
UNIVERSIDAD DE SONORA

División de Ciencias Exactas y Naturales

Departamento de Investigación en Física

# Métodos de la Física Estadística en la Economía Conductual.

TESIS

Para obtener el grado de: Doctor en Ciencias (Física)
Presenta: Andrés García Medina
Director de tesis: Dr. Leonidas Sandoval
Co-director: Dr. Efraín Urrutia Bañuelos

Miembros del comité tutorial:
Dr. Santos Jesús Castillo, Dr. Mario Flores Acosta y Dr. Rafael García Gutiérrez

Hermosillo, Sonora, México

11 de agosto de 2016

# Universidad de Sonora

Repositorio Institucional UNISON


# Aprobación de Tesis

"El saber de mis hijos hará mi grandeza"

# UNIVERSIDAD DE SONORA

## DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES

Departamento de Investigación en Física

*Programa de Posgrado en Ciencias (Física)*

El Jurado de Tesis abajo firmante, APRUEBA POR UNANIMIDAD el presente manuscrito de la TESIS intitulada *"Métodos de la Física Estadística en la Economía Conductual"* presentado por el **M.C. Andrés García Medina** bajo la dirección del Dr. Leonidas Sandoval Junior y la Codirección del Dr. Efraín Urrutia Bañuelos, dando así cumplimiento a la fase escrita del proceso de titulación del Programa de Doctorado en Ciencias (Física) de la Universidad de Sonora.

**Dr. Moisés Martínez Mares**
Profesor Titular C
Departamento de Física
Universidad Autónoma Metropolitana
Unidad Iztapalapa

**Dr. Ricardo L. Mansillas Corona**
Investigador Titular C
Centro de Investigaciones Interdisciplinarias
en Ciencias y Humanidades
Universidad Nacional Autónoma de México

**Dr. Carlos G. Pacheco González**
Profesor Asociado 2C
Departamento de Matemáticas
Centro de Investigación y de estudios
Avanzados

**Dr. Carlos Lizárraga Celaya**
Maestro de Tiempo Completo
Departamento de Física
Universidad de Sonora

**Dr. Efraín Urrutia Bañuelos**
Maestro de Tiempo Completo
Departamento de Investigación en Física
Universidad de Sonora

**Vo.Bo. Dra. Susana Álvarez García**
Coordinadora del Programa de
Posgrado en Ciencias (Física)
Universidad de Sonora

El saber de mis hijos hará mi grandeza
Posgrado en Ciencias (Física)
Departamento de Investigación en Física

*En Hermosillo, Sonora a 09 de Diciembre de 2016*

# Productos Obtenidos

## Artículos

- A. García. *Global financial indices and twitter sentiment: A random matrix theory approach*. Physica A **461** 509, 2016. (Publicado)
- A. García. *EL uso de twitter en el análisis financiero: aproximación desde la econofísica*. Actas de Economía y Complejidad II. Capítulo de Libro. Editor Ricardo Mansilla, Editorial UNAM, Septiembre 2016. (Aceptado)

## Congresos

- Ponencia en *Econophysics colloquium 2015*. Instituto de Estudios Económicos, Charles University, Praga, Republica Checa. Del 14 al 16 de Septiembre de 2015.
- Ponencia en *Seminario de Economía y Complejidad*. Centro de Investigaciones Interdisciplinarias en Ciencias y Humanidades, UNAM, Ciudad de México. 15 de Marzo de 2016.
- Ponencia en *Econophysics colloquium 2016*. International Centre for Theoretical Physics, Sao Paulo, Brazil. Del 27 al 29 de Julio de 2016.

## Estancias

- Instituto de Ciencias Físicas, UNAM, Cuernavaca, Morelos. Dr. Thomas Seligman, Investigador Emérito. Del 22 de Octubre al 14 de Diciembre de 2012.
- Centro de Investigaciones Interdisciplinarias en Ciencias y Humanidades, UNAM, Ciudad de México. Dr. Ricardo Mansilla, Coordinador del departamento de Ciencia y Tecnología. Del 12 de Enero al 18 de Marzo del 2016.
- Insper, Institute of Education and Research, Sao Paulo, Brazil. Dr. Leonidas Sandoval Junior, Profesor - Investigador (Director de Tesis). Del 1 de Mayo al 30 de Julio de 2016.

*Dedicado a*
*Rebeca, mi madre, y los fridillos*

# Agradecimientos

Primeramente, quiero agradecer a Rebeca Moreno, por acompañarme durante esta etapa de mi vida, y motivarme a no desistir de este proyecto, es realmente debido a su apoyo emocional que he logrado concluir esta tesis. Agradezco asimismo a mi familia, ya que aún en la distancia siempre estuvieron presentes en todo momento.

Agradezco a las personas que me han guiado durante mi desarrollo científico. A Leonidas Sandoval Junior y Efrain Urrutia, por aceptarme como su estudiante y creer en mi, aún y cuando no me conocían en persona. A Susana Álvarez por darme la oportunidad de desarrollar y concluir mi trabajo de tesis, aún y cuando mi situación académica era de lo más atípica. A los miembros de mi comité de seguimiento: Santos Jesús Castillo, Mario Flores, y Rafael García; sus observaciones y comentarios lograron que pudiera situar y aclarar los fundamentos de mi trabajo de investigación dentro de los margenes de la física estadística. Agradezco a los investigadores que me han ayudado e influenciado con sus comentarios y sugerencia a aclarar y desarrollar puntos cruciales de mi tesis: Thomas Seligman, Vinayak Vinayak, Sunil Kumar, Ricardo Mansilla, Moisés Martínez. Así como los compañeros que he encontrado a lo largo de este proyecto, y que de una u otra manera me apoyaron a realizar mi trabajo: Ángel Martínez, David de la Rosa, Anupam Mukherjee, Ángel Reyes, Diana Kusters, Nahuel Moreno, Estefania Duran, Manuel Sanchez, Braulio Rojas, Alejandro Mitrani, Leonardo Baez, Moisés Chávez, Alberto Riesgo, Charly, entro muchos otros que harían la lista interminable, pero no por ello menos importantes.

Agradezco a los miembros y trabajadores del Departamento de Investigación en Física, y de la División de Ciencias Exactas y Naturales de la Universidad de Sonora, por el apoyo y facilidades presentadas para llevar a cabo mi formación académica. Por último, agradezco a CONACYT por el apoyo económico aportado para realizar esta tesis a través de la beca 339264.

# Resumen

Se han utilizado distintas técnicas matemáticas provenientes de la física estadística, los sistemas complejos, y teoría de la información, para analizar datos textuales de Twitter y *The New York Times* (NYT) en el contexto de los mercados financieros globales, con implicaciones en la economía conductual. Para esto, se han analizado dos periodos de tiempo entre 2014 y 2016. El primer periodo fue de 7 meses para Twitter, y el segundo de 10 meses para el caso de NYT, considerado los retornos de 20 índices financieros globales para comparar los resultados. La información textual se logro extraer mediante el ensamblaje de distintos lenguajes de programación, construyendo series de tiempo de polaridad mediante dos técnicas distintas de análisis de sentimiento. Mediante el análisis de taxonomía y jerarquización fue posible visualizar ciertas estructuras de correlación geográficas de los datos empíricos. Sin embargo, los resultados más interesantes se observan al aplicar técnicas modernas de la Teoría de Matrices Aleatorias (RMT), revelándonos que existen correlaciones verdaderas entre los indices financieros, las polaridades y la mezcla entre ambos indicadores en ambos periodos de estudio. Además, se encontró una buena concordancia entre el comportamiento temporal de los eigenvalores extremos de los retornos y polaridades, con resultados similares para el cociente de participación inversa en los dos periodos de estudio, lo cual nos da información acerca de la emergencia de factores comunes en la información financiera global, sin importar si estamos utilizando polaridades o retornos como fuente de datos. Finalmente, el test de causalidad de Granger y el análisis de transferencia de entropía nos revelan que la información fluye de las polaridades hacia los retornos para ciertos valores críticos en ambos casos. Nuestros resultados sugieren que al utilizar las polaridades de Twitter y NYT como un nuevo indicador financiero, proveen de información relevante acerca del comportamiento colectivo e incluso individual de los índices financieros globales. Esto genera una fuerte y novedosa evidencia en contra de la hipótesis de mercado eficiente, y apoya la tendencia de la economía conductual, en la que los precios de los mercados se ven afectados por las decisiones irracionales de los inversionistas, siendo estos influenciados por la tendencia de las noticias y redes sociales.

# Abstract

Has been used different mathematical techniques from statistical physics, complex systems, and information theory to analyze textual data of Twitter and *The New York Times* in the context of global financial markets, whose results have implications in the field of behavioral finance. For this purpose, has been analyzed two periods of time between 2014 and 2016. The first period is of 7 months for Twitter, and the second period of 10 months for NYT, where also has been taking into account the returns of 20 global financial indices as comparative values. The textual information was extracted by assembly various programming languages, and was constructed polarity time series trough two different sentiment analysis techniques. It was possible to visualize some geographical correlation structures from the empirical data trough taxonomy and hierarchical analysis. However, the most interesting results are observed when applying the very new techniques from Random Matrix Theory (RMT), showing up true correlation between returns, polarities, and the mixture between them, in both periods of study. Furthermore, it has been found a good agreement between temporal behavior of the extreme eigenvalues of returns and polarities, with similar results for the inverse participation ration (IPR) in both cases of study, which give us information about the emergence of common factors in the global financial information, regardless if we are using polarities or returns as data source. Finally, Granger causality test and transfer entropy analysis reveal that information flows from polarities to returns for certain critical values in both studies. Our results suggest that using polarity as a new financial indicator provide of useful information about collective and even individual behavior of global financial indices. This builds a strong and novel evidence against the efficient market hypothesis, and supporting the school of behavioral finance where the market prices are affected by the irrational decisions of investors, which are influenced for trending news and social networks sources.

# Índice general

# Índice de cuadros

# Índice de figuras

*"Today a person can become rich or poor without doing anything, without lifting a finger, without an occurrence of nature taking place, without anyone giving anyone anything, or physically robbing anything. Price fluctuations are like secret movements directed by an invisible agency behind the back of society, causing continuous shifts and fluctuations in the distribution of social wealth. This movement is observed as atmospheric pressure read on a barometer, or temperature on a thermometer. And yet commodity prices and their movements manifestly are human affairs and not black magic..."*

Rosa de Luxemburgo

# Capítulo 1

# Introducción

Durante los últimos años ha surgido una enorme contribución de la física teórica hacia el entendimiento de los sistemas económicos, dando lugar a la emergencia de un nuevo campo de estudio denominado *Econofísica* [1, 2, 3]. Dentro de esta área, entender la estructura de las correlaciones entre diferentes mercados financieros es una de las lineas de investigación que mas rápidamente crece debido a la gran importancia en el contexto de la optimización de portafolios [4].

Un nuevo enfoque para entender este tipo de correlaciones viene de la Teoría de Matrices Aleatorias (RMT, por sus siglas en inglés). Muchos fenómenos de la física teórica han sido resueltos exitosamente utilizando el formalismo de RMT, pero no fue hasta la aparición de los trabajos casi simultáneos de Stanley et. al.[5] y Bouchoud et. al.[6] que se incrementó considerablemente la cantidad de estudios dedicados a entender la estructura de los mercados financieros a través de la aplicación de métodos provenientes de RMT [7, 8, 9, 10, 11]. En el contexto de los indices financieros globales, existen investigaciones desde este marco de referencia para entender efectos de la globalización, así como de correlaciones de largo alcance[12, 13]. Asimismo, Sandoval y Franca[14] han analizado correlaciones globales en tiempos de crisis, mientras que Kumar y Deo[15] han estudiado las propiedades de las redes complejas de índices mundiales durante la crisis financiera de 2008.

Por otro lado, la influencia de las noticias financieras y de los redes sociales no han sido exploradas exhaustivamente debido a la Hipótesis del Mercado Eficiente (EMH, por sus siglas en inglés). De acuerdo a la EMH el precio de la acción de un mercado dado

instantáneamente incorpora toda la información disponible del mercado, y su valor no depende del precio en el pasado[16]. No obstante, recientemente una serie de trabajos han comenzado a investigar la influencia de los redes sociales en los movimientos de los mercados [17, 18, 19, 20, 21], mostrando que la la información extraída de Twitter, Google Trends, y revistas financieras dan indicaciones tempranas que pueden ayudar a predecir cambios en la bolsa de valores. Estos nuevos resultados están construyendo un fuerte sustento en contra del tan aceptado paradigma de mercado eficiente, apoyando la aproximación de la economía conductual [22]. En ese sentido, los resultados de este trabajo intentan mostrar una evidencia más en contra de este paradigma.

La aproximación que se seguirá aquí para analizar los mercados financieros globales se sustenta en los resultados de RMT y la teoría de la información. Pero ademas a manera ilustrativa, se muestran resultados relacionados con la teoría de redes complejas para el análisis de la taxonomía y jerarquización de los indicadores financieros mediante la matriz de correlación, así como de la econometría para estudiar la intensidad de la causalidad de los series de tiempo de polaridad sobre las de retornos. Asimismo, en este estudio se incorpora la influencia de la red social *Twitter* y de noticias del *The New York Times* (NYT) mediante el análisis de sentimiento, el cual se basa en asignar un puntaje a los datos textuales de acuerdo a su contenido. Aquí se utilizó la polaridad como medida del estado de ánimo, relacionado el puntaje obtenido con el precio al cierre de las bolsas de valores estudiadas. Cabe resaltar que este es el primer trabajo en la literatura que muestra el uso la Teoría de Matrices Aleatorias en el estudio de los mercados financiero globales mediante series de tiempo de polaridad.

Este trabajo está organizado como sigue. En el capítulo 2 se describen brevemente las corrientes económicas donde se sitúan los resultados de este trabajo. En el capítulo 3 se describen los datos analizados, así como la metodología para extraer la colección pública de *tweets* y las noticias del NYT. En el capítulo 4 se explica como se llevo a cabo el análisis de sentimiento, y se describe el método seguido para construir las series de tiempo de polaridad. En el capítulo 5 se muestran los resultados y análisis de los mismos, comenzando con la construcción de la matriz de correlación, explicando enseguida como transformarla en una red compleja que nos permita determinar la taxonomía y jerarquización de los indicadores financieros involucrados. En el capítulo 6 se presenta un exhaustivo estudio de los resultados más relevante y recientes de la teoría de matrices aleatorias en el contexto del análisis multivariante. En el capítulo 7 se explican los fundamentos y resultados de los métodos de causalidad, desde el test de Granger hasta la Transferencia de Entropía.

Finalmente en el capítulo 8 se da una conclusión general del trabajo.

# Capítulo 2

# Marco Conceptual desde la Economía

Nuestro método de análisis se desarrolla desde el marco conceptual de la física estadísticas y de los sistemas complejos, ya que intentamos descubrir propiedades emergentes que escapan a los analistas financieros en econometría, sin embargo cuando se intentan crear paralelismos entre la física estadística y los mercados financieros, una cuestión importante que se debe tener en cuenta siempre es la complejidad del comportamiento humano, el cual es el origen de toda estrategia de comercio en la bolsa de valores [3]. Es por ello que en este capítulo se describen las dos corrientes económicas donde caen los resultados de este trabajo, con la intención de contextualizar las implicaciones de nuestros resultados.

En la primera sección se enuncia la hipótesis de mercado eficiente (HME), ya los resultados de este trabajo muestran evidencia en contra de este paradigma. Por lo que comenzamos con la descripción de las tres versiones que existen para HME, mientras que en la segunda sección se describe brevemente el enfoque de las finanzas del comportamiento o conductuales, donde desde el punto de vista económico se sitúa nuestro trabajo. Para cumplir este propósito hemos seguido los textos [23, 22, 24], tomando extractos de ellos, y adecuándolos al contexto de nuestro estudio.

## 2.1. Hipótesis de Mercado Eficiente

La teoría de que los mercados financieros son eficientes, y la extensa investigación acerca de ello, forman la idea tan aceptada por la mayoría de los economista tradicionales de que los mercados no se ven excesivamente afectados por las burbujas económicas o decisio-

nes irracionales de los agentes individuales. La hipótesis del mercado eficiente establece que los precios del mercado incorporan a cada instante y con gran precisión toda la información publica disponible acerca de ellos. En otras palabras, los mercados financieros siempre reflejan el precio correcto de sus acciones dada la información publica conocida. Desde esta perspectiva, el que los precios en ciertos instantes aparenten ser muy altos o muy bajos es simplemente una efecto pasajero o en todo caso una ilusión [23].

Es común distinguir tres versiones de esta hipótesis: la débil, la semi-fuerte, y la fuerte. Estas versiones difieren por la noción del significado cuando nos referimos a *toda la información pública disponible* [24].

La forma débil de esta hipótesis afirma que los precios de las acciones reflejan toda la información que puede ser derivada mediante el análisis de datos financieros, como son los precios históricos, el volumen de transacción, y los intereses. Esta versión implica que el análisis predictivo es imposible, ya que si los datos históricos fueran capaces de proveer indicios acerca del comportamiento futuro de los precios, todos los inversionistas ya habrían aprendido a explotar estas señales. La forma débil establece además que si tales señales existieran, perderían su valor predictivo al hacerse públicas, ya que una señal de compra resulta en un incremento instantáneo del precio.

La versión semi-fuerte de HME declara que toda la información pública disponible, sin importar las intenciones de la empresa, es reflejada en los precios de las acciones. Tal información incluye, además de los precios históricos, datos fundamentales de los productos de la empresa, como es el manejo administrativo, las hojas de balance, las patentes previstas, las ganancias, y las practicas contables. Por lo que si los inversionistas tuvieran acceso a tal información mediante alguna fuente pública disponible, se esperaría que esta información ya este incorporada en los precios de las acciones, por lo que desde este enfoque es una perdida de tiempo buscar una estrategia para generar ganancias en la bolsa a partir de esta información.

Finalmente, la versión fuerte de HME declara que los precios de las acciones reflejan toda la información relevante de la firma, incluso aquella información disponible solamente a los socios de la compañía. Esta versión de HME es muy extrema. Pocos refutarán la proposición de que los funcionarios corporativos tienen acceso a información exclusiva con suficiente tiempo de anticipación de la liberación al público, permitiendoles generar ganancias con esta información privilegiada.

## 2.2. Economía Conductual

La economía conductual es una escuela del pensamiento económico relativamente nueva la cual argumenta que la extensa literatura acerca de las estrategias de inversión en la bolsa han olvidado el importante factor humano por enfocar su análisis desde el marco de referencia de HME. La escuela conductual declara que incluso si los precios de las acciones son incorrectos, es decir que no reflejan el valor del bien o producto, generar ganancias a partir de ellos es aún una tarea difícil y, por lo tanto, el fracaso para descubrir estrategias obvias de comercio en la bolsa no puede ser considerada una prueba de la eficiencia del mercado. En otras palabras, esta teoría establece que [24]:

- Sí los precios son correctos $\Rightarrow$ no se puede generar ganancias del comercio en la bolsa.

pero la implicación inversa no es verdadera:

- Sí no se puede generar ganancias del comercio en la bolsa $\not\Rightarrow$ los precios son correctos.

Mientras que las teorías tradicionales asumen que los inversionistas son racionales, la economía conductual comienza con la premisa de que no lo son.

Los modelos conductuales asumen frecuentemente formas especificas de irracionalidad. Estas ideas vienen de los psicólogos conductuales, que observan que las decisiones de las personas están basadas en creencias y preferencias, más de allá de juicios racionales [22].

Los inversionistas forman expectativas basados en la confianza de su juicio o de sus asesores al tomar una decisión, ellos generan muchas veces apreciaciones de sus habilidades alejadas de la realidad, sobrestimando su capacidad de tomar decisiones óptimas. Además de esto una ves que las personas forman sus opiniones, suelen extrapolarlas a situaciones fueran del rango de validez, y es difícil que busquen evidencia que contradiga su tendencia y su manera de proceder, viendo con escepticismo otros paradigmas económicos [22].

Un factor muy importante que influencia a la hora de generar creencias y preferencias son las noticias e información proveniente de las redes sociales. Los medios de comunicación se siente naturalmente atraídos por los mercados financieros ya que al final de

cuentas ellos proveen información continua en forma de cambios diarios en el precio de acciones, lo cual puede ser interpretado como noticias, generando una fuente inagotable de información que ayuda a mantener sus editoriales en el mercado. Estos medios están siempre en constate competencia por atraer la atención del público, ya que de estos dependen sus ganancias. Este por esta razón que siempre están buscando encontrar, e incluso definir nuevos tópicos de interés, enfocándose en las noticias que tiene un gran poder de ser transmitidas de boca en boca [1], para de esta manera expandir y mantener su influencia en un amplio grupo de personas [23].

Por otro lado, eventos significativos en la bolsa de valores sólo ocurren sí existen un pensamiento común entre un gran número de personas, por lo que las noticias y redes sociales son un vehículo esencial para difundir las ideas y generar tendencias [23]. Es por todo esto que nosotros esperamos que un análisis cuidadoso de esta información pueda revelarnos el estado del mercado y muestre evidencias de su influencia en las fluctuaciones de los índices financieros.

[1] Conocidas en inglés como noticias *word-of-mouth*

# Capítulo 3

# Datos Analizados

Nuestro análisis se llevo a cabo para 20 países alrededor del mundo durante dos periodos de tiempo distinto. El primer periodo de tiempo comprendió del 22 de febrero al 13 de octubre de 2014, dando un total de $L_1 = 166$ días de negociación, es decir, sin considerar los fines de semana; mientras que el segundo periodo de tiempo fue del 1 de Julio de 2015 al 1 de Mayo del 2016, abarcando en este caso $L_2 = 217$ días de negociación. En ambos periodos de tiempo se consideraron los precios diarios al cierre de 20 índices financieros alrededor del mundo. Los países donde cotizan estos índices, así como los símbolos correspondientes están listados en las dos primeras columnas de las tabla 3.1. Estos datos fueron obtenidos de la base de datos de Bloomberg, y siguieron el mismo preprocesamiento que en [14].

Para el primer periodo de tiempo, un tercer grupo de datos fue obtenido de la red social Twitter mediante la extracción de *tweets* asociados a cada uno de los países listados en la tabla 3.1 (tercera columna). Por otro lado, para el segundo periodo de tiempo, se generó un cuarto grupo de datos mediante la extracción de noticias de NYT relacionadas nuevamente con cada uno de los países de la tabla 3.1 (cuarta columna). La extracción de *tweets*, así como de noticias de NYT fueron hechas en el horario universal (UTC), mientras que la consulta de los precios al cierre varían de acuerdo a la zona horaria donde cotizan los mercados involucrados de cada índice.

Cuadro 3.1: Lista de los indices financieros analizados en este trabajo. Primera columna: países donde cotizan los índices. Segunda columna: símbolo correspondiente en el sistema Bloomberg. Tercera columna: palabra clave para hacer la búsqueda en Twitter. Cuarta columna: palabra clave para hacer la búsqueda en NYT.

| **País** | **Símbolo de Bloomberg** | **keyword Twitter** | **keyword NYT** |
|---|---|---|---|
| México | MEXBOL | IPC_MEXICO | Mexico |
| Estados Unidos | SPX | SP500 | United States |
| Argentina | MERVAL | MERVAL | Argentina |
| Brasil | IBOV | IBOVESPA | Brazil |
| Reino Unido | UKX | FTSE_UK_INDEX | England |
| Francia | CAC | CAC_40 | France |
| Suiza | SMI | SWISS_MARKET_INDEX | Switzerland |
| Alemania | DAX | DAX_INDEX | Germany |
| Austria | ATX | ATX_INDEX | Austria |
| Egipto | CASE | EGX_EGYPT | Egypt |
| Israel | TA-25 | TEL_AVIV_STOCK | Israel |
| India | SENSEX | BSE_SENSEX | India |
| Indonesia | JCI | JAKARTA_STOCK | Indonesia |
| Malasia | FBMKLCI | BURSA_MALAYSIA | Malaysia |
| Singapur | FSSTI | STRAITS_TIMES_INDEX | Singapore |
| Hong Kong | HSI | HANG_SENG_INDEX | Hong Kong |
| Taiwan | TWSE | TAIWAN_STOCK | Taiwan |
| Korea del Sur | KOSPI | KOSPI | South Korea |
| Japón | NKY | NIKKEI_INDEX | Japan |
| Australia | AS51 | ALL_ORDINARIES | Australia |

## 3.1. Extracción de la Red Social *Twitter*

Todos los *tweets* fueron extraídos de la base de datos de Twitter mediante la interface *Twitter Search API* [1]. Además, se utilizó un código *wrapper* [2] en el lenguaje PYTHON para manejar de manera más eficiente los métodos del API de Twitter. Los parámetros y valores utilizados para obtener la colección de datos de Twitter se muestran en la tabla 3.2.

En esta tabla, el primer parámetro *term*, es la palabra o frase que queremos buscar dentro de la base de datos de Twitter. A este parámetro se le pasa el valor *keyword Twitter*, el cual hace referencia a cada elemento en la tercera columna de la tabla 3.1. El siguiente parámetro *geocode* especifica el radio geográfico de búsqueda alrededor del usuario, el cual se especificó como *none*, ya que nuestra intención es obtener la información más diversificada posible. Asimismo, a los parámetros *since_id* y *max_id* se les asignó también el valor

[1] En general un API (del inglés: Application Programming Interface) es un conjunto de métodos, protocolos y rutinas para construir software y aplicaciones. En particular, el API de Twitter nos permite acceder de manera estructurada a su base de datos.

[2] Código envolvente, que ayuda a traducir una interface dada en otra compatible, siendo generalmente más simple.

Cuadro 3.2: Parametros y valores utilizados en el *Twitter Search API*.

| **Parámetro** | **Valor** |
|---|---|
| term | keyword |
| geocode | none |
| since_id | none |
| max_id | none |
| until | current |
| count | 100 |
| lang | en |
| locale | none |
| result_type | mixed |

de *none* debido a que no estamos interesados en tener control de los *tweets* a través de su número identificador, sino más bien a través de la fecha en que fueron publicados. De esta manera, decidimos asignarle al parámetro *until* el valor de *current*, para lograr así obtener todos los *tweets* disponible hasta la fecha corriente. El parámetro *count* fija el número de *tweets* devueltos por petición, siendo su valor máximo de 100 el que se utiliza aquí. El lenguaje de búsqueda se elige en Inglés al asignar el valor *en* en el parámetro *lang*. Además, el parámetros *locale* especifica el lenguaje de la consulta que estamos enviando, pero dado que estamos investigando palabras claves relacionadas con los símbolos de Bloomberg, el lenguaje de búsqueda es irrelevante en nuestras consultas por lo que fue fijado como *none*. Finalmente, al parámetro *result_type* se le asignó el valor de *mixed* para obtener como respuesta una combinación de los resultados más relevantes y actuales.

Al trabajar con el API de Twitter se tuvieron que superar principalmente tres obstáculos para la adquisición de datos históricos. El primero, se debe a que solamente es posible hacer peticiones de datos dentro de los últimos siete días respecto a la fecha de consulta. El segundo obstáculo se debió a que en cada consulta sólo se puede obtener como respuesta un máximo 100 *tweets*, lo que no es suficiente para tener una muestra representativa de los *tweets* diarios para cada *keyword*. El tercer obstáculo se debe a que está limitado el número de consultas dentro de un rango de tiempo, y si este límite es alcanzado no es posible seguir haciendo más consultas hasta los siguientes quince minutos por lo menos.

Para lograr superar estas dificultades y obtener así una muestra representativa de *tweets* historicos, se realizó el siguiente procedimiento. En un servidor se automatizaron las consultas de cada *keyword* hora por hora. Además, se accedió al API de Twitter con distintas credenciales, asegurando de esta manera no alcanzar el límite de peticiones permitidas por

Cuadro 3.3: Parámetros y valores utilizados en el *Article Search API* de NYT.

| **Parámetro** | **Valor** |
|---|---|
| q | keyword |
| fq | source: The New York Times |
| begin date | current |
| end date | current |
| sort | oldest |
| page | vary |

usuario. Al programar peticiones hora por hora, también se logró incrementar el número de *tweets* recolectados, obteniendo así hasta 2400 *tweets* diarios para cada *keyword*. No obstante, este número puede ser mucho menor en periodos de tiempo en los que el *keyword* consultado no es un *trending topic*, es decir, no es el tema del momento. Por otra parte, con la ayuda del servidor remoto fue posible construir una base de datos (ver apéndice A) con la información obtenida de Twitter, y como resultado logramos hacer nuestro análisis para un periodo de tiempo mucho más extenso ($L_1 = 165$ días) que los siete días permitidos por el API de Twitter.

## 3.2. Extracción de *The New York Times*

Las noticias de *The New York Times* fueron extraídas con ayuda de la interface *Article Search API*. La cual nos permitió acceder a su base de datos de manera estructurada, y con esto obtener los identificadores para extraer el texto completo de las noticias en una segunda etapa. Asimismo, se hizo uso del código *wrapper* NYTIMESARTICLE [3] escrito en PYTHON para volver compatible la extracción de los identificadores con los códigos de preprocesamiento del texto obtenido en la segunda fase. Los parámetros y valores utilizados para obtener la colección de identificadores para nuestro periodo de estudio se muestran en la tabla 3.3.

En esta tabla, el primer parámetro *q*, se refiere a la palabra o frase que queremos buscar dentro de la base de datos de NYT. A este parámetro se le pasa el valor *keyword NYT*, el cual hace referencia a cada elemento de la cuarta columna de la tabla 3.1. El parámetro *fq* es un filtro de búsqueda que utiliza la sintaxis *Lucene* [4], en el cual sólo se ha especificado

[3] https://pypi.python.org/pypi/nytimesarticle
[4] Sintaxis creada para facilitar la búsqueda de información a través de páginas web.

que la búsqueda se hiciera para fuentes del NYT, ya que el API también permite extraer información de las fuentes de noticias *Reuters* y *American Press*, sin embargo estás fuentes presentan restricciones para acceder a su contenido. Los parámetros *begin date* y *end date* toman ambos el valor de *current*, siendo este valor la fecha corriente dentro de un *loop* que recorre el rango de tiempo de estudio, es decir, del 1 de Julio de 2015 al 1 de Mayo de 2016. Por otro lado, al asignarle al parámetro *sort* el valor de *oldest*, la información de cada consultas es devuelta en orden cronológico; del identificador más antiguo al más reciente. Finalmente, el parámetro *page* hace referencia al número de resultados (diarios en nuestro caso) que se desean obtener, esté valor se varió dependiendo del país, siendo el valor *3* el que se utilizó para la mayoría de ellos, para los cuales se obtuvieron 30 identificadores relacionados con 30 noticias diarias, sin embargo para *England* se utilizó un valor de *6*, mientras que para *United States* fue de *12*, pues estos países tienen un mayor volumen de noticias por día.

Los identificadores que se pueden obtener mediante el *Article Search API* son el *id*, *abstract*, *headline*, *paragraph*, *desk*,*date*,*section*, *snippet*, *source*,*type*,*url*,*word count*,*locations*, y *subjects*. Sin embargo, no es posible obtener el texto completo de cada una de las noticias, lo cual representó el obstáculo principal en la extracción de noticias de NYT. El segundo problema se debió a que, al igual que Twitter, el número de consultas está limitado. Esté último problema fue rápidamente resuelto al crear 10 cuentas distintas y automatizar las consultas, lo que nos tomó aproximadamente una semana de computo para poder extraer todos los identificadores de los 20 países para el periodo de tiempo de especificado, logrando recabar un total de 44229 noticias.

El problema con respecto a la extracción del texto completo de las noticias se resolvió utilizando los identificadores *date* y *url* [5], así como los módulos de extracción de datos de PYTHON: *urllib* y *cookielib*; para lidiar con los protocolos de acceso. De esta manera se creó una base de datos conteniendo las noticias de cada día (*date*) del periodo de estudio, y de cada *Keyword NYT* listado en la cuarta columna de la tabla 3.1.

[5]Uniform Resource Locator (url): es el término genérico para todos los nombres y direcciones que se refieren a objetos en la *World Wide Web* (www), y que utilizan el protocolo HTTP o HTTPS.

# Capítulo 4

# Análisis de Sentimiento

El análisis de sentimiento es un campo de estudio del procesamiento de lenguaje natural, minería de opiniones, y lingüística computacional [25, 26]. Las dos principales aproximaciones para llevar a cabo el análisis de sentimiento son la basada en diccionarios o léxicos, y la de aprendizaje automático. En esta tesis se seguirá la basada en diccionarios o léxicos, debido a que el procesamiento de los textos se pude realizar con relativamente poco tiempo de gasto computacional en comparación con el gran costo de los métodos de aprendizaje automático.

Dentro de la aproximación basada en léxicos, el método de estados emocionales y el de polaridad son las dos aproximaciones principales para calificar numéricamente los datos textuales. Nosotros utilizamos la aproximación de polaridad debido a que se puede asociar de manera más directa con los movimientos positivos y negativos de los indices financieros.

A continuación se describen detalladamente las metodologías seguidas para hacer el análisis de sentimiento de Twitter y de NYT, desde el preprocesamiento del texto hasta la construcción de las series de tiempo de polaridad.

## 4.1. Twitter

La red social Twitter permite a sus usuarios mandar y recibir mensajes cortos de hasta 140 caracteres, a los cuales se les conoce como *tweets*. Un *tweet* es una combinación de caracteres intercalados con espacios en blancos, donde cada cadena de caracteres está compuesta de una secuencia alfanumérica, mezclada con caracteres especiales como son: @, $, #, %, &, etc. Por lo que cada *tweet* necesita de un preprocesamiento para eliminar los caracteres no deseados, los cuales introducen ruido para interpretar el significado de la información contenida en su texto.

El primer paso es dividir cada uno de los *tweets* en las cadenas de caracteres que lo componen, evitando los símbolos no alfanuméricos. Como resultado se obtiene una colección de palabras individuales. A esta operación se le conoce como *tokenization*. Una vez que el *tweet* es *tokenizado*, cada elemento de la colección de palabras se simplifica o reduce mediante el método de *stemming*, el cual consiste en eliminar los afijos morfológicos dejando solamente la raíz de la palabra. De esta manera, una vez dividido el *tweet* en una colección de palabras, y reducida cada una de estas a su raíz morfológica, se encuentra ya listo para realizar el cálculo de sentimiento.

El análisis de sentimiento de Twitter fue realizado con ayuda de la librería de código abierto PYSENTIMENT [1]. La principal contribución de la librería PYSENTIMENT es la implementación de diccionarios [27]. Aquí se eligió el diccionario Harvard IV puesto que ha tenido éxito en predecir el desempeño de los mercados de valores [28, 29]. Este diccionario está compuesto de más de 8000 palabras y 182 categorías. Al considerar solamente las categorías *positive* y *negative* fue posible obtener la polaridad de cada elemento de la colección publica de *tweets*.

En PYSENTIMENT, La polaridad se obtiene al calcular la diferencia entre el número de palabras positivas y negativas encontradas en el texto. Este puntaje está dado por la fórmula

$$polarity = \frac{p - n}{p + n}, \tag{4.1}$$

donde $p$ y $n$ se refiere al total de palabras positivas y negativas, respectivamente.

[1] Han Zhichao, disponible en https://pypi.python.org/pypi/pysentiment

Cuadro 4.1: Ejemplo del proceso de cálculo de polaridad para un *tweet* individual.

| **Proceso** | |
|---|---|
| tweet | "#SP500 closed near the lows for the week. Expecting more downside Monday. $SPX" |
| tokenization | {SP500, closed, near, the, lows, for, the, week, Expecting, more, downside, Monday, SPX} |
| stemming | {sp500, close, near, the, low, for, the, week, expect, more, downsid, monday, spx} |
| categorización | {0, 0, 0, 0, negative, 0, 0, 0, 0, 0, negative, 0, 0} |
| polaridad | -1 |

De está manera fue posible obtener series de tiempo asociadas a cada *keyword Twitter* de la tabla 3.1. Estas series de tiempo se construyeron como sigue. Primeramente, la colección pública de *tweets* se clasificó por *keyword* y fecha en una base de datos. En seguida, a cada *tweet* de la base de datos se le calculó la polaridad mediante la ec. 4.1. Después, se calculó el promedio de todas la polaridades para cada día y *keyword* dado. Finalmente, a este promedio se le consideró la polaridad $P_k(t)$ del *keyword* $k$ al tiempo $t$, donde las unidades de tiempo se eligieron en días. En la tabla 4.1 se muestra esquemáticamente el proceso de cálculo de polaridad para un *tweet* individual. En la primera línea de este cuadro se muestra un *tweet* típico, en la segunda se hace la *tokenización*, en la siguiente el *stemming*, después la *categorización* mediante el diccionario, y finalmente se calcula la *polaridad*.

## 4.2. The New York Times

La fuente de noticias *The New York Times* es un periódico publicado principalmente en la ciudad de Nueva York, que se distribuye en más de 160 países alrededor del mundo. En su versión impresa circulan más de 1 millón de ejemplares diariamente, mientras que a la versión *online* puede acceder ilimitadamente cualquier persona que cuente con una suscripción. Lo que lo convierte en uno de los diarios más accesibles y con mayor circulación alrededor del mundo.

Una ventaja al utilizar nuestros códigos de extracción, es que se logró evitar el tener una suscripción, ya que se descargaron las noticias en su versión HTML, sin embargo esto trajo como consecuencias el que la información pasara por un minucioso preprocesamiento para eliminar el texto no deseado, como son los delimitadores, saltos de pagina, y demás caracteres propios de la sintaxis HTML (ver apéndice A). Esta etapa de limpieza se llevo a cabo mediante el *Kit de Herramientas de Lenguaje Natural* (NLTK) de PYTHON, siendo esta una plataforma para trabajar con datos provenientes del lenguaje humano. Además

se construyo un *script* [2] en el lenguaje BASH del SHELL de LINUX, para eliminar los patrones de caracteres que podrían causar ruido al momento de calcular el sentimiento.

El análisis de sentimiento para los noticias de NYT se realizó con la ayuda de la librería de código abierto VADER (Valence Aware Dictionary and sEntiment Reasoner) [3], la cual es a la vez un léxico y una herramienta de análisis de sentimiento basado en reglas. Esta librería ha sido ajustada para capturar el sentimiento expresado en las redes sociales, mas ha sido probado en otros contextos, mostrando excelentes resultados al trabajar con NYT [30].

El código de VADER implementa reglas sintácticas y gramaticales, incorporando cuantificadores derivados empíricamente para medir la intensidad de sentimiento presente en los textos. Este código toma en cuenta las características y propiedades del texto como son: puntuación, el uso de mayúsculas, modificadores de grado o adverbios, la conjunción *but*, y otros elementos que pueden cambiar el sentido del enunciado como son: *is not* and *do not*. VADER además toma en cuenta expresiones comunes en *microblogs*, incluyendo una lista completa de emoticones [4] occidentales, acrónimos, iniciales y palabras informales: nah, meh, giggly, LOL, WTF, :-); etcétera.

El léxico creado por VADER consta de 7500 elementos, cada uno de los cuales fueron evaluados por 10 calificadores del *Amazon Mechanical Turk* (AMT), asignándoles un puntaje en la escala del -4 al 4. Los resultados arrojados por la librería constan de cuatro valores: NEGATIVE, NEUTRAL, POSITIVE, y COMPOUND. Para nuestro análisis se decidió tomar en cuenta únicamente COMPOUND como el valor de polaridad, debido a que este puntaje es el promedio de los 10 calificadores, así como de los intensificadores o *boosters* derivados empíricamente, los cuales como se menciono anteriormente incorporan reglas sintácticas y gramaticales. De está manera, de aquí en adelante cuando se hable del sentimiento o polaridad de los textos de NYT, nos estaremos refiriendo al puntaje dado por COMPOUND.

El cálculo de la polaridad diaria se hizo de la misma manera que en la sección anterior, es

[2] Lenguaje de programación que se utiliza principalmente para el entorno de ejecución de tareas automatizada.

[3] C.J. Hutto and Eric Gilbert, disponible en https://pypi.python.org/pypi/vaderSentiment

[4] Es una representación pictórica que representa expresiones faciales utilizando símbolos de puntuación, números y letras; siendo generalmente escritas para expresar emociones o estados de animo.

decir, se calculó el promedio de las polaridades considerando todas las noticias de un día para un *keyword* dado, y a este promedio se le consideró la polaridad $P_k(t)$ del *keyword* $k$ al tiempo $t$, donde las unidades de tiempo son días.

# Capítulo 5

# Taxonomía y Jerarquización

En este capitulo se presentan algunas técnicas matemáticas utilizadas en la física estadística y sistemas complejos que nos ayudarán a entender la estructura de las matrices de correlación asociadas a nuestros datos empíricos, es decir, los datos provenientes de Twitter, NYT de y de los indices financieros globales. Para Twitter y NYT, en lugar de los datos textuales se utilizaron las series de tiempo de polaridad, mientras que para los indices se usaron los retornos diarios al cierre. Además, puesto que en general los mercados financieros no cotizan los fines de semana, se ajustaron las series de tiempo de polaridad a los días de cotización de los indices financieros, desfasando además sus valores por un día.

Nuestro interés en esta sección será investigar los coeficientes de la matriz de correlación para detectar la taxonomía y organización jerárquica de los índices financieros, así como de los indicadores de polaridad. Para lograr esto haremos uso de la teoría de redes y de la distancia ultramétrica, la cual ha sido una medida importante en física para entender la topología de los vidrios de espín de en el espacio de estados [31, 32, 33].

## 5.1. Matrices de Correlación

Denotemos por $S_k(t)$ el precio al cierre del índice $k$ al día $t$. Los retornos $R_k(t)$ para cada índice $k = 1, \ldots, 20$ al tiempo $t$ se obtienen mediante

$$R_k(t) = \frac{S_k(t + \Delta t) - S_k(t)}{S_k(t)}, \tag{5.1}$$

donde se eligió $\Delta t = 1$, tal que el intervalo de retorno sea de un día. Además, con el propósito de comparar nuestros datos empíricos con los resultados provenientes de la física estadística, las series de tiempo de polaridad y retornos son normalizadas. El retorno normalizado para el indice $k$ al tiempo $t$ esta dado por

$$r_k(t) = (R_k(t) - \langle R_k \rangle)/\sigma_k, \tag{5.2}$$

donde $\sigma_k$ es la desviación estándar $R_k$, y $\langle \dots \rangle$ denota el promedio temporal sobre el periodo estudiado. La polaridad se normalizó de las misma manera, y es denotada como $p_k(t)$ para el indice $k$ al tiempo $t$.

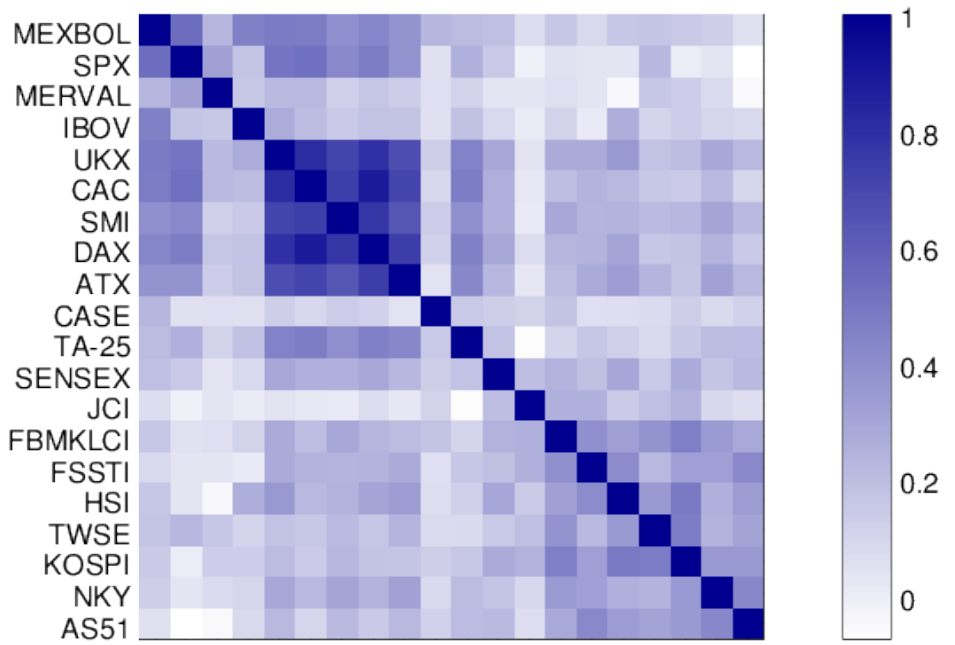
La forma más simple de caracterizar las coeficientes de correlación entre series de tiempo normalizadas es mediante el cálculo de los elementos de matriz de Pearson [34]

$$c_{k,l}^{(x)} = \langle x_k(t) x_l(t) \rangle, \tag{5.3}$$

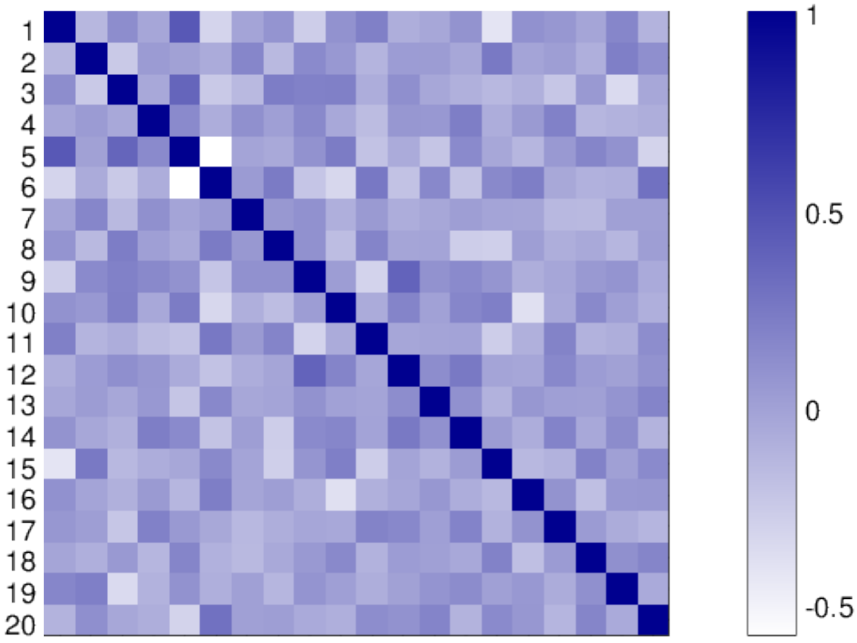
donde el superíndice $x$ es para denotar el tipo de serie de tiempo con la que se está trabajando, de tal manera que $c_{k,l}^{(p)}$ y $c_{k,l}^{(r)}$ son los elementos de la matriz de correlación $k, l$, construidos a partir de las series de tiempo de polaridad y retornos, respectivamente.

En las figs. 5.1(Twitter) y 5.2(NYT) se muestran las matrices de correlación de las polaridades y retornos como mapas de calor, para los dos periodos de estudio respectivamente. En la figs. 5.1(b) y 5.2(b) las etiquetas del *keyword* son reemplazadas por el número que representa la posición del indice financiero asociado a las series de polaridad. En esta representación los cuadros de color más obscuro denotan las correlaciones fuertes, mientras que los cuadros más claros representan las correlaciones débiles o anti-correlaciones. Los casos extremos son $c_{k,l} = 1(-1)$, lo que corresponde a una perfecta correlación (anti-correlación), mientras que $c_{k,l} = 0$ significa que la correlación es nula entre los elementos $k$ y $l$.

Se puede ver que para ambos periodos de tiempo emergen patrones en los datos de retorno (figs. 5.1(a) y 5.2(a)), como es la fuerte correlación entre el sector Europeo, los Estados Unidos de Norteamérica y México, así como Norteamérica con Europa, y el sector asiático, lo cual refleja la codependencia de las economías debido a su posición geográfica. No obstante, si seguimos la posición de número de estos índices en las fig. 5.1(b) y 5.2(b), no encontramos la misma estructura, aunque los datos de NYT muestran ciertos

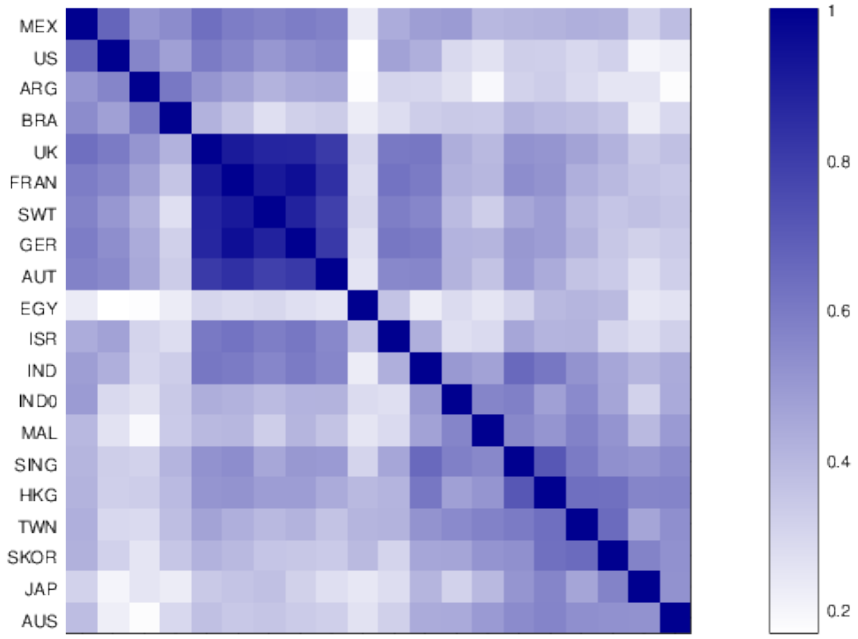



(a)

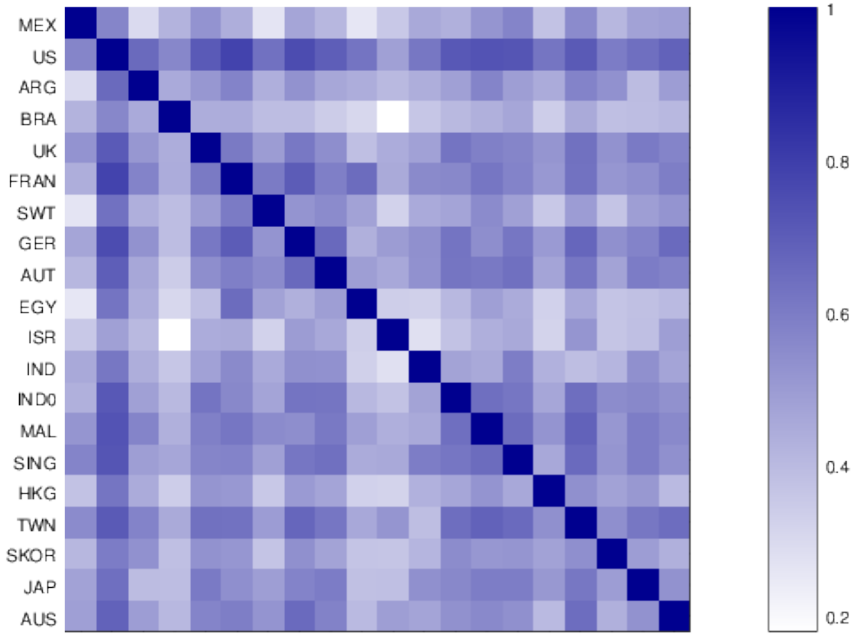



(b

Figura 5.1: Primer periodo de tiempo: Twitter e índices financieros. (a) elementos de la matriz de correlación para datos de retorno.(b) elementos de la matriz de correlación para datos de polaridad. En esta escala de representación de colores, el valor mínimo corresponde al blanco, mientras que los valores más grandes corresponden a colores azul intenso.




(a)




(b)

Figura 5.2: Segundo periodo de tiempo: NYT e índices financieros. (a) elementos de la matriz de correlación para datos de retorno.(b) elementos de la matriz de correlación para datos de polaridad. En esta escala de representación de colores, el valor mínimo corresponde al blanco, mientras que los valores más grandes corresponden a colores azul intenso.

trazos similares a los datos de retorno asociados, sin embargo no es totalmente claro. De aquí surge la necesidad de realizar un análisis más profundo para averiguar si existe una estructura de correlación oculta en la series de tiempo de polaridad.

## 5.2. Matriz de Distancia

Los coeficientes de correlación no pueden ser utilizados como una medida de distancia ya que no satisfacen los tres axiomas que definen a una métrica. No obstante, esta puede ser definida utilizando una función de los coeficientes de correlación. Una expresión apropiada para esta función es [35]

$$d(k,l) = \sqrt{2(1 - c_{k,l})}. \tag{5.4}$$

Con esta elección de distancia se satisfacen lo tres axiomas que definen una métrica:

- $d(k,l) = 0$ si y sólo si $k = l$
- $d(k,l) = d(l,k)$
- $d(k,l) \leq d(k,m) + d(m,l)$

El primer axioma es válido debido a que $d(k,l) = 0$ si y sólo si la correlación es máxima ($c = 1$), es decir, solamente si los dos indicadores realizan el mismo proceso estocástico. El segundo axioma es válido debido a que la matriz de coeficientes de correlación $C$, así como la correspondiente matriz de distancias $D$ son simétricas por definición. El tercer axioma es válido debido a que la eq. 5.4 es equivalente a la distancia Euclidiana entre dos vectores $\tilde{V}_k$ y $\tilde{V}_l$, los cuales se pueden asociar a las series de tiempo $x_k$ y $x_l$, considerando cada medición como un componente de los vectores asociadas. De esta manera tenemos que la matriz de distancias está dada por:

$$D = \begin{pmatrix} 0 & d(1,2) & d(1,3) & \dots & d(1,n) \\ & 0 & d(2,3) & \dots & d(2,n) \\ & & \ddots & \dots & \vdots \\ & \mathbf{0} & & 0 & d(n-1,n) \\ & & & & 0 \end{pmatrix} \tag{5.5}$$

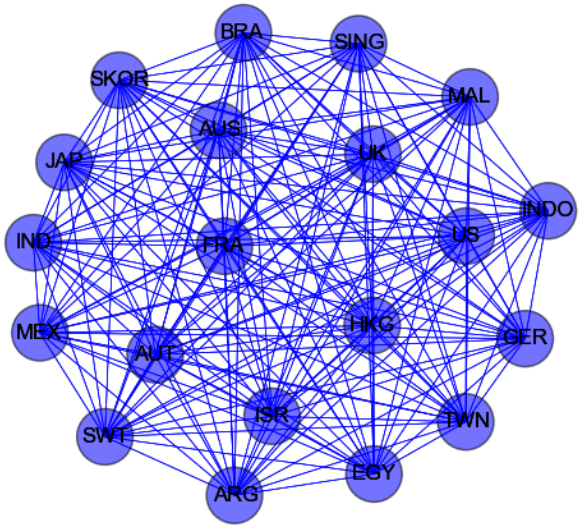



Figura 5.3: Red asociada a la matriz de distancia de los 20 indicadores globales dados en la segunda columna de la tabla 3.1.

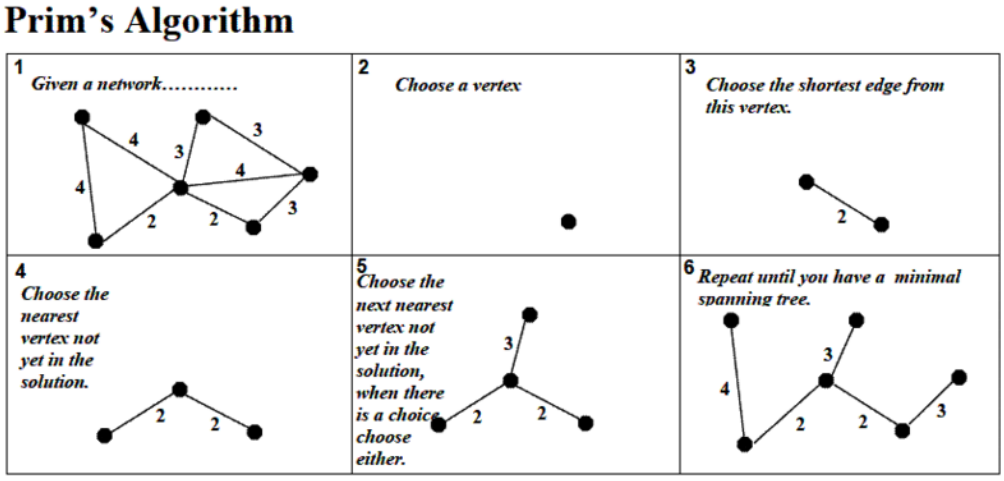
## 5.3. Arbol de Expansión Mínima (MST) y la Taxonomía

Un grafo es un par ordenado $G(V, E)$, donde $V$ representa las aristas, y E los nodos. En nuestro caso, a la matriz de distancia definida en la ec. 5.5 se le puede asociar de manera única la estructura de una gráfica, en donde los nodos vienen a ser los indicadores financieros, mientras que las aristas son los coeficientes $D_{k,l}$. En la fig. 5.3 se muestra de manera ilustrativa la gráfica o red asociada a los 20 indicadores que estamos estudiando.

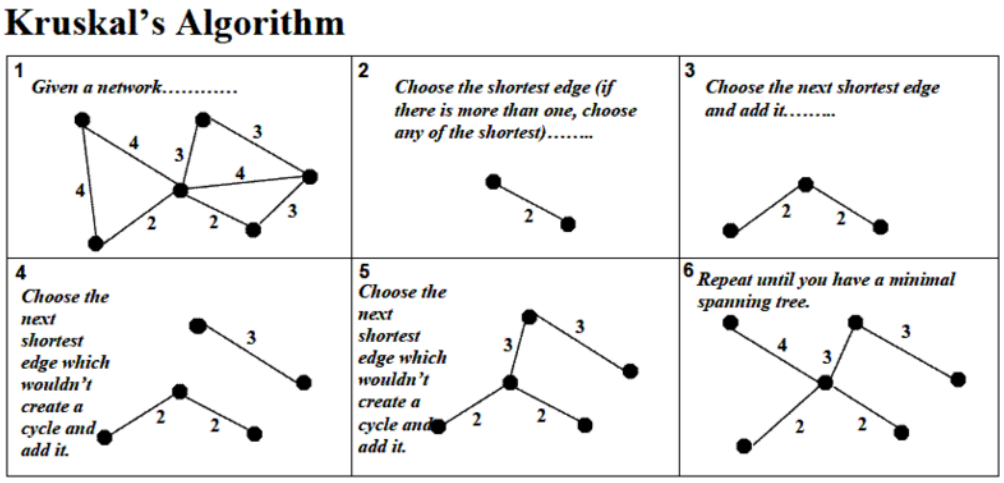
El árbol recubridor mínimo o MST [1] es un concepto proveniente de la teoría de gráficas [36, 37]. En una grafo ponderado de $n$ nodos, el MST es un árbol de $n - 1$ aristas, el cual minimiza la suma de las distancias de las aristas. La manera usual de calcular MST es a través del algoritmo de Prim y el de Kruskal [38, 39], representados esquemáticamente en la fig. 5.4.

Al aplicar cualquiera de estos dos Algoritmo en las redes de polaridad y retornos del primer periodo de tiempo, es decir, para Twitter y los índices financieros, se obtienen

[1] minimum spanning tree por sus siglas en inglés




(a)




(b)

Figura 5.4: Representación esquemática de los dos principales algoritmos para calcular el MST. (a): Algoritmo de Prim. (b): algoritmo de Kruskal.

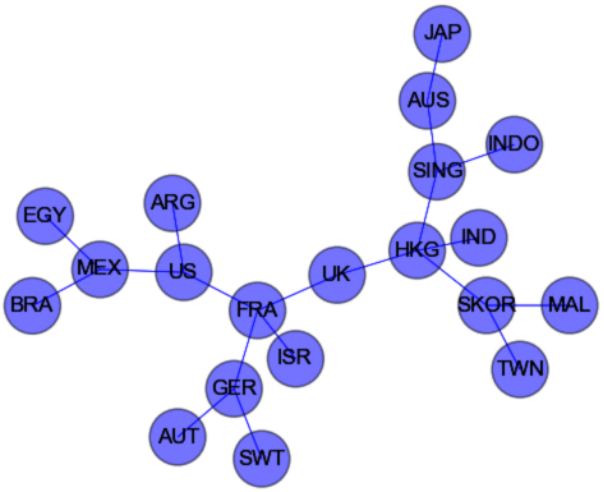
los MST mostrados en la fig. 5.5. Se puede ver que los MST de ambos datos difieren en su estructura. El MST de los retornos captura la taxonomía esperada por la posición geográfica de las bolsas de valores donde cotizan los mercados de cada uno de los índices, mientras que el MST de las polaridades muestra un comportamiento más desordenado lo cual puede deberse a que la información de Twitter está desfasada en el tiempo, es decir, está información puede estar influyendo en el valor del precio de los índices por lo que el MST asociado a las polaridades está en un estado anterior o transitorio en relación al MST de los retornos.

Para el segundo periodo de tiempo, que involucra al NYT y los índices financieros, se obtienen los MST mostrados en la fig. 5.6. Se puede ver que los MST de ambos datos difieren drásticamente en su estructura. De nuevo el MST de los retornos captura la taxonomía esperada por la posición geográfica de las bolsas de valores donde cotizan los mercados de cada uno de los índices, mientras que el MST de las polaridades muestra como nodo central a las noticias de Estados Unidos, lo cual es de esperarse dado que el NYT tiene su basa principal en la ciudad de Nueva York, por lo que la información esta centralizada en Norteamérica. Con la intención de evitar este posible sesgo se decidió remover el efecto de los Estados Unidos. Para ello, en la fig. 5.7 se muestra ahora el mismo MST, pero sin tomar en cuenta a los Estados Unidos, por lo que la red esta compuesta solamente de 19 nodos. Se puede ver ahora cierta estructura geográfica, al menos mucho más ordenada que la encontrada para los datos de Twitter, sin embargo no corresponde totalmente a la de los índices, por lo que también la hipótesis sigue siendo que esta información está influyendo en el valor del precio de los índices, por lo que el MST asociado a las polaridades está en un estado anterior o transitorio en relación al MST de los retornos.

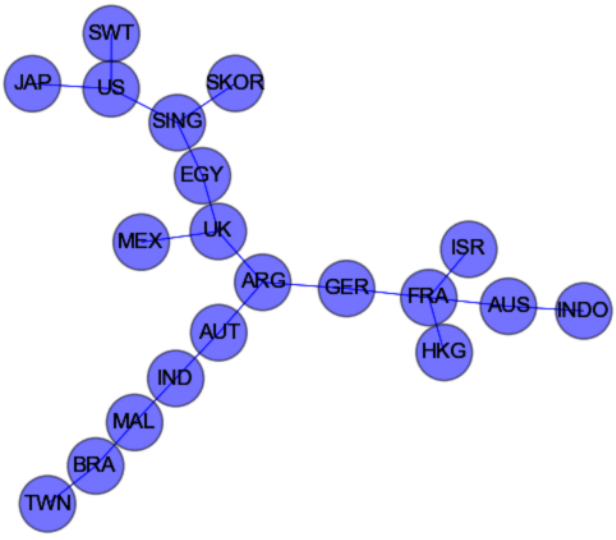
## 5.4. Espacio ultramétrico

Un concepto matemático muy útil para ligar $n$ acciones, mercados, o indicadores financieros es a través del espacio ultramétrico. Esto ha sido corroborado por el hecho de que la estructura geométrica que emerge en este espacio contiene información de gran valor desde el punto de vista económico [35]. En este espacio la distancia entre objetos está dada por la distancia ultramétrica [32], la cual esta dada por la regla fuerte del triangulo o la desigualdad ultramétrica

$$d(k,l) \leq \text{máx}\{d(k,m), d(m,l)\}. \tag{5.6}$$

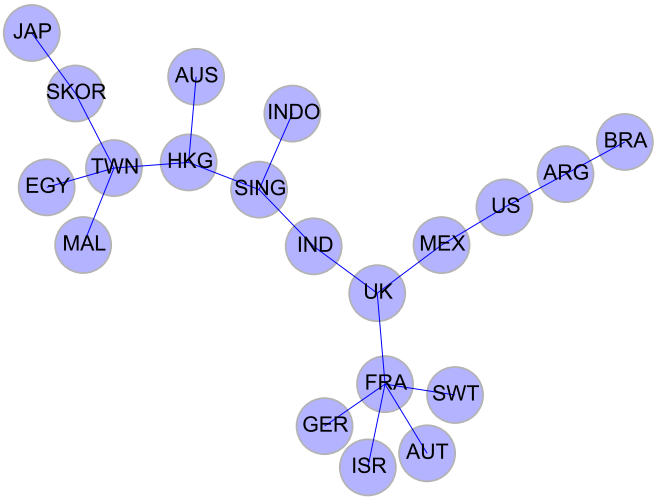



(a)

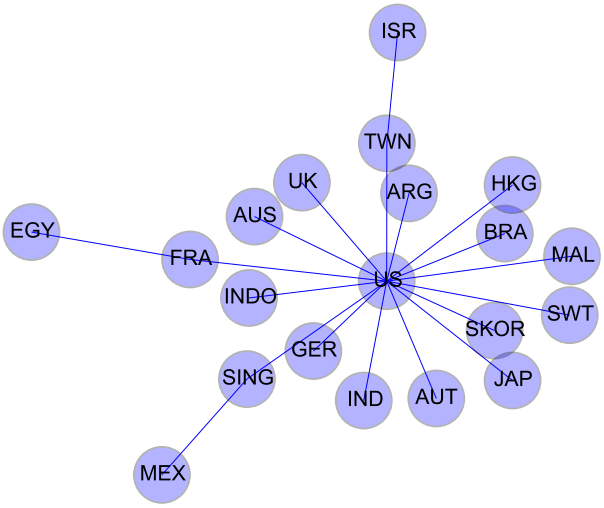



(b)

Figura 5.5: Arboles de expansión mínima del primer periodo: Twitter e índices financieros. (a) Retornos. (b) Polaridades.

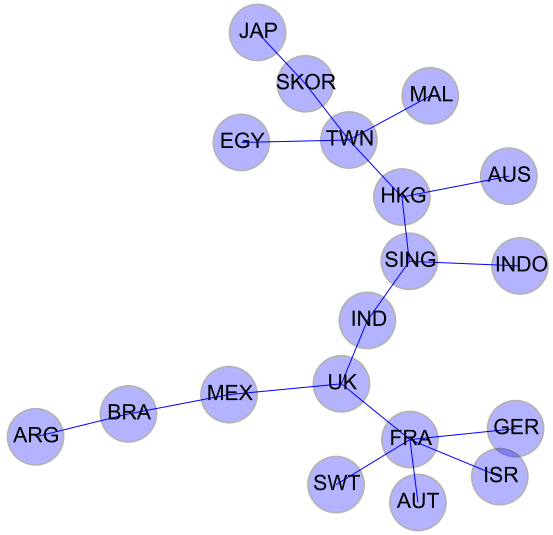



(a)

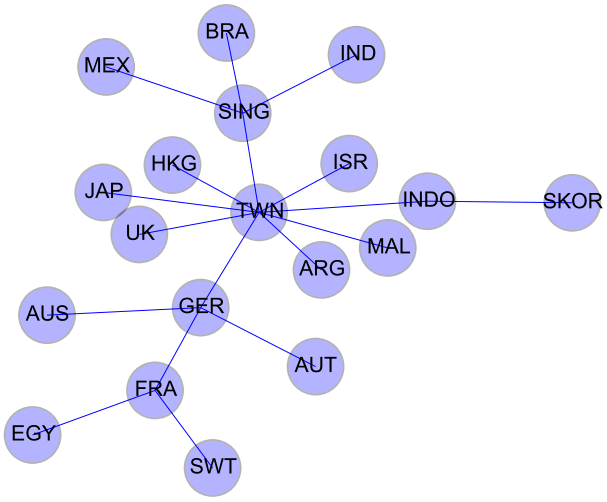



(b)

Figura 5.6: Arboles de expansión mínima del segundo periodo: NYT e índices financieros. (a) Retornos. (b) Polaridades.




(a)




(b)

Figura 5.7: Arboles de expansión mínima del segundo periodo al remover el efecto de los Estados Unidos. (a)Retornos. (b) Polaridades.

## 5.4. Espacio ultramétrico

Un espacio ultramétrico provee de manera natural de una estructura jerárquica para describir a los mercados financieros ya que el concepto de ultrametricidad está directamente conectado con el concepto de jerarquía. La medida ultramétrica subdominante es el objeto especifico que utilizaremos, este sobresale de entre todas las posibles estructuras ultramétricas asociadas con la distancia $d(k,l)$ debido a su simplicidad y propiedades singulares que se adecuan perfectamente a una estructura jerárquica. Esta medida esta dada por la siguiente expresión

$$d^{<}(A,B) = \text{máx}\{d(w_i, w_{i+1}), 1 \leq i \leq n-1\}, \tag{5.7}$$

donde $W = \{(w_1, w_2), \ldots, (w_{n-1}, w_n)\}$ denota el camino único del MST entre los nodos $A$ y $B$ $(w_1 = A, w_n = B)$. De esta manera, en presencia de una espacio métrico en el cual $n$ objetos se encuentran conectados, la ultramétrica subdominante se puede obtener determinando el MST que conecta los $n$ objetos [40, 41].

Mientras que el MST nos da la taxonomía asociada a los mercados involucrados (nodos), el espacio ultramétrico subdominante asociado provee de una buena definición topológica, asociando una única jerarquía indexada al MST. De manera ilustrativa en la fig. 5.8(a) se muestra un MST de 5 nodos etiquetados como $\{a, b, c, d, e\}$ con sus respectivas aristas ponderadas. Estos nodos podrían representar 5 mercados financieros distintos, y las aristas los valores obtenidos por la eq. 5.5. La fig. 5.8(b) representa la ultramétrica subdominante asociada una vez que se le ha aplicado la eq. 5.7 al MST. De esta manera podemos construir el dendrograma de la fig 5.8(c), donde los nodos ligados con los valores más bajos representan los mercados más dominantes e interrelaciones, mientras que los nodos que se van ligando con valores más altos representan los menos dominantes. Hemos entonces obtenido una jerarquización de los nodos (en nuestro caso de los indicadores financieros globales), donde aquellos que se conectan con valores bajos son los que están en la escala jerárquica más alta, es decir, son los más correlacionados.

Al aplicar esta metodología en los datos de retornos y polaridades se obtienen los dendrogramas mostrados en la figs. 5.9 y 5.10, para el primer y segundo periodo respectivamente, donde se removió nuevamente el efecto de los Estados Unidos. Los colores se refieren al grado de agrupamiento encontrado, donde el color rojo representa el *cluster* más dominante, el verde el intermedio, y el azul el menos dominante. Se puede observar en ambos periodos un comportamiento similar al mostrado en las figuras de MST, es decir, los retornos están jerarquizados por su posición geográfica, mientras que las polaridades no

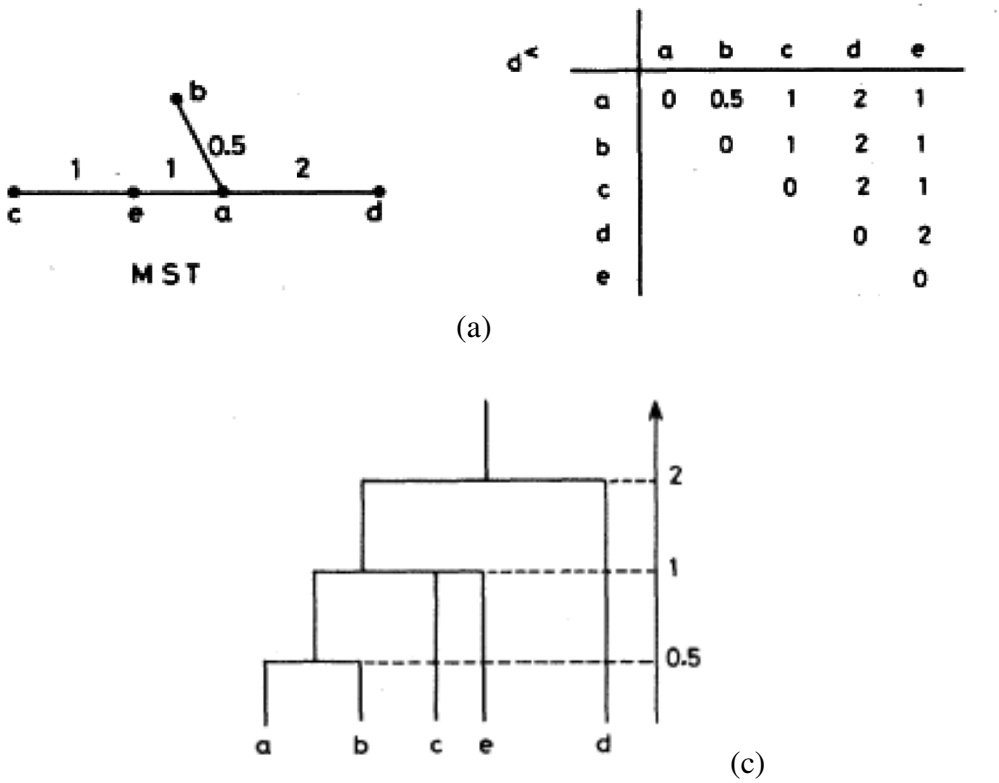
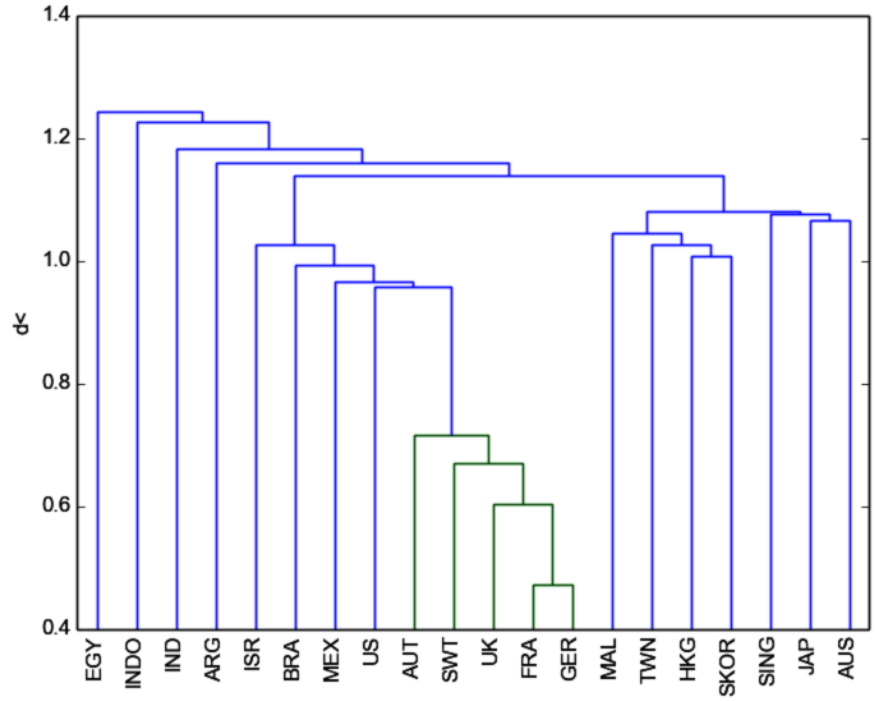


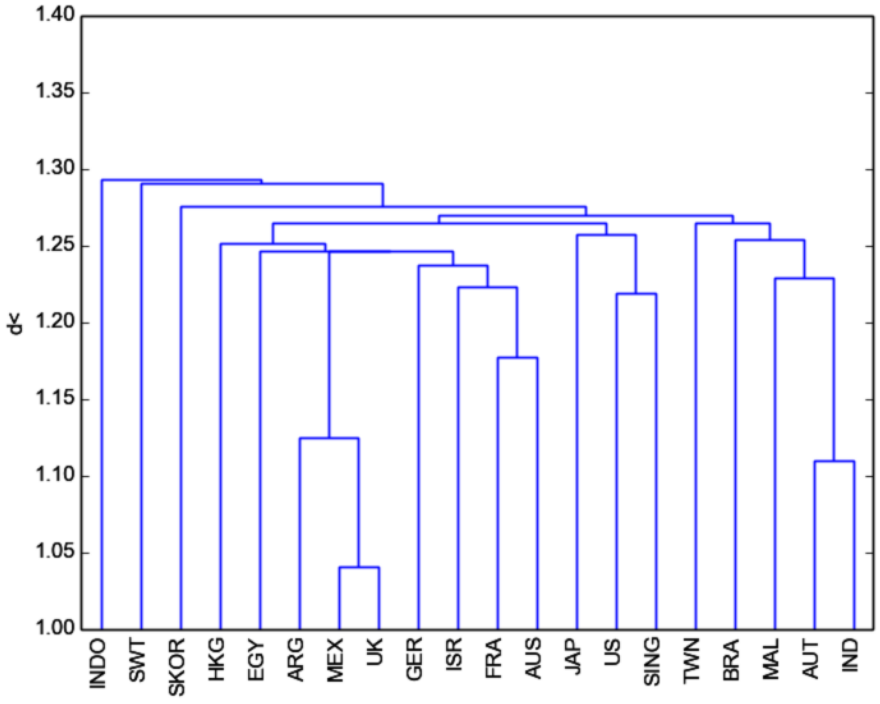

Figura 5.8: (a): MST de 5 nodos. (b): Ultramétrica subdominante asociada. (c): Dendrograma correspondiente.

muestran un patrón tan claro posiblemente a que se encuentran en un estado transitorio como ya hemos argumentado.

Para finalizar está sección, cabe resaltar que los datos de NYT muestran ligeramente una estructura más definida con respecto a los de datos Twitter. Sin embargo, las polaridades no tienen por que comportarse como los retornos, estas desviaciones que hemos encontrado pueden servir de indicios para construir una evidencia en contra de la hipótesis de mercado eficiente, al menos en alguna de sus versiones.

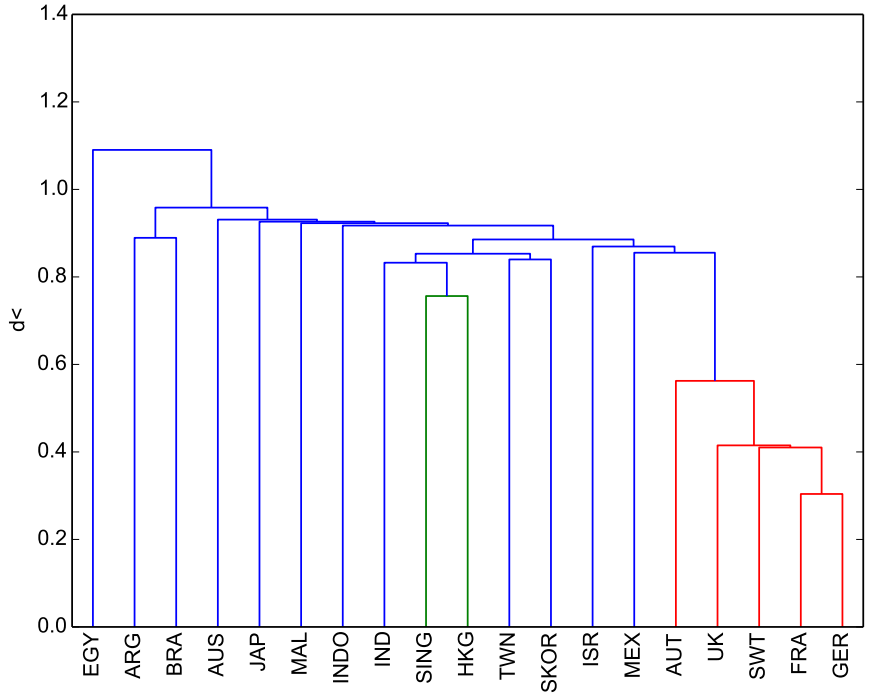



(a)

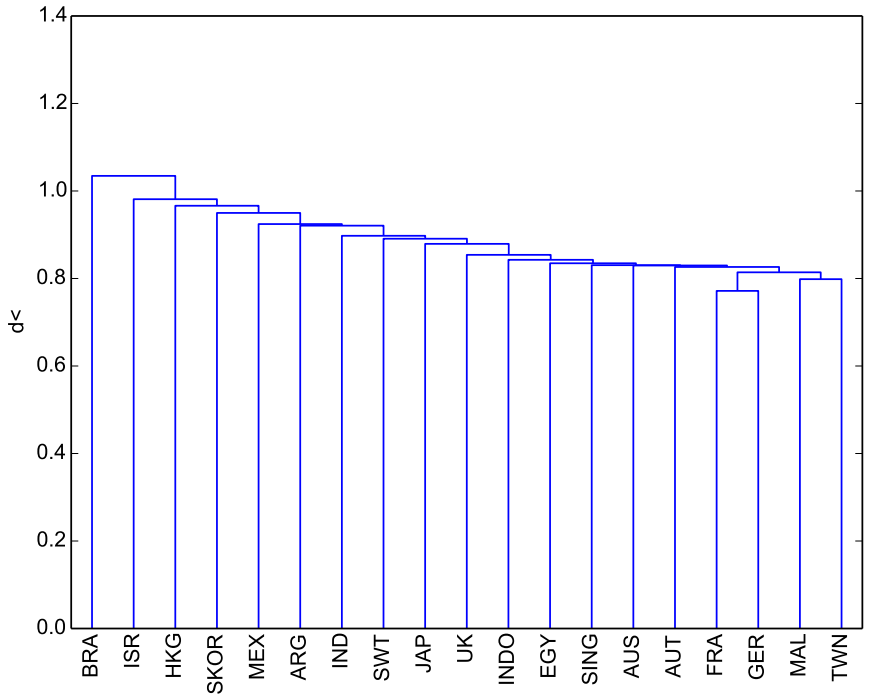



(b)

Figura 5.9: Dendrogramas. (a) Retornos. (b) Polaridades.




(a)




(b)

Figura 5.10: Dendrogramas. (a) Retornos. (b) Polaridades.

# Capítulo 6

# Teoría de Matrices Aleatorias

En este capítulo presentamos los fundamentos de la teoría de matrices aleatorias, así como sus resultados más relevantes en el contexto de la econofísica.

## 6.1. Fundamentos

La teoría de matrices aleatorias (RMT) es un nuevo tipo de mecánica estadística, donde en vez de tener un ensemble de estados gobernados por el mismo hamiltoniano [1], se tiene un ensemble de hamiltonianos gobernados por la misma simetría.

Esta teoría fue introducida en estadística matemática por Wishart en 1928 [42]. Muchos matemáticos trabajaron después en esta área por interés puramente teórico. En 1935 Élie Cartan clasifica los ensembles según la simetría que se conserva [43], mientras que el primer libro que contiene los resultados matemáticos más relevantes lo publicó L. K. Hua en 1959 [44]. En la década de 1950, Wigner utiliza RMT para lidiar con la estadística de eigenvalores y eigenvectores de sistemás complejos de muchos cuerpos en el contexto de la física nuclear [45, 46, 47, 48]. En ese dominio, RMT ha tenido aplicaciones exitosas para la descripción de las fluctuaciones espectrales de los núcleos atómicos, así como de átomos y moléculas complejas [49]. F.J. Dyson y M.L. Mehta hicieron cálculos analíticos detallados en la década de 1960, Mehta publicó en 1967 un libro en donde describe las principales técnicas desarrolladas hasta ese momento [46].

[1] En física, el hamiltoniano representa la energía del sistema, del cual pueden extraerse todas las propiedades dinámicas del mismo.

Pasaron varios años sin aplicaciones relevantes en el área de la física matemática, hasta que resurgió el interés por RMT cuando Casati et al., influenciado por las ideas de Michael Berry, publicaron la primera evidencia de que los resultados de RMT pueden ser aplicables a los espectros de los sistemás caóticos [50]. Sin embargo, la formulación más aceptada y general de este resultado, conocida como conjetura de caos cuántico, se publicó en un artículo de Bohigas, Giannoni y Schmidt en 1984 [51]. Actualmente las aplicaciones en física son inmensas, las cuales van desde la información cuántica hasta fenómenos de transporte en nanomateriales, surgiendo cada día nuevos aplicaciones en áreas emergentes, como es el caso que nos compete, es decir, el de la econofísica.

En lo que ahora se conoce como la versión clásica, la teoría de matrices aleatorias maneja tres ensembles gaussianos de matrices, cada uno para un distinto grupo de transformaciones canónicas. Estos ensembles se definen en términos de las propiedades de simetría del hamiltoniano [48]:

- Para sistemas con invariancia ante inversión temporal y simetría rotacional, la matriz hamiltoniana $H$ se puede elegir real y simétrica

$$H = H^T. \tag{6.1}$$

- Para los sistemas sin invariancia ante inversión temporal, las matriz $H$ es hermitiana:

$$H = H^\dagger. \tag{6.2}$$

- Para los sistemas con invariancia ante inversión temporal con espín $1/2$ y sin simetría rotacional, el hamiltoniano se escribe en términos de las matrices de Pauli $\sigma_\gamma$

$$H^0_{nm} I_2 - i \sum_{\gamma=1}^{3} H^{(\gamma)}_{nm} \sigma_\gamma, \tag{6.3}$$

  donde $H^0$ es simétrica y $H^\gamma$ son antisimétricas.

La densidad de probabilidad de encontrar una matriz particular dentro de uno de estos ensembles está dada por:

$$P_{N\beta}(H) \propto \exp\left(-\frac{\beta N}{\lambda^2} trH^2\right), \tag{6.4}$$

en donde las propiedades de simetría y las funciones de peso $P_{N\beta}(H)$ son invariantes bajo transformaciones ortogonales ($\beta = 1$), unitarias ($\beta = 2$) y simplécticas ( $\beta = 4$) del hamiltoniano, respectivamente. Es por ello que a estos ensembles se les conoce como $GOE$, $GUE$ y $GSE$ [2]. Por otro lado, cada miembro $H_\beta$ de estos ensembles puede representarse en la forma:

$$H_\beta = WEW^{-1}, \tag{6.5}$$

## 6.2. Ensemble de Wishart

Deseamos ahora introducir una herramienta fundamental para el análisis multivariante proveniente de RMT. Sea $A$ una matriz de dimensión $N$x$T$, cuyos elementos son variables gaussianas estadísticamente independientes con media cero y varianza fija. La matriz $H = AA^\dagger$ es conocida en RMT como matriz de Wishart y al ensemble generado con estas matrices como ensemble de Wishart (WE). Por construcción, estas matrices están formadas con series de tiempo no correlacionadas de longitud finita. La distribución conjunta de probabilidad del espectro de eigenvalores $\{\lambda_1, \ldots, \lambda_N\}$ de este ensemble esta dada por

$$P[\{\lambda_i\}] = C_{N,T} \exp\left[-\frac{\beta}{2}\sum_i^N \lambda_i\right] \prod_{i=1}^N \lambda_i^{\alpha\beta/2} \prod_{j<k} |\lambda_j - \lambda_j|^\beta, \tag{6.6}$$

donde $\alpha = (1 + T - N) - 2\beta$, y $C_{N,T}$ es una constante de normalización que puede ser calculada exactamente. Además si $N > T$, existen exactamente $N - T$ eigenvalores iguales a cero, y la distribución de los $T$ eigenvalores restantes esta dada por la misma distribución, pero intercambiando $N$ por $T$.

La densidad de probabilidad del espectro de eigenvalores para el caso $\beta = 1$ se puede resolver analíticamente en el límite $N \to \infty$, $T \to \infty$, con $Q = T/N(\geq 1)$, lo que se conoce como ley de Marčenko-Pastur [52]

[2] por sus siglas en inglés: Gaussian Orthogonal Ensemble, Gaussian Unitary Ensemble y Gaussian Symplectic Ensemble, respectivamente.

$$\rho(\lambda) = \frac{Q}{2\pi\sigma^2}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \tag{6.7}$$

la cual presenta las cotas $\lambda_- \leq \lambda \leq \lambda_+$, siendo $0$ fuera de este rango. Además, el eigenvalor más grande (más pequeño) de WE está dado por

$$\lambda_-^+ = \sigma^2(1 + 1/Q \pm 2\sqrt{1/Q}). \tag{6.8}$$

La cuestión de interés para nosotros es que ***sí no existen correlaciones entre las variables, entonces la distribución de los eigenvalores de la matriz de correlación debe de estar acotada dentro de la ley de Marčenko-Pastur***. Estas predicciones son conocidas como resultados universales de las matrices de Wishart, y constituyen la hipótesis nula de la ausencia de correlaciones entre las variables de estudio, en nuestro caso entre los indices financieros globales (ó entre las polaridades de Twitter asociadas).

Aunque los resultados universales de las matrices de Wishart son válidos únicamente para dimensiones asintóticas, compararlos con nuestros datos empíricos sigue siendo útil, ya que nos puede proporcionar indicios acerca de la presencia de correlaciones ocultas. Para este propósito, hemos construido un conjunto de matrices de correlación muestra a partir de ventanas de tiempo de $T = 80$ días de cotización, deslizándolas por un día. De esta manera, para el primer periodo de estudio hemos obtenido dos muestras de $M_1 = 86$ matrices de correlación, un conjunto para los valores de polaridad, y el otro para los de retorno; mientras que para el segundo periodo el conjunto estas muestras contiene $M_2 = 138$. Dentro de estos conjuntos, cada matriz de correlación tiene dimensiones $N\text{x}T = 20\text{x}80$, con $Q = T/N = 4$. Por lo que el espectro de eigenvalores está acotado entre los límites

$$\lambda_- = 0.25 \qquad \lambda_+ = 2.25, \tag{6.9}$$

y uno esperaría, como hipótesis nula, que la gran mayoría de los eigenvalores no presenten correlaciones y se encuentren dentro de estos límites.

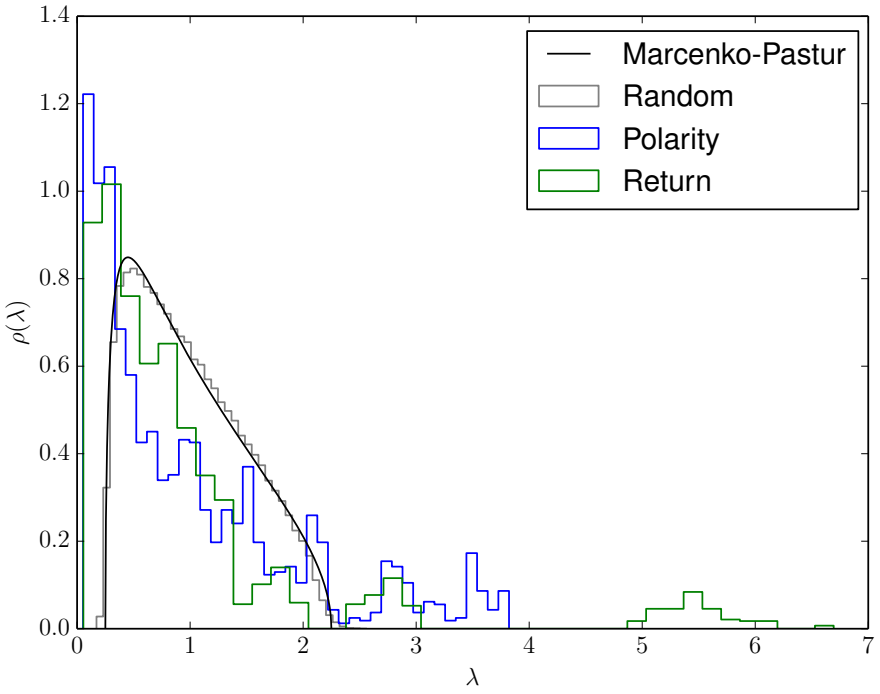
Para el primer periodo de estudio, se encontró que los eigenvalores extremos para el conjunto de matrices de correlación de polaridad son $\lambda^p_{min(max)} = 0.0526(3.8215)$, y para las de retornos $\lambda^r_{min(max)} = 0.0556(6.6942)$. Además, se encontró que solamente el $68.02\,\%$ de los eigenvalores de polaridad y $72.27\,\%$ de los de retorno caen dentro de los resultados universales de RMT, es decir, dentro de la zona asociada a ruido donde no se presentan

correlaciones. Con la intención de comparar estos resultados empíricos con datos que sabemos *a priori* que no presentan correlaciones, se realizó un cálculo numérico con 10000 miembros del ensemble de Wishart de la misma dimensión 20x80, donde se obtuvo que $\lambda^{p}_{min(max)} = 0.1716(2.5604)$, siendo ahora el $99.39\,\%$ de los eigenvalores los que caen dentro de la zona de ruido. El error cuadrático medio [3] entre los resultados teóricos y las distribuciones empíricas es de $0.3811$ y $0.4620$, para las polaridades y retornos respectivamente, mientras que para la realización numérica este error es igual a $0.1158$; mucho menor que los casos empíricos. La distribución de eigenvalores para los datos empíricos, así como del test numérico se han graficado en la fig. 6.1(a), superponiendo la ley Marčenko-Pastur en la misma figura.

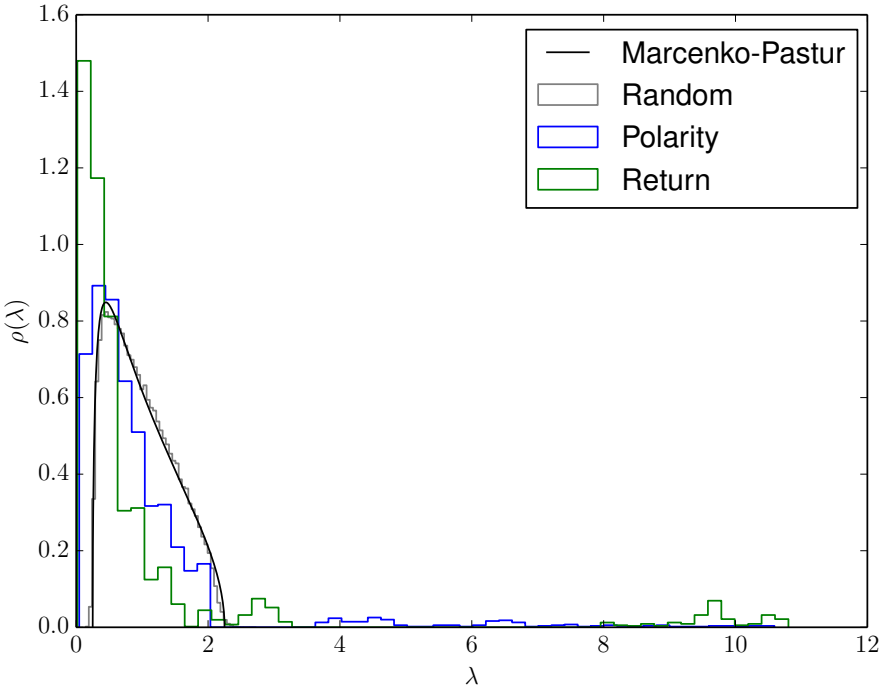
Para el segundo periodo de estudio, se encontró que los eigenvalores extremos para las correlaciones de polaridad son $\lambda^{p}_{min(max)} = 0.04569(10.5907)$, y para las de retornos $\lambda^{r}_{min(max)} = 0.0175(10.8026)$, donde ahora el $80.04\,\%$ de los eigenvalores de polaridad y $56.49\,\%$ de los de retorno caen dentro de los resultados universales de RMT. Con la simulación numérica de matrices aleatorias se obtuvo $\lambda^{p}_{min(max)} = 0.1424(2.7230)$, con $99.40\,\%$ de los eigenvalores en la zona de ruido. Se encontró un error cuadrático medio, entre la distribución de Marčenko-Pastur y las distribuciones empíricas, de $0.4931$ y $0.5511$, para las polaridades y retornos respectivamente, mientras que para la simulación este error es $0.1572$; nuevamente mucho menor que los casos empíricos. La distribución de eigenvalores de los datos de NYT, retornos, así como de la simulación, se han graficado en la fig. 6.1(b).

Más allá de los resultados que hemos encontrados, es importante recordar que la distribución de Marčenko-Pastur es válida solamente para el límite asintótico $N \to \infty$, $T \to \infty$, por lo que las distribuciones finitas siempre presentan desviaciones a este resultado. Además, entre más grande es el valor $T/N$, más confiable son los resultados, y las fluctuaciones son descritas de forma más realista por la varianza de los datos. Pero si $T/N$ es un número pequeño, los resultados serán afectados fuertemente por la finitud de la matriz de datos. En estos casos es una práctica común utilizar técnicas de *noise dressing* para omitir el ruido intrínseco de la matriz de correlación [53, 54, 55]. Sin embargo, estas técnicas funcionan bien hasta dimensiones cercanas a $N = 50$, para dimensiones más pequeñas (que es nuestro caso) debemos proceder de manera diferente.

[3] RMSE, por sus siglas en inglés




(a)



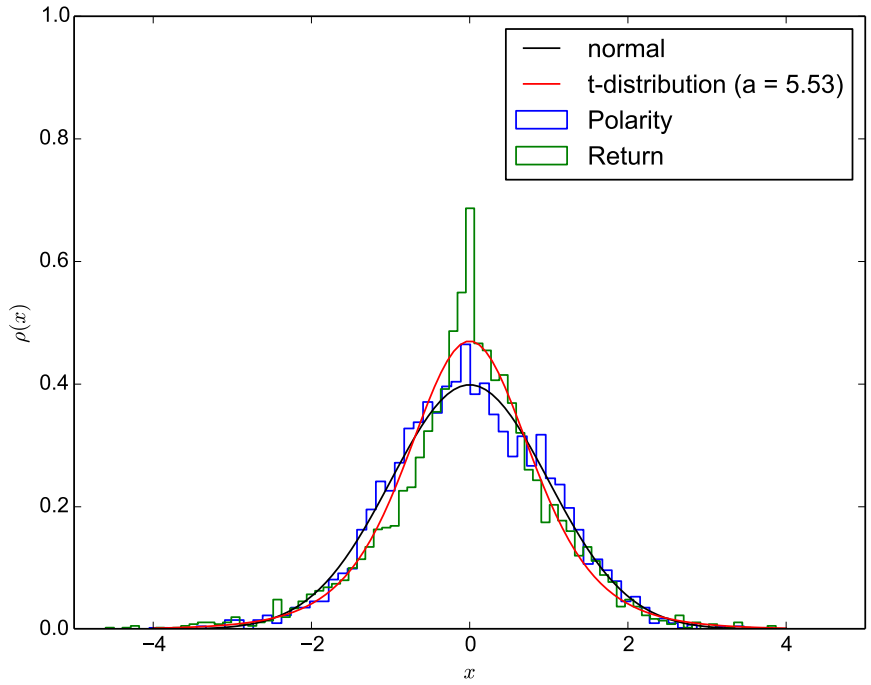

(b)

Figura 6.1: Distribución de eigenvalores de las matrices de correlación. La línea negra muestra la ley de Marčenko-Pastur. La línea gris representa los resultados numéricos para 10000 miembros de WE, la línea azul los resultados para las polaridades, y la línea verde para los retornos. (a) Resultados para Twitter e índices financieros. (b) Resultados para NYT e índices financieros.

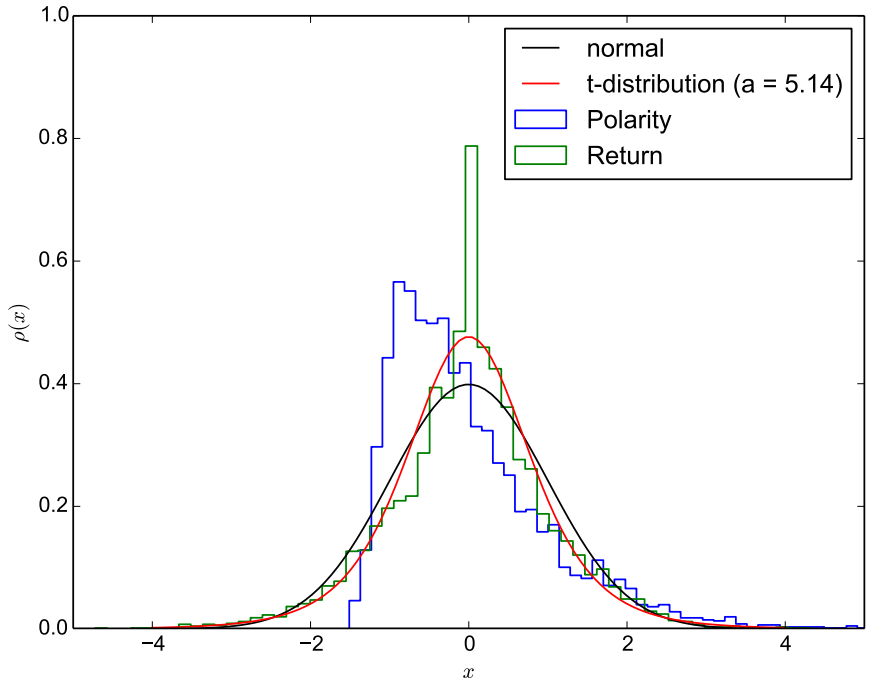
Además de esto, otro hecho que puede generar desviaciones de los resultados universales de las matrices de Wishart se debe a que la distribución de los retornos usualmente tienen colas más largas que la distribución normal [1], la cual se asume de manera idealizada en la derivación de la ley de Marčenko-Pastur. Para observar este fenómeno, en la fig. 6.2 se ha graficado la distribución de los datos empíricos junto con la distribución normal y la distribución t-Student para caracterizar la distribución de los retornos. Para el primer periodo de estudio (fig. 6.2(a)), el parámetro $a = 5.53$ fue el que mejor se ajustó al caracterizar la distribución de los retornos, mientras la distribución de las polaridades parece ajustarse mucho mejor con la distribución normal en este caso. Para el segundo periodo (fig. 6.2(b)), el ajuste de los retornos se obtuvo con el parámetro $a = 5.14$, pero en este caso las polaridades presentan una distribución asimétrica. En ambos casos, los datos de polaridad parecen romper la regla de colas largas encontrada por muchos autores para los retornos, pero será necesario más evidencia empírica para confirmar este nuevo resultado.

Nos interesa ahora aplicar el test de Kolmogorov-Smirnov para cuantificar que tan bien se ajustan las distribuciones empíricas de eigenvalores respecto a la distribución de Marčenko-Pastur. Este test en su versión *one-sample*, compara la muestra empírica con una distribución de probabilidad de referencia. El test cuantifica una medida de distancia entre la distribución de las muestras y la función de distribución acumulada de referencia. La hipótesis nula en este caso es calculada bajo el supuesto de que la muestra es generada a partir de la distribución de referencia, en nuestro caso, la distribución de Marčenko-Pastur.

Al aplicar el test de Kolmogorov-Smirnov, se ha encontrado que rechaza la hipótesis de que las distribuciones de eigenvalores empíricas se generaron a partir de la distribución de Marčenko-Pastur, esto con una confiabilidad mayor al 99.99 % en todos los datos empíricos. Por lo tanto, no podemos asegurar la existencia de correlaciones verdaderas con los resultados asintóticos de las matrices de Wishart. Las desviaciones que se observan a las cotas predichas por los resultados universales de RMT podrían no deberse a la presencia de correlaciones, sino a que el modelo no se ajusta bien a estos casos debido que las dimensiones $N$ y $T$ son muy pequeñas. Un manera de profundizar un poco más en este problema y aclarar la existencia de correlaciones ocultas es utilizando los nuevos resultados para las matrices de correlación modelo de Wishart, los cuales son aplicables a dimensiones pequeñas.




(a)




(b)

Figura 6.2: Distribución de datos empíricos, donde se ha superpuesto la distribución normal y la distribución t-Student con el parámetro $a$ que mejor ajusta a los valores de retornos. (a) Twitter e índices financieros. (b) NYT e índices financieros

## 6.3. Matrices de correlación modelo de Wishart

Un ensemble de matrices de correlación modelo $WW^{\dagger}/N$ fluctuando alrededor de la matriz de correlación empírica $C$ está dado por

$$\frac{1}{T}\langle WW^{\dagger}\rangle = C. \tag{6.10}$$

Para ajustar las matrices de datos modelo $W$, y por lo tanto las matrices de correlación modelo $WW^{\dagger}/T$, $W$ debe ser una matriz rectangular de dimensiones $N$x$T$, con elementos $W_{ij} \in R$, para que $C$ satisfaga la condición de ser una matriz real y simétrica. Los elementos de matriz $W_{ij}$ se generan aleatoriamente siguiendo una distribución gaussiana con varianza $C_{ji}$, de tal manera que la distribución de probabilidad condicional está dada por [56, 57, 58, 59]:

$$P(W|C) = \frac{\exp(\frac{-\beta}{2}TrWW^{\dagger}C^{-1})}{(\beta/2\pi)^{NT\beta/2}det^{NT/\gamma_1}(C)} \tag{6.11}$$

donde $\gamma_1 = 2$ para $\beta = 1$ (nuestro caso real y simétrico).

Mediante el uso de la técnica de supersimetrías [60, 61], Recher et. al. [62] encontraron una expresión para el caso real ($\beta = 1$) de la eq. 6.11 en términos de integrales sobre variedades de dos dimensiones y sumas finitas. Sin embargo, debido a la complejidad de esta expresión y la escasa literatura que hay acerca de ello, nuestra aproximación será numérica. Para este fin, hemos hecho una simulación de Monte Carlo [63] usando la ec. 6.10 con 1000 matrices modelos para el primer periodo de tiempo (Twitter) y 500 para el segundo periodo (NYT), esto para cada una de las matrices de correlación muestra generadas por el deslizamiento de ventana descrito en la sección anterior. En la fig. 6.3 mostramos los resultados para las matrices de correlación de retornos (figura superior) y polaridades (figura inferior), respectivamente, para el primer periodo de tiempo. Asimismo en la fig. 6.4, se muestran los resultados correspondientes para el segundo periodo de tiempo. Se puede ver ahora, a diferencia de la fig. 6.1, que hay una mucho mejor concordancia entre las distribuciones empíricas y las generadas numéricamente a partir del ensemble descrito por la ec. 6.10. Algo que vale la pena resaltar es que en el segundo periodo de estudio, se observan eigenvalores un orden de magnitud mucho más grandes que la cota superior predicha por la ley de Marčenko-Pastur, lo cual nos dice que la dinámica del sistema puede ser caracterizada por un sólo factor, es decir, aunque usemos un modelo multifactores para

caracterizar las correlaciones, persiste un modo normal del sistema dado por el eigenvalor más grande que captura toda la dinámica del mercado. Sin embargo, la razón principal de tan buen ajuste es debido a la característica del ensemble, ya que es un modelo que utiliza los eigenvalores empíricos como parámetros de entrada, por lo que es natural que se adapte muy bien en promedio a las fluctuaciones del espectro empírico. Así, aunque este método es capaz de caracterizar la densidad de eigenvalores para dimensiones pequeñas, no es posible extraer información valiosa más allá del buen ajuste de todo el espectro.

## 6.4. Aproximación no-simétrica

Nuestro propósito es discriminar si las correlaciones observadas se deben a ruido por la finitud de la matriz o si por el contrario contienen información valiosa, por lo que se debe considerar una aproximación no-simétrica del modelo correlacionado de Wishart para matrices ortogonales, el cual es mejor conocido como CWOE por sus siglas en inglés [64]. CWOE es un ensemble de matrices reales y simétricas del tipo $C = WW^t/T$ , donde ahora $\mathcal{W} = \xi^{1/2}W$, siendo $\xi$ una matriz definida positiva la cual contiene la información de las correlaciones. Nuevamente, los elementos de la matriz $W$ son variables aleatorias independientes con distribución normal, media cero y la misma varianza, las cuales caracterizan el ruido blanco.

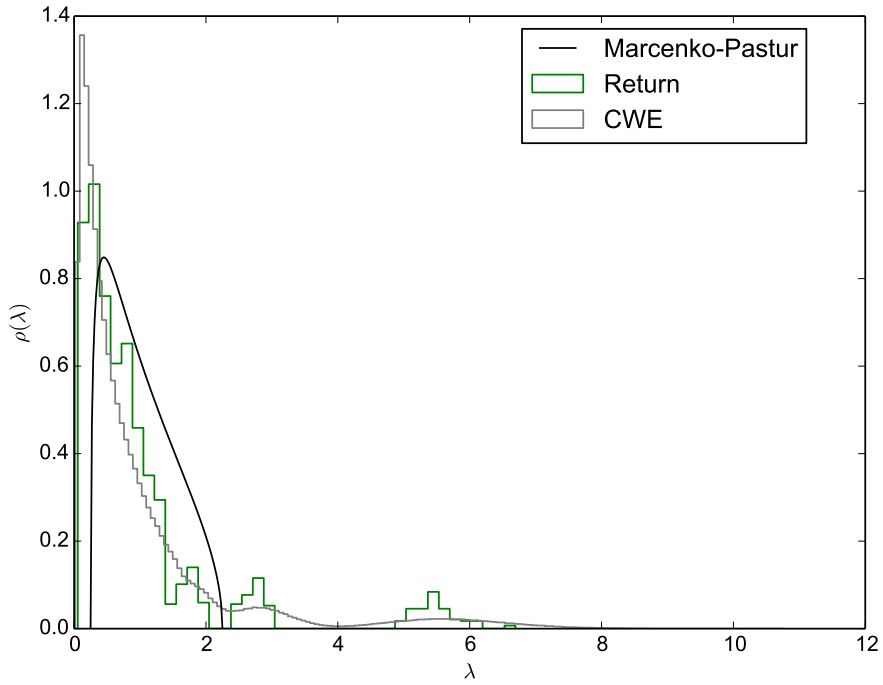
Para comenzar con esté análisis, definamos $D^{(r)}$ y $D^{(p)}$ como las matrices de datos compuestas de las series de polaridad y retornos, respectivamente. Podemos entonces construir la supermatriz empírica de datos de dimensión $2N\text{x}T$ de la siguiente manera

$$D = \begin{pmatrix} D^{(r)} \\ D^{(p)} \end{pmatrix}, \tag{6.12}$$

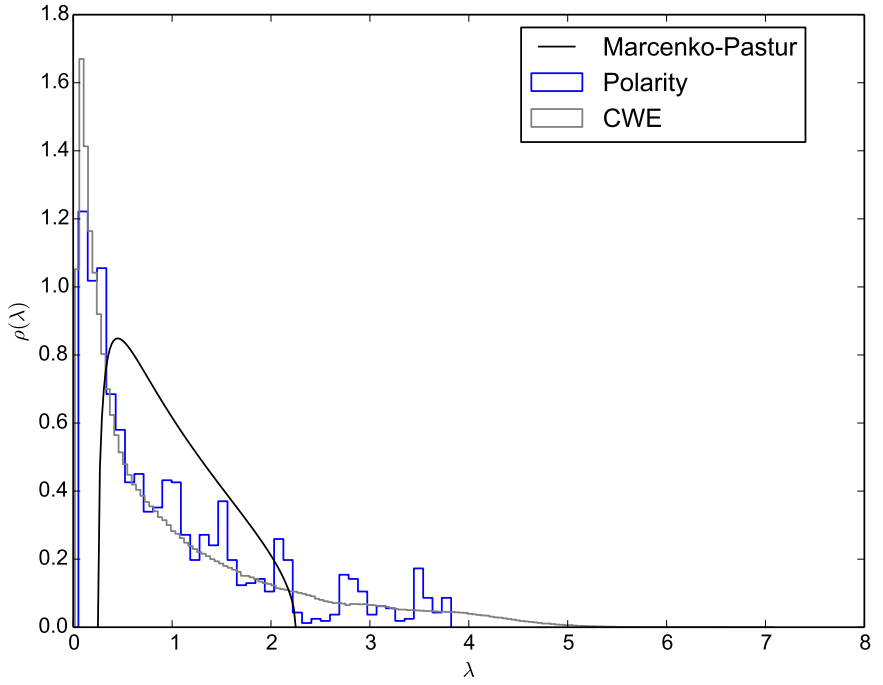
y consecuentemente tenemos que la supermatriz de correlación está dada por

$$C = \frac{1}{T}DD^T = \begin{pmatrix} C^{(r)} & C^{(r,p)} \\ C^{(p,r)} & C^{(p)} \end{pmatrix} \tag{6.13}$$

donde $C^{(r,p)} = C^{(p,r)T} = D^{(r)}D^{(p)T}$. Ahora la matriz $C$ está compuesta de cuatro bloques. Los elementos diagonales caracterizan las correlaciones de las series de tiempo de retornos y polaridades de manera independiente, mientras que los bloques fuera de la diagonal consideran las correlaciones mixtas entre los dos indicadores.

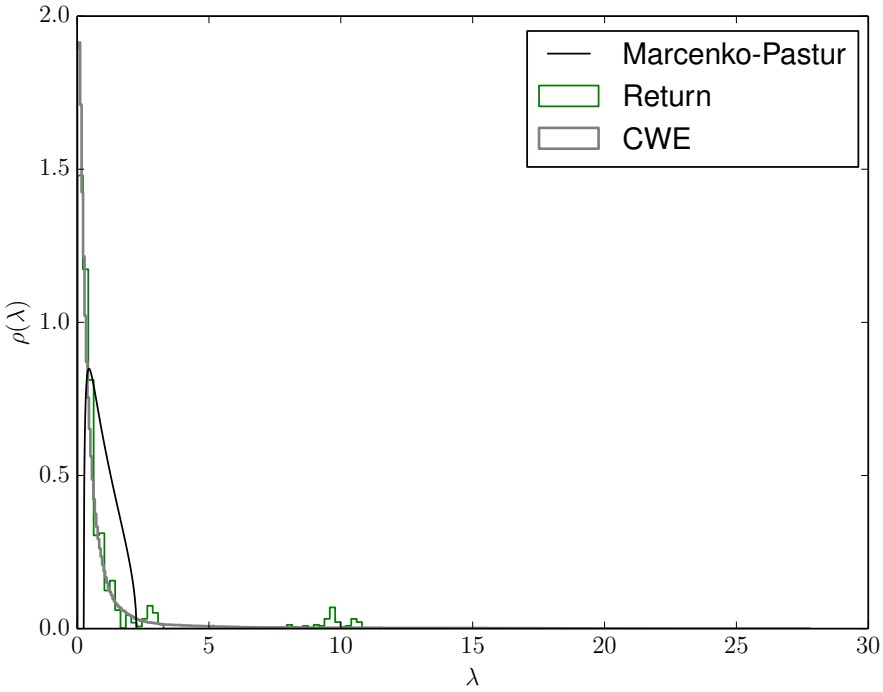



(a)

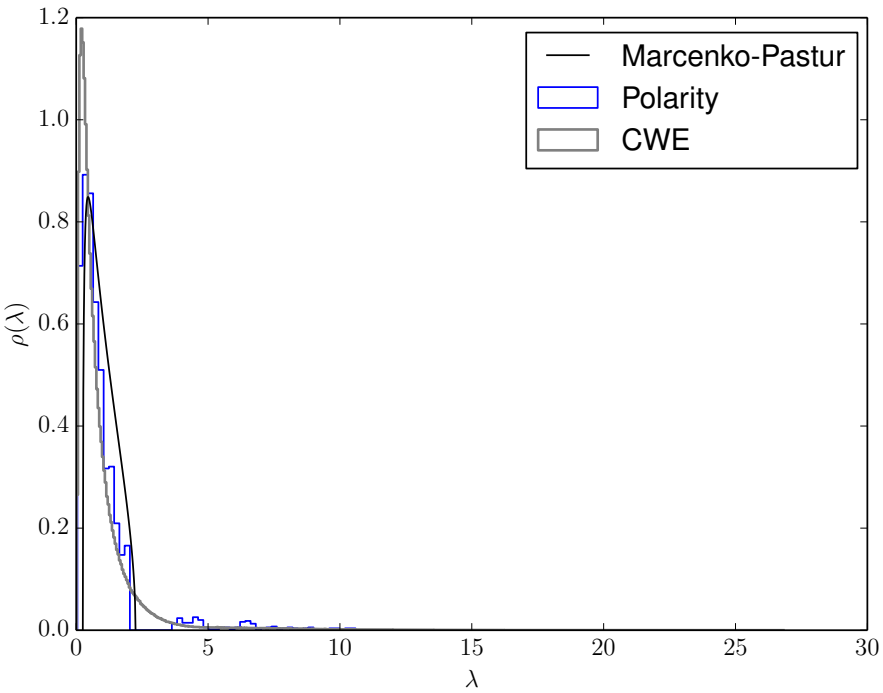



(b)

Figura 6.3: Distribución de eigenvalores para Twitter e índices financieros. (a) La línea verde representa los resultados para los retornos, mientras que la línea gris los resultados del modelo correlacionado de Wishart. (b) Se muestran los mismos resultados que arriba, pero para las polaridades. En ambas figuras la línea negra representa la ley de Marčenko-Pastur, válida para dimensiones asintóticas.




(a)




(b)

Figura 6.4: Distribución de eigenvalores para NYT e índices financieros. (a) La línea verde representa los resultados para los retornos, mientras que la línea gris los resultados del modelo correlacionado de Wishart. (b) Se muestran los mismos resultados que arriba, pero para las polaridades. En ambas figuras la línea negra representa la ley de Marčenko-Pastur, válida para dimensiones asintóticas.

Sea ahora $\xi$ alguna de nuestras matrices de correlación empírica $C$. De esta manera, $\xi^{(r)} = C^{(r)}$, $\xi^{(p)} = C^{(p)}$, y considerando que $W_1, W_2 \in \mathbf{R}^{NxT}$ sean dos elementos del ensemble de Wishart, podemos definir la supermatriz empírica de datos con ruido blanco $\mathcal{W}$ de dimensión $2N$x$T$, construida por la concatenación de los datos de retornos y polaridades, de la siguiente forma

$$\mathcal{W} = \begin{pmatrix} \sqrt{C^{(r)}}W_1 \\ \sqrt{C^{(p)}}W_2 \end{pmatrix}. \tag{6.14}$$

Por lo tanto, la supermatriz de correlación con ruido blanco resulta ser

$$C' = \frac{1}{T}\mathcal{W}\mathcal{W}^{\dagger} = \begin{pmatrix} C^{(r)}W_1W_1^t & C^{(r,p)}W_1W_2^t \\ C^{(p,r)}W_2W_1^t & C^{(p)}W_2W_2^t \end{pmatrix}. \tag{6.15}$$

Nuestro objetivo es entonces comparar las matrices $C$ y $C'$, para saber si la estructura de la supermatriz empírica es robusta ante la adición de ruido blanco, y así lograr confirmar o refutar definitivamente la presencia de correlaciones verdaderas, es decir, que no se deban a la finitud de los datos [65, 64]. Para esto, hemos tomado de nuevo ventanas de tiempo de $T = 80$ días para las matrices de correlación empíricas, y las hemos deslizado por 40 días, generando así tres muestras temporales para el primer periodo de estudio, y cuatro muestras para el segundo.

Para el primer periodo de estudio, vemos en la fig. 6.5 que no hay cambios significativos entre las matrices originales y a las que se les ha añadido ruido blanco. En la fig. 6.6 vemos los resultados correspondientes para el segundo periodo, donde se observa que la intensidad de las correlaciones aumenta a partir de la segunda ventana de tiempo, sin embargo la estructura de las correlaciones parece no alterarse. Para cuantificar la diferencia entre las supermatrices con y sin ruido, se ha calculado el promedio de la diferencia absoluta entre primeros vecinos de los elementos de matriz [4]. Al cuantificar este valor, se encontró que el resultado difiere menos del 10 % entre las matrices originales y las que se le añadió ruido blanco, cumpliéndose esto para cada ventana de tiempo de los dos periodos de estudio, es decir, para los datos de Twitter, NYT, y los indices financieros de cada periodo respectivo.

[4] Se ha aplicado aquí la misma medida utilizada por Schäfer et. al. para comparar sus matrices de correlación antes y después de aplicar su técnica de normalizacíón local [66].

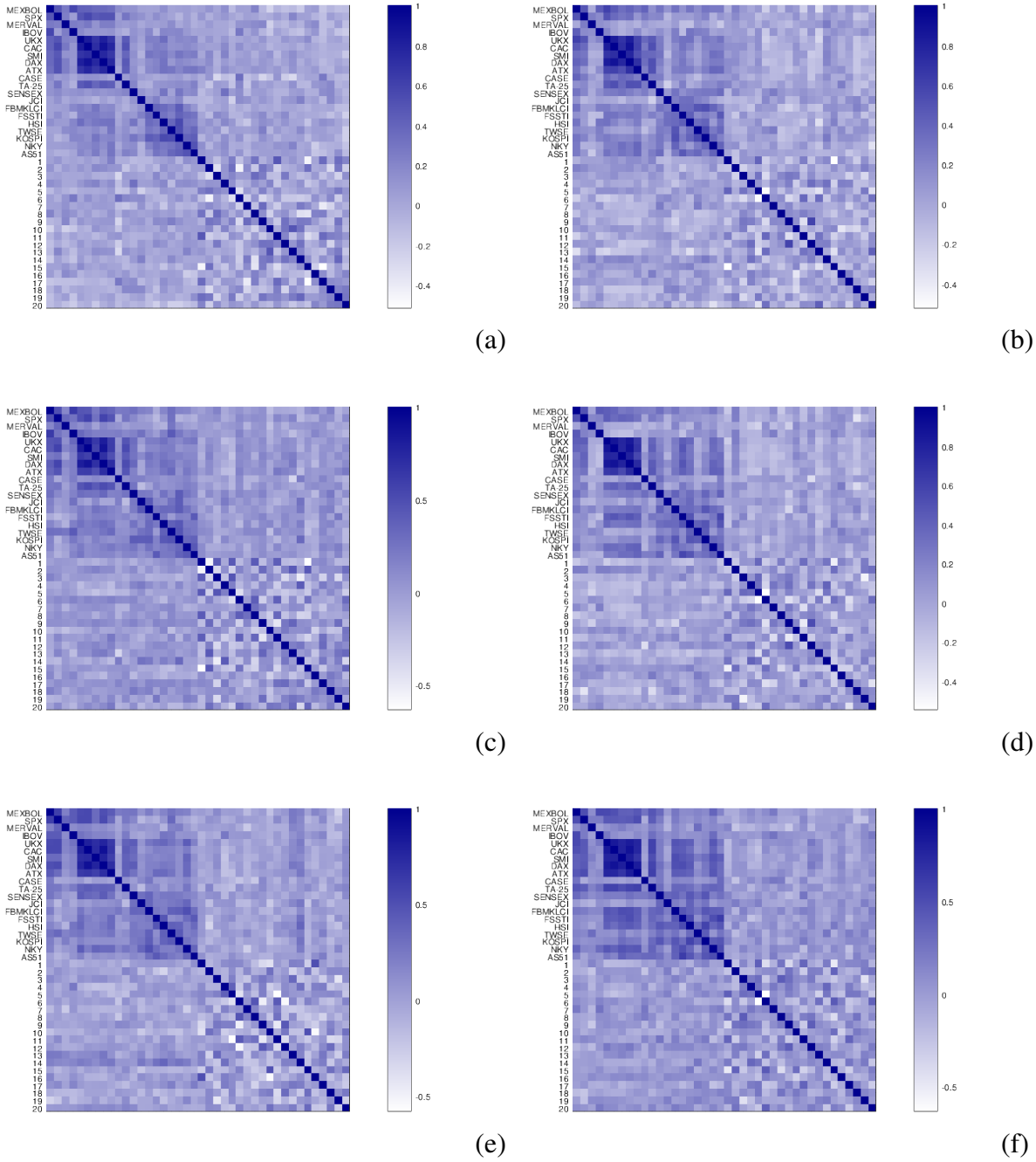



Figura 6.5: Matrices de correlación $C$ y $C'$ para las ventanas de tiempo considerados en el primer periodo de estudio: Twitter e índices financieros. Las figuras de las izquierda representan $C$, y las de la derecha $C'$. (a) y (b) son para los primeros 80 días, (c) y (d) del día 41 al 120, y la última hilera (g) y (h) del día 121 al 160.

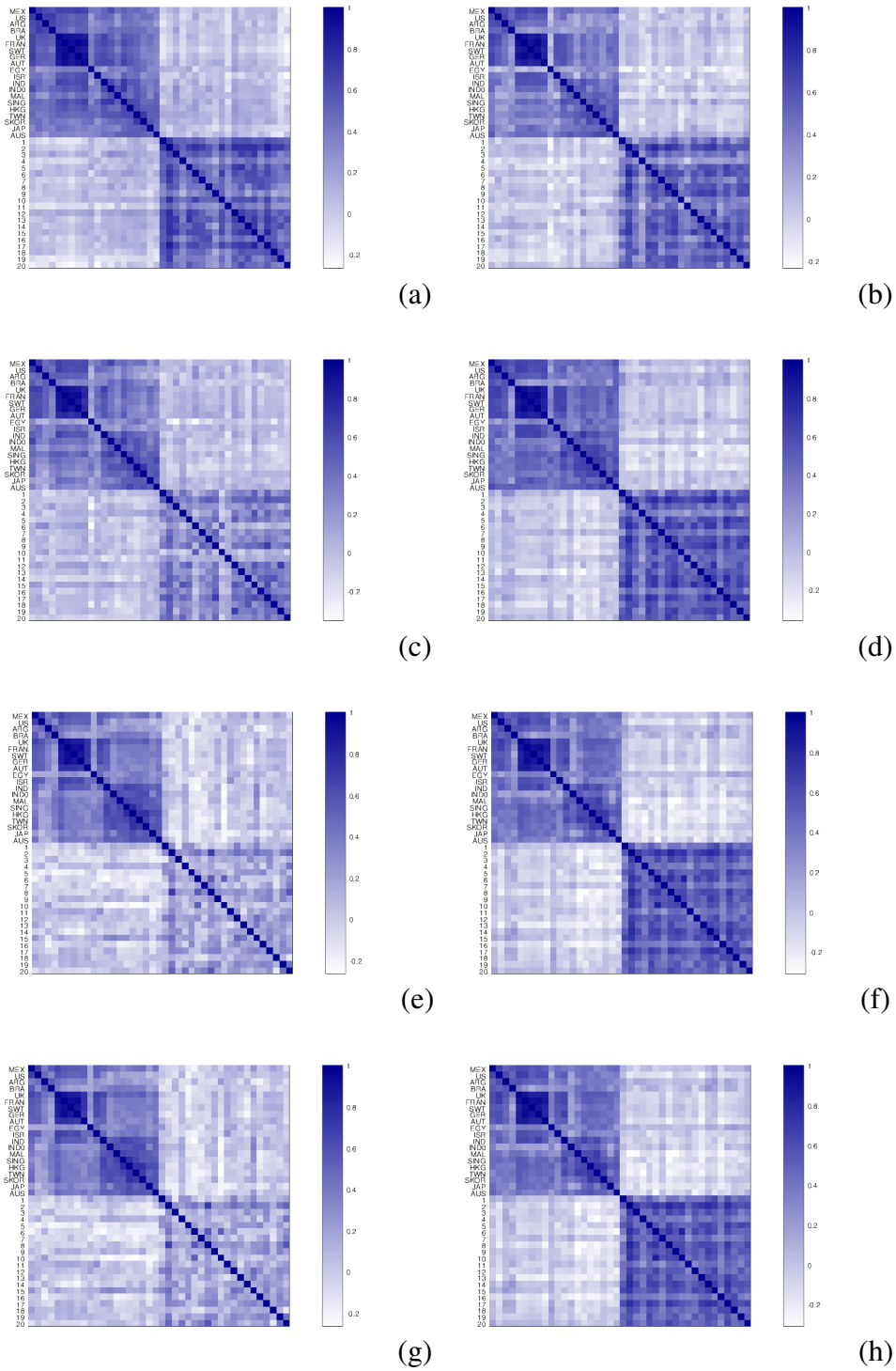


Figura 6.6: Matrices de correlación $C$ y $C'$ para las ventanas de tiempo consideradas en el segundo periodo de estudio: NYT e índices financieros. Las figuras de las izquierda representan $C$, y las de la derecha $C'$. (a) y (b) son para los primeros 80 días, (c) y (d) del día 41 al 120, (e) y (f) del día 81 al 160, y la última hilera (g) y (h) del día 161 al 200.

Estamos ahora finalmente en la posición de argumentar que la estructura de las correlaciones contienen mucha más información que la de ruido blanco, lo cual es válido para las correlaciones no-mezcladas de los bloque diagonales, así como para las correlaciones entre las series de tiempo de polaridad y retornos dadas por los bloques no-diagonales. Además estas correlaciones son robustas, pues se preservan incluso al añadir ruido blanco. Por lo tanto, estos resultados confirman la existencia de correlaciones verdaderas entre los índices financieros, las polaridades, y la mezcla de estos dos indicadores, al considerar ya sea datos de Twitter o de NYT para cada periodo de estudio respectivo.

## 6.5. Eigenvalores extremos

En el área de riesgo financiero y optimización de portafolios, los eigenvalores más grandes y pequeños representan cantidades muy importantes, ya que están asociados con los casos extremos de riego en una cartera de inversión [4]. Los eigenvalores más grandes corresponden a una mezcla arriesgada de acciones o mercados financieros, mientras que los eigenvalores más pequeños están relacionados con un portafolio de bajo riesgo [67]. El eigenvalor más grande es el factor que representa la información colectiva de los indices, y el eigenvector correspondiente es conocido como el *modo del mercado*. Este eigenvector nos dice si los indices como conjunto van a la alza o a la baja, siendo su tendencia condicionada al estado actual del mercado [67].

Estas características han sido exploradas exhaustivamente con datos provenientes de los indices financieros, sin embargo hasta el momento no se ha hecho un estudio con datos textuales. Es por ello que aquí se explora si este fenómeno también emerge al trabajar con la información proveniente de Twitter y de NYT. Para este fin se analizó el comportamiento temporal del eigenvalor más grande y más pequeño de las matrices de correlación empíricas para cada periodo de estudio, utilizando la misma muestra de matrices construidas en la sección 6.2.

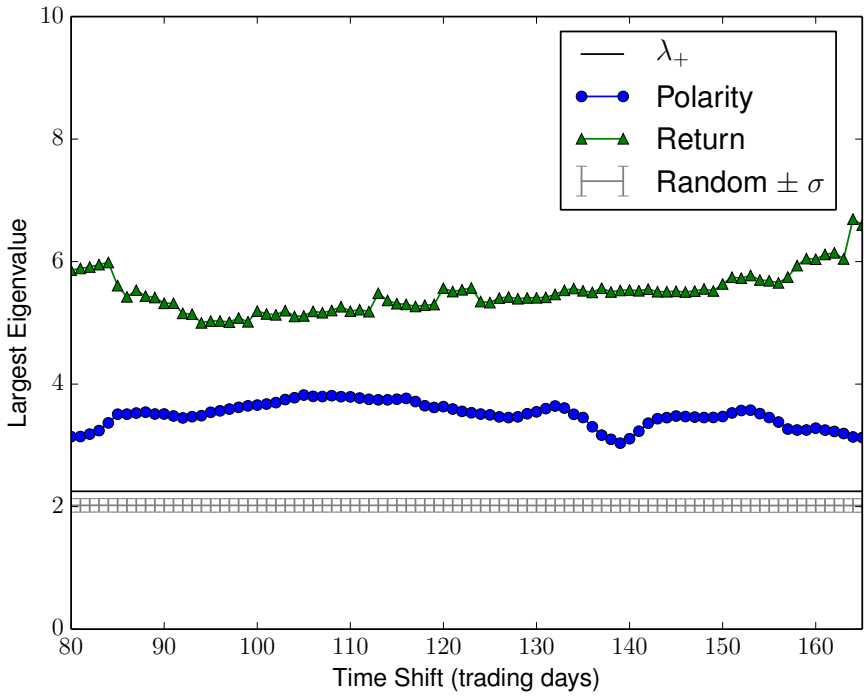
En la fig. 6.7 se muestran los resultados empíricos para el primer periodo de estudio, junto con la media y desviación estándar de una simulación numérica de 10000 miembros de WE, donde cada punto se calcula teniendo en cuenta los 80 días de transacción anteriores. Se puede observar que los resultados empíricos están lejos de los bordes teóricos, así como a más de tres desviaciones estándar de los resultados numéricos. Se encontró además

una anti-correlación fuerte entre polaridades y retornos al comparar el comportamiento temporal de sus eigenvalores más grandes. El coeficiente de Pearson [34] encontrado fue $P_c = -0.70$, el de Spearman [68] $S_c = -0.69$, ambos con valores de confidencia (valor p) menores a $1\text{x}10^{-12}$. Por el contrario, el comportamiento temporal de los eigenvalores más pequeños muestran correlaciones positivas más moderadas, con valores de $P_c = 0.45$ y $S_c = 49$.

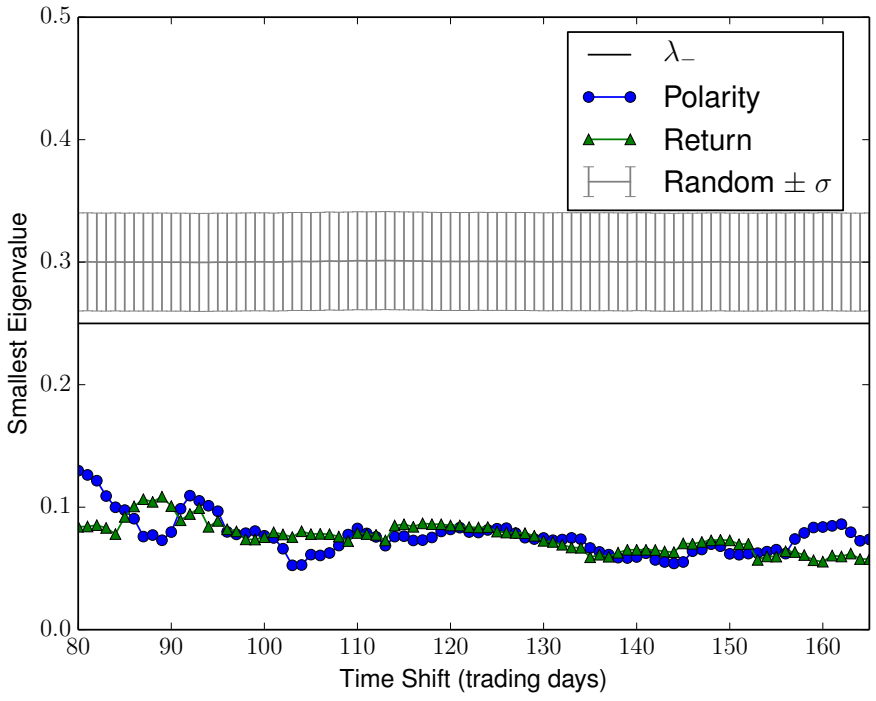
Los resultados respectivos para el segundo periodo de tiempo se pueden ver en la fig. 6.8, donde ahora la simulación numérica se realizó sólo con 1000 miembros de WE debido a que las series de tiempo son más largas. Se puede observar de nuevo que los resultados para los retornos están lejos de los bordes teóricos, así como a más de tres desviaciones estándar de los resultados numéricos, sin embargo esto no se cumple para todos los eigenvalores más pequeños en este caso. Se encontró ahora una correlación moderada en el comportamiento temporal de los eigenvalores más grandes, con un $P_c = 44$, y $S_c = 41$, ambos con valores $p$ menores a $1\text{x}10^{-7}$. Por el contrario, el comportamiento temporal de los eigenvalores más pequeños no muestra una correlación clara; el valor de $p$ es alto.

El hecho de que para el primer periodo de estudio el comportamiento temporal de los eigenvalores más grandes de los retornos y polaridades estén anti-correlacionados mutuamente, puede deberse a un retraso en la transmisión de información de Twitter hacia los precios de los mercados financieros globales. Siendo esta una evidencia más en contra de la ampliamente aceptada hipótesis de mercado eficiente. Además para este mismo periodo de estudio, el coeficiente de correlación que se encontró para el comportamiento temporal de los eigenvalores empíricos más pequeños revela que el portafolio de menor riesgo se preserva aproximadamente sin importar si usamos polaridades o retornos para su cálculo, por lo que Twitter resulta ser una fuente de información muy interesante para el análisis de portafolios. Por otro lado los resultados del comportamiento temporal de los eigenvalores más grande del NYT parecen mostrar que la polaridad de la información se refleja más rápidamente en el comportamiento global de los retornos, mientras que los portafolios de menor riesgo siguen una dinámica distinta, sin embargo la mayor parte del tiempo están lejos de un comportamiento aleatorio.

En general, estos resultados proveen evidencias acerca del surgimiento de factores comunes en la información financiera global, sin la necesidad de discriminar si los datos provienen de los retornos o de las polaridades, en otras palabras, con la información pro-

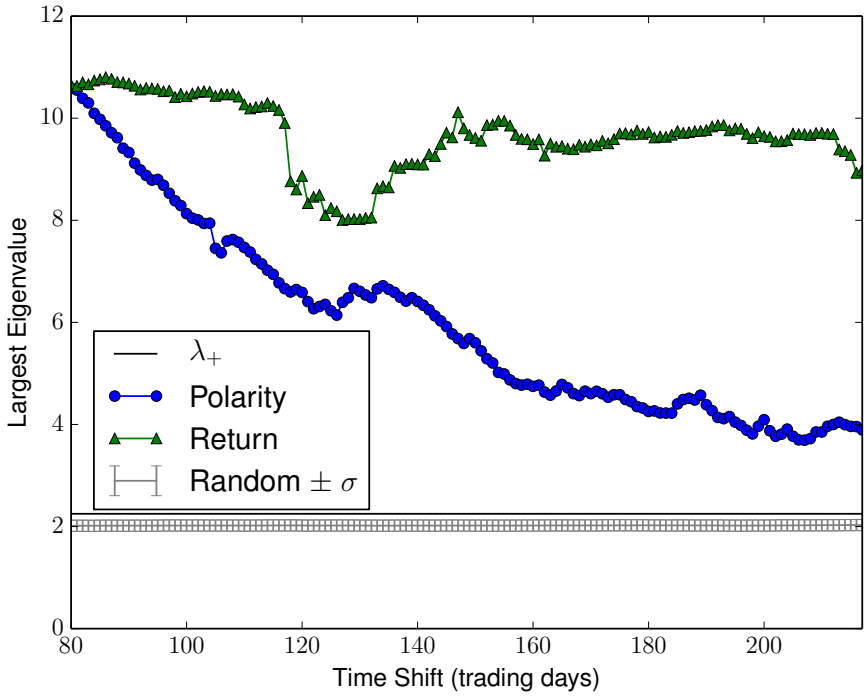



(a)

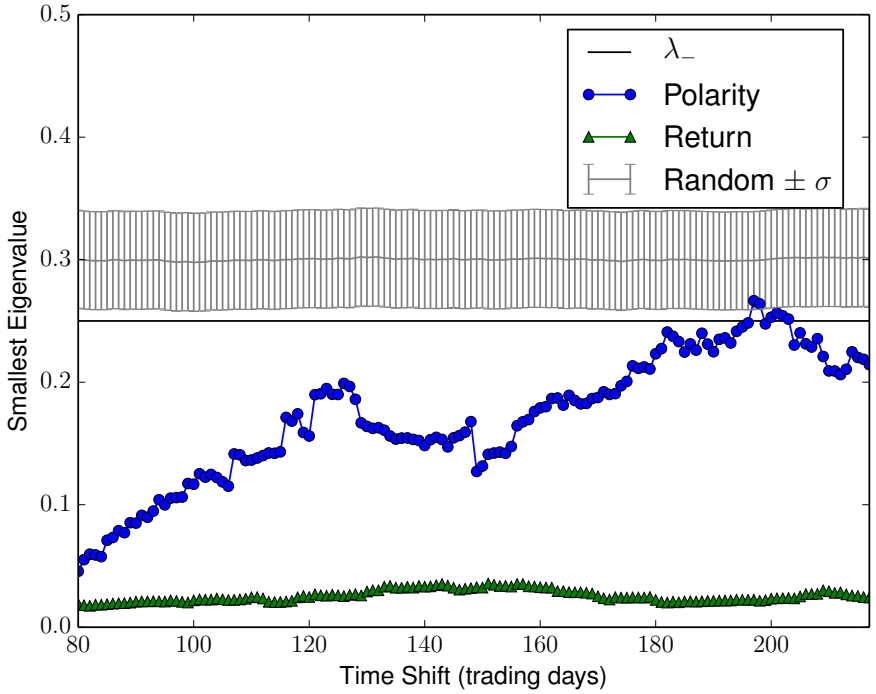



(b)

Figura 6.7: Eigenvalores extremos para Twitter e índices financieros. (a) Comportamiento temporal de los eigenvalores más grandes. (b) Comportamiento temporal de los eigenvalores más pequeños. La línea azul representa los resultados para las polaridades, la verde para retornos, y la línea negra los límites predichos por RMT para las matrices de Wishart, mientras que la línea gris representa la media y desviación estándar para una simulación numérica con 10000 miembros de WE.




(a)




(b)

Figura 6.8: Eigenvalores extremos para NYT e índices financieros. (a) Comportamiento temporal de los eigenvalores más grandes. (b) Comportamiento temporal de los eigenvalores más pequeños. La línea azul representa los resultados para las polaridades, la verde para retornos, la línea negra los límites predichos por RMT para las matrices de Wishart, mientras que la línea gris representa la media y desviación estándar para una simulación numérica con 1000 miembros de WE.

veniente de Twitter y NYT parecer ser posible caracterizar el comportamiento colectivo de los mercados financieros globales.

## 6.6. Tracy-Widow

El surgimiento de eigenvalores fuera de los bordes afilados de la distribución de Marčenko-Pastur es una fuerte indicación en contra de la hipótesis nula de la no existencia de correlaciones entre las variables de estudio. No obstante, esto es válido únicamente en el límite asintótico $N \to \infty$, $T \to \infty$. Para dimensiones grandes pero finitas, la probabilidad de encontrar un eigenvalor fuera de los bordes asintóticos es muy pequeña, pero diferente de cero [67]. Se ha demostrado que si $C$ es una matriz de Wishart real y si $\lambda_{max}$ es su eigenvalor más grande, entonces para $N, T$; grande pero finito, tal que $T/N \to \in [0, \infty]$

$$\frac{T\lambda_{max} - \mu_{\text{NT}}}{\sigma_{\text{NT}}} \xrightarrow{\mathcal{D}} \mathcal{TW}_1 \tag{6.16}$$

donde $\mathcal{TW}_1$ denota una variable aleatoria con distribución Tracy-Widom de orden uno [69, 70, 71]. Los parámetros de escalamiento y centrado están dados por

$$\begin{aligned} \mu_{\text{NT}} &= \left(\sqrt{N - 1/2} + \sqrt{T - 1/2}\right)^2 \\ \sigma_{\text{NT}} &= \sqrt{\mu_{\text{NT}}} \left(\frac{1}{\sqrt{N - 1/2}} + \frac{1}{\sqrt{T - 1/2}}\right)^{1/3}. \end{aligned} \tag{6.17}$$

De esta manera, la función de distribución acumulada Tracy-Widow de orden uno esta dada por [69, 72]

$$F_1(x) = \exp\left(-\frac{1}{2}\int_x^\infty q(y) + (y - q)q(y)^2 dy\right) \tag{6.18}$$

donde $q(y)$ es la solución única a la ecuación diferencial de Painlevé II

$$q''(y) = yq(y) + 2q^3(y) \tag{6.19}$$

La cual satisface la condición $q(y) \sim Ai(y)$, $y \to \infty$, mientras que $Ai(y)$ denota la fun-

ción de Airy. La función de distribución de probabilidad Tracy-Widow correspondiente se obtiene simplemente al derivar; $f_1(x) = F_1'(x)$.

Con el objetivo de examinar si los eigenvalores más grandes siguen el comportamiento de la distribución Tracy-Widow de orden uno (ver ec. (6.16)), utilizamos la aproximación de M. Chiani [69] para dimensiones grandes, la cual nos evita la necesidad de resolver numéricamente las ecs. (6.18) y (6.19). Esta aproximación se escribe como

$$\mathcal{TW}_1 \simeq \Gamma(k,\theta) - \alpha, \tag{6.20}$$

donde $\Gamma(k,\theta) = \int_0^\theta t^{k-1}e^{-t}dt$. Así, de la ec. (6.16) tenemos

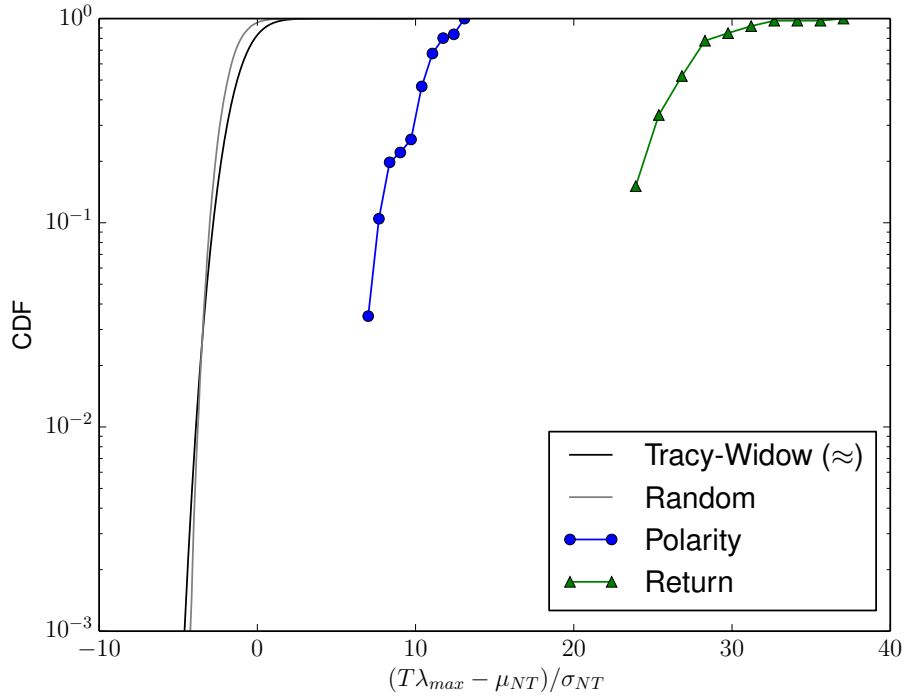
$$\frac{T\lambda_{max} - \mu_{\text{NT}}}{\sigma_{\text{NT}}} \approx \Gamma(k,\theta) - \alpha, \tag{6.21}$$

donde los parámetros $k, \theta$, y $\alpha$ se fijan con los mismos valores que en [69], es decir, 44.446, 0.1861, and 9.8480, respectivamente.

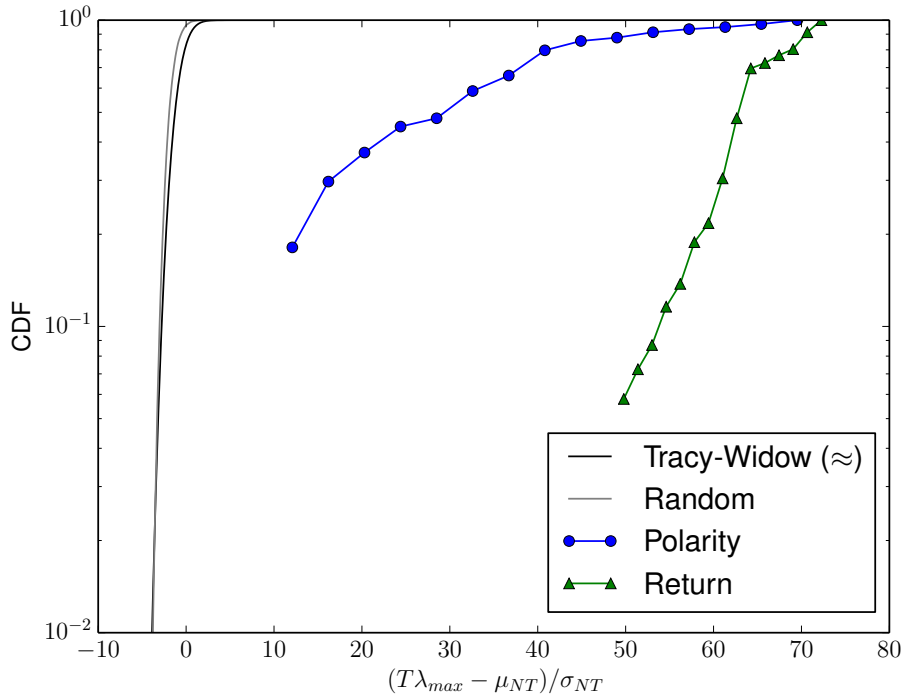
En la fig. 6.9 se grafica la función de distribución acumulada (CDF, por sus siglas en inglés) para la aproximación Tracy-Widow, y para los eigenvalores más grandes de retornos y polaridades obtenidos por el método de la ventana deslizante para ambos periodos de estudio. A manera comparativa, hemos también graficado en ambos casos los resultados de una simulación numérica con 10000 miembros de WE, de la misma dimensión que los datos financieros.

Se puede observar que la mayoría de los eigenvalores más grandes de la simulación numérica caen dentro de la región predicha por la aproximación Tracy-Widow. Además aún y cuando esta aproximación es válida únicamente para dimensiones grandes, muestra una muy buena concordancia para las dimensiones pequeñas $N = 20$, $T = 80$. Para el primer periodo de estudio, los resultados de polaridad caen más de 10 unidades a la derecha de los predicho por la aproximación Tracy-Widow, y 30 unidades para el caso de los datos de retornos. Mientras que para el segundo periodo, estos valores están entre 10 y 50 veces más lejos. Por otro lado, si tomamos en cuenta la cola de la distribución Tracy-Widow [70]

$$\sqrt{(1/Q) + \lambda_+^{2/3} N^{-2/3}} \approx 0.1165, \tag{6.22}$$




(a)




(b)

Figura 6.9: CDF para los eigenvalores más grandes. (a) Twitter e índices financieros. (b) NYT e índices financieros. En ambas figuras, la línea negra muestra la aproximación para la CDF de Tracy-Widow dado por la eq. (6.20), la línea gris representa los resultados numéricos para una muestra de 10000 miembros de WE, la línea azul representa los resultados para los datos de polaridad, y la línea verde para los de retornos. Se usó una escala semi-logaritmica en el eje vertical, así como unidades normalizadas de $\lambda_{max}$ dadas por la eq. (6.17).

este valor es aun muy pequeño para caracterizar la mayoría de las desviaciones fuera del borde superior de Marčenko-Pastur, lo que es una confirmación más del hecho de que las desviaciones en los eigenvalores más grandes no son debidas a efectos de ruido por trabajar con dimensiones pequeñas, sino más bien surgen como consecuencia de una estructura asociada a sectores económicos identificables [67].

## 6.7. Cociente de Participación Inverso

Una manera simple de extraer información a partir de los eigenvectores es estimando el cociente de participación inverso (IPR, por sus siglas en inglés), el cual nos permite conocer el número de componentes que participan significativamente en cada eigenvector (o portafolio). Este exhibe la distinción entre los eigenvectores asociados a los extremos y aquellos que pertenecen a la mayoría [5], dentro de la zona de ruido.

El IPR del eigenvector $V^k$ está dado por

$$IPR_k = \sum_{j=1}^{N} |V_j^k|^4, \tag{6.23}$$

cuyo valor siempre cae entre los límites $1/N$ y uno. Si el eigenvector $V^k$ se encuentra localizado solamente en un componente, entonces $IPR_k = 1$. Por el contrario, si se encuentra distribuido uniformemente sobre los $N$ componentes, entonces $IPR_k = 1/N$. Es de esperarse que los valores para $IPR_N$ fluctúen cerca del límite inferior $1/N$ ya que corresponde al portafolio más diversificado, mientras que para $IPR_1$ se esperan valores más altos, ya que esta asociado al eigenvalor más pequeño y por lo tanto al portafolio menos diversificado [4]. Asimismo, para valores de $k$ dentro de la región considerada como ruido, es de esperarse que surjan combinaciones aleatorias de los componentes, y por lo tanto valores de $IPR_k$ comprendidos entre los de $IPR_N$ y los de $IPR_1$.

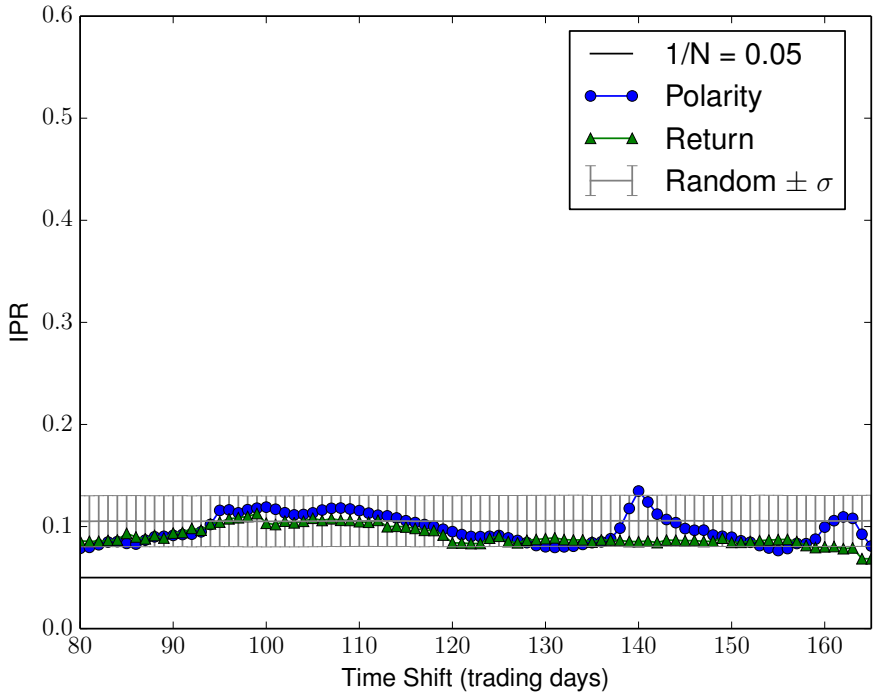
En la fig. 6.10 se muestra el comportamiento temporal de $IPR_N$ (fig. 6.10(a)) y de $IPR_1$ (fig. 6.10(b)) para los datos empíricos del primer periodo, es decir, Twitter y los índices financieros. En estas mismas figuras se muestra la media y desviación estándar de una simulación numérica de 10000 miembros de WE, donde de nuevo cada punto se calcula teniendo en cuenta los 80 días de transacción anteriores.

[5] conocido en la literatura como *bulk*.

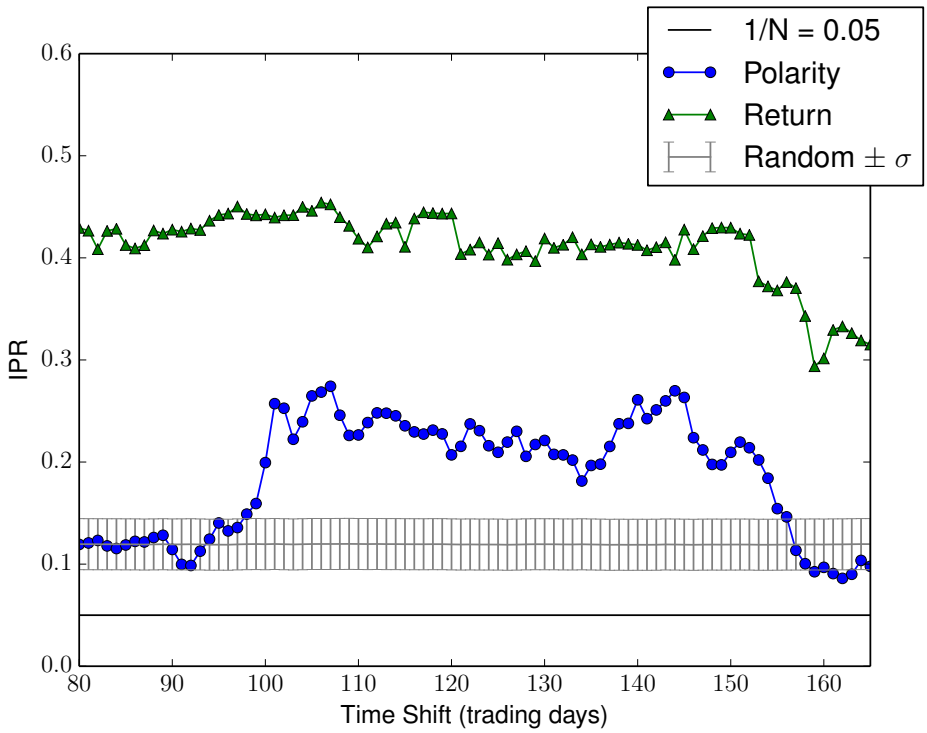
Se puede ver en la fig. 6.10(a) que ambos resultados empíricos presentan un comportamiento suave y fluctúan alrededor del límite inferior como es de esperarse, cayendo en la región de los resultados numéricos, lo cual confirma que cada uno de los indicadores financieros involucrados participa significativamente en $V_N$, y como consecuencia todos los índices se mueven como uno solo en éste eigenmodo. Es interesante observar que esta misma característica emerge cuando trabajamos con las polaridades. Para este caso se ha encontrado un $P_c = 0.6$ entre ambos comportamientos empíricos, con un valor p menor a $1\text{x}10^{-9}$.

Si ahora ponemos nuestra atención en la fig. 6.10(b), podemos observar que el $IPR_1$ se comporta de manera bastante diferente en ambos datos empíricos. Los resultados para retornos se mantienen fluctuando más de tres desviaciones estándar sobre los esperado para los resultados numéricos, mientras que los resultados de polaridad se encuentran lejos de los valores numéricos la mayor parte del tiempo, aunque hay periodos en que caen dentro de los valores de la simulación. Esto último podría implicar la presencia de ruido en la adquisición de los datos de polaridad, principalmente al comienzo del periodo de estudio. Sin embargo, incluso así se presenta una correlación positiva entre el comportamiento temporal de los datos empíricos, con $P_c = 0.49$ y valores de confidencia menores a $1\text{x}10^{-5}$.

Por otro lado los resultados para el segundo periodo de estudio se muestran en la fig. 6.11, donde ahora la simulación numérica se hizo con 1000 miembros del WE. Para $IPR_n$ se observa la misma característica encontrada para los resultados del primer periodo, es decir, un comportamiento suave fluctuando cerca del límite teórico. Sin embargo, para estos resultados no encontramos una correlación clara en el comportamiento temporal. El comportamiento de $IPR_1$ para las noticias de NYT fluctúa mucho más, sin embargo se encuentran lejos del límite teórico como es de esperarse, e incluso presenten una moderada correlación temporal de $P_c = 0.43$, con valor $p = 1x10^{-7}$.

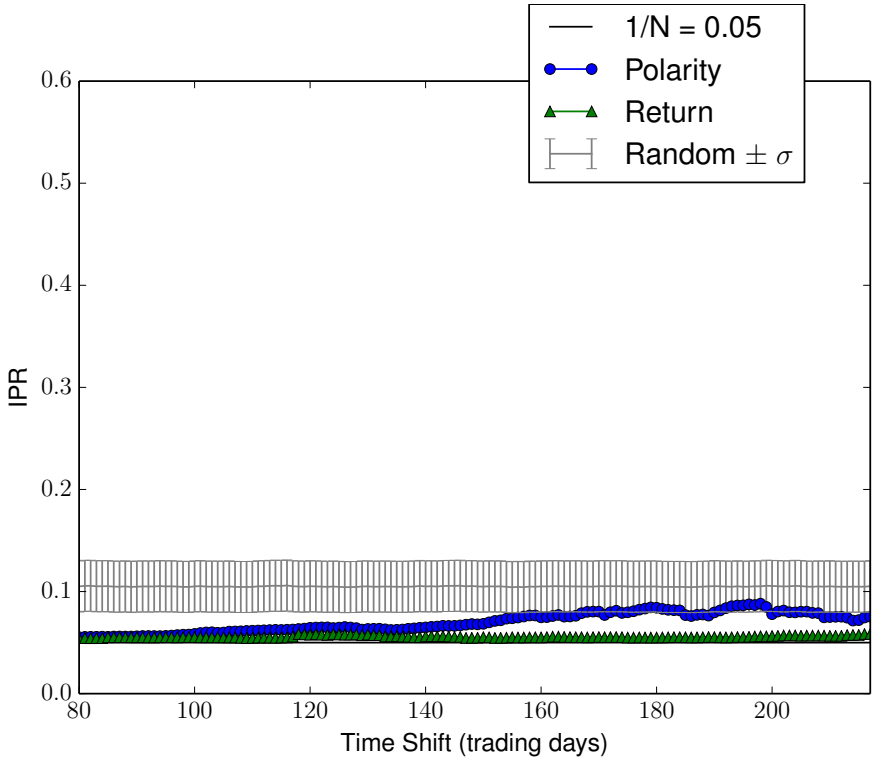



(a)

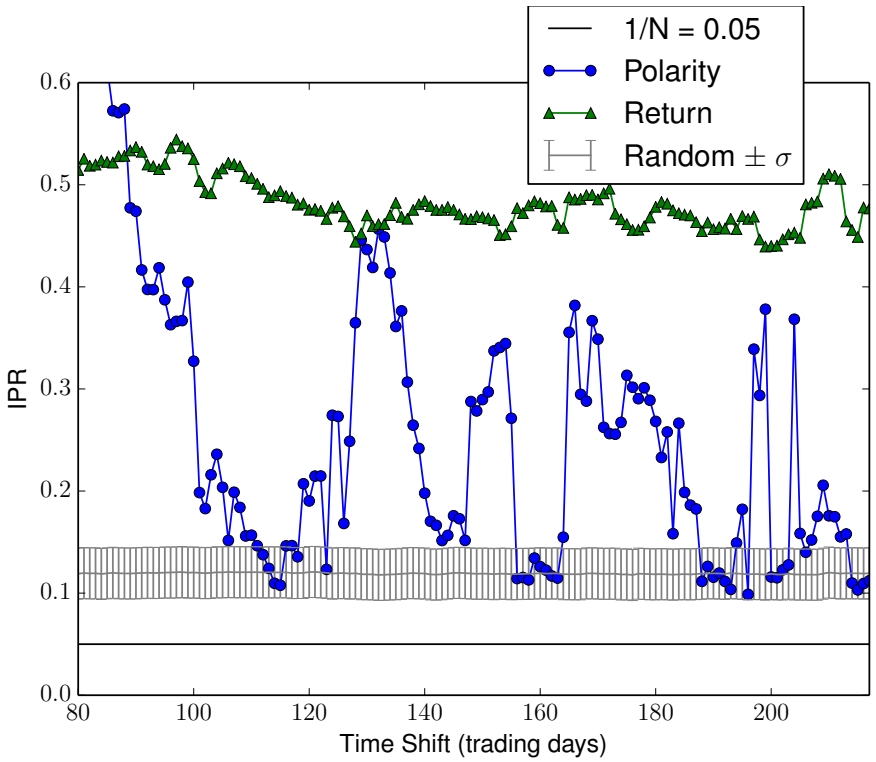



(b)

Figura 6.10: IPR para Twitter e índices financieros. (a) Comportamiento temporal de IPR correspondiente a los eigenvalores más grandes. (b) comportamiento temporal de IPR para el eigenvalor más pequeño. La línea azul representa los resultados para los retornos, la línea verde para polaridades, y la línea negra el límite inferior $1/N$. Además, la línea gris representa la media y la desviación estándar de los resultados de la simulación numérica de 10000 miembros de WE.




(a)




(b)

Figura 6.11: IPR para NYT e índices financieros. (a) Comportamiento temporal de IPR correspondiente a los eigenvalores más grandes. (b) comportamiento temporal de IPR para el eigenvalor más pequeño. La línea azul representa los resultados para los retornos, la línea verde para polaridades, y la línea negra el límite inferior $1/N$. Además, la línea gris representa la media y la desviación estándar de los resultados de la simulación numérica de 10000 miembros de WE.

# Capítulo 7

# Causalidad y Transferencia de Entropía

El estudio de correlaciones es útil para determinar cuales son los mercados que se comportan más similares, lo que nos ayuda determinar un portafolio de inversión. Sin embargo, utilizando solamente las medidas de correlación, no podemos establecer una relación de causalidad o influencia entre ellos dado que la acción de una variable sobre otra no es necesariamente simétrica. Para estudiar este problema debemos proceder de manera distinta. Por lo que en este capítulo se muestran los resultados relacionados con la causalidad y transferencia de entropía.

En la primera parte de este capítulo se describe el test de causalidad de Granger, el cual es un modelo lineal muy utilizado en econometría. En la segunda parte se presenta el enfoque de transferencia de entropía, el cual proviene de la teoría de la información, y está ligado en algunos aspectos a interpretaciones que competen a la física teórica.

Es importante señalar en este momento que [1] la interpretación de la entropía es mucho más natural desde el punto de vista subjetivo, donde una densidad de probabilidad $p$ no hace más que medir nuestro conocimiento parcial (y la correspondiente ignorancia parcial) del resultado o consecuencia de un evento. En este caso la entropía mide la incertidumbre de los observadores. Por otro lado, las leyes estadísticas nos ayudan a expresar las reglas de los comportamientos colectivos, sin importar la naturaleza física de los elementos y de sus interacciones. Estas leyes establecen propiedades matemáticas generales de cualquier sistema de dimensionalidad alta, como son los sistemas de muchos cuerpos en física o los mensajes largos en la teoría de la comunicación. La predictibilidad y simplicidad de

[1] Extracto de ideas tomadas de [73]

los fenómenos físicos macroscópicos vienen del hecho de que, al nivel macroscópico, una inmensa variedad de comportamientos surgen como consecuencia de los fenómenos de emergencia o *bottom-up integration*. La física esta envuelta solamente en prescribir la clase de universalidad del comportamiento emergente. Por lo que en general, la universalidad y robustocidad surge en física tan pronto como la estadística y geometría son suficientes para describir las características emergentes. Al momento presente, la teoría de la información no es capaz de describir como es que las características emergentes modifican el espacio de configuración, ni tampoco de generar reglas de interacción, eso sólo le compete a la física, siendo ahora la *econofísica* la encargada de contextualizar estos fenómenos dentro de la economía.

En la siguiente sección en lugar de las polaridades crudas, hemos considerado los retornos de sus polaridades, las cuales se obtuvieron mediante la ec. 5.2 de la sec. 5.1. Esto se hizo así debido a que se encontraron mejores resultados con este tipo datos. Sin embargo, en el análisis de transferencia de entropía se usaron las polaridades crudas nuevamente. De aquí en adelante nos referiremos indistintamente a los retornos de las polaridades y a las polaridades crudas simplemente como polaridades, teniendo en cuenta que difieren dependiendo de la sección en la que nos encontremos.

## 7.1. Test de Causalidad de Granger

Sean $X_t$ y $Y_t$, dos series de tiempo estacionarias a las que se les ha substraído la media. El análisis de causalidad de Granger se sustenta en asumir que si una variable $X_t$ causa $Y_t$, entonces, cambios en $X_t$ ocurrirán sistemáticamente antes de que sucedan cambios en $Y_t$. Es de interés en muchas disciplinas averiguar si valores retrasados de $X_t$ exhiben una correlación estadísticamente significativa con los valores de $Y_t$.

Una correlación, sin embargo, no prueba una causalidad, por lo que aquí realmente no estamos cuantificando causalidad, sino más bien si una serie contiene información predictiva acerca de la otra. El modelo lineal que se utiliza generalmente esta dado por [74]

$$Y_t = \sum_{j=1}^{lag} a_j X_{t-j} + \sum_{j=1}^{lag} b_j Y_{t-j} + \epsilon_t, \tag{7.1}$$

donde $\epsilon_t$ se considera una serie de tiempo de ruido blanco sin correlaciones, y $lag$ es

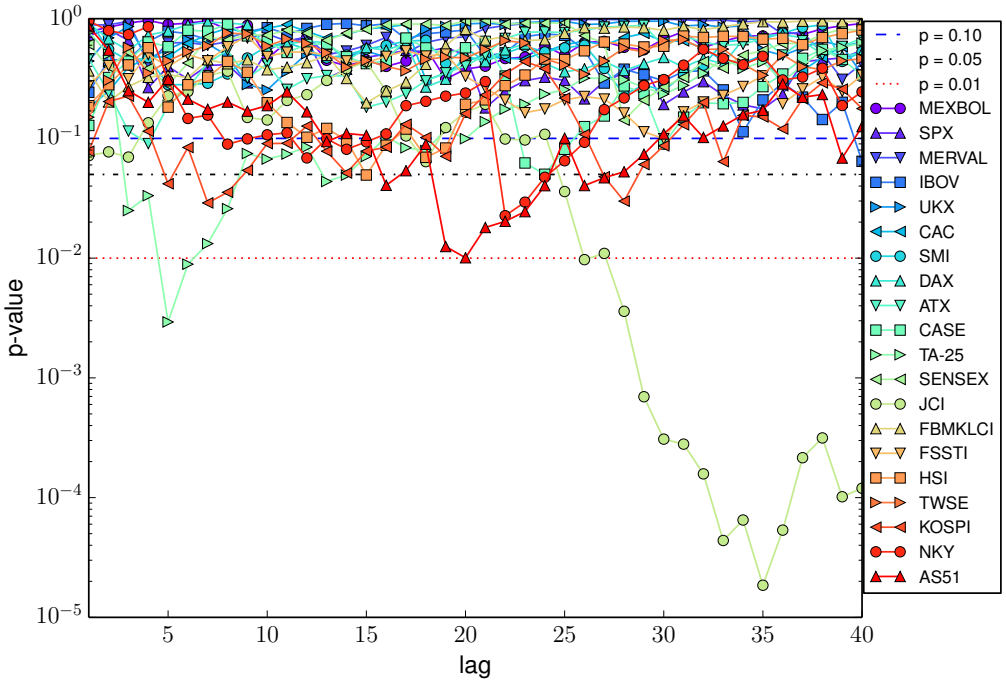
el máximo número de días que se retrasan las series para explorar si existe información predictiva de una sobre otra. Técnicamente, si la varianza de $\epsilon_t$ es reducida al incluir los términos de $X_t$, entonces se dice que $X_t$ Granger-Causa $Y_t$. En otras palabras. $X_t$ Granger-causa $Y_t$ si los coeficientes $a_j$ son conjuntamente significativamente diferentes de cero.

En resumidas cuentas, la hipótesis nula de este test asume que una serie de tiempo $X_t$ no causa la serie de tiempo $Y_t$, esta hipótesis es rechazada sí el *valor p* de confidencia de la prueba $F$ de Fisher [75] para los valores $a_j$ están por debajo de $1\text{x}10^{-1}$.

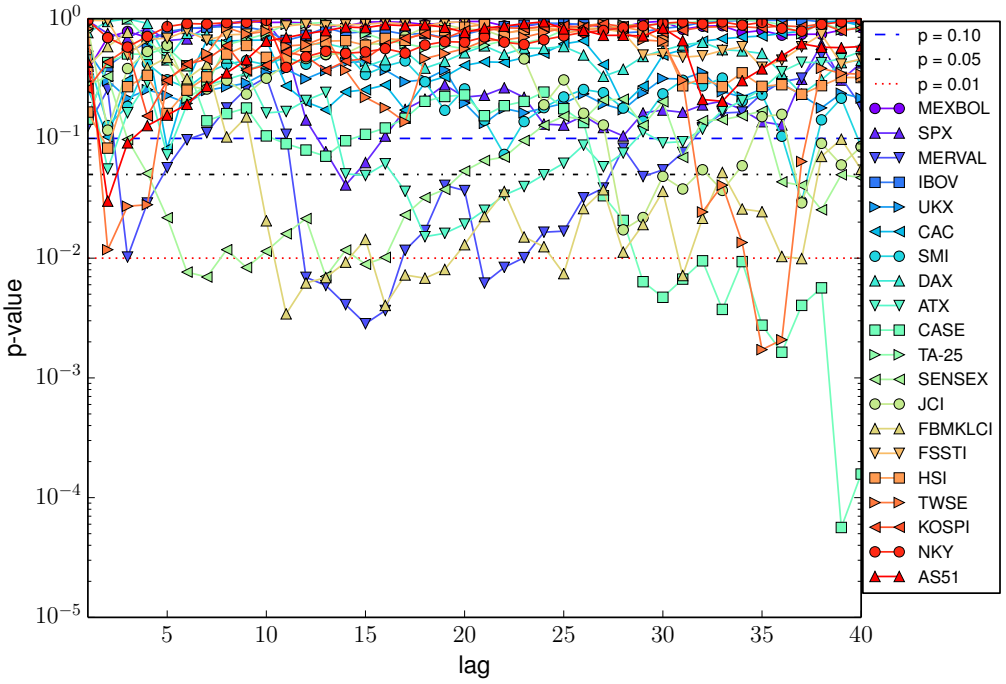
Siguiendo la metodología de este test, hemos renombrado $X_t = p_k(t)$ y $Y_t = r_k(t)$, para cada índice $k$ ($k = 1, .., N$), con la finalidad de examinar si las series de tiempo de polaridad causan (en el sentido de Granger) los valores de las series de tiempo de retornos, utilizando como modelo la ec. 7.1 dada anteriormente. En la fig. 7.1 se muestran los resultados para ambos periodos de estudio, donde para cada índice $k = 1, \ldots, N$, los valores de confidencia $p$ están dados en función del parámetro $lag$.

Hemos divido las gráficas dependiendo de su presunción acerca de la hipótesis nula. La primera sección está acotada por el borde superior (p=1) y la linea punteada (p=0.1); para los valores que caen dentro de esta región no existe presunción en contra de la hipótesis nula. La siguiente sección esta delimitada entre las lineas punteadas azul y negra (p=0.05); los valores $p$ en este rango de valores presentan baja presunción en contra de la hipótesis nula. Asimismo, los valores $p$ acotados en la sección comprendida entre las líneas negra y roja (p=0.01) muestran una presunción fuerte, mientras que los valores $p$ debajo de la línea roja son los que tienen las más fuerte presunción en contra de la hipótesis nula.

Se puede ver en la fig. 7.1(a) que los índices $TA-25$, $KOSPI$, $AS51$, $NKY$ y $JCI$ caen dentro de las dos últimas secciones para diferentes periodos o valores del parámetro $lag$, siendo el índice $JCI$ quien alcanza el valor más bajo de $p$ para un rango amplio de valores del parámetro $lag$, esto para el caso de estudio de Twitter. Para el caso de NYT, se ve en la fig. 7.1(b) que los índices $MERVAL$, $CASE$, $SENSEX$, $JCI$, $HSI$, $ATX$, $AS51$, $TWSE$ y $FBMKLCI$ caen dentro de las dos últimas secciones para diferentes periodos o valores del parámetro $lag$, siendo ahora el índice $CASE$ quien alcanza el valor más bajo de $p$ para un rango amplio de valores del parámetro $lag$.




(a)




(b)

Figura 7.1: Valor de confidencia $p$ como función del parámetro $lag$ de la hipótesis nula de que $p_i(t)$ no Granger causa $r_i(t)$. Se ha utilizado una escala semi-logarítmica para cada uno de los índices financieros bajo estudio. Las líneas punteadas azul, negra y roja delimitan los niveles de confidencia. Los valores $p$ debajo de la linea roja (0.01) presentan una fuerte presunción en contra de la hipótesis nula.(a) Resultados para Twitter e índices financieros. (b) resultados para NYT e índices financieros.

En la fig. 7.2(a), los valores de retorno del índice $JCI$ se superponen a los valores desfasados de la polaridad asociada de Twitter, con un $lag = 35$ ya que para este valor se presenta la más fuerte presunción en contra de la hipótesis nula. Asimismo, en la fig. 7.2(b) se superpone los retornos de $CASE$ con la polaridad correspondiente de NYT para el $lag = 39$, ya que para este valor se presentaron los mejores resultados en el segundo periodo de estudio.

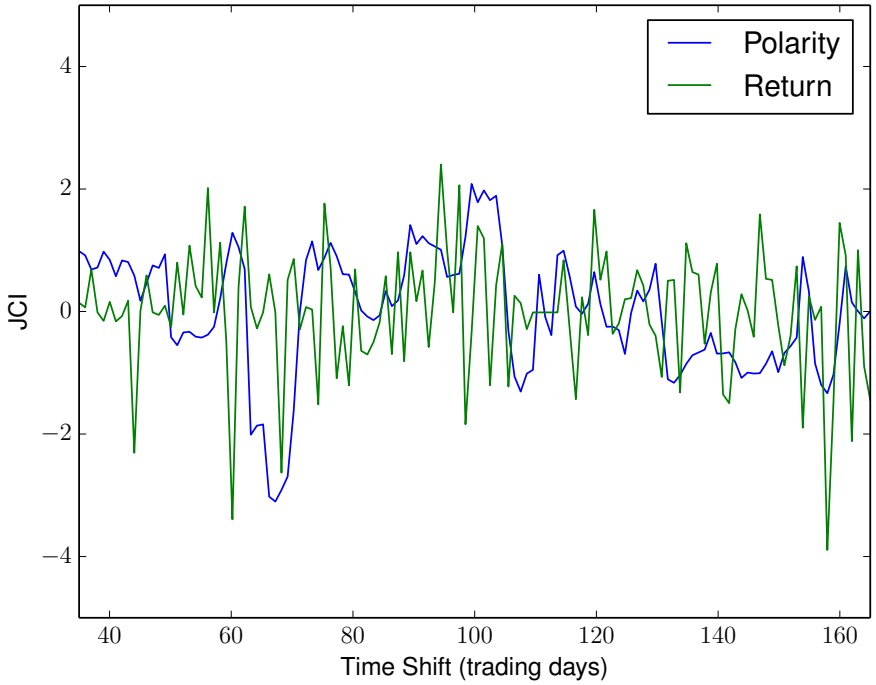
Como puede verse en ambas figuras, las series de tiempo frecuentemente se superponen o apuntan en la misma dirección. Por lo que movimientos en el pasado de $p_i(t-lag)$ podían predecir una posible subida o caída en los valores $r_i(t)$.

## 7.2. Transferencia de Entropía

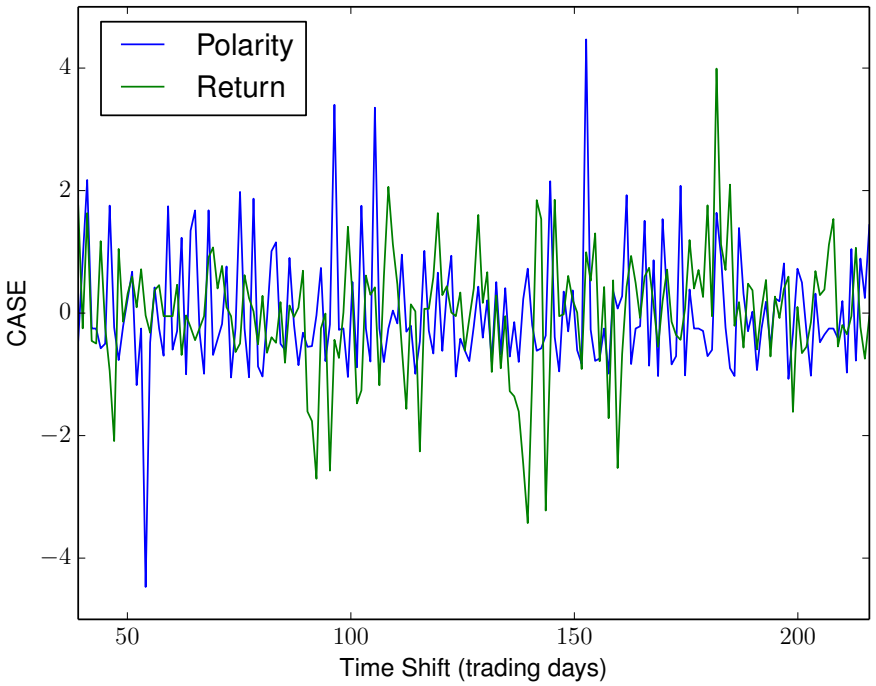
La transferencia de entropía fue desarrollada inicialmente por Schreiber [76], y está basada en conceptos relacionados con la entropía de Shannon [77], desarrollada dentro de la teoría de la información. Básicamente, la transferencia de entropía de un proceso $X$ hacia otro proceso $Y$ cuantifica la reducción en la incertidumbre de un valor futuro de $Y$ dado el conocimiento pasado de los valores simultáneos de $X$ y $Y$ [76]. La transferencia de entropía se reduce al test de causalidad de Granger para un proceso auto-regresivo [78]. Por lo que es de gran ayuda cuando el objeto de estudio no satisfacen las condiciones del modelo de causalidad de Granger, lo cual ocurre en la mayoría de los fenómenos reales, ya que estos involucran series de tiempo no lineales [79, 80].

Además, la transferencia de entropía es una medida dinámica y no simétrica, la cual se ha comprobado ser útil en una gran cantidad de problemas. Ha sido utilizado para el estudio de autómatas celulares [81, 82], en el estudio de la corteza neuronal del cerebro [83, 84, 85, 86], en el estudio de redes sociales [79], finanzas [87, 88, 89, 90], física estadística [91], y en sistemas dinámicos [92], recibiendo una interpretación termodinámica en [93].

A continuación se deriva la definición de la transferencia de entropía siguiendo la misma línea del artículo de Schreiber [76], explicando brevemente algunas definiciones de entropía desde el punto de vista de la teoría de la información, como son la entropía de Shannon, entropía de Kullback, entropía mutua, entre otros conceptos, ya que esto nos ayudará a comprender mejor las implicaciones de la también llamada transferencia de in-




(a)




(b)

Figura 7.2: (a) Traslape de los valores de retornos con los valores de polaridad desfasados 35 días para el índice $JCI$ para Twitter e índices financieros. (b) Traslape de los valores de retornos con los valores de polaridad desfasados 39 días para el índice $CASE$ para NYT e índices financieros.

formación.

### 7.2.1. Fundamentos

Claude Elwood Shannon, matemático, ingeniero electrónico y criptógrafo, fundó en 1948 la teoría de la información en su trabajo titulado: *A Mathematical Theory of Communication* [77]. De acuerdo a Shannon, el problema principal de la teoría de la información es reproducir un mensaje que es enviado desde otro punto. Si se considera un conjunto de eventos posibles cuyas probabilidades de ocurrencia sean $p_i$, $i = 1, \ldots, n$, entonces la medida de la incertidumbre de ocurrencia un evento $H(p_1, p_2, \ldots, p_n)$ dada tales distribuciones de probabilidades, debe tener las tres propiedades siguientes:

- $H(p_i)$ debe ser continuo en $p_i$.
- Sí todas las probabilidades son iguales, es decir, $p_i = 1/n$, entonces $H$ debe ser una función monótonamente creciente al variar $n$ (sí existen más opciones posibles, entonces la incertidumbre acerca del resultado debe incrementar).
- Sí la elección se desglosa en otras opciones a su vez, con probabilidades $c_j$ , $j = 1, \ldots, k$, entonces $H = \sum_{j=1}^{k} c_j H_k$, donde $H_k$ es el valor de la función $H$ para cada elección particular.

Shannon probó que la única función que satisface las tres propiedades anteriores está dada por la siguiente expresión

$$H = -\sum_{i=1}^{N} p_i \log_2(p_i), \tag{7.2}$$

donde la suma de todos los estados para los cuales $p_i \neq 0$ [2]. La base dos del logaritmo determina únicamente las unidades utilizadas para medir la información, por lo que pueden ser omitida de aquí en adelante y consideraremos que estamos midiendo siempre en bits. Cabe mencionar que esta definición de entropía guarda mucha semejanza con la entropía de Gibbs, pero es más general, ya que puede ser aplicada a cualquier sistema que almacene información.

[2] la definición original de Shannon tiene una constante $k$ multiplicando todo el término, la cual se considera igual a uno para efectos de simplicidad en la deducción de transferencia de entropía.

Consideremos ahora que el número promedio de bits necesarios para codificar las realizaciones independientes de una variable discreta $I$ dada por la distribución de probabilidad $p(i)$ está dada por la entropía de Shannon $H_I$ (ec. 7.2). Con el propósito de construir un codificador óptimo que utilice solamente el número de bits dados por la entropía, es necesario conocer la distribución de probabilidad $p(i)$. El valor excedente de bits que son codificados al utilizar una distribución diferente $q(i)$, esta dado por la *entropía de Kullback* [94]

$$K_I = \sum_i p(i) \log \frac{p(i)}{q(i)} \tag{7.3}$$

Por otro lado, la información mutua de dos procesos $I$ y $J$ con probabilidad conjunta $p_{IJ}(i,j)$, puede ser vista como el excedente de código producido erróneamente por asumir que los dos sistemas son independientes, es decir, por considerar $q_{IJ}(i,j) = p_I(i)p_J(j)$ en lugar $p_{IJ} = p(i,j)$. La entropía de Kullback correspondiente para este caso está dada por

$$M_{IJ} = \sum p(i,j) \log \frac{p(i,j)}{p(i)p(j)}, \tag{7.4}$$

cuya expresión viene a ser la conocida fórmula de *información mutua* [95]. Esta derivación muestra que la información mutua es la forma más natural de cuantificar la desviación de dos procesos independientes. Si desarrollamos el lado derecho de la ecuación anterior, obtenemos la siguiente expresión

$$\begin{aligned} M_{IJ} &= \sum_{i,j} p(i,j) \log p(i,j) - \sum_{i,j} p(i,j) \log p(i) - \sum_{i,j} p(i,j) \log p(j) \\ &= \sum_{i,j} p(i,j) \log p(i,j) - \sum_i \log p(i) \sum_j p(i,j) - \sum_j \log p(j) \sum_i p(i,j) \\ &= \sum_{i,j} p(i,j) \log p(i,j) - \sum_i p(i) \log p(i) - \sum_j p(j) \log p(j) \\ &= H_I + H_J - H_{IJ}. \end{aligned} \tag{7.5}$$

De esta manera, observamos que $M_{IJ}$ es simétrica ante el intercambio de $I$ y $J$, por lo que esta medida no contiene ningún sentido direccional en el flujo de información. No obstante, se le puede asignar un sentido direccional a la información mutua de una manera artificial al introducir un retraso en cualquiera de las variables involucradas. Una manera

de lograr esto es incorporar una estructura dinámica para estudiar probabilidades de transición en lugar de probabilidades estáticas.

Consideremos para esto un sistema que puede ser aproximado por un proceso estacionario de Markov de orden $k$, es decir, la probabilidad condicional de encontrar $I$ en el estado $i_{n+1}$ al tiempo $n+1$ es independiente del estado $i_{n-k}$

$$p(i_{n+1}|i_n, \ldots, i_{n-k+1}, i_{n-k}) = p(i_{n+1}|i_n, \ldots, i_{n-k+1}). \tag{7.6}$$

De aquí en adelante usaremos la notación $i_n^{(k)}$ para referirnos a $(i_n, \ldots, i_{n-k+1})$, para cadenas de longitud $k$.

El número promedio de bits necesarios para codificar un estado adicional del sistema sí todos los estados previos son conocidos, está dado por la *tasa de entropía*

$$h_I = \sum p(i_{n+1}, i_n^{(k)}) \log p(i_{n+1}|i_n^{(k)}). \tag{7.7}$$

Puesto que

$$p(i_{n+1}|i_n^{(k)}) = p(i_{n+1} \cap i_n^{(k)})/p(i_n^{(k)}) = p(i_{n+1}^{(k+1)})/p(i_n^{(k)}), \tag{7.8}$$

la expresión ec. 7.7 es solamente la diferencia entre las entropías de Shannon para los procesos dados por los vectores de dimensión $k+1$ y $k$, construidos a partir de $I$ [96]:

$$\begin{aligned} h_I &= \sum p(i_{n+1}, i_n^{(k)}) \log \frac{p(i_{n+1}^{(k+1)})}{p(i_n^{(k)})} \\ &= \sum p(i_{n+1}^{(k+1)}) \log p(i_{n+1}^{(k+1)}) - \sum p(i_n^{(k)}) \log p(i_n^{(k)}) \\ &= H_{I^{(k+1)}} - H_{I^{(k)}}. \end{aligned} \tag{7.9}$$

La manera más directa de construir una tasa de información mutua generalizada para dos procesos $(I, J)$ es de nuevo midiendo la desviación del resultado para un caso donde se dé la independencia. Sin embargo, la entropía de Kullback correspondiente es simétrica ante el intercambio de $I$ y $J$. Es por lo tanto preferible medir la desviación a partir del proceso generalizado de Markov

$$p(i_{n+1}|i_n^{(k)}) = p(i_{n+1}|i_n^{(k)}, j_n^{(l)}) \tag{7.10}$$

De la ecuación anterior, en la ausencia de flujo de información de $J$ a $I$, el estado de $J$ no tiene influencia sobre las transiciones de probabilidad en el sistema $I$. La suposición incorrecta puede ser cuantificada de nuevo por la entropía de Kullback, mediante la cual se obtiene finalmente la *Transferencia de Entropía* [76]

$$T_{J\to I} = \sum p(i_{n+1}, i_n^{(k)}, j_n^{(l)}) \log \frac{p(i_{n+1}|i_n^{(k)}, j_n^{(l)})}{p(i_{n+1}|i_n^{(k)})}, \tag{7.11}$$

donde la elección más natural para el índice $l$ es $l = k$ o $l = 1$. Usualmente, la última opción es preferida por razones computacionales. Podemos ver que $T_{J\to I}$ es explícitamente no-simétrica puesto que esta cantidad mide el grado de dependencia de $I$ sobre $J$ y no de manera inversa.

### 7.2.2. Resultados

Para calcular la transferencia de entropía dada por la ec. 7.11 se ha utilizado la librería JIDT [3] [97], donde las distribuciones de probabilidad que aparecen en la ecuación se han obtenido mediante la estimación de la densidad del kernel [4]. Este estimador se define como

$$p_h = \frac{1}{n} \sum_{i=1}^{n} K_h(t - t_i), \tag{7.12}$$

donde cada kernel K, está identificado por el parámetro de posición $t_i$, y el ancho de banda $h$. En nuestro caso, la función kernel cuenta el número de retornos o polaridades que caen dentro de la caja de longitud $h$ centrada en $t$. Una elección común para $h$ está dada por la regla de Silverman [98]

$$h = \left(\frac{4\sigma^5}{3n}\right)^{\frac{1}{5}}, \tag{7.13}$$

donde $\sigma$ es la desviación estándar de las series de tiempo, y $n$ su dimensión.

[3] Disponible en http://jlizier.github.io/jidt/

[4] Técnica mejor conocida como *kernel density estimation* (Véase el apéndice B para una discusión acerca de los distintos tipos de estimadores).

En las figs. 7.3(a),7.3(b), se muestran las matrices de transferencia de entropía para el primer (Twitter) y segundo (NYT) periodo de estudio, respectivamente. En ambos casos se fijo $k = l = 1$ en la ec. 7.11, y se utilizó el ancha de banda dado por la regla de Silverman, el cual resulto ser de $h = 0.38$ para el primer periodo, y de $h = 0.36$ para el segundo. Estas matrices se han ordenado de las misma manera que las supermatrices de correlación de la sección 6.4, por lo que se pueden observar 4 bloques distintos. El bloque superior izquierdo de cada matriz muestra la transferencia de entropía de los retornos hacia los retornos, el bloque inferior derecho la transferencia de las polaridades hacia las polaridades, y los bloques no-diagonales la transferencia cruzada de información entre retornos y polaridades.

Las matrices de trasferencia de entropía generalmente contienen mucho ruido debido a que las series de tiempo utilizadas en su calculo son finitas, no-estacionarias, y presentan efectos no-lineales. En el caso de los indices financieros los mercados que presentan mayor volatilidad naturalmente transfieren mayor entropía a los mercados de menor volatilidad [90]. Para eliminar algunos de estos efectos podemos calcular las matriz de transferencia de entropía aleatoria, es decir, permutado aleatoriamente los elementos de las series de tiempo, para romper cualquier relación causal entre las variables, pero manteniendo las mismas distribuciones de probabilidad de cada series de tiempo original [99]. De esta manera, la matriz de transferencia de entropía efectiva (TEE) se obtiene al substraer la matriz de transferencia de entropía aleatoria (TEA) de la matriz de transferencia de entropía (TE) de las serie originales:

$$TEE_{J\to I} = TE_{J\to I} - TEA_{J\to I} \tag{7.14}$$

En las figs. 7.3(c),7.3(d) se muestran los resultados del promedio de 25 permutaciones aleatorias distintas, es decir de 25 TEA, para el primer y segundo periodo de estudio, respectivamente. Se calcularon solamente 25 simulaciones ya que el cálculo es muy demandante computacionalmente, además de que los resultados para cada simulación son similares. Finalmente, en las figs. 7.3(e),7.3(f) se muestran los resultados para las TEE obtenidas mediante la ec. 7.14. Se puede ver que antes después de substraer el ruido, las intensidad de las TE disminuye, pero la estructura se preserva para el caso de Twitter, mientras que para NYT la discrepancia marcada entre la información hacia los retornos y hacia las polaridades se desvanece después de substraer el ruido.

En la fig. 7.4 mostramos los resultados de TEE para el primer periodo de estudio, al variar

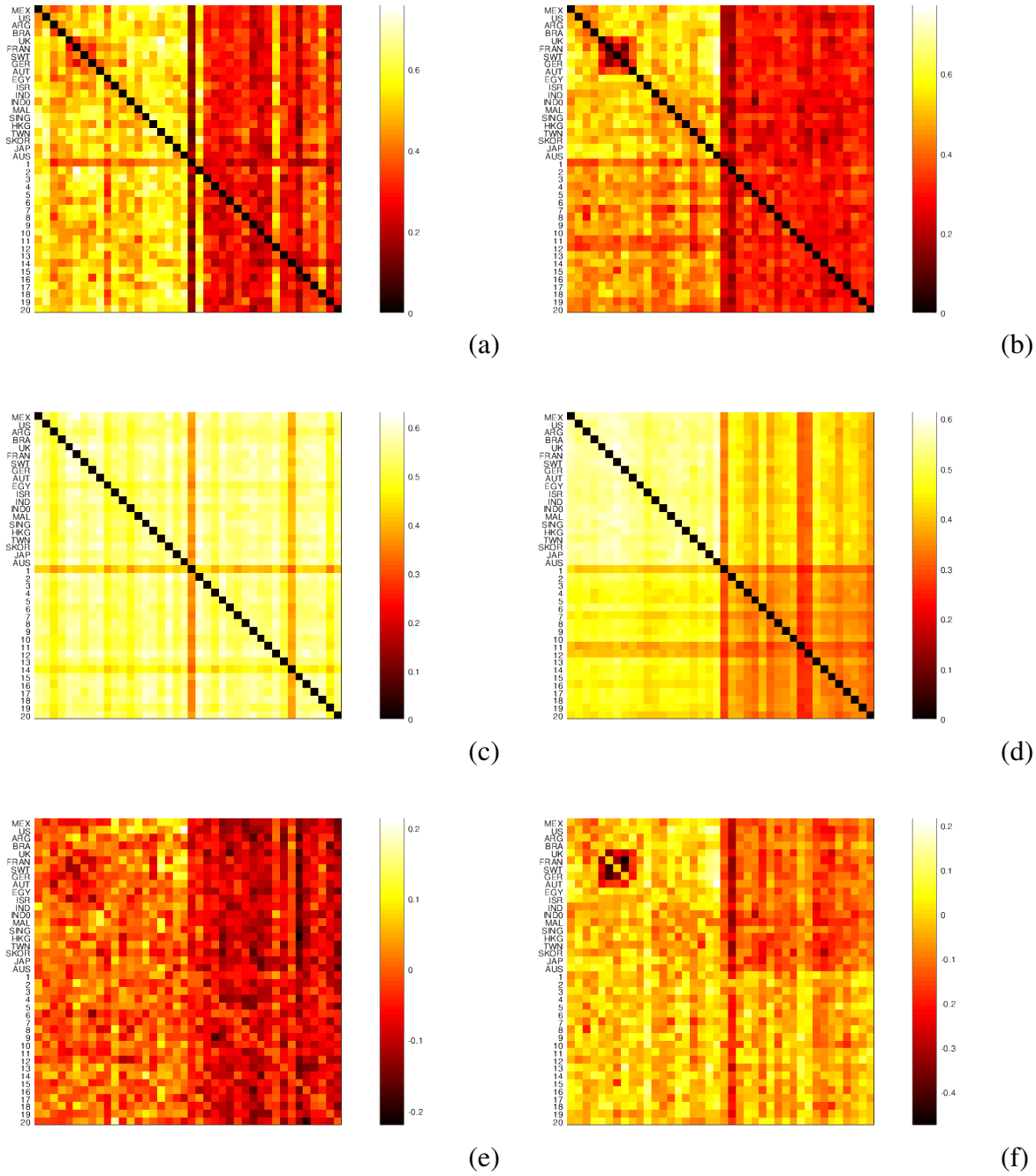



Figura 7.3: De arriba a abajo: TE, TEA, y TEE. Las figuras de la izquierda representan los resultados para Twitter e índices financieros, mientras que los de la derecha para NYT e índices financieros. Para las TEA se tomó el promedio de los resultados para 25 matrices permutaciones aleatorias de las series de tiempo originales. En todos los casos se uso $k = l = 1$ con una resolución de densidad de kernel de $h = 0.38$, para el caso de Twitter e índices, y de $h = 0.36$, para el caso de NYT e índices.

.

la resolución del kernel con los valores de $h = 0.1, 0.2, \ldots, 0.6$. Se puede observar que la estructura de TEE no cambia significativamente para valores bajos de $h$, y se vuelve más difusa la estructura después del valor $h = 0.5$, por lo que se comprueba empíricamente que una buena elección del parámetro $h$, está dado por la ec. 7.13. De la misma manera, en la fig. 7.5 se muestran las gráficas correspondientes al variar $h$ en el calculo de TEE para los datos que involucran las noticias de NYT. En este caso se puede ver que las estructuras cambian ligeramente al variar $h$, sin embargo se presenta cierto comportamiento estable entre las figuras correspondientes a $h = 0.3$ y $h = 0.4$, siendo donde cae el valor dado por la regla de Silverman (ec. 7.13).

Una vez mostrado el comportamiento de TEE al variar $h$, hemos fijado estos valores a los obtenidos por la regla de Silverman (es decir, $h = 0.38$ y $h = 0.36$ para los datos que involucran a Twitter y NYT, respectivamente), y hemos variado ahora los parámetros $k, l$ de la ec. 7.11 los cuales especifican la memoria de las series de tiempo al considerarlas como un proceso estacionario de Markov (ver sec. 7.2.1). En la fig. 7.6 se muestran los resultados de TEE para $k = l = 2, 3, 4$, en ambos periodos de estudio. Se puede observar en el caso de estudio de Twitter que para $k = 2$ (fig. 7.6(a)), la transferencia de entropía es marcadamente mayor hacia los retornos, sin embargo al aumentar el valor al $k = l = 4$ (fig. 7.6(e)) la transferencia ocurre hacia las polaridades, donde en el valor de $k = l = 3$ (7.6(c)) se da la transición del flujo de información. Por el contrario, en el caso de estudio de $NYT$ parece ser que la información siempre fluye hacia los retornos, siendo los valor de $k = l = 3$ donde se observa este comportamiento de manera más uniforme.

Puesto que nos interesa explorar el fenómeno donde la información fluye hacia los retornos, ya que esto abre una inmensa gama de posibilidades de generar estrategias de *trading* [5] fuera de la hipótesis de mercado eficiente, hemos concentrado nuestra atención en los resultado de TEE donde se muestra un mayor flujo de información hacia los retornos exclusivamente. Se han elegido los casos $k = l = 2, h = 0.38$ y $k = l = 3$, para el primer y segundo periodo de estudio, respectivamente, como los más representativos de este fenómeno.

Para estos últimos casos se ha hecho uso de la teoría de redes para transformar la TEE en un grafo dirigido y se ha hecho el análisis de los agrupamientos que surgen a diferentes cotas de transferencia de información, a los que denominaremos $t_c$, donde ahora

[5] Conjunto de estrategias que consiste en generar ganancias a partir de las variaciones a corto plazo de los precio de las acciones.

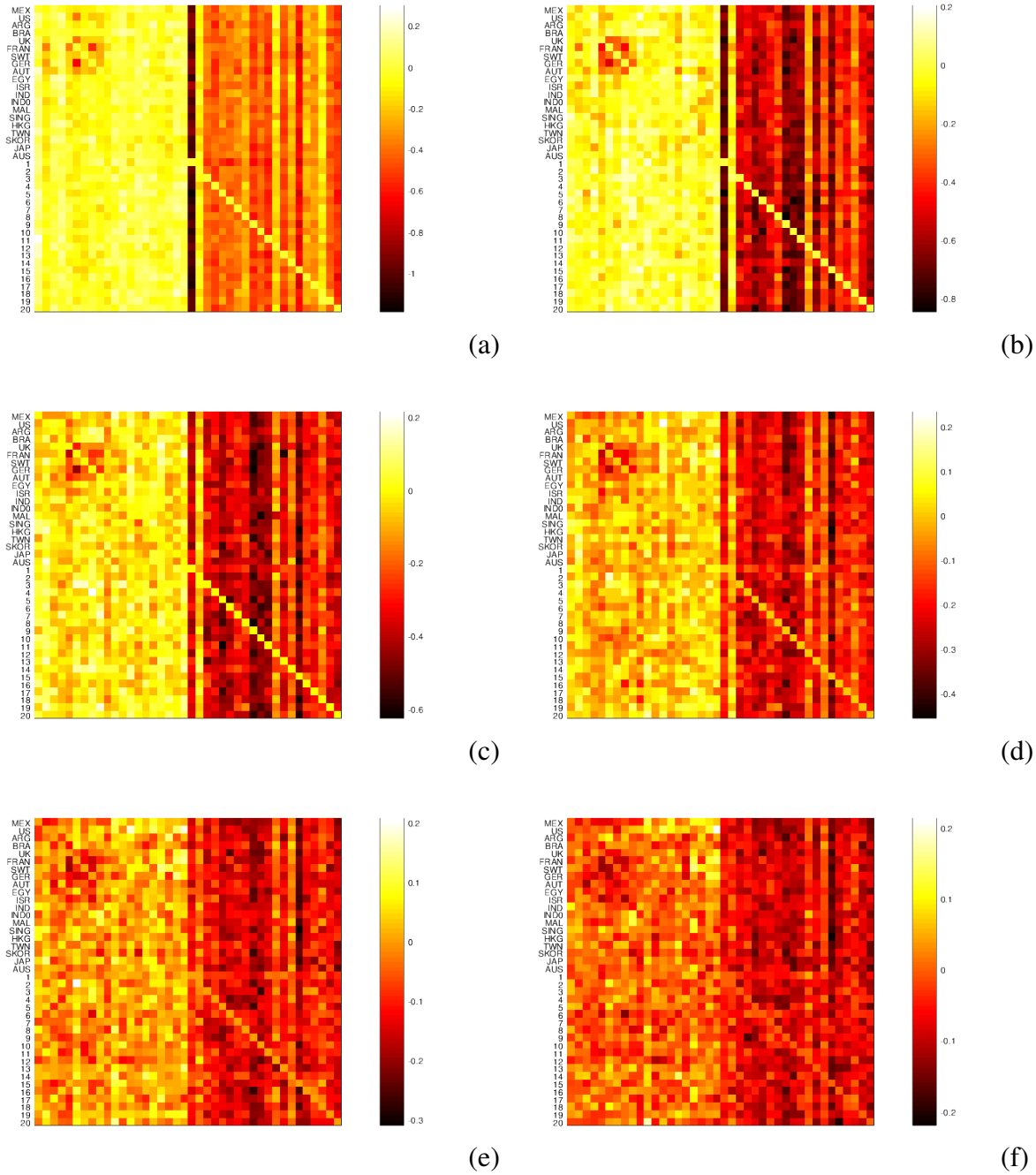



Figura 7.4: TEE al variar el parámetro de resolución $h$ en el periodo de estudio de Twitter. (a) $h = 0.1$. (b) $h = 0.2$. (c) $h = 0.3$. (d) $h = 0.4$. (e) $h = 0.5$. $h = 0.6$

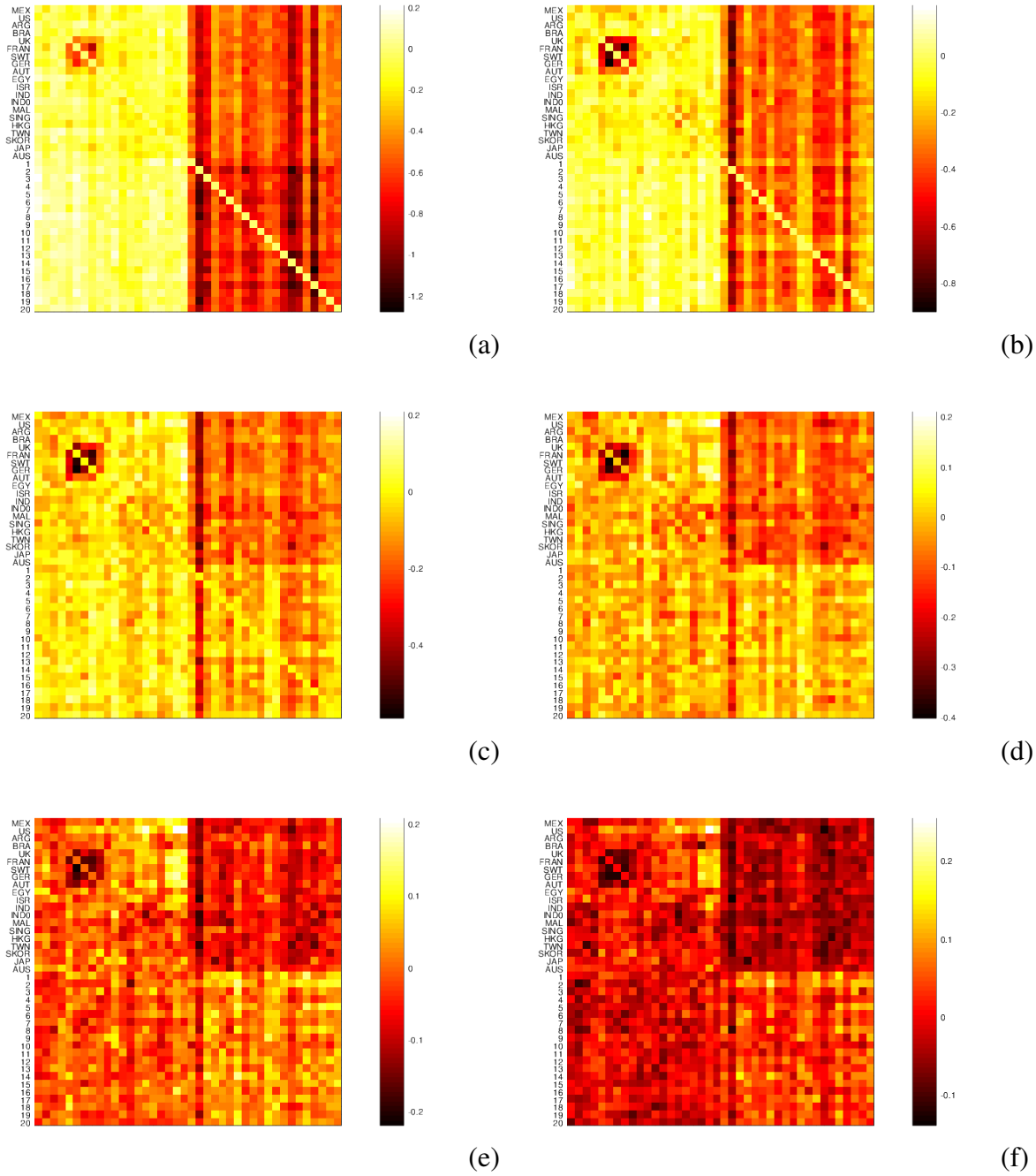



Figura 7.5: TEE al variar el parámetro de resolución $h$ en el periodo de estudio de NYT. (a) $h = 0.1$. (b) $h = 0.2$. (c) $h = 0.3$. (d) $h = 0.4$. (e) $h = 0.5$. $h = 0.6$

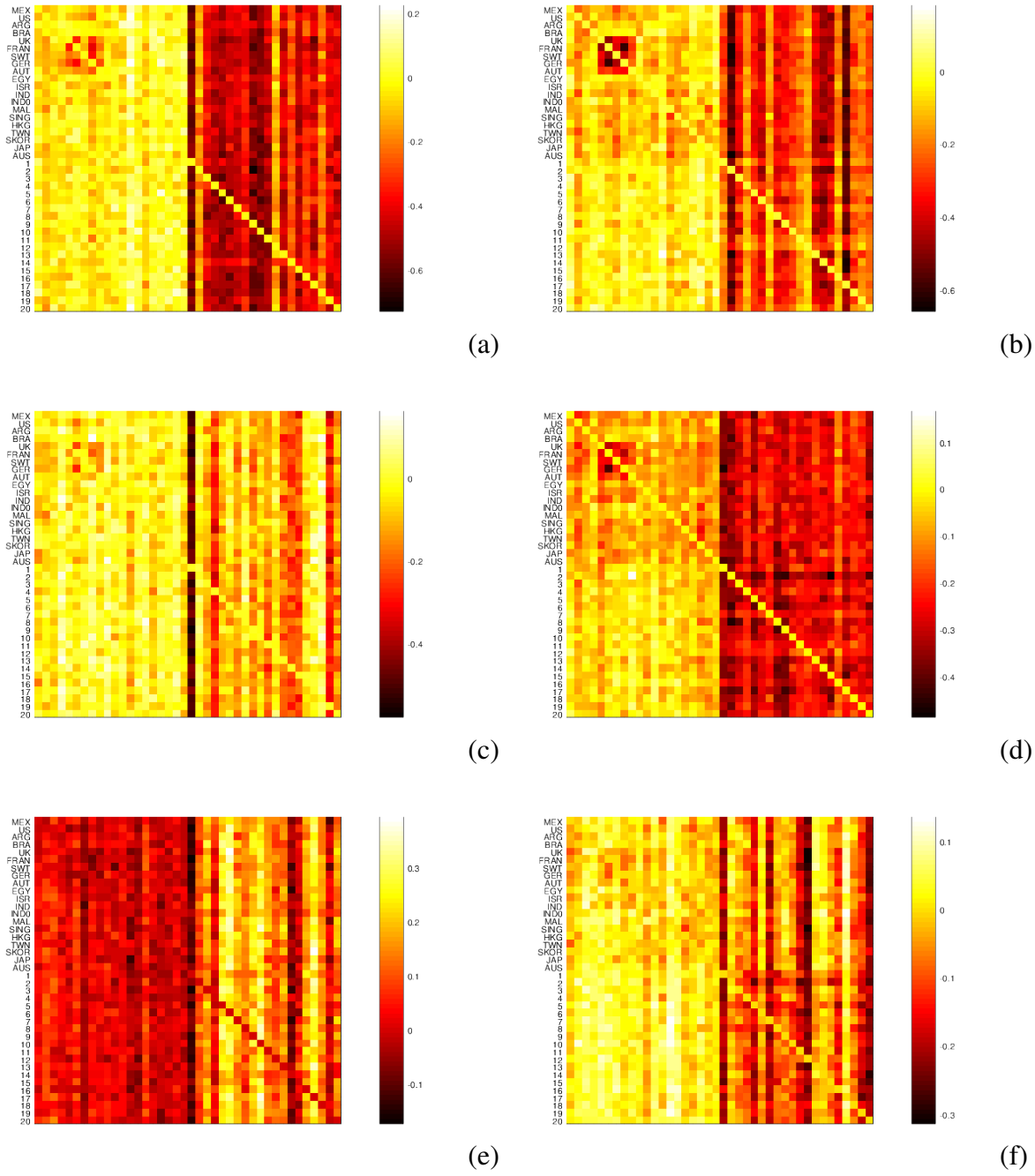


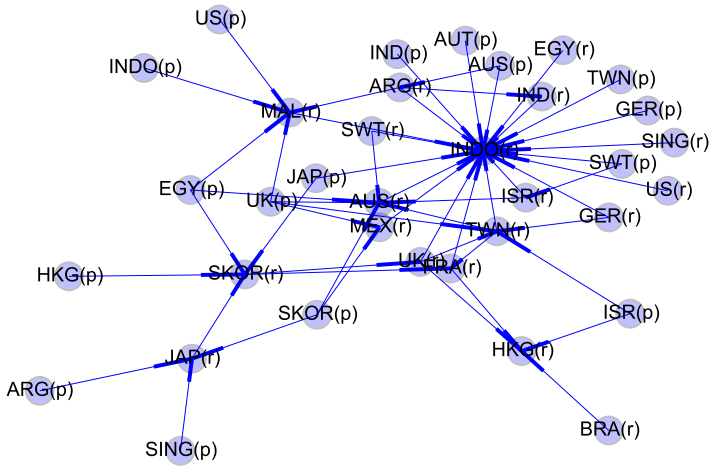

Figura 7.6: TEE al variar los parámetro de dependencia $k, l$ de la ec. 7.11. Figuras superiores $k = l = 2$, (a) Caso Twitter (b) Caso NYT. Figuras intermedias $k = l = 3$, (c) Caso Twitter (d) Caso NYT. Figuras inferiores $k = l = 4$, (e) Caso Twitter (f) Caso NYT.

para construir las matrices de estos casos se le ha substraído a las TE el promedio de 100 permutaciones aleatorias para los datos que involucran a Twitter, y 75 para los de NYT, esto se hizo así buscando que la estructura de la red fuera más estable en el análisis de formación de agrupamientos.

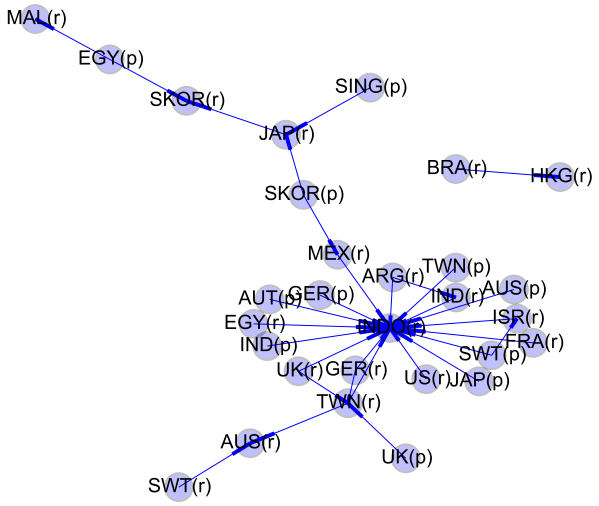
Para el caso de estudio de Twitter, en las figs. 7.7 y 7.8, se muestran las conexiones que permanecen arriba de los límites $t_c = 0.10, 0.13, 0.16$ y $0.19$, donde las conexiones con puntas gruesas implican que se está transfiriendo información hacia el nodo donde apuntan. Se puede observar en la fig. 7.7(a) que para un $tc = 0.10$ los retornos de Indonesia son los que mayor información recibe de los demás indicadores. Para el $t_c = 0.13$ emerge otra subred (fig.7.7(b)) donde los retornos de Brazil transfieren información hacia hacia los retornos de Hong Kong. Después en la fig. 7.8(a) ($t_c = 0.16$) desaparece esta subred, pero surgen otras dos, prevaleciendo como la subred con más conexiones la que transfiere información a los retornos de Indonesia y a los retornos de Taiwan. Finalmente para un valor de $t_c = 0.19$ (fig. 7.8(b)) sólo se mantiene 2 subredes, las cuales mandan información a los retornos de Indonesia y Taiwan. Se puede ver que en todos estos caso, la información fluye hacia los retornos y no de manera inversa.

En las figs. 7.9 y 7.10, se muestra para el caso de estudio de NYT, las conexiones que permanecen arriba de $t_c = 0.05, 0.06, 0.07$ y $0.08$. Se puede observar en la fig. 7.9 que los nodos que más reciben entropía de distintos indicadores son los retornos de India y Austria, esto prevalece para las cotas $t_c = 0.05$ y $t_c = 0.06$. Para el valor de $t_c = 0.07$ (fig. 7.10(a)) comienzan ha disgregarse las subredes formando 5 aglomerados con pocas conexiones. Finalmente cuando $t_c = 0.08$ sólo sobreviven tres aglomerados. Al igual que en el caso de Twitter, aquí también la información solamente fluye hacia los retornos para los cotas de $t_c$ seleccionadas.

En general, se observa una gran cantidad de información que fluye hacia los retornos, la cual proviene de las polaridades de Twitter y de NYT. Todo esto es conforma una evidencia fuerte en contra de la hipótesis de mercado eficiente. Lo que abre las posibilidades a generar nuevas estrategias de *trading* utilizando información cuantitativa de noticias financieras y de las redes sociales, en particular de Twitter y NYT.

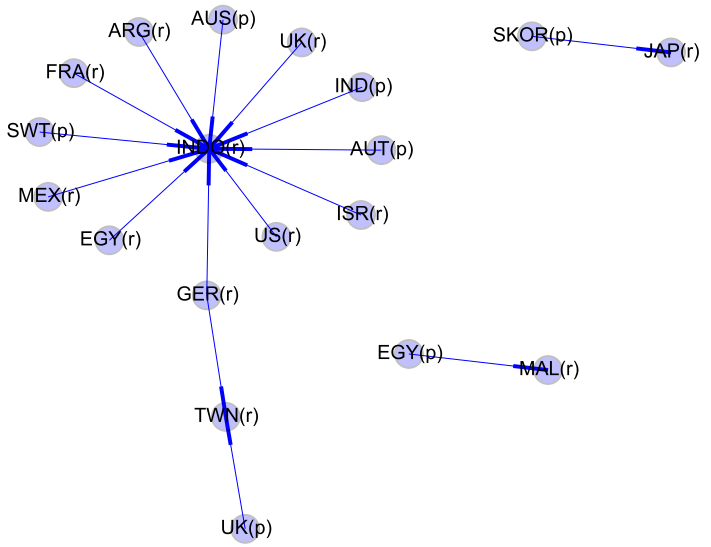



(a)

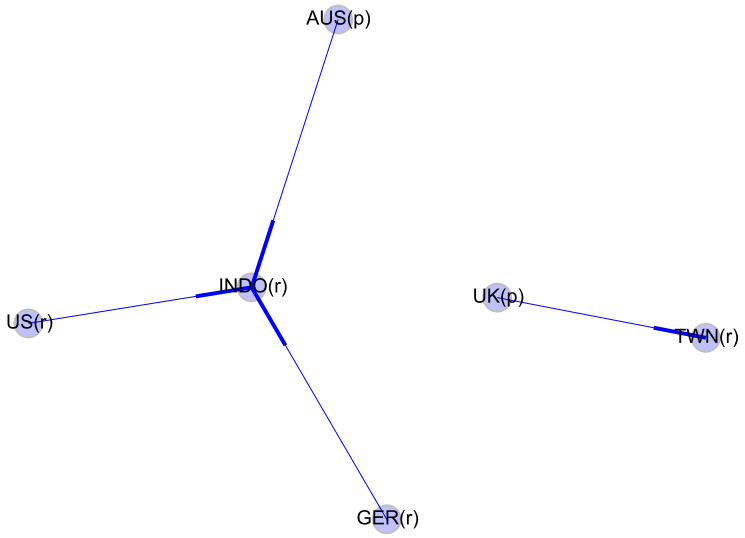



(b)

Figura 7.7: Conexiones que prevalecen en el periodo de estudio de Twitter para (a) $t_c = 0.10$ (b) $t_c = 0.13$.

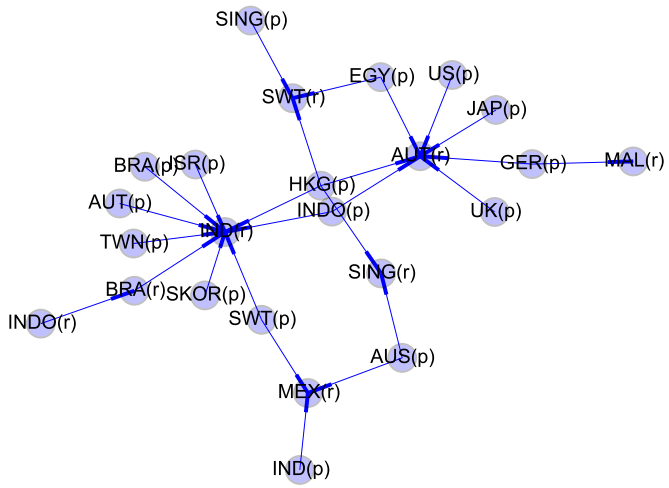



(a)

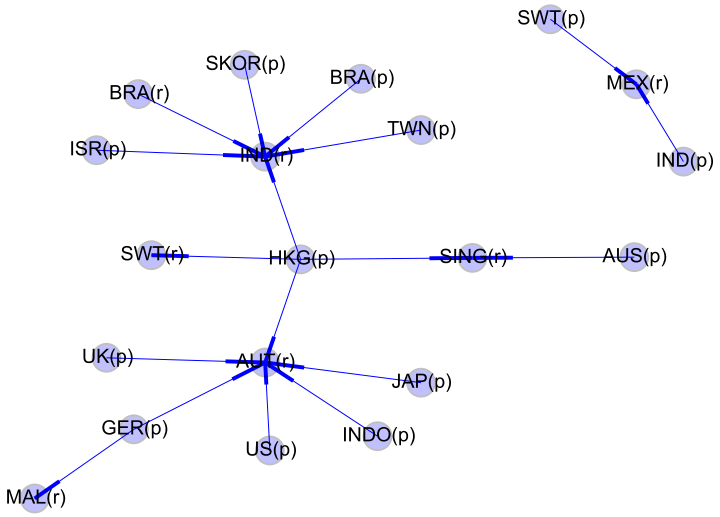



(b)

Figura 7.8: Conexiones que prevalecen en el periodo de estudio de Twitter para (c) $t_c = 0.16$ (d) $t_c = 0.19$

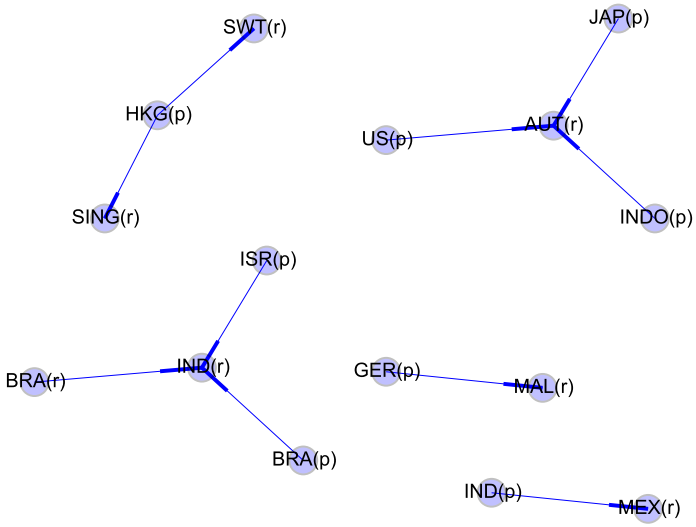



(a)

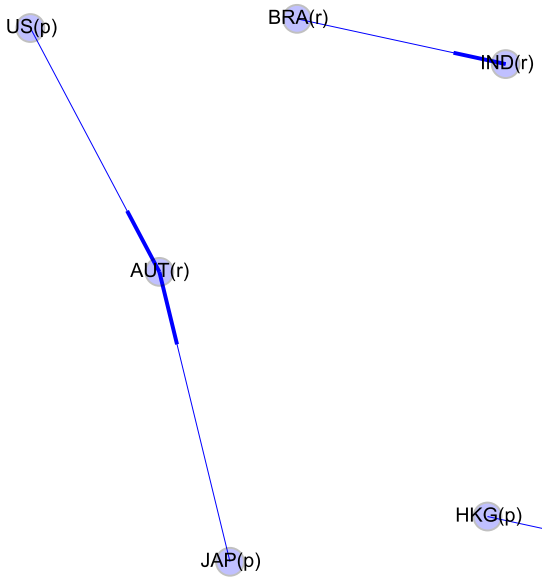



(b)

Figura 7.9: Conexiones que prevalecen en el periodo de estudio de NYT para (a) $t_c = 0.05$ (b) $t_c = 0.06$




(a)




(b)

Figura 7.10: Conexiones que prevalecen en el periodo de estudio de NYT para (a) $t_c = 0.07$ (b) $t_c = 0.08$

# Capítulo 8

# Conclusión

Se logró extraer información de las fuentes de Twitter y de NYT mediante el ensamblaje de distintos lenguajes de programación, cuantificando su contenido mediante técnicas de análisis de sentimiento. Donde las técnicas matemáticas provenientes de la física estadística mostraron que los datos extraídos de estas fuentes contienen información relevante para el análisis de los índices financieros globales estudiados aquí.

El análisis de taxonomía y jerarquía del capítulo 5, reveló la formación de cierta estructura geográfica, la cual sin embargo difiere de la encontrada para los retornos en ambos periodos de estudio. No obstante, la formación de ciertos agrupamientos o *clusters* se preserva, como son el de Alemania y Francia para el caso de NYT, lo que nos hace pensar que la información de las fuentes informativas se encuentra desfasada de los precios de los índices financieros. Estos resultados son interesantes ya que pueden ser una primera evidencia de nuestro estudio en contra de la hipótesis de mercado eficiente, pero hace falta realizar un estudio temporal más extenso para comprobar esta idea.

Más allá de si se pueden predecir o no los mercados financieros, con la ayuda del análisis de RMT se ha podido descubrir que la información textual presenta correlación verdaderas, es decir, que escapan de la zona de ruido. Por lo que el hecho de que en este momento no seamos capaces de generar estrategias de compra en las bolsas de valores o de prevenir posibles crisis financieras a partir de la información textual, no imposibilita que en un futuro podamos hacerlo.

Siendo más específicos, mediante el análisis de RMT del capítulo 6, se ha encontrado que los datos de Twitter y NYT comparten la misma estructura de correlaciones encontrada con los datos de retorno asociados a cada periodo de estudio. Asimismo, se han encontrado desviaciones largas de los eigenvalores, más allá de los límites predichos por la ley de Marčenko-Pastur. Sin embargo, fue necesario aplicar técnicas recién desarrolladas para los modelos correlacionados de Wishart, así como resultados para matrices no-simétricas (CWE) para asegurarnos que están desviaciones no surgían por efectos de la finitud de las matrices de correlación empírica. Con los resultados de este análisis se pudo concluir que existen correlaciones verdaderas entre los indices financieros, las polaridades y la mezcla de ambos indicadores, para ambos periodos de estudio, es decir, el de Twitter y el de NYT.

Además de esto, se han encontrado una correlación moderada entre los comportamientos temporales de los eigenvalores extremos de las polaridades y retornos en ambos periodos de estudio. Esto implica que la información colectiva de los índices financieros globales emerge también al analizar las polaridades, y por lo tanto el portafolio de inversión óptimo o más diversificado es preservado al utilizar este tipo de información. Asimismo, con el análisis de Tracy-Widom se pudieron confirmar estos resultados, observando desviaciones significativas de lo predicho por la teoría. Por otro lado, se encontró para todos los casos, que los valores de $IPR_N$ fluctúan cerca del límite inferior $1/N$, mientras que para $IPR_1$ se obtuvieron los valores más altos como es de esperarse para el caso del portafolio más y menos diversificado, respectivamente. Es notable observar que esta característica surge cuando trabajamos con las polaridades de ambas fuentes textuales. Ello implica que la estructura de las correlaciones globales puede ser preservada independientemente de si estamos trabajando con los las fuentes de noticias (Twitter y NYT) o la información financiera (retornos). Este conjunto de resultados obtenidos por medio de RMT sugieren que los retornos y polaridades comparten una estructura de correlaciones común para los países y periodos de tiempo estudiados aquí.

Por otro lado, en el capítulo 7 se ha hecho un análisis de causalidad y transferencia de entropía. En la primera parte el test de causalidad de Granger revela que algunas series de tiempo de polaridad contienen información predictiva sobre las series de tiempo de retornos, siendo el índice de Indonesia el que mostró recibir mayor información predictiva para una amplio número de días en el caso del primer periodo de estudio (Twitter), mientras que para el segundo periodo (NYT), Egipto mostró los mejores resultados. Por otro lado, el análisis de transferencia de entropía mostró, después de calcular la trans-

ferencia de entropía efectiva (TEE) y transformar la matriz asociada a una red dirigida, la formación de agrupamientos o *clusters* donde para ciertos valores críticos de TEE la información sólo es transmitida hacia los retornos desde las polaridades. Estos resultados nuevos constituyen ya una fuerte evidencia en contra de la hipótesis de mercado eficiente y se espera publicarlos a la brevedad.

En suma, estos nuevos resultados apoyan el paradigma de la economía conductual, es decir, de que las decisiones de los inversionistas se ven influenciados por las información de los medios de comunicación (NYT), e incluso por información divulgada por las redas sociales (Twitter) lo cual influye para generar juicios o estrategias precipitadas de comercio en el mercado de valores, influyendo finalmente estas decisiones en el precio final de las acciones.

# Apéndice A

# Códigos para manejo de información

A continuación se muestran dos códigos para lidiar con los datos textuales. El primero es para generar una base de datos en SQL de la colección pública de *tweets*, mientras que el segundo muestra un script en BASH (SHELL de Linux) para limpiar las noticias de NYT, el cual elimina el texto que podría causar ruido en el cálculo de sentimiento.

## A.1. Base de datos de Twitter

```
import MySQLdb
import numpy as np
import os
import subprocess
from datetime import datetime
from time import gmtime, strftime

db = MySQLdb.connect(host="localhost", # your host, usually localhost
                     user="root",      # your username
                     passwd="root",    # your password
                     db="Sentiment")   # name of the database

# you must create a Cursor object.
# It will let you execute all the query you need
cur = db.cursor()
current = strftime("%Y-%m-%d %H:%M", gmtime())
```

```
#configure PATH of current directory
path="my-current-directory"

#collect & query
cur.execute("SELECT * FROM Market")
for Market in cur.fetchall() :
    cur2 = db.cursor()
    command = "SELECT * FROM Keywords WHERE idMarket =,\
    %d ORDER BY idKeywords DESC LIMIT 1" %  Market[0]
    cur2.execute(command)
    for Keywords in cur2.fetchall() :
        print Market[0] , Keywords[2]
        subprocess.call([path + 'collect.sh', str(Keywords[2])])
        subprocess.call([path + 'query.sh', str(Keywords[2])])
        sentiment = open(path + 'sentiment.txt', "r")
        subprocess.call('rm '+path+'sentiment.txt', shell=True)
values = sentiment.readlines()
sentiment.close()
values = [item.rstrip('\n') for item in values]
polarity = values[0]
subjectivity = values[1]
print values
        cur3 = db.cursor()

        command = "INSERT INTO Data_twi (Date,Polarity,Subjectivity,\
        idKeyword) VALUES ('%s', %s, %s, %d)" % (current, polarity,\
        subjectivity, Keywords[0])
        cur3.execute(command)
        db.commit()
```

# A.2. Limpieza de texto de NYT

```
#!/bin/bash

declare -a array=("Advertisement" "Sections" "Home" "Search"\
"Skip to content" "Skip to navigation" "View mobile version"\
"Subscribe Now" "Log In" "Settings" "Close search"  "search \
sponsored by" "Clear this text input" "Loading..." "See next\
articles" "See previous articles" "Site Navigation" "Site \
Mobile Navigation" "Supported by" "LEARN MORE »" "Share This\
Page" "Related Coverage" "The New York Times")
```

```
folder=$1 #folder where news are contained
current=$(pwd)
FILES=${current}/world/${folder}/*
for f in $FILES; do
  echo $f
  for i in "${array[@]}"; do
    sed -i "/${i}/d" $f
  done
  sed -i '/NYTimes.com no longer supports Internet Explorer 9/d' $f
  sed -i '/http/d' $f
  sed -i 's/Continue reading the main story//' $f
  sed -i 's/             0//' $f
  sed -i 's/     Go//' $f
  sed -i '/''What'\''s Next/,$d' $f
  sed -i '/^\s*$/d' $f
done
```

# Apéndice B

# Estimadores

En esta sección seguiremos los textos [100, 101] para hablar de los estimadores numéricos de la transferencia de entropía. No obstante la formulación matemática de la sección anterior es directa, en la practica estimar la transferencia de entropía a partir de un número finito $N$ de series de tiempo empíricas puede llegar a ser un proceso muy complejo, ya que es dependiente del tipo de datos que estemos trabajando así como de sus propiedades. Los estimadores son típicamente sujetos a sesgo y varianza debido al uso de muestras de tamaño pequeño.

Un estimador es una función o regla que toma los datos observados como entrada y salida a una estimación de un parámetro desconocido o variable [102]. Cualquier estimador puede ser caracterizado en términos del sesgo y varianza de sus estimaciones, es decir, su desviación sistemática del valor verdadero, así como su variabilidad a través de diferentes realizaciones de pruebas. No importa que tan grande sea el número de muestras, un número finito nunca determina completamente una densidad de probabilidad continua arbitraria. Una primera división usual entre tipos de estimadores consiste en la separación entre aquellos que son paramétricos de los que no lo son.

Los estimadores paramétricos asumen que la densidad de probabilidad en cuestión pertenece a cierta categoría o familia de distribuciones, y siempre comienzan infiriendo aquellos parámetros de la familia que mejor se ajusten a la distribución empírica. En la mayoría de los casos reales no existe una simple familia de distribuciones que se ajuste a los datos. También se da el caso de que las distribuciones se tengan que ajustar en la variable temporal debido a las diferentes condiciones empíricas. De esta manera, las series de tiempo

en los regímenes no estacionarios, no se son adecuadamente descritas por una sola familia de distribuciones. En estas situaciones, la aproximación paramétrica para estimar la transferencia de entropía o cualquier función en general, no es recomendado.

Las aproximaciones no-paramétricas hacen solamente asunciones débiles acerca de la continuidad y diferenciabilidad de la distribución, y en todo caso asumen que los datos pertenezcan a alguna familia de distribuciones de probabilidad en particular. Es por ello que nos enfocaremos exclusivamente en este tipo de estimadores, para este propósito seguiremos el trabajo de Hlaváčková et al. [103], describiendo los dos principales tipos de estimadores no-paramétricos.

## B.1. Estimación basada en particiones

El método más intuitivo para estimar densidades de probabilidad es posiblemente el basado en histogramas. La idea se basa simplemente en estimar densidades de probabilidad mediante el conteo del número de cajas que caen dentro de cada división de cierta partición del espacio de estados. Incluso, dicho procedimiento corresponde al estimador más probable para una distribución de probabilidad. Por lo que en principio es posible estimar la transferencia de entropía mediante el computo de las frecuencias de visita a cada uno de los estados, esto como una aproximación para cada una de las probabilidades involucradas en la eq. 7.11. No obstante, debido a la concavidad de la función logarítmica incluso un estimador con sesgo asintótico puede resultar en un sesgo significativo en el cálculo de la transferencia de entropía e incluso en la entropía de Shannon.

## B.2. Estimadores Plug-in

La idea detrás de este tipo de estimadores es el encontrar una estimación consistente para la densidad de probabilidad y *enchufarla* (conocido en inglés como plug-in) en el funcional correspondiente. Sin embargo, en contraste con la aproximación paramétrica, no asume a priori la forma general de las densidades. Así, las densidades no son forzadas a pertenecer a ninguna familia de distribuciones. En vez de ello, las densidades son estimadas tipicamente utilizando técnicas más flexibles como son el estimador de densidad de kernel (o estimador Parzen) [104, 105, 76].

En los estimadores de kernel una densidad es escrita como la suma de funciones de kernel que decaen, como lo son el kernel gaussiano o el de caja, los cuales se centran en las observaciones empíricas $x_1, x_2, \ldots, x_N$. Estas expresiones se justifican teóricamente puesto que puede demostrarse que es equivalente a estimar la función de densidad a partir de la transformada inversa de Fourier de su función característica $\frac{1}{N}\sum_{t=1}^{N}\exp(i\lambda x_t)$.

El ancho de banda o suavidad de la ventana del kernel local $h$, compromete el sesgo contra un error estadístico de balance. Los estimadores de densidad de kernel son una solución muy popular para superar aquellos problemas donde la aproximación por histogramas tiene limitaciones, como la sensibilidad al ruido cerca de los bordes y el problema de la localización arbitraria de los bordes. Para variables aleatorias continuas, la suma de kernel diferenciables converge más rápido a la densidad que las técnicas basadas en particiones [106]. En la fig. B.1(a) se superponen los resultados de 3 distintos valores de $h$ (en la figura escrito como $bw$) a un histograma dado, se puede observar que al usar valores mas pequeños de $h$, la distribución asociada representa con más detalle aquella dada por el histograma, sin embargo, si no necesitamos tanto detalle o resolución optaríamos por valores más altos de $h$. En la fig. B.1(b) se muestran las distintas opciones de kernel que PYTHON tiene a nuestra disposición.

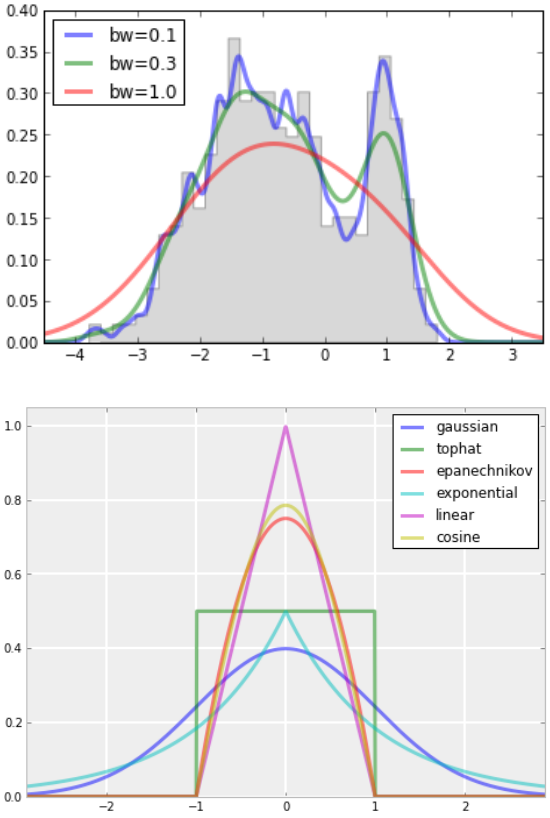



Figura B.1: Figura (a): Distintas elecciones de ancho de banda $bw$ (en el texto referido como $h$) para un histograma dado. (b) opciones de kernel dadas por PYTHON.

# Bibliografía


[1] R. N. Mantenga y H. E. Stanley. *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, Cambridge, 2000.

[2] J. P. Bouchaud y M. Potters. *Theory of Financial Risks: from Statistical Physics to Risk Management*. Cambridge University Press, Cambridge, 2000.

[3] J. Voit. *The Statistical Mechanics of Financial Markets*. Springer-Verlag, Berlin, 2005.

[4] H. Markowitz. *Portfolio Selection: Efficient Diversification of Investment*. John Wiley & Sons, New York, 1959.

[5] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. Nunes Amaral y H. E. Stanley. En: *Phys. Rev. Lett.* 83 (1999), pág. 1471.

[6] L. Laloux, P. Cizeau, J. P. Bouchaud y M. Potters. En: *Phys. Rev. Lett.* 83 (1999), pág. 1467.

[7] V. Plerou, P. Gopikrishman, B. Rosenow, L. A. N. Amaral y H. E. Stanley. En: *Physica A* 287 (2000), pág. 374.

[8] V. Plerou, P. Gopikrishman, B. Rosenow, L. A. N. Amaral, T. Guhr y H. E. Stanley. En: *Phys. Rev. E* 65 (2002), pág. 066126.

[9] M. Potters, J.-P. Bouchaud y L. Laloux. En: *Acta. Phys. Pol. B* 36 (2005), pág. 2767.

[10] L. Medina y R. Mansilla. En: *Rev. Admon. Fzas. Ecom.* 2 (2008), pág. 117.

[11] M. C. Münnix, R. Schäfer y T. Guhr. En: *Physica A* 389 (2010), pág. 767.

[12] S. Maslov. En: *Physica A* 301 (2001), pág. 397.

[13] D. Wang, B. Podobnik, D. Horvatić y H.E. Stanley. En: *Phys. Rev. E* 83 (2011), pág. 046121.

[14] L. Sandoval Jr. e I. D. P. Franca. En: *Physica A* 391 (2012), pág. 187.

[15] S. Kumar y N. Deo. En: *Phys. Rev. E* 86 (2012), pág. 026101.

[16] E. F. Fama. En: *J. Bus.* 38 (1965), pág. 34.

[17] X. Zhang, H. Fuehres y P.A. Gloor. En: *Procedia Soc. Behav. Sci.* 26 (2010), pág. 55.

[18] J. Bollen, H. Mao y X. Zeng. En: *J. Comput. Phys.* 2 (2011), pág. 1.

[19] T. Preis, H. S. Moat y H. E. Stanley. En: *Sci. Rep.* 3 (2013), pág. 1684.

[20] M. Alanyali, H. S. Moat y T. Preis. En: *Sci. Rep.* 3 (2013), pág. 3578.

[21] I. Zheludev, R. Smith y T. Aste. En: *Sci. Rep.* 4 (2014), pág. 4213.

[22] Nicholas Barberis. *Handbook of the Economics of Finance, pag. 1051*. Elsevier Science B.V., North-Holland, 2003.

[23] Robert J. Shiller. *Irrational Exuberance*. Princeton University Press Princeton, New Jersey, 2002.

[24] Zvi Bodie, Alex Kane y Alan J. Marcus. *Investments, 10th Edition*. McGraw-Hill Education, New York NY, 2014.

[25] B. Pang y L. L. Found. En: *Trends. Network.* 2 (1986), pág. 1.

[26] D. Jurafsky y J. H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Prentice-Hall, Englewood Cliffs NJ, 2000.

[27] H. Zhichao. *M.Sc. thesis: Data and Text Mining of Financial Markets Using News and Social Media*. University of Manchester, Manchester, 2012.

[28] P. C. Tetlock. En: *J. Finance* 62 (2007), pág. 1139.

[29] F. Lillo, S. Miccichè, M. Tumminello, J. Piilo y R. N. Mantegna. En: *Quant. Finance* 15 (2015), pág. 2013.

[30] *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*. Vol. 14. Ann Arbor, Michigan, US, 2014.

[31] M. Mézard, G. Parisi, N. Sourlas, G. Toulouse y M. Virasoro. En: *Phys. Rev. Lett.* 52 (1984), pág. 1165.

[32] R. Rammal, G. Toulouse y M.A. Virasoro. En: *Rev. Mod. Phys.* 58 (1986), pág. 765.

[33] A. Cacciuto, E. Marinari y G Parisi. En: *J. Phys. A* 30 (1996), pág. 263.

[34] K. Pearson. En: *Proceedings of the Royal Society of London* 58 (1895), pág. 240.

[35] R. N. Mantegna. En: *Eur. Phys. J. B* 11 (1999), pág. 193.

[36] Papadimitriou, C. H. y K. Steiglitz. *Combinatorial Optimization*. Prentice-Hall, Englewood Cliffs, 1982.

[37] D. B. West. *Introduction to Graph Theory*. Prentice-Hall, Englewood Cliffs NJ, 1996.

[38] R. C. Prim. En: *Bell Syst. Tech. J.* 36 (1957), pág. 1389.

[39] J. B. Kruskal. En: *Proc. Amer. Math. Soc.* 7 (1956), pág. 48.

[40] A. D. Gordon. *Classification*. Chapman y Hall, London, 1981.

[41] F. Murtagh. En: *Comput. J.* 26 (1986), pág. 354.

[42] J. Wishart. En: *Biometrika* 20A (1928), pág. 32.

[43] E. Cartan. En: *Abh. Math. Sem. Hamburgischen Univ.* 11 (1928), pág. 116.

[44] L.K. Hua. *Harmonic analysis of functions of several complex variables in the classical domains*. Science Press, Pekin, 1958.

[45] E. Wigner. En: *Ann. Math.* 62 (1955), pág. 548.

[46] M. L. Mehta. *Random matrices and the statistical theory of energy levels*. Academic Press, New York, 1967.

[47] T. A. Brody, J. Flores, J. B. French, P. A. Mello, A. Pandey S. S. M. y Wong. En: *Rev. Mod. Phys.* 53 (1981), pág. 385.

[48] T. Guhr, A. Müller-Groeling y H. A. Weidenmüller. En: *Phys. Rep.* 299 (1998), pág. 189.

[49] R.U. Haq O. Bohigas y A. Pandey. *In Nuclear Data for Science and Technology*. K.H. Böchhoff edition. Reidel, Dordrecht, 1983.

[50] G. Casati, F. Valz-Gris e I. Guarnieri. En: *Lett. Nuovo Cim.* 28 (1980), pág. 8.

[51] O. Bohigas, M.J. Giannoni y C. Schmit. En: *Phys. Rev. Lett.* 52 (1984), pág. 1.

[52] V. A. Marčenko y L. A. Pastur. En: *Sb. Math.* 72 (1967), pág. 507.

[53] L. Laloux, P. Cizeau, J. P. Bouchaud y M. Potters. En: *Int. J. Theor. Appl. Finance.* 3 (2000), pág. 391.

[54] B. Rosenow, V. Plerou, P. Gopikrishnan y H.E. Stanley. En: *Europhys. Lett.* 59 (2002), pág. 500.

[55] L. Sandoval, A. B. Bortoluzzo y M. K. Venezuela. En: *Physica A* 410 (2014), pág. 94.

[56] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, 2003.

[57] R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley Intersience, 2005.

[58] P. J. Forrester. *Log-Gases and Random Matrices*. Princeton University Press, 2010.

[59] T. Wirtz. *M.Sc. thesis: Aspects of Spectral Statistics in the Correlated Wishart Model*. Universität Duisburg-Essen, 2015.

[60] J. Verbaarschot, H. Weidenmüller y M. Zirnbauer. En: *Phys. Rep* 129 (1985), pág. 367.

[61] K. Efetov. *Supersymmetry in Disorder and Chaos*. Cambridge University Press, 1997.

[62] C. Recher, M. Kieburg y T. Guhr. En: *Phys. Rev. Lett.* 105 (2010), pág. 244101.

[63] D. P. Kroese, T. Brereton, T. Taimre y Z. I. Botev. En: *WIREs Comput. Stat.* 6 (2014), pág. 386.

[64] Vinayak. En: *Phys. Rev. E* 88 (2013), pág. 042130.

[65] G. Livan y L. Rebecchi. En: *Eur. Phys. J. B* 85 (2012), pág. 213.

[66] Rudi Schäfer y Thomas Guhr. En: *Physica A* 389 (2010), pág. 3856.

[67] J. P. Bouchaud y M. Potters. *The Oxford Handbook of Random Matrix p. 824*. Oxford University Press, Oxford, 2011.

[68] C. Spearman. En: *Am. J. Psychol.* 15 (1904), pág. 72.

[69] M. Chiani. En: *J. Multivar. Anal.* 129 (2014), pág. 69.

[70] I. M. Johnstone. En: *Ann. Stat.* 29 (2001), pág. 295.

[71] C. Tracy y H. Widom. *New Trends Mathematical Physics p.753*. Springer, Rio de Janeiro, 2009.

[72] C. Tracy y H. Widom. En: *Commun. Math. Phys.* 177 (1996), pág. 727.

[73] Annick Lesne. En: *Math. Structures Comput. Sci.* 24 (2014), e240311.

[74] C. W. J. Granger. En: *Econometrica* 37 (1969), pág. 424.

[75] R. A. Fisher. En: *J. R. Stat. Soc.* 85 (1922), pág. 87.

[76] T. Schreiber. En: *Phys. Rev. Lett.* 85 (2000), pág. 461.

[77] C. E. Shannon. En: *Bell Syst. Tech. J.* 27 (1948), págs. 379,623.

[78] L. Barnett. En: *Phys. Rev. Lett.* 103 (2009), pág. 238701.

[79] G. V. Steeg y A. Galstyan. En: *Proceedings of the 21st International Conference on World Wide Web, Lyon, France* 21 (2012), pág. 509.

[80] M. Lungarella, K. Ishiguro e Y. Kuniyoshi y N. Otsu. En: *Int. J. Bifurcat. Chaos* 17 (2012), pág. 903.

[81] J. T. Lizier, M. Prokopenko y A. Y. Zomaya. En: *Phys. Rev. E* 77 (2008), pág. 026110.

[82] J. T. Lizier y M. Prokopenko. En: *Eur. Phys. J. B* 73 (2010), pág. 605.

[83] A. Papana y D. Kugiumtzis. En: *Phys. Rev. E* 83 (2011), pág. 036207.

[84] W. L. Shew, H. Yang, S. Yu, R. Roy y D. Plenz. En: *J. Neurosci.* 31 (2011), pág. 55.

[85] R. Vicente, M. Wibral, M. Lindner y G. Pipa. En: *J. Comput. Neurosci.* 30 (2011), pág. 45.

[86] L. Faes, G. Nollo y A. Porta. En: *Entropy* 15 (2013), pág. 198.

[87] S. K. Baek, W. S. Jung y H. T. Moon O. Kwon. En: *ArXiv.org* physics (2005), pág. 0509014v2.

[88] O. Kwon y J. S. Yang. En: *Euro. Phys. Lett.* 82 (2005), pág. 68003.

[89] P. Jizba, H. Kleinert y M. Shefaat. En: *Physica A* 391 (2012), pág. 2971.

[90] L. Sandoval Jr. En: *Entropy* 16 (2014), pág. 4443.

[91] L. Barnett. En: *Phys. Rev. Lett.* 109 (2012), pág. 138105.

[92] X. S. Liang. En: *Entropy* 15 (2013), pág. 327.

[93] M. Prokopenko y J. T. Lizier. En: *Entropy* 15 (2013), pág. 524.

[94] S. Kullback. *Information Theory and Statistics*. Wiley, 1959.

[95] C. E. Shannon y W. Weaver. *The Mathematical Theory of Information*. University of Illinois Press, Urbana, IL, 1963.

[96] H. Kantz y T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge, MA, 1997.

[97] Joseph T. Lizier. En: *arXiv:1408.3270 [cs.IT]* (2014).

[98] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London, 1998.

[99] R. Marschinski y H. Kantz. En: *Euro. Phys. J. B* 30 (2002), pág. 275.

[100] J. T. Lizier. En: *Front. Robot. AI* 1 (2014), pág. 11.

[101] M. Wibral, R. Vicente y J. T. Lizier. *Directed Information Measures in Neuroscience*. Springer-Verlag Berlin Heidelberg, 2014.

[102] S.M. Kay. *Fundamentals of statistical signal processing. In: Estimation Theory, vol. 1*. Prentice Hall, New Jersey., 1993.

[103] K. Hlaváčkoá-Schindler, M. Palus amd M. Vejmelkaa y J. Bhattacharya. En: *Phys. Rep.* 441 (2007), pág. 1.

[104] B. W. Silverman. *Density estimation for statistics and data analysis*. CRC Press, 1986.

[105] M. Young-Il, B. Rajagopalan y U. Lall. En: *Phys. Rev. E* 52 (1995), pág. 2318.

[106] R. Steuer, J. Kurths, C. O. Daub, J. Weise y J. Selbig. En: *Bioinformatics* 18 (2002), pág. 231.