/2016$ . La distribución de los elementos no diagonales de esta matriz se muestra en la figura C.2.

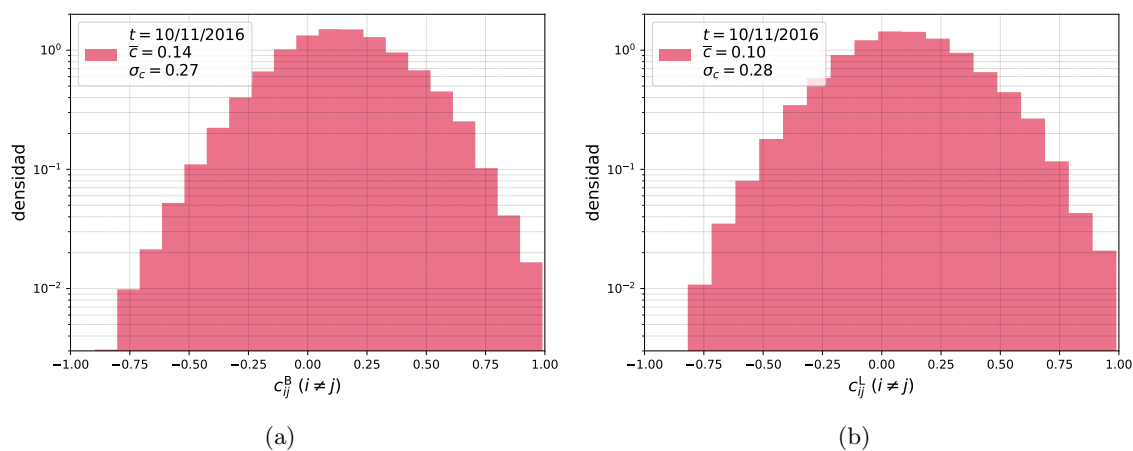


Figura C.2: Distribución de los elementos no diagonales de las matrices de correlación de la figura C.1. Dentro de cada marco se despliega el promedio y desviación estándar de los datos correspondientes. Nótese la elección de la escala vertical. Los *bins* de este histograma se calcularon con el método de Doane.

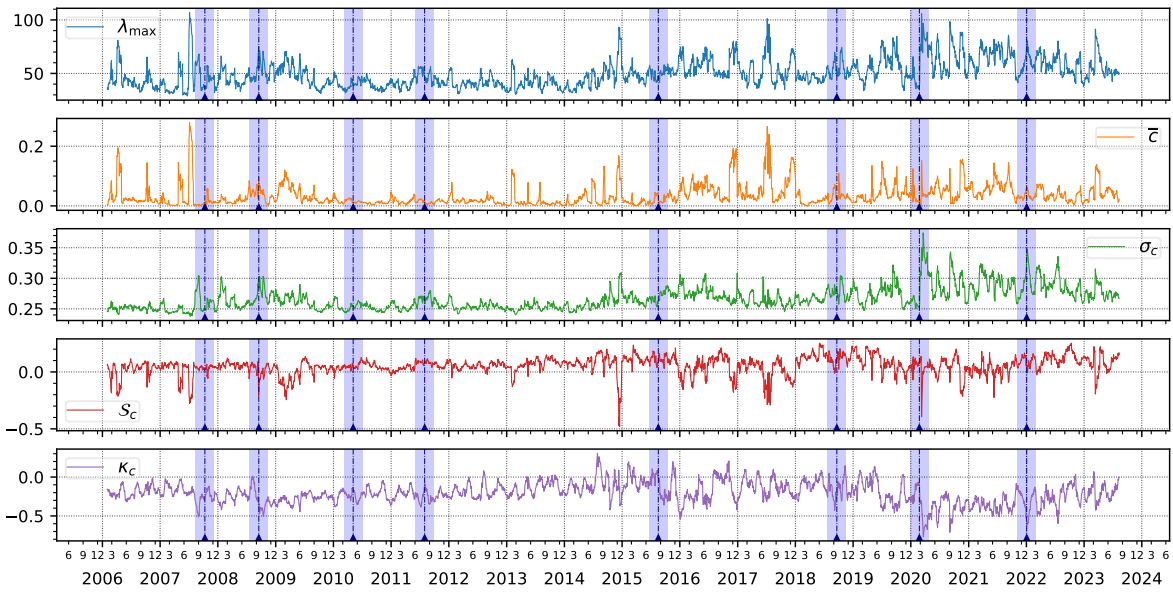


Figura C.3: Series de tiempo del mayor eigenvalor  $\lambda_{\text{máx}}$ , la correlación promedio  $\bar{c}$ , la desviación estándar  $\sigma_c := \sqrt{c^2 - \bar{c}^2}$ , así como su asimetría (*skewness*)  $\mathcal{S}_c$  y kurtosis  $\kappa_c$ , de las matrices de correlación proyectadas de Guhr locales en el tiempo, con el método de covarianza. Las franjas verticales traslúcidas de color azul marcan períodos de crisis; véase la tabla A.2.

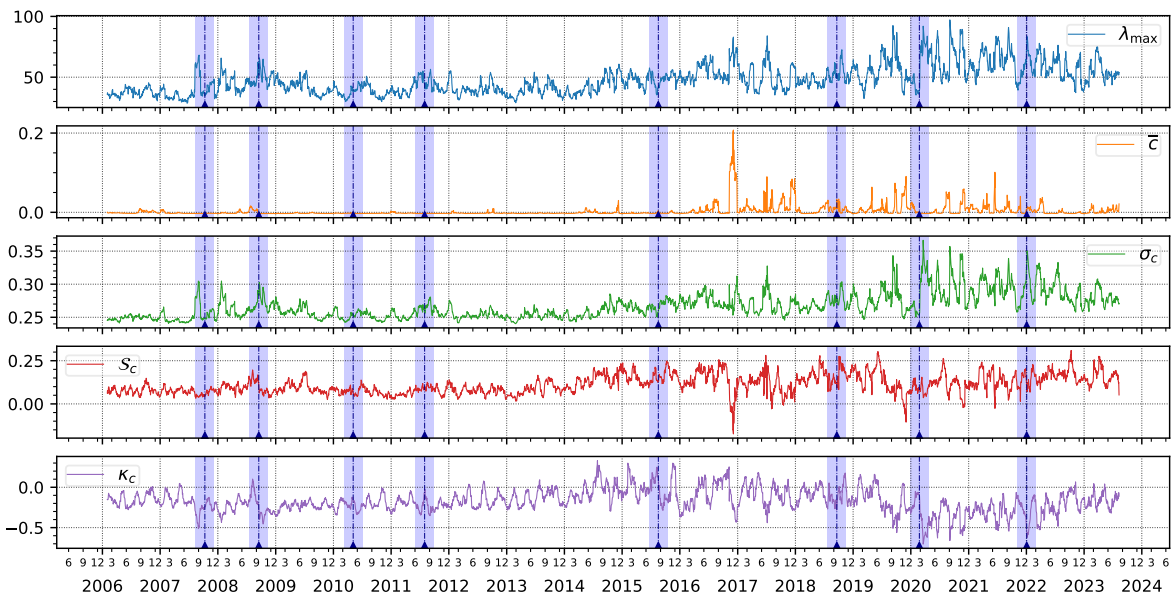
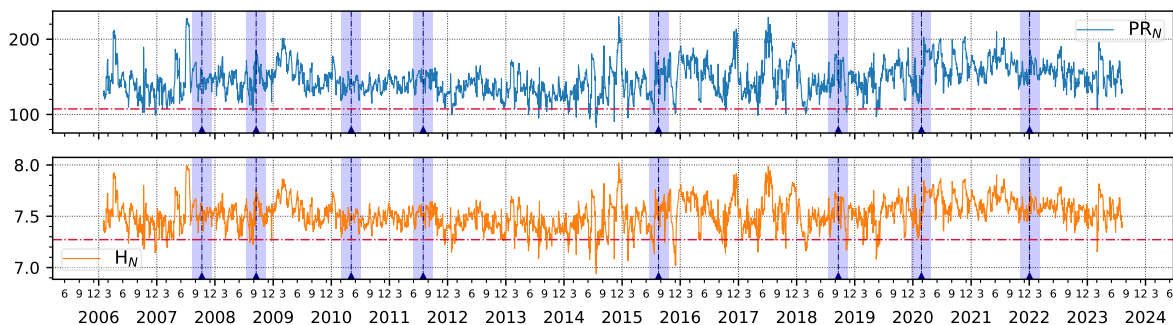
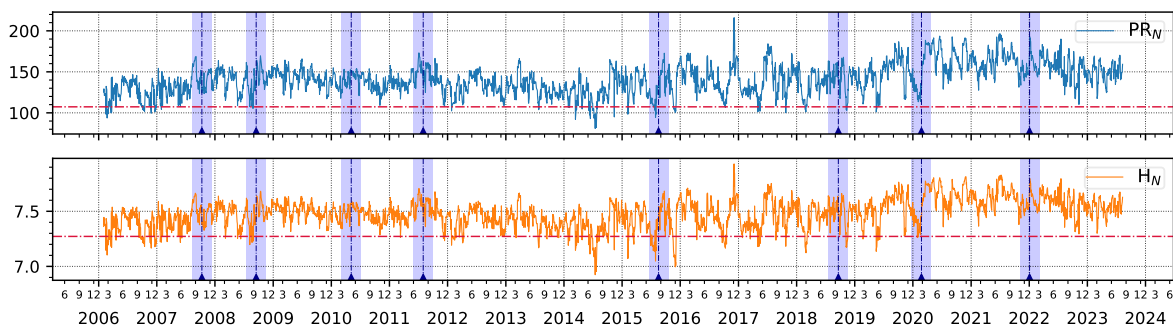


Figura C.4: Series de tiempo del mayor eigenvalor  $\lambda_{\text{máx}}$ , la correlación promedio  $\bar{c}$ , la desviación estándar  $\sigma_c := \sqrt{c^2 - \bar{c}^2}$ , así como su asimetría (*skewness*)  $\mathcal{S}_c$  y kurtosis  $\kappa_c$ , de las matrices de correlación proyectadas de Guhr locales en el tiempo, con el método de correlación. Las franjas verticales traslúcidas de color azul marcan períodos de crisis; véase la tabla A.2.



(a)



(b)

Figura C.5: Series de tiempo de la razón de participación del eigenvector principal  $PR_N$  y entropía de Shannon del (módulo al cuadrado de las componentes del) eigenvector principal  $H_N$ , para las correlaciones proyectadas de Guhr locales con los métodos de covarianza (subfigura C.5(a)) y correlación (subfigura C.5(b)). Las franjas verticales traslúcidas de color azul marcan períodos de crisis; véase la tabla A.2. Las líneas horizontales discontinuas (color rojo) corresponden a los valores del WOE.



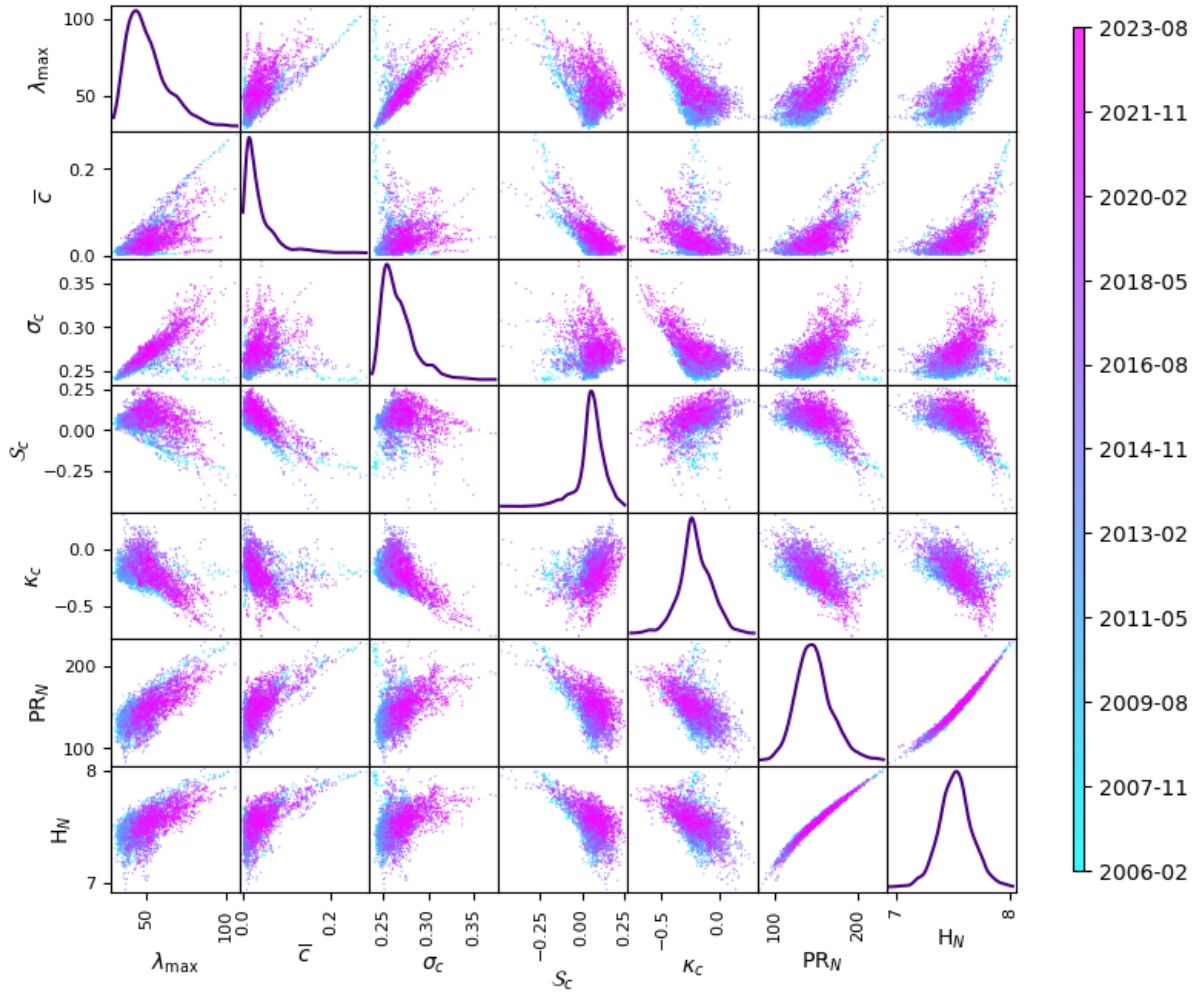


Figura C.6: Gráficos de dispersión entre las series de tiempo del eigenvalor máximo  $\lambda_{\max}$ , correlación promedio  $\bar{c}$ , desviación estándar de las correlaciones  $\sigma_c$ , asimetría  $S_c$ , kurtosis  $\kappa_c$ , razón de participación del eigenvector principal  $PR_N$  y entropía de Shannon del (módulo cuadrado de las componentes del) eigenvector principal  $H_N$ , calculadas a partir de las correlaciones proyectadas de Guhr locales (método de covarianza). En la diagonal de esta gráfica se representa la distribución empírica de cada serie de tiempo sobre el horizonte completo, calculada con el método KDE.

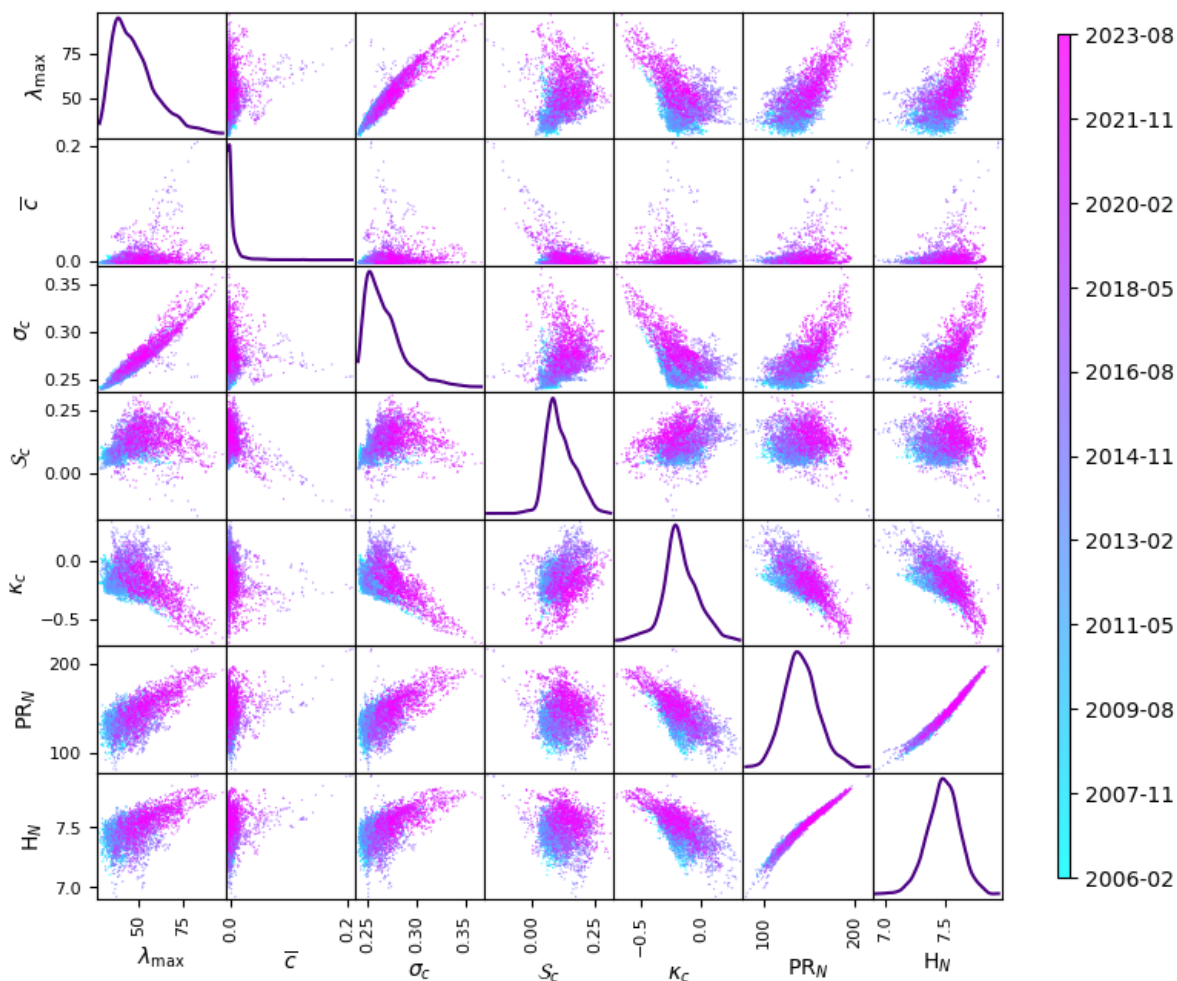


Figura C.7: Gráficos de dispersión entre las series de tiempo del eigenvalor máximo  $\lambda_{\max}$ , correlación promedio  $\bar{c}$ , desviación estándar de las correlaciones  $\sigma_c$ , asimetría  $S_c$ , kurtosis  $\kappa_c$ , razón de participación del eigenvector principal  $PR_N$  y entropía de Shannon del (módulo cuadrado de las componentes del) eigenvector principal  $H_N$ , calculadas a partir de las correlaciones proyectadas de Guhr locales (método de correlación). En la diagonal de esta gráfica se representa la distribución empírica de cada serie de tiempo sobre el horizonte completo, calculada con el método KDE.

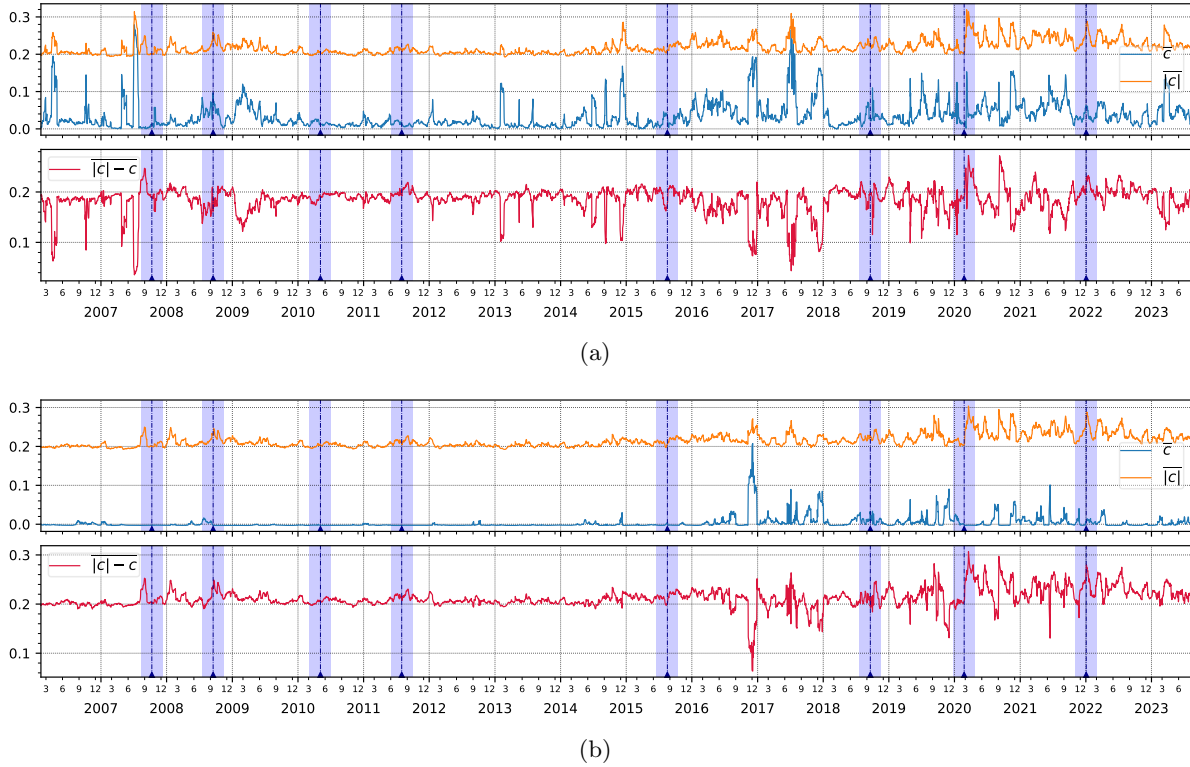


Figura C.8: Para medir la presencia de correlaciones negativas, calculamos  $\overline{|c|} = \overline{|c|}(t)$ ,  $\overline{c} = \overline{c}(t)$  (primer renglón), así como su diferencia (segundo renglón). Tomando en cuenta las  $T - q + 1 = 4411$  matrices de correlación proyectadas, los valores promedio de  $\overline{|c|} - \overline{c}$  sobre  $t$  son de aproximadamente 0.184 (método de covarianza, C.8(a)) y 0.212 (método de correlación, C.8(b)), con desviaciones estándar de 0.027 (método de covarianza) y 0.019 (método de correlación), y cuartiles de 0.177, 0.189 (mediana) y 0.198 (método de covarianza), y 0.202, 0.209 (mediana) y 0.221 (método de correlación).

### C.3. Eigenvalores en el S&P 500

Las figuras C.9 y C.10 muestran, en distintas escalas, los promedios temporales de los eigenvalores dominantes de las matrices de correlación de Pearson y relativas al índice.

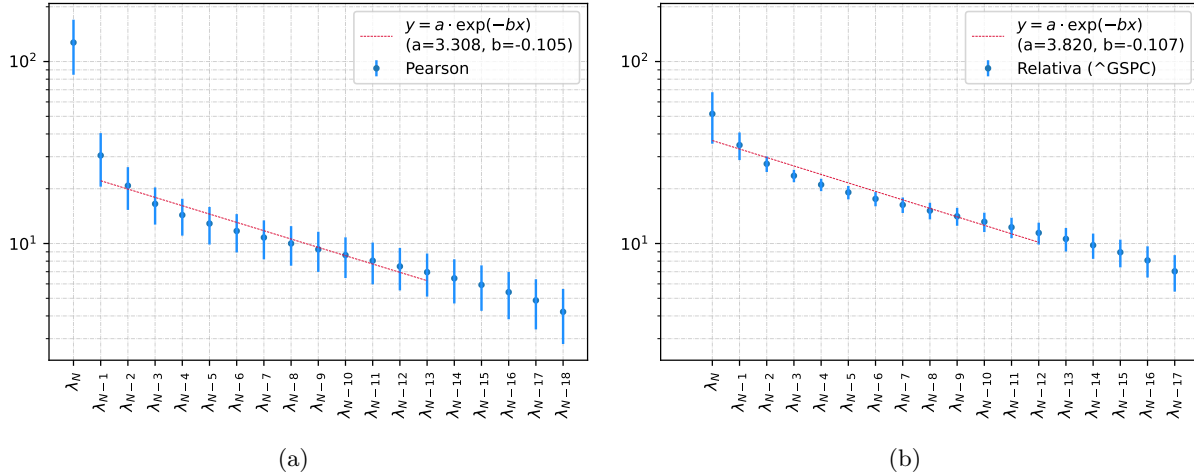


Figura C.9: Diagramas de pedregal (*scree diagrams*) en escala lin-log, para los eigenvalores principales, promediados sobre todo el horizonte temporal, de las matrices locales de correlación de Pearson (subfigura C.9(a)) y relativas al índice (subfigura C.9(b)). Las barras de error marcan simétricamente las desviaciones estándar de cada eigenvalor promediado. Solo se tomaron los  $q - 1$  (Pearson) o  $q - 2$  (relativa) eigenvalores principales, pues las matrices de correlación se calcularon sobre series de tiempo cortas en épocas de  $q = 20$  días, y la normalización, así como la aplicación de la fórmula 2.25, reducen el rango de las matrices.

Mostramos escalas lin-log y log-log para contrastar la calidad de los ajustes exponencial y por ley de potencias de algunos de los eigenvalores. El ajuste por ley de potencias  $\lambda_N \sim N^b$ , correspondiente a la figura C.10, parece ser el más adecuado, y muestra cómo el exponente característico  $b$  es similar en las correlaciones de Pearson y relativas al índice.

#### C.3.1. Proyecciones de Guhr

Covarianza	$\lambda_N$	$\lambda_{N-1}$	$\lambda_{N-2}$	$\lambda_{N-3}$	Correlación	$\lambda_N$	$\lambda_{N-1}$	$\lambda_{N-2}$	$\lambda_{N-3}$
$\lambda_N$	1	0.386	0.129	-0.232	$\lambda_N$	1	0.495	0.188	-0.146
$\lambda_{N-1}$	0.386	1	0.382	0.017	$\lambda_{N-1}$	0.495	1	0.366	-0.015
$\lambda_{N-2}$	0.129	0.382	1	0.440	$\lambda_{N-2}$	0.188	0.366	1	0.443
$\lambda_{N-3}$	-0.232	0.017	0.440	1	$\lambda_{N-3}$	-0.146	-0.015	0.443	1

Cuadro C.1: Correlación entre las series de tiempo de los cuatro mayores eigenvalores de las matrices proyectadas de Guhr locales en el tiempo, con los métodos de covarianza (izquierda) y correlación (derecha).

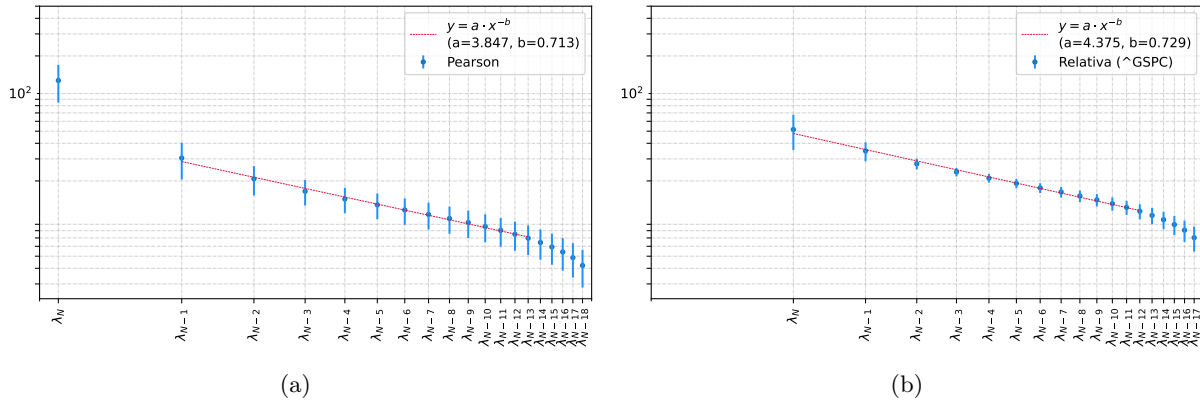
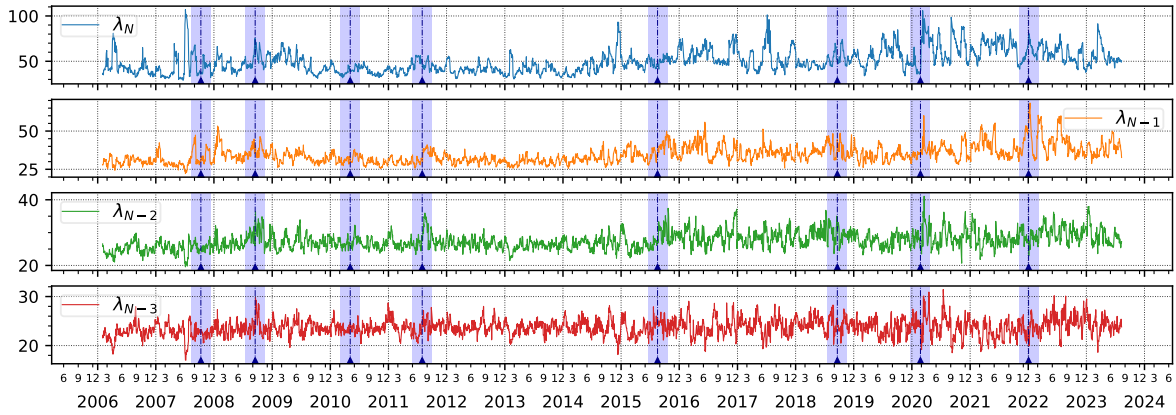
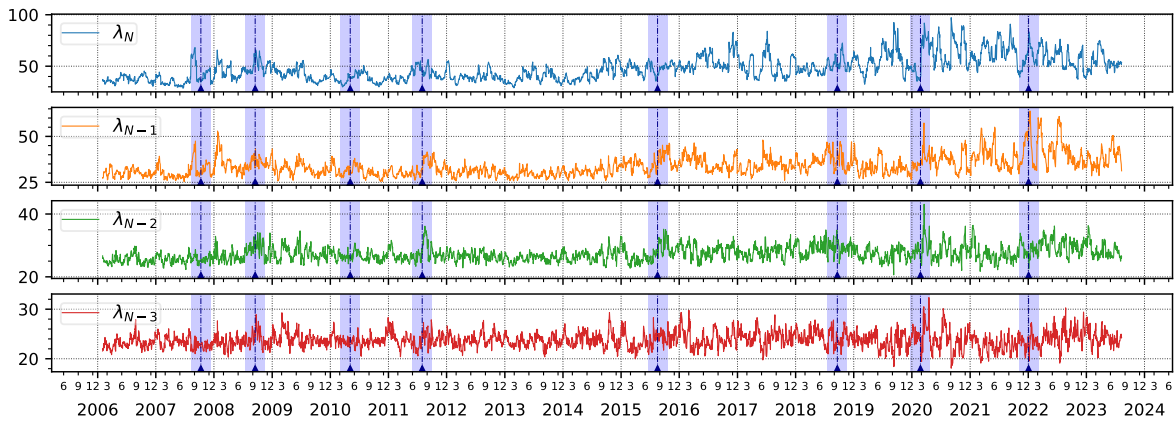


Figura C.10: Diagramas de pedregal (*scree diagrams*) en escala log-log, para los eigenvalores principales, promediados sobre todo el horizonte temporal, de las matrices locales de correlación de Pearson (subfigura C.9(a)) y relativas al índice (subfigura C.9(b)). Las barras de error marcan simétricamente las desviaciones estándar de cada eigenvalor promediado. Solo se tomaron los  $q - 1$  (Pearson) o  $q - 2$  (relativa) eigenvalores principales, pues las matrices de correlación se calcularon sobre series de tiempo cortas en épocas de  $q = 20$  días, y la normalización, así como la aplicación de la fórmula 2.25, reducen el rango de las matrices.



(a)



(b)

Figura C.11: Series de tiempo de los cuatro mayores eigenvalores  $\lambda_N = \lambda_{\text{máx}}$ ,  $\lambda_{N-1}$ ,  $\lambda_{N-2}$  y  $\lambda_{N-3}$  de las matrices de correlación proyectadas de Guhr locales en el tiempo, con los métodos de covarianza (subfigura C.11(a)) y de correlación (subfigura C.11(b)). Las franjas verticales traslúcidas de color azul marcan períodos de crisis; véase la tabla A.2. Compárense estas figuras con la tabla C.1 y la figura C.12.

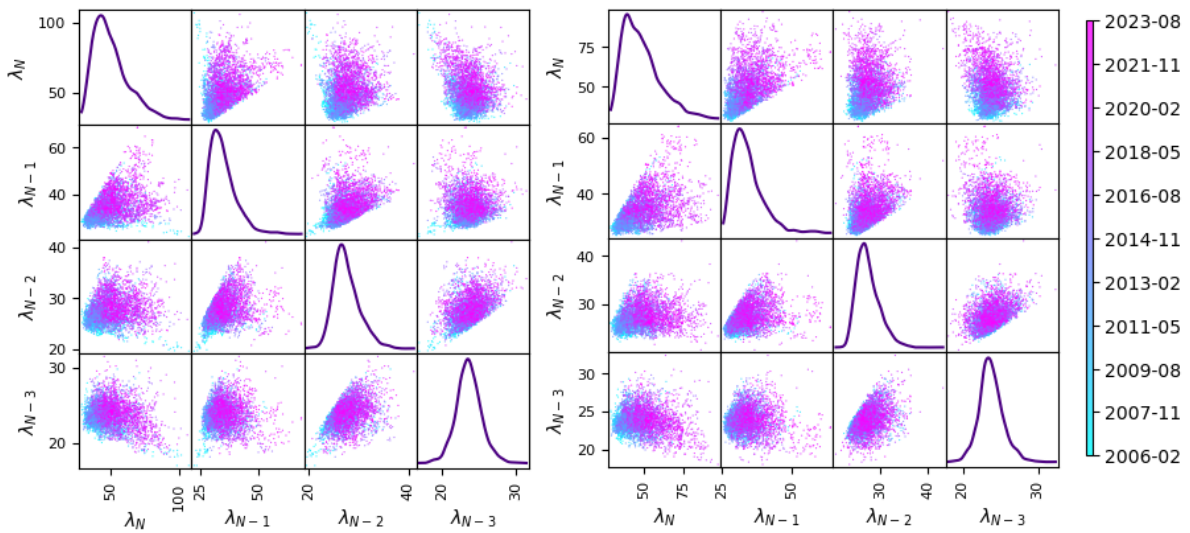


Figura C.12: Gráficos de dispersión entre las series de tiempo de los mayores eigenvalores  $\lambda_N = \lambda_{\text{máx}}, \lambda_{N-1}, \lambda_{N-2}$  y  $\lambda_{N-3}$  de las matrices de correlación proyectadas de Guhr locales en el tiempo, con los métodos de covarianza (izquierda) y de correlación (derecha). Las anticorrelaciones mostradas por la dispersión se reflejan en la tabla izquierda del C.1. En la diagonal de esta gráfica se representa la distribución empírica de cada serie de tiempo sobre el horizonte completo, calculada con el método KDE.

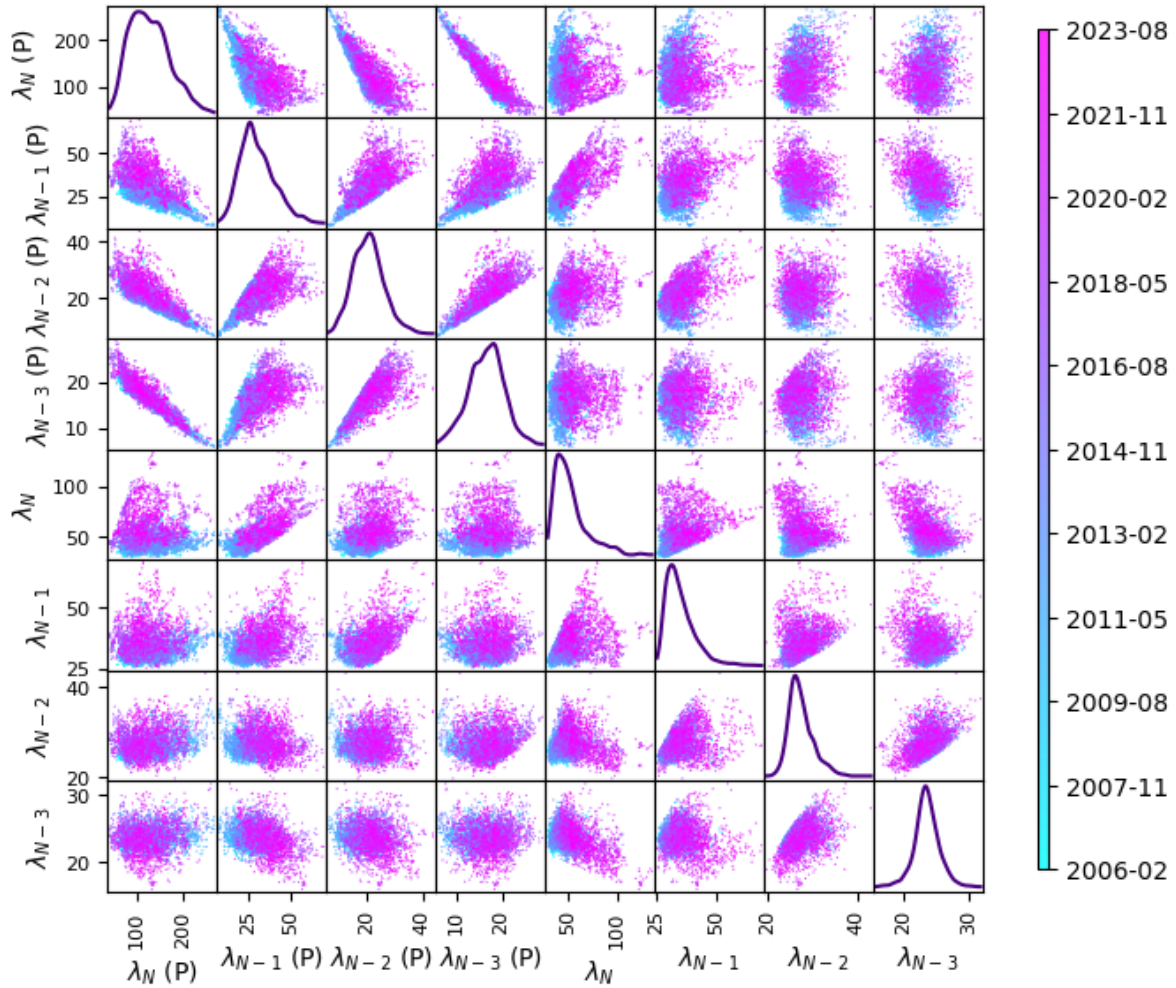


Figura C.13: Gráficos de dispersión entre las series de tiempo de los mayores eigenvalores  $\lambda_N = \lambda_{\text{máx}}$ ,  $\lambda_{N-1}$ ,  $\lambda_{N-2}$  y  $\lambda_{N-3}$  de las matrices de correlación locales en el tiempo de Pearson y relativas al índice. Las etiquetas  $(P)$  distinguen a las series de tiempo de Pearson. En la diagonal de esta gráfica se representa la distribución empírica de cada serie de tiempo sobre el horizonte completo, calculada con el método KDE. Esta figura es una extensión de 3.13, que incluye las correlaciones cruzadas de las estadísticas de Pearson y relativas al índice.



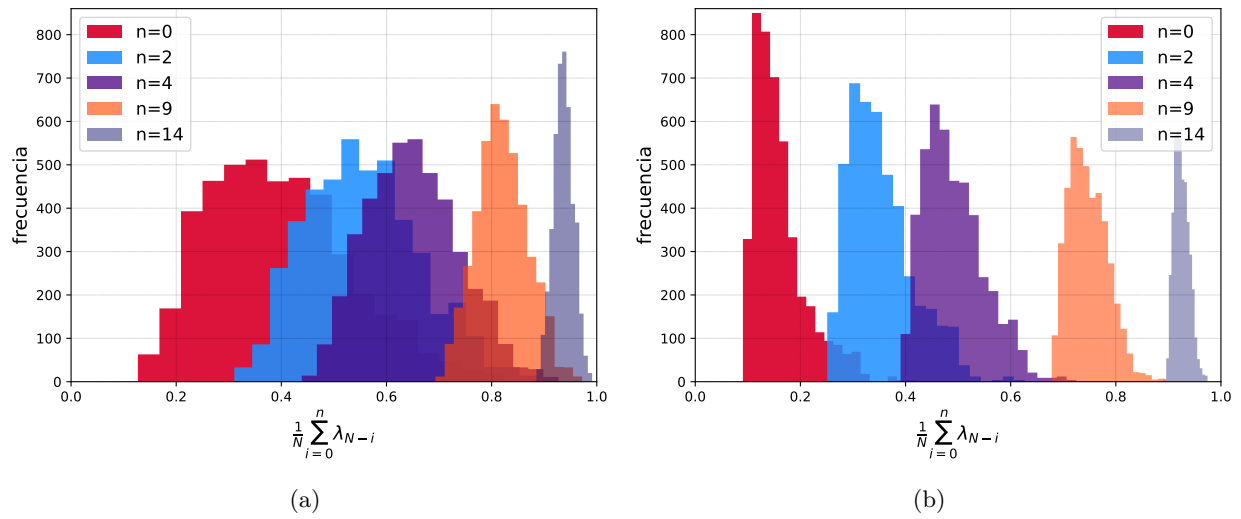
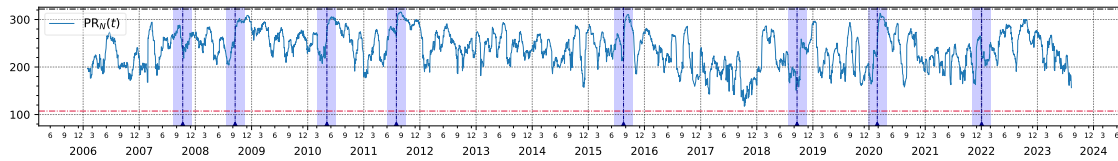
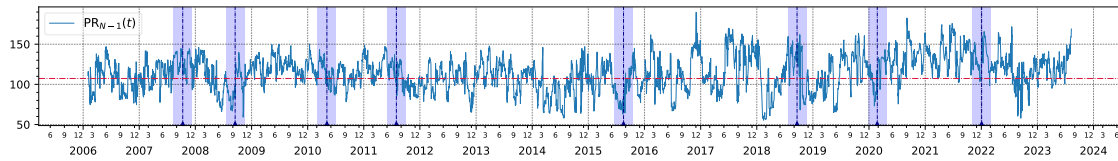


Figura C.14: Distribución de  $f(n) := \frac{1}{N} \sum_{i=0}^n \lambda_{N-i}$  para distintos valores de  $n$ , sobre todo el horizonte temporal, donde  $\{\lambda_{ij}\}_{i=0}^N$  son los eigenvalores de las matrices locales de correlación de Pearson (subfigura C.14(a)) y relativas al índice (subfigura C.14(b)) ordenados de forma ascendente. A medida que  $n$  se acerca a  $q-1$  o  $q-2$ ,  $f(n)$  se concentra en 1, debido a la invariancia de la traza de una matriz cuadrada bajo cambio de coordenadas, y porque las matrices de correlación se calcularon sobre series de tiempo cortas en épocas de  $q = 20$  días, y la normalización, o la aplicación de la fórmula 2.25, reducen el rango de las matrices en 1 (Pearson) o 2 unidades (correlación relativa al índice).

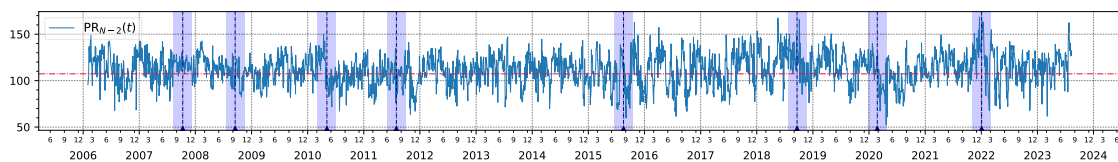
## C.4. Eigenvectores en el S&amp;P 500



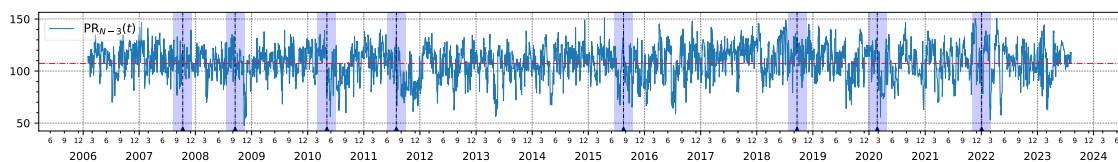
(a)



(b)



(c)



(d)

Figura C.15: De arriba abajo: series de tiempo de la razón de participación  $PR_n(t)$  para  $n = N, N - 1, N - 2, N - 3$ . Las líneas horizontales marcan las cotas superior (color negro, solamente en la subfigura C.15(a)) e inferior (color rojo, en todas las subfiguras) correspondientes a los valores de correlación constante y WOE, respectivamente.

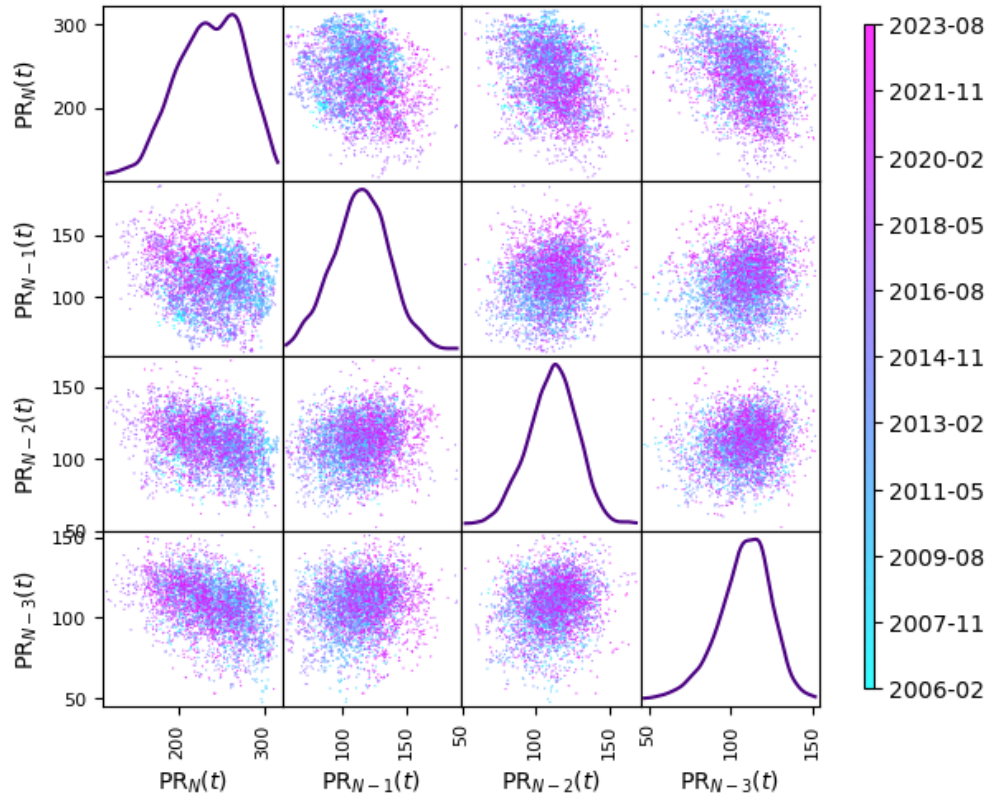


Figura C.16: Gráficos de dispersión entre las series de tiempo de la figura C.15, correspondientes a la razón de participación de los cuatro eigenvectores principales de  $\hat{\mathbf{C}}^{(\epsilon)}(t; q)$ . En la diagonal de esta gráfica se representa la distribución empírica de cada serie de tiempo sobre el horizonte completo, calculada con el método KDE.

	$PR_N(t)$	$PR_{N-1}(t)$	$PR_{N-2}(t)$	$PR_{N-3}(t)$
Media	238.066341	114.066548	111.674535	108.434453
Desviación estándar	37.028645	21.639393	16.368411	15.627272
Mínimo	117.391093	55.430496	52.368632	47.584891
25 %	211.311902	99.593914	100.943752	98.900495
50 %	239.585383	114.611709	112.406715	109.732312
75 %	267.239549	129.007325	122.969531	119.261812
Máximo	315.962994	189.695363	168.484933	151.271158

Cuadro C.2: Estadísticas básicas de las series de tiempo de  $PR_n(t)$  para  $n = N, N - 1, N - 2, N - 3$  (ver figura C.15). Conforme  $n$  disminuye, el valor de  $PR_n$  se concentra en el del GOE,  $PR_{\text{GOE}} = N/3 \approx 107.3$ .

## C.5. Dinámica de los estados

### C.5.1. Proyecciones de Guhr

#### Distribución de los elementos de matriz en los $k=5$ centroides

Como complemento a la gráfica 3.33, de los centroides de las matrices proyectadas de Guhr del Capítulo 3, se muestra a continuación la figura C.17 con la distribución, en escala semilogarítmica, de sus elementos no diagonales. Aparentemente, hay una bimodalidad que se difunde conforme incrementa el índice de cada cúmulo; véanse los números reportados en la tabla C.3

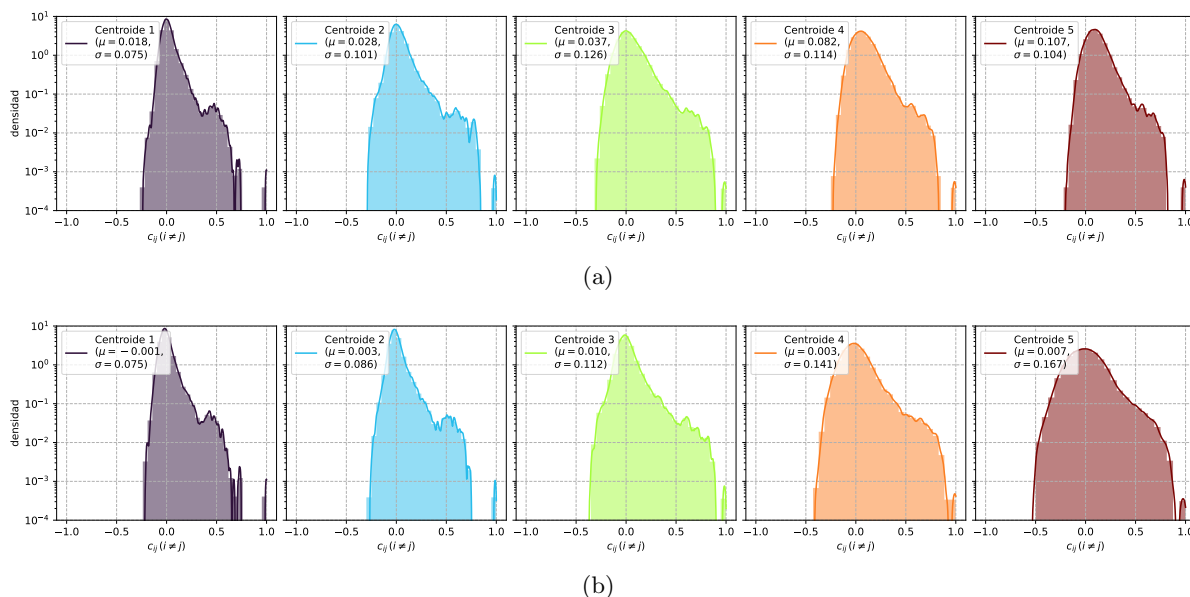


Figura C.17: KDE de las entradas no diagonales de los centroides de los cúmulos para las correlaciones proyectadas de Guhr  $\mathcal{C}^B$  (método de covarianza, subfigura C.17(a)) y  $\mathcal{C}^L$  (método de correlación, subfigura C.17(b)), cuando  $k = 5$  y no se aplica el *power map*.

#### $k=4$

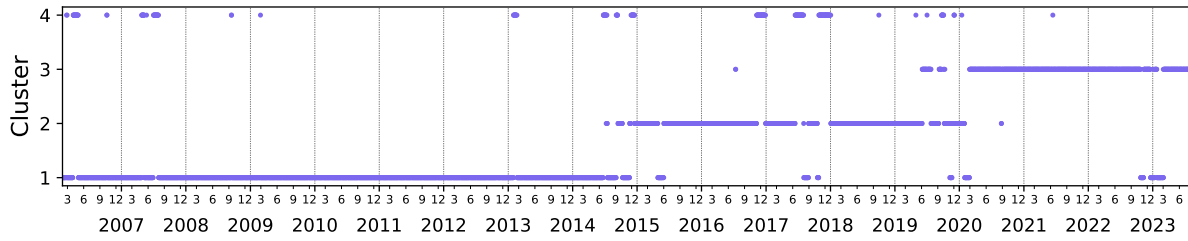
En esta subsección nos concentraremos en mostrar los resultados de las matrices de correlación proyectadas de Guhr, para  $k = 4$  y  $\epsilon = 0$ . Esta elección de parámetros, de acuerdo con lo discutido en el Capítulo 3, parece ser la más adecuada.

$\mathcal{C}^{\mathbf{B}}$	Centroide 1	Centroide 2	Centroide 3	Centroide 4	Centroide 5
$\bar{c}$	0.018	0.028	0.037	0.082	0.107
$\sigma_c$	0.075	0.101	0.126	0.114	0.104
$\mathcal{S}_c$	2.751	2.433	1.604	1.360	1.451
$\kappa_c$	13.956	11.388	5.248	4.094	5.036

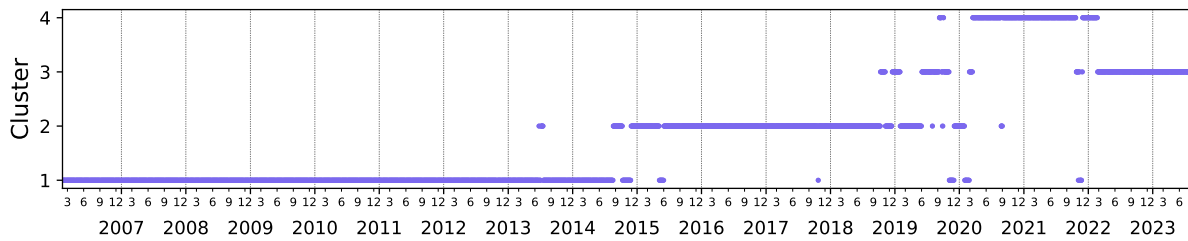
  

$\mathcal{C}^{\mathbf{L}}$	Centroide 1	Centroide 2	Centroide 3	Centroide 4	Centroide 5
$\bar{c}$	-0.001	0.003	0.010	0.003	0.007
$\sigma_c$	0.075	0.086	0.112	0.141	0.167
$\mathcal{S}_c$	2.774	2.921	2.311	1.403	0.757
$\kappa_c$	14.049	14.861	10.328	4.543	1.668

Cuadro C.3: Primeros momentos centrales de las entradas no diagonales de las matrices de correlación que representan los centroides de los  $k = 5$  clusters de las matrices de correlación proyectadas de Guhr con los métodos de covarianza (tabla en la parte superior) y correlación (tabla inferior), representados en la figura 3.33.



(a)



(b)

Figura C.18: Evolución de los estados de mercado para las correlaciones proyectadas de Guhr  $\mathcal{C}^{\mathbf{B}}$  (método de covarianza) y  $\mathcal{C}^{\mathbf{L}}$  (método de correlación), cuando  $k = 4$  y no se aplica el *power map*. El ordenamiento de los estados es creciente con el promedio del eigenvalor máximo de los miembros de cada estado o cúmulo (relación 3.1).

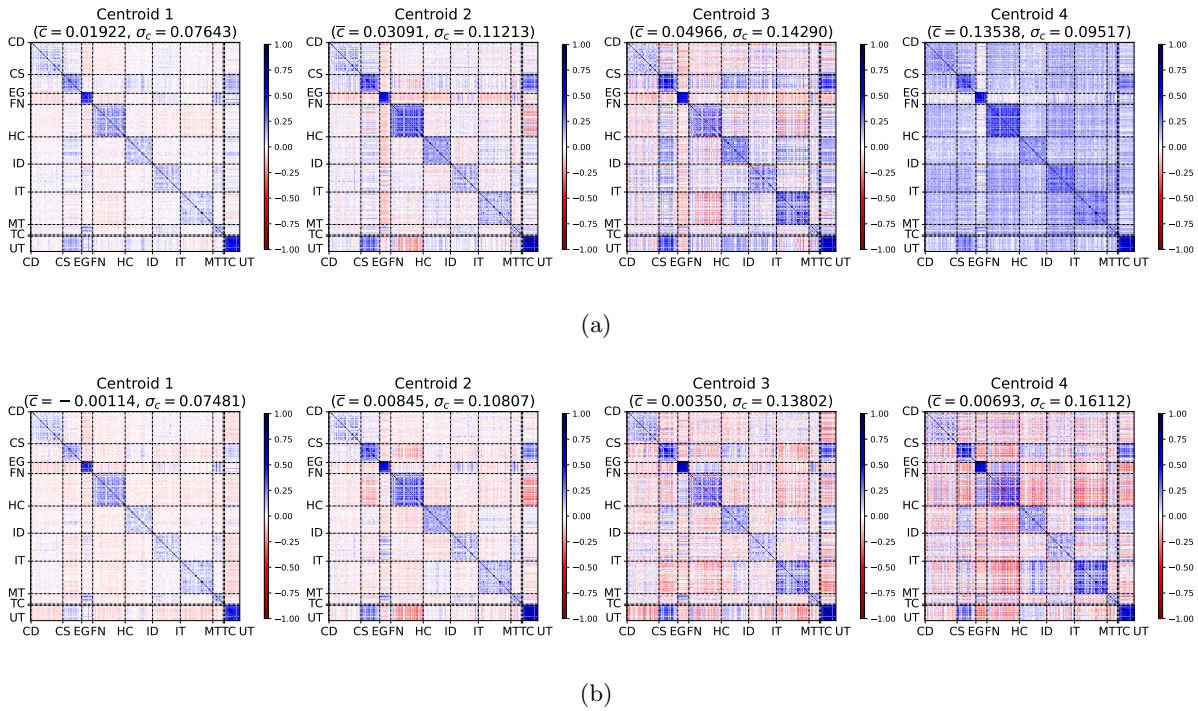


Figura C.19: Centroides de los estados de mercado para las correlaciones proyectadas de Guhr, cuando  $k = 4$  y no se aplica el *power map*.

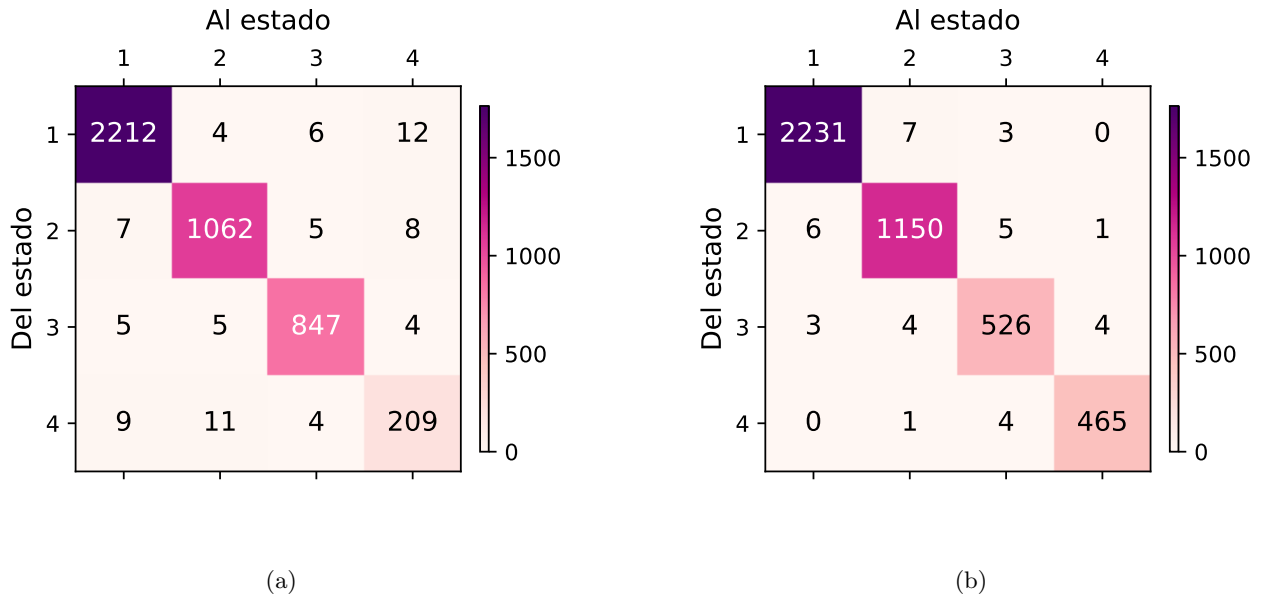


Figura C.20: Matrices de transición para las correlaciones proyectadas de Guhr  $\mathcal{C}^B$  (método de covarianza) y  $\mathcal{C}^L$  (método de correlación), cuando  $k = 4$  y no se aplica el *power map*. La dinámica de los estados de la figura C.18 se traduce en matrices de transición cuasi-simétricas, pero no tridiagonales como en 3.17. La subfigura C.20(a) corresponde a los cúmulos visitados por las correlaciones  $\mathcal{C}^B$ , mientras que la subfigura C.20(b) corresponde a los *clusters* ocupados por las correlaciones  $\mathcal{C}^L$ .

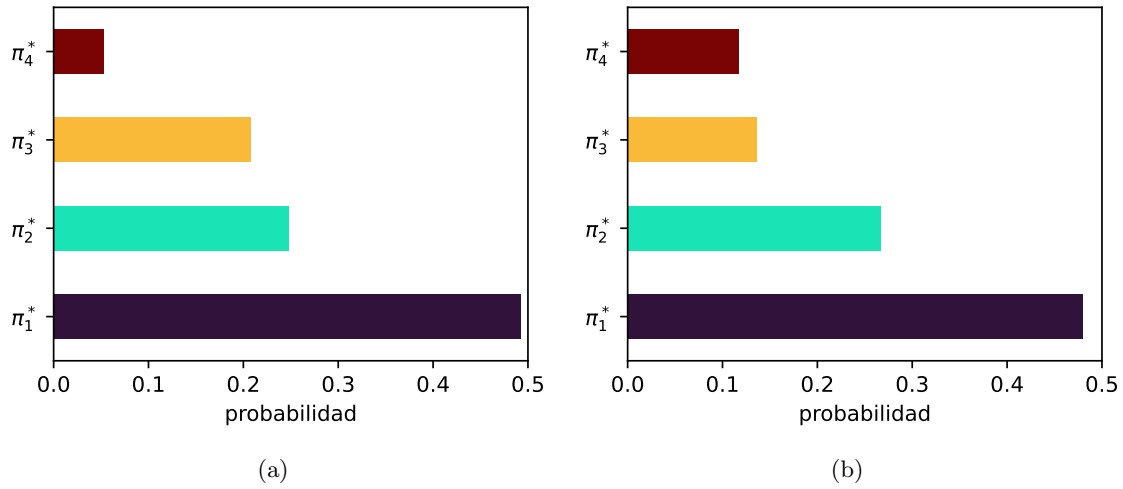


Figura C.21: Distribuciones de probabilidad estacionarias, calculadas mediante la condición 2.42, a partir de las matrices de transición mostradas en la figura C.20, cuando aquéllas se normalizan de acuerdo con la fórmula 2.41. Los valores son: método de covarianza:  $\pi^* = (0.4921, 0.2477, 0.2074, 0.0528)$ ; método de correlación:  $\pi^* = (0.4797, 0.2673, 0.1361, 0.1169)$ .

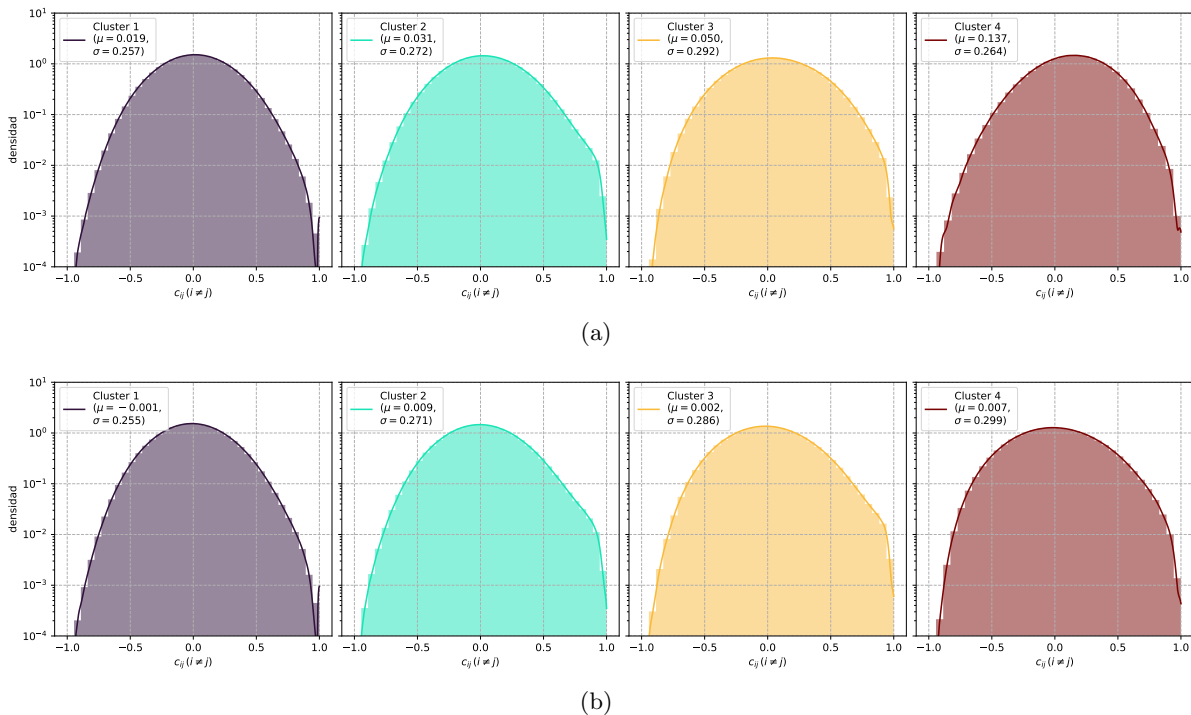


Figura C.22: KDE de las entradas no diagonales de las matrices que conforman cada uno de los  $k = 4$  cúmulos de las correlaciones proyectadas de Guhr  $\mathcal{C}^B$  (método de covarianza, subfigura C.22(a)) y  $\mathcal{C}^L$  (método de correlación, subfigura C.22(b)). No se aplica el *power map*.

$\mathcal{C}^{\mathbf{B}}$	Cúmulo 1	Cúmulo 2	Cúmulo 3	Cúmulo 4
$\bar{c}$	0.019	0.031	0.050	0.137
$\sigma_c$	0.257	0.272	0.292	0.264
$\mathcal{S}_c$	0.055	0.101	0.064	-0.123
$\kappa_c$	-0.194	-0.134	-0.325	-0.199

$\mathcal{C}^{\mathbf{L}}$	Cúmulo 1	Cúmulo 2	Cúmulo 3	Cúmulo 4
$\bar{c}$	-0.001	0.009	0.002	0.007
$\sigma_c$	0.255	0.271	0.286	0.299
$\mathcal{S}_c$	0.084	0.146	0.161	0.120
$\kappa_c$	-0.166	-0.071	-0.184	-0.332

Cuadro C.4: Primeros momentos centrales de las entradas no diagonales de las matrices que conforman cada uno de los  $k = 4$  cúmulos de las matrices de de correlación proyectadas de Guhr con los métodos de covarianza (tabla en la parte superior) y correlación (tabla inferior).

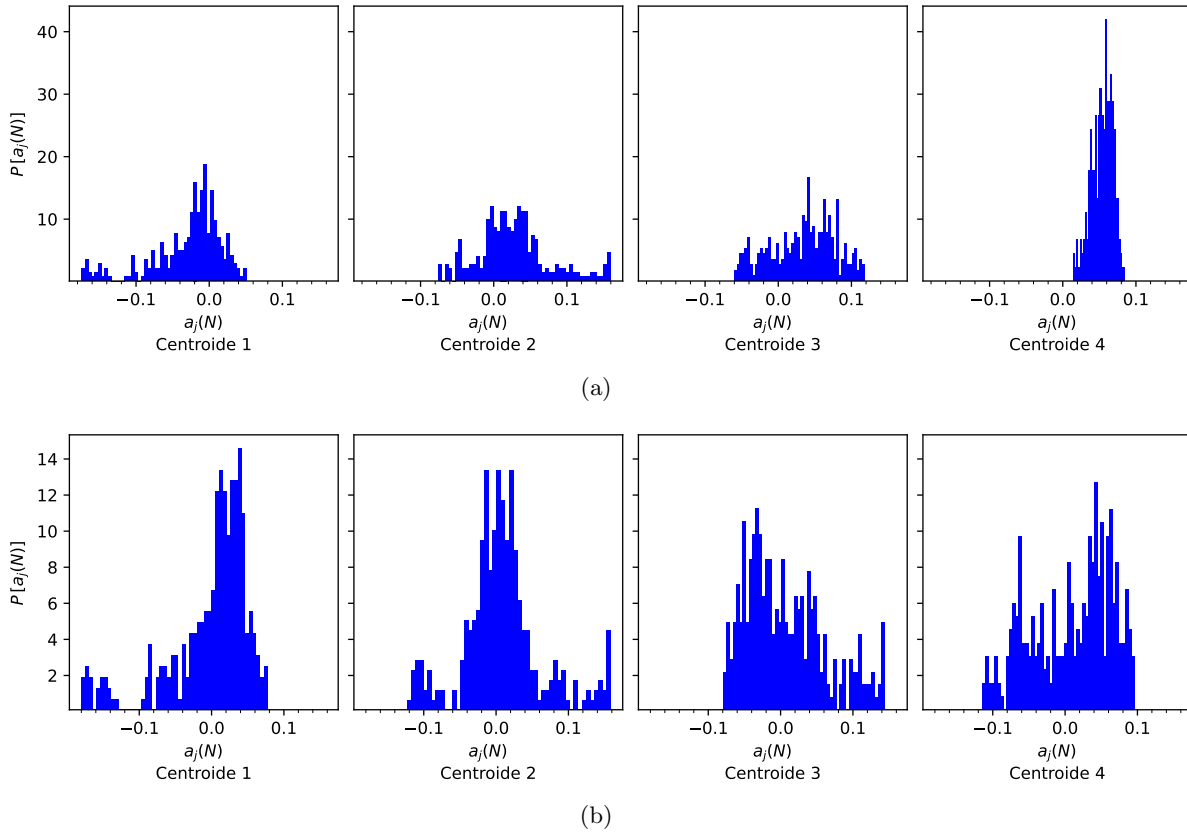


Figura C.23: Histograma de las componentes del eigenvector principal  $|N\rangle$  de cada centroide de los cúmulos de las correlaciones proyectadas de Guhr  $\mathcal{C}^{\mathbf{B}}$  (método de covarianza, subfigura C.23(a)) y  $\mathcal{C}^{\mathbf{L}}$  (método de correlación, subfigura C.23(b)), cuando  $k = 4$  y no se aplica el *power map*. Los eigenvectores están normalizados, y se orientan de tal manera que  $a_1(N) \geq 0$  (véase la relación 2.18).



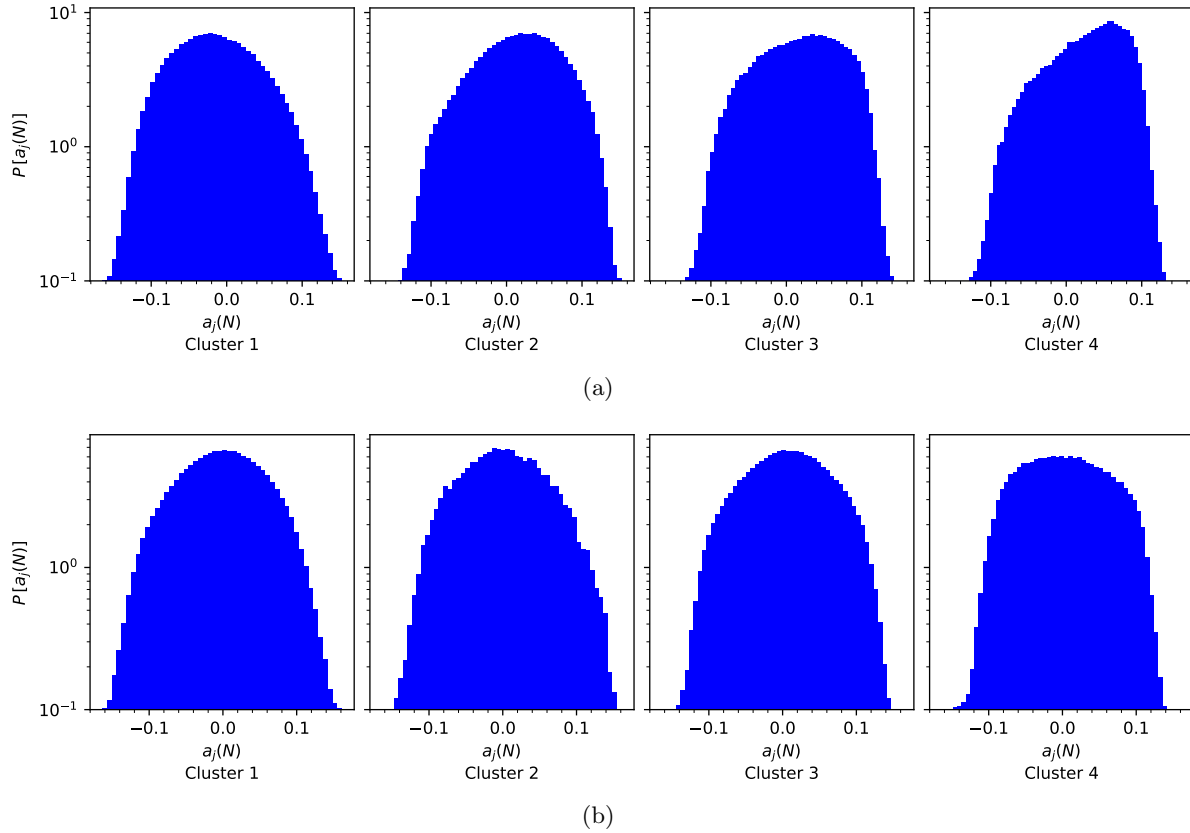


Figura C.24: Histograma de las componentes del eigenvector principal de los elementos de  $\mathcal{C}^B$  (subfigura C.24(a)) y  $\mathcal{C}^L$  (subfigura C.24(b)), cuando  $k = 4$  y no se aplica el *power map*. Los eigenvectores están normalizados, y se orientan en el mismo semiplano que el eigenvector principal  $|N\rangle$  del centroide correspondiente.

$\mathcal{C}^B$	Centroide 1	Centroide 2	Centroide 3	Centroide 4
$\overline{a(N)}$	0.028	-0.026	-0.034	0.054
$\sigma_{a(N)}$	0.048	0.049	0.044	0.014
$\mathcal{S}_{a(N)}$	1.281	-0.613	0.280	-0.472
$\kappa_{a(N)}$	1.404	0.584	-0.771	-0.304

$\mathcal{C}^L$	Centroide 1	Centroide 2	Centroide 3	Centroide 4
$\overline{a(N)}$	0.002	0.009	-0.007	-0.009
$\sigma_{a(N)}$	0.056	0.055	0.055	0.055
$\mathcal{S}_{a(N)}$	1.516	0.437	-0.719	0.433
$\kappa_{a(N)}$	1.953	1.082	-0.294	-0.986

Cuadro C.5: Primeros momentos centrales de las componentes del eigenvector principal  $|N\rangle$  de cada centroide de los cúmulos para las correlaciones proyectadas de Guhr  $\mathcal{C}^B$  (subfigura C.23(a)) y  $\mathcal{C}^L$  (subfigura C.23(b)), cuando  $k = 4$  y no se aplica el *power map*.

## C.6. Figuras complementarias del MDS

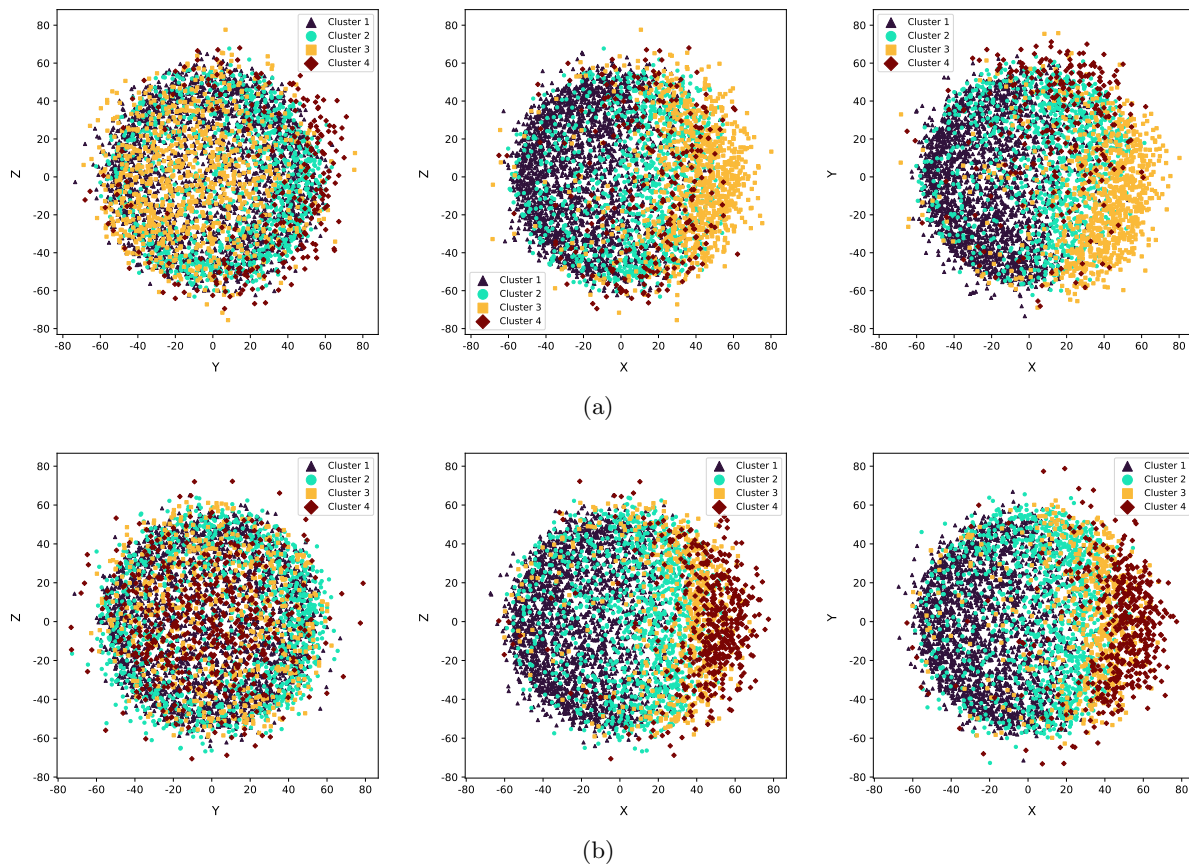
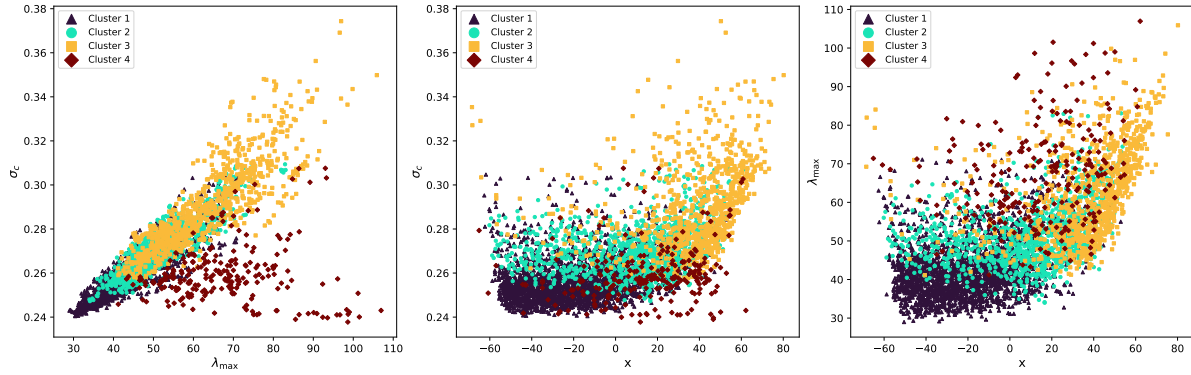
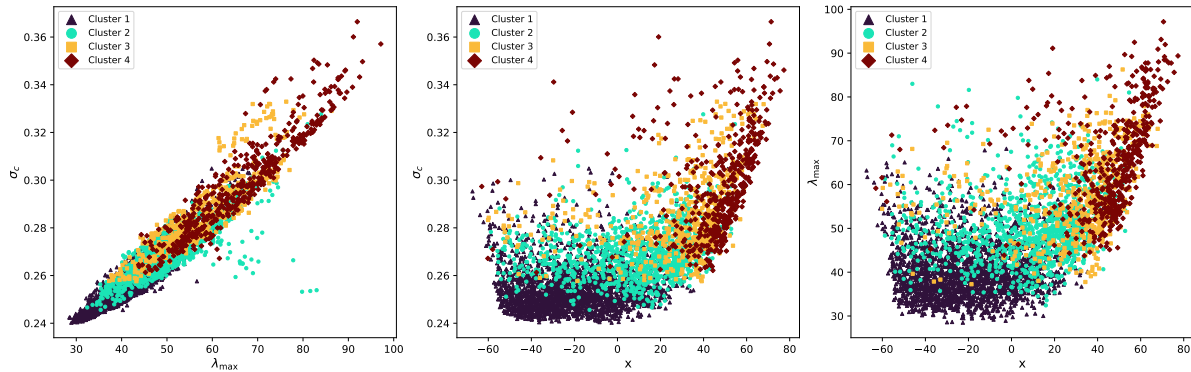


Figura C.25: MDS de los cúmulos para las correlaciones proyectadas de Guhr  $\mathcal{C}^B$  (método de covarianza, subfigura C.25(a)) y  $\mathcal{C}^L$  (método de correlación, subfigura C.25(b)), cuando  $k = 4$  y no se aplica el *power map*. El k-means se hace en el espacio  $N(N - 1)/2$ -dimensional, para luego obtenerse las coordenadas 3D con el MDS. Además, el sistema de coordenadas  $x, y, z$  del espacio reducido se alinea con los ejes de mayor varianza (es decir, los ejes principales), con el origen en el centro geométrico de la proyección.



(a)



(b)

Figura C.26: Diagramas de dispersión de la coordenada  $x$  del MDS, el mayor eigenvalor  $\lambda_{\max}$  y la desviación estándar  $\sigma_c$ . El código de color corresponde a los *clusters* obtenidos mediante el k-means para las correlaciones proyectadas de Guhr  $\mathcal{C}^B$  y  $\mathcal{C}^L$ .

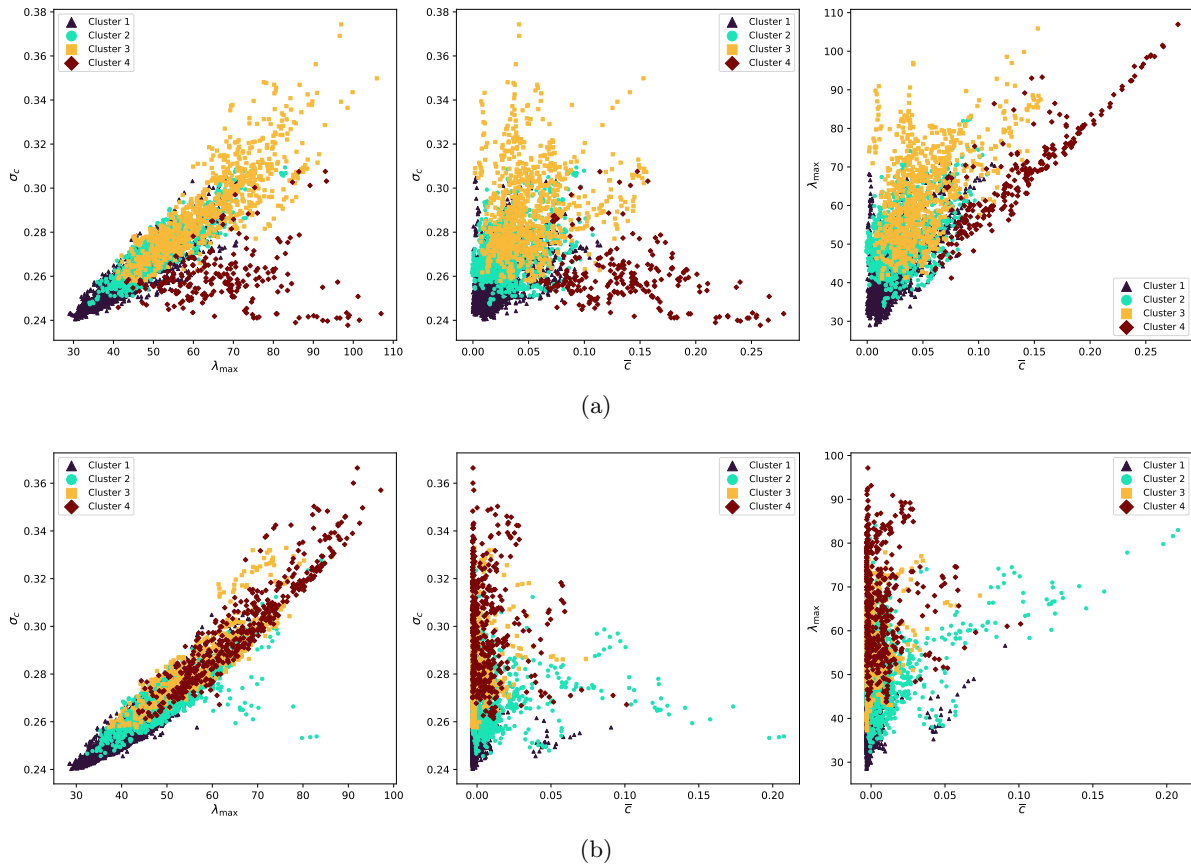


Figura C.27: Diagramas de dispersión de la correlación promedio  $\bar{c}$ , el mayor eigenvalor  $\lambda_{\text{máx}}$  y la desviación estándar  $\sigma_c$  de las correlaciones proyectadas de Guhr  $\mathcal{C}^{\text{B}}$  (subfigura C.27(a)) y  $\mathcal{C}^{\text{L}}$  (subfigura C.27(b)). El código de color corresponde a los *clusters* obtenidos mediante el k-means.



# Referencias

1. Regnault, J. *Calcul Des Chances et Philosophie de La Bourse* 1.<sup>a</sup> ed. (Mallet-Bachelier et Castel, Paris, 1863).
2. Quetelet, A. *Du système social et des lois qui le régissent* (Guillaumin et compagnie, Paris, 1848).
3. Jovanovic, F. y Schinckus, C. *Econophysics and Financial Economics: An Emerging Dialogue* 230 págs. ISBN: 978-0-19-020503-4 (Oxford University Press, New York City, 2017).
4. Bachelier, L. Théorie de La Spéculation. *Annales Scientifiques de l'École Normale Supérieure* **17**, 21-86. ISSN: 0012-9593, 1873-2151. (2024) (1900).
5. Chakraborti, A., Toke, I. M., Patriarca, M. y Abergel, F. Econophysics Review: I. Empirical Facts. *Quantitative Finance* **11**, 991-1012. ISSN: 1469-7688. (2023) (1 de jul. de 2011).
6. Requião da Cunha, C. *Introduction to Econophysics: Contemporary Approaches with Python Simulations* 1st edition. ISBN: 978-0-367-64845-9 978-0-367-65128-2 (CRC Press, Boca Raton, FL, 2022).
7. Brown, R. XXVII. A Brief Account of Microscopical Observations Made in the Months of June, July and August 1827, on the Particles Contained in the Pollen of Plants; and on the General Existence of Active Molecules in Organic and Inorganic Bodies. *The Philosophical Magazine* **4**, 161-173. ISSN: 1941-5850. (2024) (1 de sep. de 1828).
8. Richmond, P., Mimkes, J. y Hutzler, S. *Econophysics and Physical Economics* First edition. 243 págs. ISBN: 978-0-19-967470-1 (Oxford University Press, Oxford, 2013).
9. Einstein, A. Über Die von Der Molekularkinetischen Theorie Der Wärme Geforderte Bewegung von in Ruhenden Flüssigkeiten Suspensierten Teilchen. *Annalen der Physik* **322**, 549-560. ISSN: 1521-3889. (2024) (1905).
10. Von Smoluchowski, M. Zur Kinetischen Theorie Der Brownschen Molekularbewegung Und Der Suspensionen. *Annalen der Physik* **326**, 756-780. ISSN: 1521-3889. (2024) (1906).
11. Perrin, J. L'agitation Moléculaire et Le Mouvement Brownien. *Annales de Chimie et de Physique*, 1-114 (1909).
12. Hoel, P. G., Port, S. C. y Stone, C. J. *Introduction to Stochastic Processes* 203 págs. ISBN: 978-0-395-12076-7 (Houghton Mifflin, Boston, 1972).
13. Mazo, R. M. *Brownian Motion: Fluctuations, Dynamics, and Applications Oxford Science Publications* **112**. 289 págs. ISBN: 978-0-19-851567-8 (Clarendon Press ; Oxford University Press, Oxford : New York, 2002).

14. Paul, W. y Baschnagel, J. *Stochastic Processes: From Physics to Finance* ISBN: 978-3-319-00326-9 978-3-319-00327-6. (2023) (Springer International Publishing, Heidelberg, 2013).
15. Mantegna, R. N. y Stanley, H. E. *An Introduction to Econophysics: Correlations and Complexity in Finance* 148 págs. ISBN: 978-0-521-62008-6 (Cambridge University Press, Cambridge, UK; New York, 2000).
16. Herlin, P. *Finance: Le Nouveau Paradigme: Comprendre La Finance et l'économie Avec Mandelbrot, Taleb* 207 págs. ISBN: 978-2-212-54657-6 (Organisation, Paris, 2010).
17. Hammerstein, P. y Selten, R. en *Handbook of Game Theory with Economic Applications* 929-993 (Elsevier, 1994). ISBN: 978-0-444-89427-4. (2024).
18. Fisher, I. *Mathematical Investigations in the Theory for Value and Prices* Monograph (1892). 125 págs.
19. Bergersen, B. Statistical Mechanics of Utility and Equilibrium: Analogies between Economics and Statistical Physics. *Physics in Canada* **65**, 205-210 (2009).
20. Kolmogorov, A. N., Bharucha-Reid, A. T. y Morrison, N. *Foundations of the Theory of Probability* Second English edition, Dover edition. 84 págs. ISBN: 978-0-486-82159-7 (Dover Publications, Inc, Muneola, New York, 2018).
21. Mandelbrot, B. y Hudson, R. L. *The (Mis)Behavior of Markets: A Fractal View of Risk, Ruin, and Reward* ISBN: 978-0-465-00468-3 (Basic Books, 2004).
22. Caticha, A. y Giffin, A. Updating Probabilities. *AIP Conference Proceedings* **872**, 31-42. ISSN: 0094-243X. (2023) (29 de nov. de 2006).
23. Caticha, A. *Entropic Physics: Probability, Entropy, and the Foundations of Physics* (2023) ().
24. Flam, F. D. The Odds, Continually Updated. *The New York Times. Science*. ISSN: 0362-4331. (2023) (29 de sep. de 2014).
25. Jaynes, E. T. Information Theory and Statistical Mechanics I. *Physical Review* **106**, 620-630. (2023) (15 de mayo de 1957).
26. Jaynes, E. T. Information Theory and Statistical Mechanics II. *Physical Review* **108**, 171-190. (2023) (15 de oct. de 1957).
27. Bunge, M. *La Ciencia, Su Método y Su Filosofía* 1.<sup>a</sup> ed. ISBN: 968-39-0176-X (Nueva Imagen, 2000).
28. Popper, K. R. *Conjeturas y refutaciones: el desarrollo del conocimiento científico Paidós Básica* **7**. 513 págs. ISBN: 978-84-7509-146-4 (Paidós, Barcelona, 1991).
29. Karmeshu. *Entropy Measures, Maximum Entropy Principle, and Emerging Applications* ISBN: 978-3-540-36212-8 (Springer, Berlin, 2003).
30. Boccaletti, S., Pisarchik, A. N., Del Genio, C. I. y Amann, A. *Synchronization: From Coupled Systems to Complex Networks* 255 págs. ISBN: 978-1-107-05626-8 (Cambridge University Press, Cambridge, 2018).
31. Fabozzi, F. J. y Peterson Drake, P. *The Basics of Finance: An Introduction to Financial Markets, Business Finance, and Portfolio Management* ISBN: 978-0-470-60971-2 (Wiley, Hoboken, NJ, 2010).
32. *How to Calculate a Stock's Adjusted Closing Price* Investopedia. (2024).

33. Fabozzi, F. J. y Jones, F. J. *Foundations of Global Financial Markets and Institutions* Fifth edition. 1075 págs. ISBN: 978-0-262-03954-3 (MIT Press, Cambridge, Massachusetts, 2019).
34. Bouchaud, J.-P., Bonart, J., Donier, J. y Gould, M. *Trades, Quotes and Prices: Financial Markets Under the Microscope* 1.<sup>a</sup> ed. ISBN: 978-1-107-15605-0 978-1-316-65933-5. (2024) (Cambridge University Press, 22 de mar. de 2018).
35. Fama, E. F. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance* **25**, 383. ISSN: 00221082. JSTOR: [2325486](#). (2024) (mayo de 1970).
36. Kullmann, L., Kertész, J. y Mantegna, R. Identification of Clusters of Companies in Stock Indices via Potts Super-Paramagnetic Transitions. *Physica A: Statistical Mechanics and its Applications* **287**, 412-419. ISSN: 03784371. (2024) (dic. de 2000).
37. Slanina, F. *Essentials of Econophysics Modelling* 1. ed. 411 págs. ISBN: 978-0-19-929968-3 (Oxford Univ.Press, Oxford, 2014).
38. Mandelbrot, B. The Variation of Certain Speculative Prices. *The Journal of Business* **36**, 394. ISSN: 0021-9398, 1537-5374. JSTOR: [2350970](#). (2024) (ene. de 1963).
39. *Market Index: Definition, How Indexing Works, Types, and Examples* Investopedia. (2023).
40. Münnix, M. C. *et al.* Identifying States of a Financial Market. *Scientific Reports* **2**, 644. ISSN: 2045-2322. (2022) (1 10 de sep. de 2012).
41. Pharasi, H. K. *et al.* Identifying Long-Term Precursors of Financial Market Crashes Using Correlation Patterns. *New Journal of Physics* **20**, 103041. ISSN: 1367-2630. (2023) (oct. de 2018).
42. Pharasi, H. K., Seligman, E., Sadhukhan, S., Majari, P. y Seligman, T. H. Dynamics of Market States and Risk Assessment. *Physica A: Statistical Mechanics and its Applications* **633**, 129396. ISSN: 0378-4371. (2024) (1 de ene. de 2024).
43. May, R. M. Will a Large Complex System Be Stable? *Nature* **238**, 413-414. ISSN: 1476-4687. (2023) (5364 ago. de 1972).
44. Fuchs, A. *Nonlinear Dynamics in Complex Systems: Theory and Applications for the Life-, Neuro- and Natural Sciences* ISBN: 978-3-642-33551-8 978-3-642-33552-5. (2024) (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013).
45. Arnold, V. I., Vogtmann, K. y Weinstein, A. *Mathematical Methods of Classical Mechanics* 2nd ed. *Graduate Texts in Mathematics* **60**. ISBN: 978-0-387-96890-2 978-5-409-68900-1 (Springer-Verlag, New York Berlin Paris [etc.], 1989).
46. Wigner, E. P. Characteristic Vectors of Bordered Matrices With Infinite Dimensions. *Annals of Mathematics* **62**, 548-564. ISSN: 0003-486X. JSTOR: [1970079](#). (2023) (1955).
47. Mehta, M. L. *Random Matrices* 3.<sup>a</sup> ed. *Pure and Applied Mathematics* v. **142**. ISBN: 978-0-12-088409-4 (Elsevier/Academic Press, Amsterdam San Diego, CA, 2004).
48. Weisstein, E. W. *Wigner's Semicircle Law* (2023).
49. Majumdar, S. N. y Schehr, G. Top Eigenvalue of a Random Matrix: Large Deviations and Third Order Phase Transition. *Journal of Statistical Mechanics: Theory and Experiment* **2014**, P01012. ISSN: 1742-5468. (2023) (ene. de 2014).
50. Kuter, K. *7.2: Sample Variance* Statistics LibreTexts. (2023).
51. Muirhead, R. J. *Aspects of Multivariate Statistical Theory* 673 págs. ISBN: 978-0-471-76985-9 (Wiley-Interscience, Hoboken, NJ, 2005).



52. Bun, J., Bouchaud, J.-P. y Potters, M. Cleaning Large Correlation Matrices: Tools from Random Matrix Theory. *Physics Reports* **666**, 1-109. ISSN: 03701573. (2023) (ene. de 2017).
53. Laloux, L., Cizeau, P., Bouchaud, J.-P. y Potters, M. Noise Dressing of Financial Correlation Matrices. *Physical Review Letters* **83**, 1467-1470. (2023) (16 de ago. de 1999).
54. Plerou, V., Gopikrishnan, P., Rosenow, B., Nunes Amaral, L. A. y Stanley, H. E. Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series. *Physical Review Letters* **83**, 1471-1474. (2023) (16 de ago. de 1999).
55. Plerou, V. *et al.* Random Matrix Approach to Cross Correlations in Financial Data. *Physical Review E* **65**, 066126. (2023) (27 de jun. de 2002).
56. Wishart, J. The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population. *Biometrika* **20A**, 32-52. ISSN: 0006-3444. JSTOR: [2331939](#). (2023) (1928).
57. Guhr, T., Müller-Groeling, A. y Weidenmüller, H. A. Random-Matrix Theories in Quantum Physics: Common Concepts. *Physics Reports* **299**, 189-425. ISSN: 0370-1573. (2024) (1 de jun. de 1998).
58. Weidenmüller, H. A. *Random Matrix Theory: Introduction and Overview* en *Proceedings of The IV Wigner Symposium* The IV Wigner Symposium (eds. Atakishiyev, N. M., Seligman, T. H. y Wolf, K. B.) (World Scientific, Guadalajara, México, 1996), 274-289. ISBN: 978-981-02-2663-3.
59. Kota, V. *Embedded Random Matrix Ensembles in Quantum Physics* ISBN: 978-3-319-04566-5 978-3-319-04567-2. (2024) (Springer International Publishing, Cham, 2014).
60. Marčenko, V. A. y Pastur, L. A. Distribution of Eigenvalues for Some Sets of Random Matrices. *Mathematics of the USSR-Sbornik* **1**, 457-483. ISSN: 0025-5734. (2023) (30 de abr. de 1967).
61. Vinayak y Pandey, A. Correlated Wishart Ensembles and Chaotic Time Series. *Physical Review E* **81**, 036202. (2023) (4 de mar. de 2010).
62. Füredi, Z. y Komlós, J. The Eigenvalues of Random Symmetric Matrices. *Combinatorica* **1**, 233-241. ISSN: 1439-6912. (2024) (1 de sep. de 1981).
63. Noh, J. D. Model for Correlations in Stock Markets. *Physical Review E* **61**, 5981-5982. (2024) (1 de mayo de 2000).
64. Recher, C., Kieburg, M. y Guhr, T. Eigenvalue Densities of Real and Complex Wishart Correlation Matrices. *Physical Review Letters* **105**, 244101. ISSN: 0031-9007, 1079-7114. (2024) (6 de dic. de 2010).
65. Wirtz, T., Kieburg, M. y Guhr, T. Asymptotic Coincidence of the Statistics for Degenerate and Non-Degenerate Correlated Real Wishart Ensembles. *Journal of Physics A: Mathematical and Theoretical* **50**, 235203. ISSN: 1751-8121. (2024) (mayo de 2017).
66. Schäfer, R. y Guhr, T. Local Normalization: Uncovering Correlations in Non-Stationary Financial Time Series. *Physica A: Statistical Mechanics and its Applications* **389**, 3856-3865. ISSN: 0378-4371. (2023) (15 de sep. de 2010).
67. Anderson, T. W. *An Introduction to Multivariate Statistical Analysis* 3rd ed. 721 págs. ISBN: 978-0-471-36091-9 (Wiley-Interscience, Hoboken, N.J., 2003).
68. Pandey, A. y Ghosh, S. Skew-Orthogonal Polynomials and Universality of Energy-Level Correlations. *Physical Review Letters* **87**, 024102. (2023) (21 de jun. de 2001).

69. Ghosh, S. y Pandey, A. Skew-Orthogonal Polynomials and Random-Matrix Ensembles. *Physical Review E* **65**, 046221. (2023) (8 de abr. de 2002).
70. Vinayak y Seligman, T. H. Time Series, Correlation Matrices and Random Matrix Models. *AIP Conference Proceedings* **1575**, 196. ISSN: 0094-243X. (2022) (17 de feb. de 2015).
71. Kapetanios, G. Testing for Strict Stationarity in Financial Variables. *Journal of Banking & Finance* **33**, 2346-2362. ISSN: 0378-4266. (2023) (1 de dic. de 2009).
72. Guhr, T. Non-Stationarity in Financial Markets: Dynamics of Market States Versus Generic Features. *Acta Physica Polonica B* **46**, 1625. ISSN: 0587-4254, 1509-5770. (2023) (2015).
73. Guhr, T. y Kälber, B. A New Method to Estimate the Noise in Financial Correlation Matrices. *Journal of Physics A: Mathematical and General* **36**, 3009. ISSN: 0305-4470. (2023) (mar. de 2003).
74. Vinayak, Schäfer, R. y Seligman, T. H. Emerging Spectra of Singular Correlation Matrices under Small Power-Map Deformations. *Physical Review E* **88**, 032115. (2023) (10 de sep. de 2013).
75. Vinayak, Prosen, T., Buča, B. y Seligman, T. H. Spectral Analysis of Finite-Time Correlation Matrices near Equilibrium Phase Transitions. *Europhysics Letters* **108**, 20006. ISSN: 0295-5075. (2023) (oct. de 2014).
76. Brody, T. A. *et al.* Random-Matrix Physics: Spectrum and Strength Fluctuations. *Reviews of Modern Physics* **53**, 385-479. (2024) (1 de jul. de 1981).
77. Tao, T. y Vu, V. The Wigner-Dyson-Mehta Bulk Universality Conjecture for Wigner Matrices. *Electronic Journal of Probability* **16**, 2104-2121. ISSN: 1083-6489, 1083-6489. (2024) (none ene. de 2011).
78. Malevergne, Y. y Sornette, D. Collective Origin of the Coexistence of Apparent Random Matrix Theory Noise and of Factors in Large Sample Correlation Matrices. *Physica A: Statistical Mechanics and its Applications* **331**, 660-668. ISSN: 0378-4371. (2024) (15 de ene. de 2004).
79. Izenman, A. J. *Modern Multivariate Statistical Techniques* (reds. Casella, G., Fienberg, S. y Olkin, I.) ISBN: 978-0-387-78188-4 978-0-387-78189-1. (2023) (Springer New York, New York, NY, 2008).
80. De Lange, O. L. y Pierrus, J. *Solved Problems in Classical Mechanics: Analytical and Numerical Solutions with Comments* 599 págs. ISBN: 978-0-19-958252-5 978-0-19-958251-8 (Oxford University Press, Oxford ; New York, 2010).
81. Drożdż, S., Grümmner, F., Górski, A. Z., Ruf, F. y Speth, J. Dynamics of Competition between Collectivity and Noise in the Stock Market. *Physica A: Statistical Mechanics and its Applications* **287**, 440-449. ISSN: 0378-4371. (2024) (1 de dic. de 2000).
82. Izrailev, F. M. Simple Models of Quantum Chaos: Spectrum and Eigenfunctions. *Physics Reports* **196**, 299-392. ISSN: 0370-1573. (2024) (1 de nov. de 1990).
83. Casella, G. y Berger, R. L. *Statistical Inference* 2nd ed. 660 págs. ISBN: 978-0-534-24312-8 (Thomson Learning, Australia ; Pacific Grove, CA, 2002).
84. Heckens, A. J., Krause, S. M. y Guhr, T. Uncovering the Dynamics of Correlation Structures Relative to the Collective Market Motion. *Journal of Statistical Mechanics: Theory and Experiment* **2020**, 103402 (oct. de 2020).

85. Lloyd, S. P. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory* **28**, 129-137. ISSN: 1557-9654 (mar. de 1982).
86. MacQueen, J. en *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics* 281-298 (University of California Press, 1 de ene. de 1967). (2023).
87. Jin, X. y Han, J. en *Encyclopedia of Machine Learning* (eds. Sammut, C. y Webb, G. I.) 563-564 (Springer US, Boston, MA, 2010). ISBN: 978-0-387-30164-8. (2023).
88. Chakraborti, A. *et al.* *Characterization of Catastrophic Instabilities: Market Crashes as Paradigm* arXiv: [1801.07213 \[physics, q-fin\]](https://arxiv.org/abs/1801.07213). (2023). Prepublicado.
89. Johnson, R. A. y Wichern, D. W. *Applied Multivariate Statistical Analysis* 6. ed., new internat. ed. 770 págs. ISBN: 978-1-292-02494-3 (Pearson Education Limited, Harlow, 2014).
90. *2.2. Manifold Learning* scikit-learn. (2024).
91. Eckmann, J.-P., Kamphorst, S. O. y Ruelle, D. Recurrence Plots of Dynamical Systems. *Europhysics Letters* **4**, 973. ISSN: 0295-5075. (2024) (nov. de 1987).
92. Chetalova, D., Schäfer, R. y Guhr, T. Zooming into Market States. *Journal of Statistical Mechanics: Theory and Experiment* **2015**, P01029. ISSN: 1742-5468. (2023) (ene. de 2015).
93. Rinn, P., Stepanov, Y., Peinke, J., Guhr, T. y Schäfer, R. Dynamics of Quasi-Stationary Systems: Finance as an Example. *Europhysics Letters* **110**, 68003. ISSN: 0295-5075. (2024) (jul. de 2015).
94. MacMahon, M. y Garlaschelli, D. Community Detection for Correlation Matrices. *Physical Review X* **5**, 021006. (2023) (14 de abr. de 2015).
95. Park, H.-J. y Friston, K. Structural and Functional Brain Networks: From Connections to Cognition. *Science* **342**, 1238411. (2024) (nov. de 2013).
96. Gast, H. *et al.* Epileptic Seizures as Condensed Sleep: An Analysis of Network Dynamics from Electroencephalogram Signals. *Journal of Sleep Research* **23**, 270-275. ISSN: 1365-2869. (2024) (2014).
97. Daniel Arzate-Mena, J. *et al.* Stationary EEG Pattern Relates to Large-Scale Resting State Networks – An EEG-fMRI Study Connecting Brain Networks across Time-Scales. *NeuroImage* **246**, 118763. ISSN: 1053-8119. (2024) (1 de feb. de 2022).
98. Martínez-Ramos, M. M., Majari, P. y Vyas, M. *Correlations beyond the Trend: A Financial Market Analysis (in Prep)* 2024.
99. Martínez-Ramos, M. M., Vyas, M., Majari, P. y Seligman, T. H. COVID Anomaly in the Correlation Analysis of S&P 500 Market States. *PLOS ONE* **19**, e0301238. ISSN: 1932-6203. (2024) (18 de abr. de 2024).
100. Frazier, L. *The Coronavirus Crash Of 2020, And The Investing Lesson It Taught Us* Forbes. (2024).
101. Heckens, A. J. y Guhr, T. New Collectivity Measures for Financial Covariances and Correlations. *Physica A: Statistical Mechanics and its Applications* **604**, 127704. ISSN: 0378-4371. (2023) (15 de oct. de 2022).
102. Heckens, A. J. y Guhr, T. A New Attempt to Identify Long-Term Precursors for Endogenous Financial Crises in the Market Correlation Structures. *Journal of Statistical Mechanics: Theory and Experiment* **2022**, 043401. ISSN: 1742-5468. (2023) (abr. de 2022).

103. Martínez-Ramos, M. M., Majari, P., Cruz-Hernández, A. R., Pharasi, H. K. y Vyas, M. Coarse Graining Correlation Matrices According to Macrostructures: Financial Markets as a Paradigm. *Physica Scripta*. ISSN: 0031-8949, 1402-4896. (2024) (25 de jun. de 2024).
104. Kutner, R. *et al.* Econophysics and Sociophysics: Their Milestones & Challenges. *Physica A: Statistical Mechanics and its Applications* **516**, 240-253. ISSN: 0378-4371. (2023) (15 de feb. de 2019).
105. An, L. *et al.* Challenges, Tasks, and Opportunities in Modeling Agent-Based Complex Systems. *Ecological Modelling* **457**, 109685. ISSN: 0304-3800. (2024) (1 de oct. de 2021).
106. Bollobás, B. *Random Graphs* 2nd ed. *Cambridge Studies in Advanced Mathematics* **73**. 498 págs. ISBN: 978-0-521-80920-7 978-0-521-79722-1 (Cambridge University Press, Cambridge ; New York, 2001).
107. Nishimori, H. *Statistical Physics of Spin Glasses and Information Processing: An Introduction International Series of Monographs on Physics* **111**. 243 págs. ISBN: 978-0-19-850940-0 978-0-19-850941-7 (Oxford University Press, Oxford ; New York, 2001).
108. Dotsenko, V. *Introduction to the Replica Theory of Disordered Statistical Systems* ISBN: 978-0-521-77340-9 (Cambridge university press, Cambridge (GB), 2001).
109. Vyas, M., Guhr, T. y Seligman, T. H. Multivariate Analysis of Short Time Series in Terms of Ensembles of Correlation Matrices. *Scientific Reports* **8**, 1-12 (oct. de 2018).
110. *Referencia del Lenguaje Python* Python documentation. (2024).
111. Harris, C. R. *et al.* Array Programming with NumPy. *Nature* **585**, 357-362. ISSN: 1476-4687. (2024) (sep. de 2020).
112. McKinney, W. *Data Structures for Statistical Computing in Python* en. Python in Science Conference (Austin, Texas, 2010), 56-61. (2024).
113. Team, T. P. D. *Pandas-Dev/Pandas: Pandas ver. v2.1.1*. Zenodo, 20 de sep. de 2023. (2024).
114. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90-95. ISSN: 1521-9615. (2024) (2007).
115. Team, T. M. D. *Matplotlib: Visualization with Python ver. v3.8.0*. Zenodo, 15 de sep. de 2023. (2024).
116. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261-272. ISSN: 1548-7091, 1548-7105. (2024) (2 de mar. de 2020).
117. Pedregosa, F. *et al.* Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825-2830. ISSN: 1533-7928. (2024) (2011).
118. *Introduction — KDEpy 1.1.9 Documentation* (2024).
119. *Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data* (eds. Ivezić, Ž., Connolly, A. J., VanderPlas, J. y Gray, A.) 540 págs. ISBN: 978-0-691-15168-7 (Princeton Univ. Press, Princeton, N.J., 2014).
120. Wasserman, L. *All of Statistics: A Concise Course in Statistical Inference* 442 págs. ISBN: 978-0-387-40272-7 (Springer, New York, 2004).
121. Silverman, B. W. *Density Estimation for Statistics and Data Analysis* 1.<sup>a</sup> ed. ISBN: 978-1-315-14091-9. (2024) (Routledge, 19 de feb. de 2018).
122. Scott, D. W. y Terrell, G. R. Biased and Unbiased Cross-Validation in Density Estimation. *Journal of the American Statistical Association* **82**, 1131-1146. ISSN: 0162-1459, 1537-274X. (2024) (dic. de 1987).

123. Scott, D. W. *Multivariate Density Estimation: Theory, Practice, and Visualization* 1.<sup>a</sup> ed. ISBN: 978-0-471-69755-8 978-1-118-57557-4. (2024) (Wiley, 27 de mar. de 2015).
124. Heidenreich, N.-B., Schindler, A. y Sperlich, S. Bandwidth Selection for Kernel Density Estimation: A Review of Fully Automatic Selectors. *AStA Advances in Statistical Analysis* **97**, 403-433. ISSN: 1863-818X. (2024) (1 de oct. de 2013).
125. Bashtannyk, D. M. e Hyndman, R. J. Bandwidth Selection for Kernel Conditional Density Estimation. *Computational Statistics & Data Analysis* **36**, 279-298. ISSN: 0167-9473. (2024) (28 de mayo de 2001).
126. Tommy *et al.* *Tommyod/KDEpy: V1.1.8* ver. v1.1.8. Zenodo, 9 de oct. de 2023. (2024).
127. Doane, D. P. Aesthetic Frequency Classifications. *The American Statistician* **30**, 181-183. ISSN: 0003-1305, 1537-2731. (2024) (nov. de 1976).
128. Dieng, M. y Tracy, C. A. en *Random Matrices, Random Processes and Integrable Systems* (ed. Harnad, J.) 443-507 (Springer New York, New York, NY, 2011). ISBN: 978-1-4419-9513-1 978-1-4419-9514-8. (2023).
129. Izenman, A. J. Random Matrix Theory and Its Applications. *Statistical Science* **36**. ISSN: 0883-4237. (2024) (1 de ago. de 2021).
130. Bai, Z., Fang, Z. y Liang, Y.-C. *Spectral Theory of Large Dimensional Random Matrices and Its Applications to Wireless Communications and Finance: Random Matrix Theory and Its Applications* 220 págs. ISBN: 978-981-4579-05-6 (World Scientific Publishing Company, Singapore, 2014).
131. Bai, Z. D., Silverstein, J. W. y Yin, Y. Q. A Note on the Largest Eigenvalue of a Large Dimensional Sample Covariance Matrix. *Journal of Multivariate Analysis* **26**, 166-168. ISSN: 0047-259X. (2024) (1 de ago. de 1988).
132. Bai, Z. D. y Yin, Y. Q. Necessary and Sufficient Conditions for Almost Sure Convergence of the Largest Eigenvalue of a Wigner Matrix. *The Annals of Probability* **16**, 1729-1741. ISSN: 0091-1798, 2168-894X. (2024) (oct. de 1988).
133. Bai, Z. D. y Yin, Y. Q. Limit of the Smallest Eigenvalue of a Large Dimensional Sample Covariance Matrix. *The Annals of Probability* **21**, 1275-1294. ISSN: 0091-1798, 2168-894X. (2024) (jul. de 1993).