



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

POSGRADO EN CIENCIAS FÍSICAS
INSTITUTO DE CIENCIAS FÍSICAS

**PROPIEDADES ESTADÍSTICAS
DEPENDIENTES DEL TIEMPO EN UN
LIBRO DE ÓRDENES DE BITCÓIN.**

TESIS

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

PRESENTA:

ROBERTO MOTA NAVARRO

TUTOR PRINCIPAL

DR. FRANCOIS ALAIN LEYVRAZ WALTZ
INSTITUTO DE CIENCIAS FÍSICAS, UNAM.

MIEMBROS DEL COMITÉ TUTOR

DR. HERNÁN LARRALDE RIDAURA,
INSTITUTO DE CIENCIAS FÍSICAS, UNAM.

DR. THOMAS GORIN
CENTRO UNIVERSITARIO DE CIENCIAS EXACTAS E INGENIERÍAS, UdG.

CUERNAVACA, MORELOS. JUNIO 2023.



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Índice general

1	Introducción	3
§1.1	Descripción de la contribución de este trabajo	3
§1.2	Resumen de la tesis	5
2	Antecedentes teóricos	7
§2.1	Mercados financieros	7
§2.1.1	Mercados Primarios y Secundarios	8
§2.1.2	Estructura de un mercado financiero	9
§2.1.3	Estudios previos de la estructura del libro de órdenes	12
§2.2	Herramientas estadísticas no estándar	20
§2.2.1	Función de autocorrelación parcial	20
§2.2.2	Técnicas de descubrimiento causal empírico	21
3	Resultados	35
§3.1	Estadística de gaps	35
§3.1.1	Preprocesamiento de los datos	35
§3.1.2	Distribución de tamaños	37
§3.1.3	Estructura temporal de los gaps	40
§3.1.4	Correlaciones entre gaps, rendimientos y volatilidades	43
§3.1.5	Relaciones causales entre gaps y rendimientos	48
§3.2	Estadística de volúmenes	62
§3.2.1	Distribuciones de volumen	63
§3.2.2	Gaussianidad Agregacional	64
§3.2.3	ACF y PACF de rendimientos brutos	66
§3.2.4	ACF de rendimientos absolutos	68
§3.2.5	Cuantificación directa de la aglomeración	70
§3.2.6	Tamaños de volumen relativos	72
§3.2.7	Rachas de volumen constante	73
4	Conclusiones	78

Capítulo 1

Introducción

1.1. Descripción de la contribución de este trabajo

La creación de modelos de predicción de la volatilidad en los mercados financieros es uno de los temas de investigación más activos en las áreas de la econometría y la econofísica. Por un lado, la generación de modelos que reproduzcan las propiedades cualitativas de las series de tiempo de la volatilidad ayuda al descubrimiento y comprensión de los procesos subyacentes responsables de esas propiedades. Por otro lado, la reproducibilidad y predictibilidad de las propiedades cuantitativas de la volatilidad es sumamente útil para los usuarios de los mercados financieros. Esto último es así porque poseer modelos que permitan predecir con razonable precisión la magnitud de la volatilidad futura aun en escalas muy cortas de tiempo expande el conjunto de estrategias de inversión que uno puede usar con confianza, ya sea para focalizar la actividad financiera en las épocas en que se tenga contemplado un menor riesgo o para aprovechar épocas de mayor riesgo en espera de mejores rendimientos.

La fuente de datos más común que alimenta a los modelos de predicción de volatilidad suele ser el historial de precios, con los cuales se calculan a su vez los historiales de rendimientos y volatilidades. Con esos datos históricos, se puede aprovechar la autocorrelación de la volatilidad para generar proyecciones de la volatilidad futura, como ocurre por ejemplo con los modelos ARCH y GARCH.

El uso casi exclusivo de la historia de los precios en los modelos de volatilidad hace

muy natural preguntarse si es posible mejorar su poder predictivo utilizando la información contenida en los libros de órdenes. Después de todo, la historia de los precios se extrae directamente de las mejores ofertas almacenadas en los libros de órdenes; los libros pueden extender la información útil disponible con las ofertas que no terminan en una transacción exitosa.

Varios trabajos han sugerido que existe una relación de causalidad entre la densidad de ofertas en los libros de órdenes y la magnitud de los rendimientos en el sentido de que la presencia de grandes diferencias entre las ofertas disponibles en un libro, llamadas *gaps*, tiene el efecto de producir rendimientos de gran magnitud, mientras que diferencias pequeñas inducen la generación de rendimientos de menor magnitud.

Las metodologías empleadas en la literatura que ha sugerido causalidad entre *gaps* y rendimientos tienen en común ignorar el orden temporal con el que se generan los *gaps* y los rendimientos y no utilizar técnicas de descubrimiento causal para alcanzar la conclusión de que los *gaps* de gran tamaño son la causa de los rendimientos de gran tamaño.

Esta decisión metodológica tiene dos inconvenientes mayores: primero, vuelve estadísticamente insuficientes las asociaciones reportadas para defender la hipótesis de causalidad entre *gaps* y rendimientos, después de todo, correlación no implica causalidad. Segundo, ignora el hecho fundamental de la precedencia de una causa frente a sus efectos, sin importar la definición formal de causalidad de la que se haga uso.

En esta tesis estudiamos datos históricos del libro de órdenes del mercado de cambio Bitcon/dólar para medir las correlaciones entre los *gaps* y rendimientos a lo largo del tiempo y conducimos diversas mediciones de causalidad utilizando herramientas estadísticas específicamente construidas para la detección de relaciones causales, contrastando las propiedades de las correlaciones con las mediciones de causalidad. Estas mediciones son importantes porque la confirmación y cuantificación de una relación causal por medios directos y que considere el orden cronológico puede ayudar en la creación de mejores modelos para la predicción de la volatilidad.

Adelantando de manera temprana los resultados que se exponen en las subsecuentes secciones: no pudimos detectar causalidad en un sentido estricto (dependiente del tiempo) pero pudimos replicar los resultados de los trabajos que así lo sugieren al considerar el

caso físicamente irreal de causalidad instantánea, es decir, tiempos nulos de respuesta entre la causa (gaps) y el efecto (rendimientos). Esto alude a que en realidad solo existe una asociación entre ambas variables y es consistente con los resultados que detectaron los autores de los estudios previos. Nuestros resultados ponen en duda la utilidad del acceso a los estados de los gaps para la predicción de la volatilidad futura en las escalas temporales que estudiamos y en el mercado de intercambio bitc6in/d6lar, y sugieren que aun agregando informaci6n de la historia de los libros de 6rdenes, no se gana poder predictivo de la dinámica de precios en mercados financieros.

1.2. Resumen de la tesis

Esta tesis est conformada por tres captulos: antecedentes te6ricos, resultados y conclusiones.

La secci6n de antecedentes te6ricos tiene el objetivo de describir las entidades financieras que estudiamos y situar nuestro trabajo en el contexto de los trabajos previos.

En la primera parte de este captulo exponemos el funcionamiento bsico de los mercados financieros actuales, principalmente el mecanismo de doble subasta que opera utilizando la estructura llamada libro de 6rdenes, en la cual los participantes del mercado introducen sus ofertas de compra o venta (llamadas 6rdenes) en espera de encontrar contraofertas que sean compatibles con los precios que estipulan en sus 6rdenes.

Despu6s presentamos los principales estudios de las propiedades estadsticas de los libros de 6rdenes, la mayora de los cuales se enfocan en propiedades no dependientes del tiempo, como la distribuci6n de los precios de las 6rdenes y la distribuci6n de los volmenes de las 6rdenes, integrados sobre el tiempo. Dentro de esos estudios enfatizamos aquellos que han sugerido una relaci6n causal entre los gaps y los rendimientos; estos estudios son de fundamental relevancia para el trabajo desarrollado en esta tesis, siendo la mayor parte de nuestros esfuerzos dedicada a la detecci6n directa y cuantificaci6n de esas supuestas relaciones causales con herramientas diseadas para discriminar efectos causales de simples correlaciones, de acuerdo a diversas definiciones de efecto causal. Las tcnicas de descubrimiento causal se presentan en seguida, dos de las cuales definen causalidad

de manera estadística y otra utilizando una definición inspirada en la teoría de sistemas dinámicos.

En el capítulo de resultados describimos la metodología que seguimos para la implementación de las pruebas de causalidad y la medición de las funciones de correlación entre los gaps y los rendimientos. Para obtener un número significativo de pruebas que nos permitiera realizar las pruebas de hipótesis apropiadas, dividimos el intervalo de tiempo total que engloba a los datos estudiados en múltiples ventanas de tiempo disjuntas y sobre cada una de esas ventanas efectuamos las diversas pruebas de causalidad, tomando como controles versiones permutadas de los datos. Las pruebas de hipótesis entonces se realizaron sobre las versiones permutadas de los datos, y del ensemble de esas permutaciones se obtuvieron las distribuciones empíricas correspondientes a la hipótesis nula de ausencia de efecto causal. Como resultado colateral, presentamos también la observación de correlaciones de largo rango, aglomeración y reversión a la media en las series de volúmenes disponibles en las órdenes más competitivas, además de otras propiedades que en conjunto sugieren que el proceso de selección de tamaño de orden puede ser más complejo que el de selección de precio de la orden.

Finalmente, está el capítulo de conclusiones, en el que concluimos que no existen relaciones causales entre los gaps y los rendimientos en un sentido estricto de causalidad que considere que las causas ocurren antes que sus efectos. Sin embargo, cuando tenemos en cuenta la posibilidad de tiempos de respuesta nulos, podemos recuperar los resultados que se mencionan en la literatura previa. Esto nos hace pensar que aquello que los demás trabajos presentan como detección de relaciones causales entre los gaps y los rendimientos es en realidad un fuerte acoplamiento instantáneo, que posiblemente refleja causalidad proveniente de otros observables financieros aun no identificados.

Capítulo 2

Antecedentes teóricos

2.1. Mercados financieros

Los sistemas financieros son un conjunto de entidades económicas que permiten el intercambio de fondos entre distintas partes de la sociedad. De manera general, los sistemas financieros están formados por dos tipos de subsistemas: mercados financieros e instituciones financieras. Los mercados financieros son como cualquier otro mercado en el que hay individuos que compran y venden distintos tipos de bienes y regatean los precios; la diferencia con respecto a un mercado normal es que en un mercado financiero se compran y venden instrumentos financieros como acciones, bonos, opciones, futuros, etc. Las instituciones financieras, por otro lado, son firmas (bancos comerciales, fondos de pensión, fondos de inversión, etc.) que proveen de servicios financieros a sus consumidores e invierten sus fondos en instrumentos financieros en lugar de invertirlos en bienes básicos como oro o bienes raíces. Los sistemas financieros son importantes porque ayudan a canalizar dinero de individuos que cuenten con un excedente (inversionistas) de este a individuos que lo necesitan [1].

El flujo de dinero entre los inversionistas y los individuos que solicitan una inversión ocurre por medio de dos mecanismos: (1) financiamiento directo, en el cual los fondos pasan directamente a los mercados financieros y (2) financiamiento indirecto, donde los fondos pasan primero por un intermediario que contacta a los inversionistas con aquellos que soliciten fondos. Independientemente del método de financiamiento, el objetivo es con-

tactar a ambas partes al menor costo posible y de la manera más conveniente. Un sistema financiero eficiente es entonces muy importante porque garantiza la formación adecuada de capital para el crecimiento económico. De esta forma, si el sistema financiero trabaja apropiadamente, las unidades económicas que ofrecen buenas oportunidades de inversión pueden recibir fondos; por ejemplo, si General Motors necesita dinero para pagar una nueva planta de manufactura de carros, podría obtenerlos de manos de inversionistas al venderles bonos, que son un instrumento financiero en el cual General Motors promete hacer pagos recurrentes por un periodo de tiempo especificado. Así, los inversionistas obtienen ganancias con los pagos de los bonos y General Motors consigue los fondos necesarios para construir su nueva planta.

2.1.1. Mercados Primarios y Secundarios

La venta de instrumentos financieros (bonos, acciones, etc.) se realiza inicialmente en los llamados mercados financieros primarios. Para cada tipo de instrumento existe un mercado financiero. Un ejemplo de transacción en un mercado primario es la venta de acciones recientemente creadas. En los mercados financieros secundarios se realiza la compra y venta de instrumentos con los que se ha comerciado previamente en algún mercado primario. Cuando un individuo compra un instrumento en el mercado secundario, el individuo que le vendió el instrumento recibe dinero a cambio de este, pero la compañía que lo expidió no adquiere nuevos fondos. Una corporación adquiere nuevos fondos solo cuando sus instrumentos se venden por primera vez en el mercado primario. Sin embargo, los mercados secundarios cumplen dos importantes funciones [2]: primero, hacen más rápida y fácil la venta de instrumentos para adquirir dinero; a esta facilidad de transformar un bien en dinero se le llama liquidez. La liquidez hace más deseable a un instrumento y eso facilita a la compañía que lo expide venderlo en el mercado primario. Segundo, los mercados secundarios determinan el precio al que se coloca el instrumento por primera vez en los mercados primarios. Los inversionistas que compran instrumentos en el mercado primario no pagarán a las compañías que los expiden más del precio al que piensen que se venderán en el mercado secundario. A mayor precio en el mercado secundario, mayor precio se le asignará a un instrumento nuevo que se venda en el mercado primario y, por lo tanto,

mayores serán los fondos que la compañía pueda recabar. Es por esto que las condiciones en el mercado secundario son importantes al momento de crear nuevos instrumentos financieros.

Los mercados secundarios pueden organizarse de dos formas. Un método es instituir bolsas, donde compradores y vendedores de instrumentos se encuentran en una locación central para realizar transacciones o acceden a sistemas electrónicos en los cuales se pueden ejecutar transacciones mediante el uso de computadoras, este último método de acceso ha sido el más utilizado en los últimos años. El New York Stock Exchange (NYSE), el American Stock Exchange (ASE) y la Bolsa Mexicana de Valores (BMV) son ejemplos de bolsas. El otro método de organización es crear un mercado extrabursátil, en el cual individuos situados en diferentes locaciones ofrecen realizar transacciones de compra y venta sobre un inventario de instrumentos a cualquiera que esté dispuesto a aceptar sus precios [3]. Muchas compañías venden sus acciones a través de los mercados extrabursátiles, aunque la mayoría de las corporaciones grandes cotizan sus instrumentos en la bolsa de valores [3].

2.1.2. Estructura de un mercado financiero

La mayoría de los mercados financieros operan con un sistema llamado mecanismo de doble subasta. En este, cada participante ingresa ofertas de compra o venta al precio que desee, las cuales son visibles a cualquier otro participante del mercado. Cuando un comprador inserta una oferta a un precio que algún vendedor esté dispuesto a pagar, se ejecuta la transacción entre esos dos participantes. Ejemplos de este tipo de sistema son el New York Stock Exchange y el American Stock Exchange. A una oferta de compra o venta se le llama orden. Si una nueva orden no es ejecutada inmediatamente al entrar al mercado por falta de participantes que estén dispuestos a comerciar al precio estipulado, dicha nueva orden es guardada en una lista llamada libro de órdenes. Las órdenes almacenadas en el libro de órdenes permanecen ahí hasta que sean canceladas o arriben otras ofertas compatibles con su precio. En las siguientes secciones se describe con más detalle cada una de estas estructuras para realizar transacciones.

Órdenes

Una orden es una instrucción que un *trader* (el agente que desea comprar o vender) le da a un *bróker* (la persona que se encarga de realizar la transacción). Estas instrucciones especifican con qué instrumento se desea comerciar, en qué cantidad, si se desea comprar o vender, y también pueden incluir información sobre las condiciones a las que se desea que se cumpla la orden. Las condiciones más comunes limitan los precios que el trader está dispuesto a aceptar, otras condiciones pueden especificar por cuánto tiempo estará viva la orden o si es aceptable ejecutarla de manera parcial (que se venda/compre solo una fracción del total indicado en la orden). Las órdenes son necesarias porque la mayoría de los traders no se encargan de ejecutar sus transacciones personalmente. Aquellos traders que sí lo hacen tienen ventaja sobre los que utilizan órdenes porque pueden responder inmediatamente a los cambios del mercado, mientras que los otros deben anticipar dichos posibles cambios e incluir instrucciones que consideren apropiadas para enfrentar los cambios de mercado que les sean perjudiciales.

Al mayor precio entre los precios de las órdenes de compra se le llama *mejor compra* y al menor precio entre los precios de las órdenes de venta se le llama *mejor venta* y a la diferencia entre la mejor compra y la mejor venta se le llama *margen* o *spread*.

Tipos de Órdenes

Existe una gran cantidad de tipos de órdenes, pero la gran mayoría de las transacciones que se registran en los mercados financieros caen en una de dos categorías [4]: órdenes límite y órdenes a mercado. Las órdenes a mercado son instrucciones para realizar transacciones al mejor precio disponible en el mercado. Estas órdenes se ejecutan rápidamente en mercados líquidos, pero pueden tener precios distintos al último precio reportado y son normalmente utilizadas por individuos que desean inmediatez en la ejecución de sus órdenes, incluso si existe el riesgo de obtener un precio desfavorable [5]. Una orden de compra a mercado se ejecutaría al precio de la mejor venta y una orden de venta a mercado se ejecutaría al precio de la mejor compra. Por el otro lado, las órdenes límite son instrucciones para realizar transacciones al mejor precio disponible, pero solo si este no es peor que un *precio*

límite dado, especificado por el trader que introduce la orden: para las órdenes límite de compra, el precio al que se ejecute la transacción debe ser igual o menor que el precio límite; para las órdenes límite de venta el precio de ejecución debe ser igual o mayor que el precio límite.

Libro de órdenes

Se le llama libro de órdenes a la colección de todas las órdenes límite de compra y venta vigentes y este representa el volumen total de bienes disponibles con los cuales se puede comerciar. Cuando una nueva orden llega al mercado, se almacena en el libro si no existen órdenes previamente almacenadas compatibles con ella. La dinámica de las transacciones, y por ende de los precios, depende entonces del flujo de órdenes que arriban al libro. Como las órdenes de mejor compra y mejor venta pueden ejecutarse con el arribo de nuevas órdenes, el spread es una cantidad dinámica que tiene la tendencia de crecer conforme llegan más órdenes de mercado y de hacerse más pequeño con el arribo de órdenes límite. La figura 2.1 presenta un escenario hipotético en el que dos personas insertan órdenes límite a un libro de órdenes. En todos los mercados financieros existe un tamaño mínimo en la diferencia de precios que puede haber entre dos órdenes. Por ejemplo, en el mercado de intercambio bitcoin/dólar que estudiamos en este trabajo, la diferencia mínima de precio entre dos niveles de órdenes límite es de 1 centavo de dólar, de manera que toda pareja de órdenes en el libro se encuentra a un número entero de centavos de diferencia de precios. A la distancia entre dos órdenes consecutivas (sin más órdenes almacenadas entre ellas) se le denomina *gap*. El spread se puede ver entonces como el gap que existe entre la orden de venta de menor precio y la orden de compra de mayor precio.

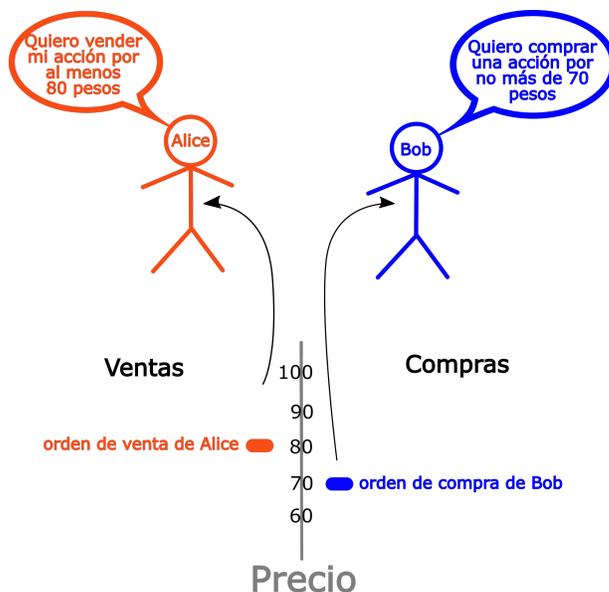


Figura 2.1: Ejemplo de inserción de órdenes límite en el libro. Alice tiene una acción de Google y desea venderla para obtener ganancias; al mismo tiempo, Bob quiere comprar una acción de Google porque cree que sus acciones subirán de precio en el futuro. Alice no desea vender su acción sin obtener por lo menos 80 pesos por la venta, y Bob está dispuesto a pagar a lo sumo 70 pesos por la acción que busca comprar. Para asegurarse de que sus transacciones se realicen a los precios que les parecen adecuados, ambos ingresan una orden límite al libro: Alice una orden de venta a 80 pesos y Bob una de compra a 70 pesos. Debido a que existe una diferencia de 10 pesos entre los precios que satisfacen las condiciones de Alice y de Bob, no se realizará una transacción entre ellos. Sin embargo, eso no implica que no podrán obtener lo que desean; sus órdenes se mantendrán en el libro esperando a que futuros participantes del mercado inserten órdenes que sean compatibles con los precios que Alice y Bob estipularon en sus órdenes límite. A la distancia entre dos órdenes límite almacenadas en el libro se le denomina gap.

2.1.3. Estudios previos de la estructura del libro de órdenes

Como se mencionó anteriormente, en los mercados que implementan un mecanismo de compra/venta de doble subasta continua, todos los intentos de transacciones que no son exitosos quedan almacenados en el libro de órdenes.

Al estar almacenados todos los intentos de transacción, el libro proporciona la mejor descripción microscópica de los mercados financieros, y es posible observar en este las expectativas de los participantes del mercado y de una manera indirecta y aproximada, las estrategias que siguen para cumplir dichas expectativas.

Por el contenido, el estudio de los libros de órdenes se torna relevante para la compren-

sión de la dinámica de los precios.

A continuación, se dará una breve enumeración de los principales trabajos que se han realizado en el estudio de mercados financieros utilizando datos de los libros de órdenes.

En un estudio de la Bolsa de París realizado por Bouchaud et al. [6] se encontró que la distribución de las distancias entre los precios a los que se comercia un activo y los precios de las órdenes límite (es decir, los gaps) que arriban al libro son idénticas para las compras y las ventas, salvo por fluctuaciones estadísticas. La forma de la distribución se aproxima muy bien por la expresión

$$P(\Delta) \propto \frac{\Delta_0^\mu}{(\Delta_1 + \Delta)^\mu}, \quad \Delta \geq 1$$

con un exponente $\mu \approx 0.6$ en un conjunto de acciones del mercado financiero de París. Esta distribución se muestra en la figura 2.2.

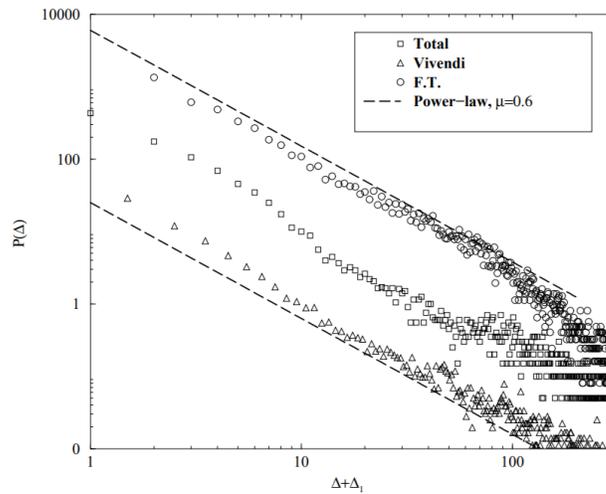


Figura 2.2: Distribución de los precios de las órdenes límite que arriban a los libros de órdenes de tres compañías: Total, Vivendi y F.T., La distribución se presenta como función de la diferencia entre el precio de las órdenes y el precio de la acción Δ . Las líneas rectas corresponden a $\mu = 0.6$. Imagen obtenida de [6].

En un estudio del perfil de volúmenes en el libro realizado por Potters et al. [7], se encontró que el flujo de órdenes entrantes es máximo alrededor del precio más reciente del activo, pero una orden que esté muy cercana al precio más reciente tiene una probabilidad mayor de ejecutarse o cancelarse, y desaparecer así del libro, lo cual hace que no se pueda estimar de manera clara el perfil de volúmenes del libro promedio. Teniendo en cuenta esta

dificultad , el perfil promediado en el tiempo parece ser asimétrico, al tener su máximo por encima del spread, como se puede observar en la figura 2.3

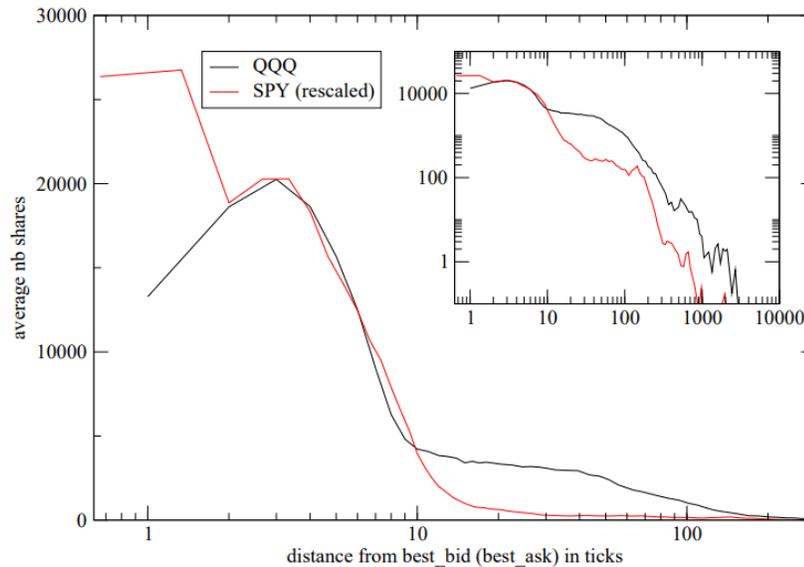


Figura 2.3: Distribución promedio de volumen de los libros de órdenes de los fondos de inversión QQQ and SPY (los cuales siguen, respectivamente, a los índices Nasdaq y SP500) como función de la distancia Δ con respecto al mejor precio de compra (o venta). El eje vertical se reescaló en el caso de SPY. Imagen obtenida de [7].

En un modelo de agentes, propuesto por Farmer et al. [8], se logró producir leyes sencillas que relacionan la llegada de órdenes al libro con algunas propiedades estadísticas del mercado, y comprobaron que esas leyes son válidas al utilizarlas para explicar variaciones en un grupo de 11 acciones reales. El modelo, que solo tiene un parámetro libre, logra explicar el 96 % de la varianza de los spreads y el 76 % de la varianza de la tasa de difusión del precio.

Biais et al. [9] mostraron que la mayoría de las órdenes nuevas que arriban al libro se depositan en niveles de precio muy cercanos al spread mientras que la mayoría del volumen no se acumula en el spread, sino en vecindades cercanas, además de encontrar que los libros que tienen poco volumen total acumulado suelen alentar a los participantes del mercado a ingresar órdenes límite que quedan en espera, mientras que un libro con mucho volumen se asocia con transacciones exitosas. Goldstein et al. [10] reportan que como consecuencia de una disminución en la diferencia mínima entre precios que ocurrió en el New York Stock

Exchange (que pasaron de $1/8$ a $1/16$ de dólar), los spreads y los volúmenes acumulados a lo largo de todo el libro también redujeron sus tamaños. Estos cambios tendieron al efecto combinado de favorecer a los participantes que consumen liquidez del libro con órdenes de pequeño volumen, pero dejando sin ventajas aparentes a los participantes que ingresaban órdenes con grandes volúmenes, especialmente con acciones cotizadas a precios bajos.

Otro estudio realizado por Gao-Feng Gu et al. [11] en el cual se considera a los gaps como series de tiempo, confirma la observación previamente realizada por Bouchaud de que los gaps presentan distribuciones con colas en ley de potencias, además de detectar memoria de larga duración con un exponente de escalamiento mayor a $1/2$ a través del método DFA (detrended fluctuation analysis por sus siglas en inglés [12]), siendo esta memoria independiente de la distribución de los gaps.

Lillo et al. [13] encontraron en un análisis de acciones correspondientes a las 1000 compañías de mayor tamaño del New York Stock Exchange, las cuales poseen diversas capitalizaciones de mercado y liquidez, que es posible colapsar todas las curvas de impacto de precio (el impacto de precio es el cambio de precio condicionado al volumen de una orden de mercado que arriba al libro) en una sola curva maestra al reescalar el cambio de precio promedio y el volumen de cada acción individual. Esto sugiere que las fluctuaciones en el equilibrio de la demanda y oferta de diferentes activos financieros en distintos mercados y con diversas capitalizaciones de mercado siguen en realidad la misma regla estadística.

En uno de los pocos estudios que consideran el orden cronológico en el que ocurren los cambios de la estructura del libro de órdenes como consecuencia del arribo de nuevas órdenes, transacciones y cancelaciones, Zheng et al. [14] lograron mostrar que los balances de liquidez en el spread portan información que permite predecir la dirección de futuras órdenes de mercado que arriben al libro. También se encontró en este estudio que el signo de las transacciones presentes y el volumen almacenado en el spread son las variables que aportan más información para la predicción del signo de la siguiente transacción. Al mismo tiempo y de manera contraintuitiva, la agresividad de las órdenes de mercado, medida por su capacidad para consumir volumen a través de varios niveles de precio, no resultó ser tan importante para el poder predictivo.

De principal relevancia en el contexto de los resultados obtenidos en esta tesis son los

estudios de la relación entre la estadística de los gaps y la estadística de eventos extremos de los rendimientos, ya que uno esperaría que al estar almacenadas todas las ofertas de venta y compra en un libro de órdenes, las propiedades estadísticas de estas ofertas, tales como la separación promedio de precios o el volumen por nivel de precio, tuvieran un efecto sobre la dinámica de los precios. A continuación describimos los trabajos más relevantes que estudian la relación entre los gaps y los rendimientos.

Schmitt et al. [15] examinaron el efecto que tienen los tiempos de vida media de las órdenes sobre las distribuciones de cambios de precio (rendimientos) y encontraron que cuando la vida media de las órdenes es pequeña, la distribución de los rendimientos desarrolla colas pesadas, aunque no en ley de potencias, mientras que con vidas medias de larga duración esta tiende a colapsar a una distribución gaussiana, como se puede ver en la figura 2.4.

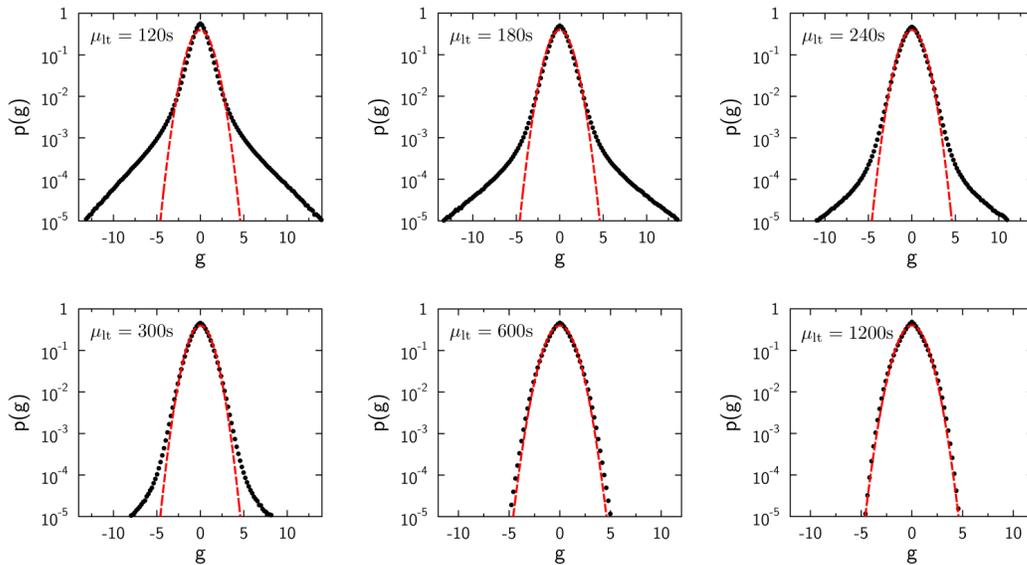


Figura 2.4: Distribuciones de probabilidad de los rendimientos normalizados sobre una escala temporal de 1 minuto para seis escenarios de simulación que difieren solamente en los tiempos de vida media μ_{lt} . La línea roja punteada corresponde a una distribución gaussiana. Imagen obtenida de [15]

Esto se debe a que cuando las órdenes viven poco tiempo antes de desaparecer del libro, los espaciamentos entre los distintos niveles de precios (gaps) se vuelven también de gran magnitud promedio, lo que lleva a grandes saltos de precio conforme las transacciones

consumen los niveles disponibles. Se puede observar en la figura 2.5 la densidad de niveles en el libro de órdenes para dos valores distintos de vida media de la simulación. Cuando la vida media es de 1200 segundos, se forman gaps mucho mayores que cuando es de 120 segundos.

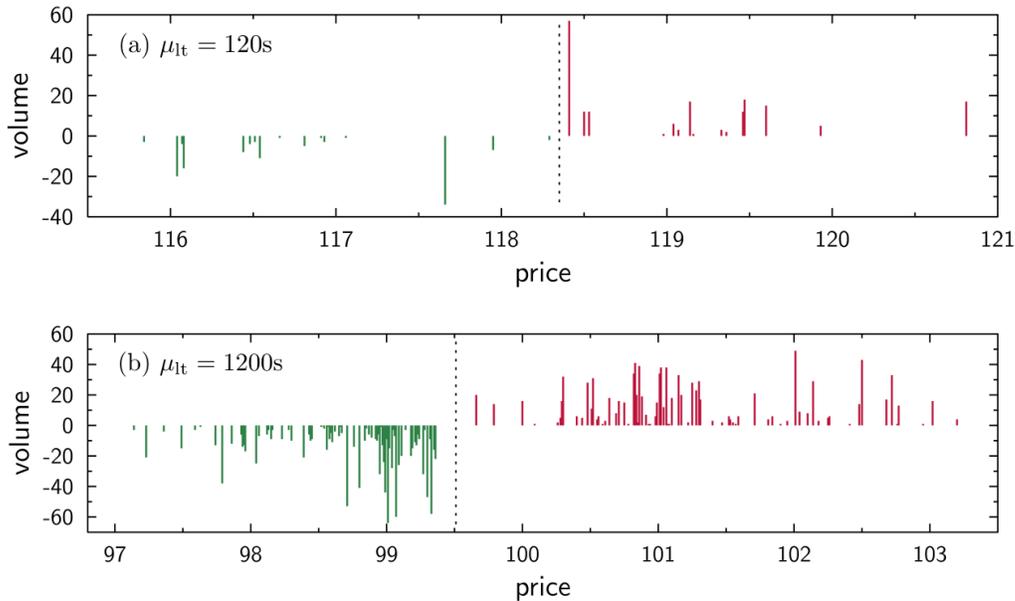


Figura 2.5: Dos estados típicos del libro de órdenes generados por el modelo de Schmitt et al. En el estado (a) se tienen menos niveles discretos de precios ocupados, que forman gaps de gran tamaño. Se asignaron signos negativos a los volúmenes de órdenes de compra para distinguirlos mejor de las órdenes de venta. El precio medio se indica con la línea negra punteada. En el estado (b) el libro presenta una densidad mucho mayor que en el caso (a). Imagen obtenida de [15].

Siguiendo un modelo similar, Cristelli et al. [16] mostraron con un modelo microscópico en el que se organiza de tres maneras distintas al flujo de órdenes que cuando se genera una crisis de liquidez en el libro por un flujo pobre de entrada, se incrementa el tamaño de los gaps, lo que se asocia con una amplificación de las variaciones de precio, y de nuevo se logra que la distribución de rendimientos presente colas pesadas pero en este modelo sí se recuperan colas que decaen en ley de potencias (figura 2.6), como ocurre en mercados verdaderos.

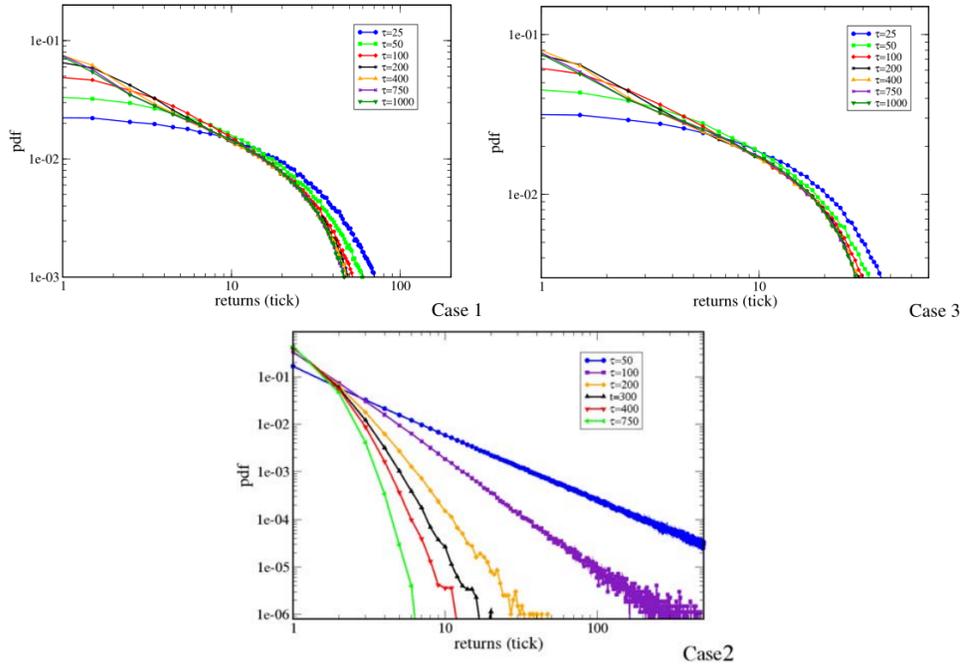


Figura 2.6: Distribuciones de probabilidad de los rendimientos en los tres escenarios de simulación del modelo de Cristelli et al. [17]. En uno de los escenarios que estudiaron (el caso 2), logran recuperar colas que decaen como leyes de potencia con valores realistas. Imagen tomada de [17]

El trabajo más relevante en el estudio de los efectos que los gaps tienen sobre los rendimientos fue realizado por Farmer et al. [18], quienes muestran que las fluctuaciones más grandes de los precios no están relacionadas con transacciones de grandes volúmenes —suposición intuitiva— sino con las fluctuaciones en el número de ofertas presentes en el libro; es decir, con la presencia de grandes gaps entre los niveles de precios ocupados por órdenes límite, como se ilustra en la figura 2.7. De esta manera, los cambios más grandes ocurren cuando una orden de mercado elimina todo el volumen de ofertas situado en uno de los mejores precios, lo cual crea un cambio en el precio proporcional al tamaño del gap entre la mejor oferta de venta (compra) y la siguiente mejor orden de venta (compra).

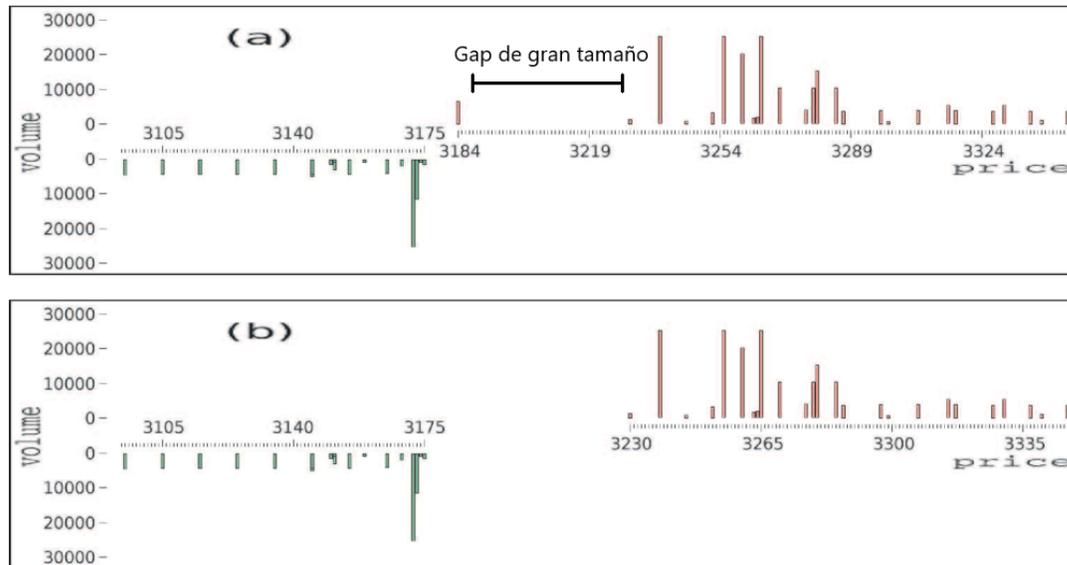


Figura 2.7: Estado del libro de órdenes para la compañía Amazon, justo antes de una fluctuación grande de precio, debida a la desaparición de un gap de gran magnitud en la primera posición. En el panel (a) existe un gap de gran magnitud entre la mejor oferta de venta y la segunda mejor oferta de venta. Tiempo después arribó una orden de mercado para comprar, la cual eliminó el volumen almacenado en la mejor oferta de venta del panel (a), lo que provocó que la nueva mejor oferta de venta esté ahora a un precio mucho mayor que antes, como se muestra en el panel (b). Imagen tomada de [18].

Uno de los resultados que Farmer et al. utilizan para concluir la relación de causalidad entre gaps y rendimientos se muestra en la figura 2.8, en la que se puede observar cómo los exponentes de las colas de las distribuciones de rendimientos y gaps están correlacionados linealmente, y todas las parejas caen muy cerca de su línea de regresión.

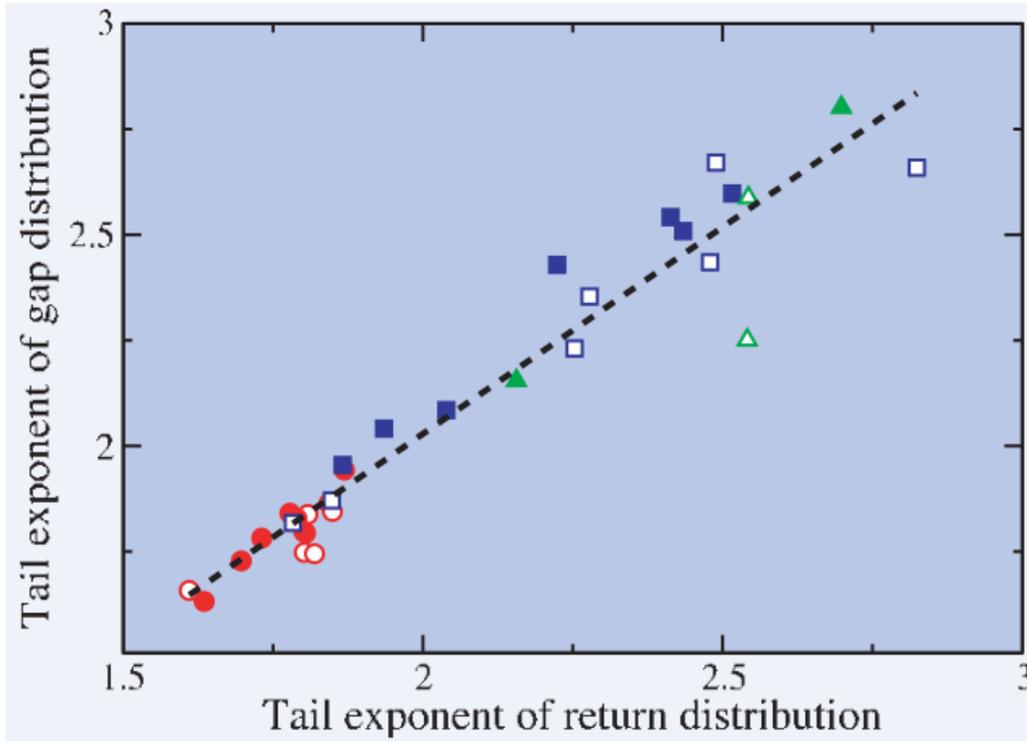


Figura 2.8: Comparación de los exponentes de las colas de las distribuciones de rendimientos (eje horizontal) y los gaps (eje vertical). Los cuadrados azules, triángulos verdes y círculos rojos corresponden a compañías de alta, mediana y baja liquidez, respectivamente. Los símbolos vacíos corresponden a órdenes de mercado, y los rellenos, a órdenes límite. Es evidente que existe una fuerte correlación entre los exponentes de los gaps y los de los rendimientos. Imagen tomada de [18].

2.2. Herramientas estadísticas no estándar

2.2.1. Función de autocorrelación parcial

La función de autocorrelación parcial (PACF por sus siglas en inglés: Partial Autocorrelation Function), está definida como

$$\phi(k) = \begin{cases} C(1) & k = 1 \\ \rho(r(t+k) - \hat{r}(t), r(t) - \hat{r}(t)) & k \geq 2 \end{cases} \quad (2.1)$$

donde

$$\hat{r}(t+k) = \sum_{j=1}^{k-1} \beta_j r(t+k-j) \quad (2.2)$$

$$\hat{r}(t) = \sum_{j=1}^{k-1} \beta_j r(t+j) \quad (2.3)$$

Aquí, $\hat{r}(t+k)$ es la regresión lineal de $r(t+k)$ en $r(t+k-1), r(t+k-2), \dots, r(t+1)$ y similarmente $\hat{r}(t)$ es la regresión de $r(t)$ en $r(t+1), r(t+2), \dots, r(t+h-1)$. Definida de esta manera, la PACF es una medida de la correlación entre $r(t)$ y $r(t+k)$ con cualquier otra correlación lineal intermedia eliminada. Mediante el uso conjugado de la función de autocorrelación (ACF) y la PACF es posible identificar si una serie de tiempo determinada se puede modelar como un proceso autorregresivo (AR), un proceso de promedio móvil (MA), o una combinación de ambos (ARMA).

2.2.2. Técnicas de descubrimiento causal empírico

Se le llama descubrimiento causal empírico al estudio de relaciones entre conjuntos de datos que, en un sentido muy amplio, implican relaciones más fuertes que asociaciones estadísticas y pueden permitir la predicción de los efectos que perturban al sistema estudiado, lo cual no es posible conociendo solo asociaciones estadísticas.

En el estudio de este tipo de relaciones, existe siempre un modelo estadístico o dinámico en el que los conjuntos de datos analizados se representan como variables con una función de respuesta entre ellas. En la mayoría de las técnicas esta relación de respuesta se asume unidireccional.

Las definiciones matemáticamente precisas de causalidad son muy variadas, lo que dificulta la labor de describir de manera general las propiedades de las pruebas y sus interpretaciones. En las siguientes secciones se describen a detalle las definiciones de causalidad y algoritmos de medición para las tres técnicas de descubrimiento causal que utilizamos en este trabajo.

Causalidad de Granger

Se dice que una señal (serie de tiempo) $X(t)$ *Granger causa* una señal $Y(t)$ si valores pasados de $X(t)$ contienen información que permite predecir valores futuros de $Y(t)$ [19]. La formulación matemática de la causalidad de Granger se basa en proponer dos modelos autorregresivos distintos para $Y(t)$. En uno de esos modelos solo se incluyen observaciones pasadas de Y , es decir, se tiene que para algún retraso $k \in \mathbb{N}$:

$$Y(t) = f(Y(t-1), Y(t-2), \dots, Y(t-k))$$

Mientras que en el otro modelo, también se incluyen observaciones pasadas de X , es decir:

$$Y(t) = f(Y(t-1), Y(t-2), \dots, Y(t-k), X(t-1), \dots, X(t-k))$$

Finalmente, se comparan los poderes predictivos de ambos modelos, y, dependiendo de la calidad de las predicciones de estos, se concluye si X causa a Y o no.

Antes de definir los modelos autorregresivos en los que se basa la prueba de Granger, explicaremos cómo se realiza una evaluación de modelos lineales a través de la prueba F de Fisher.

Supóngase que se tienen dos conjuntos de observaciones de observables arbitrarios X y Y , de manera que se desconoce si existe alguna relación entre estas dos variables.

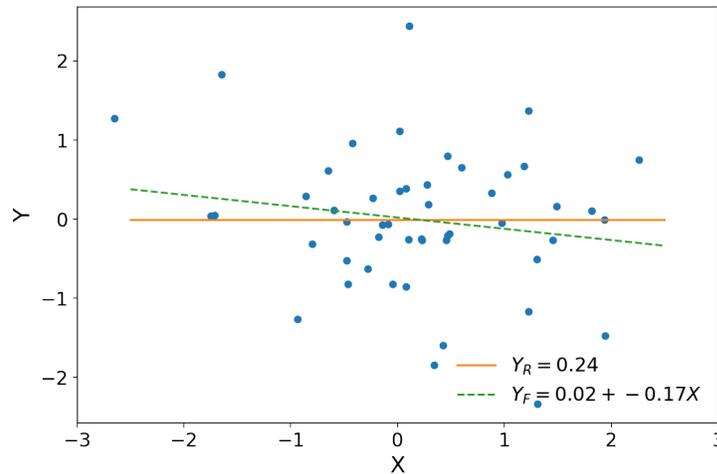


Figura 2.9: Dos modelos de regresión distintos para explicar la relación entre las variables aleatorias X y Y . Un modelo asume una relación nula (línea naranja) y otro asume una dependencia lineal (línea punteada verde). La diferencia entre la capacidad descriptiva de ambos modelos es pequeña, de manera que parece razonable seleccionar al modelo más sencillo.

Para determinar si una relación lineal es buena descripción, se proponen los siguientes modelos:

$$Y_R = \beta_0 + \epsilon_R$$

$$Y_C = \beta_0 + \beta_1 X + \epsilon_C$$

Donde a Y_R se le denomina modelo restringido por el hecho de no incluir a X en la regresión, y a Y_C se le llama modelo completo, puesto que incluye a X . Los términos ϵ_R y ϵ_C son los residuales de la regresión. Estos términos aditivos representan errores en los datos que no se pueden observar directamente y no se les impone una distribución de probabilidad particular.

Un ejemplo de lo que se propone en este escenario se muestra en la figura 2.9, en el que X y Y aparentan no tener relación alguna entre ellas. Como se observa en esa figura, el modelo restringido, que no asume una dependencia entre Y y X , y el modelo completo, que sí la asume, ofrecen una descripción muy parecida de la relación entre estas variables, lo que es de esperarse dada la aparente independencia entre ellas.

Considérese ahora el escenario mostrado en la figura 2.10. Al realizar el ajuste de los parámetros de los modelos restringido y completo, se obtiene el resultado mostrado en la figura 2.10.

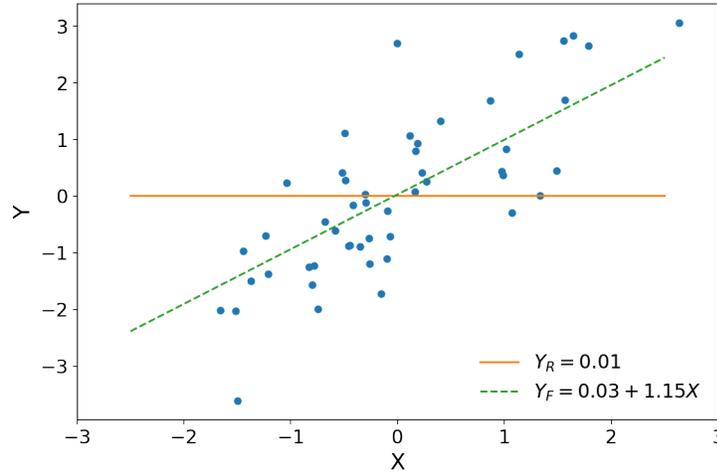


Figura 2.10: Misma situación que en la figura 2.9, pero ahora el modelo que asume una relación lineal hace un mucho mejor trabajo describiendo la dependencia entre Y y X .

Es ahora evidente que el modelo completo ofrece una mejor descripción de la relación entre X y Y , en relación con la descripción que ofrece el modelo restringido.

Una consecuencia de tener grandes diferencias en los ajustes de los datos ofrecidos por ambos modelos es que las sumas de residuales cuadrados de los dos modelos deben ser también muy diferentes entre sí (y viceversa), lo cual se confirma al comparar las sumas de residuales (SSR , por sus siglas en inglés) para el ejemplo con variables independientes, con residuales que resultan muy similares, tomando valores de $SSR_R = 38.7$ y $SSR_F = 37.16$. Mientras que para el caso en el que las variables sí están relacionadas, se tiene que $SSR_R = 104.7$ y $SSR_F = 40.16$.

Esta diferencia de tamaños entre los errores de ambos modelos es la base de la prueba F de Fisher, la cual permite descartar o no al modelo restringido como el más adecuado para modelar la relación entre X y Y .

La comparación efectuada en la prueba F depende del estadístico F^* definido como:

$$F^* = \frac{SSR_R - SSR_C}{d_{f_R} - d_{f_C}} \bigg/ \frac{SSR_C}{d_{f_C}} \quad (2.4)$$

Donde d_{f_R} y d_{f_C} son los grados de libertad de los modelos restringido y completo, respectivamente. Se puede ver que esta definición tiene sentido de manera intuitiva al ser F^* una función de la diferencia entre los errores de ambos modelos.

Ahora, la suma de residuales cuadrados tiene la misma forma que una varianza, y el estadístico F^* es el cociente de dos sumas de residuales, de manera que se puede efectuar una prueba de hipótesis de cociente de varianzas, en la que se supone que los residuales (varianzas) de ambos modelos presentan diferencias solamente debidas a los grados de libertad de los modelos y no a una genuina relación entre las variables.

Bajo la hipótesis nula mencionada, se tiene que la distribución del estadístico F^* es una distribución F, con grados de libertad d_{f_R} y d_{f_C} [20], de manera que dado un valor crítico p , se puede estimar la probabilidad de haber observado una diferencia entre los residuales tan grande como la observada asumiendo que no hay diferencia genuina subyacente. Si la probabilidad está por debajo del valor crítico, se rechaza al modelo restringido; si es mayor, se concluye que no es posible rechazar al modelo nulo como potencial explicación de la relación entre X y Y .

Una vez descrita la metodología para realizar una prueba F, podemos explicar cómo se conduce la prueba de causalidad de Granger.

Supóngase que se tienen dos series de tiempo $X(t)$ y $Y(t)$ con la sospecha de que $Y(t)$ presenta una respuesta a valores pasados de $X(t)$, como la situación que se presenta en la figura 2.11.

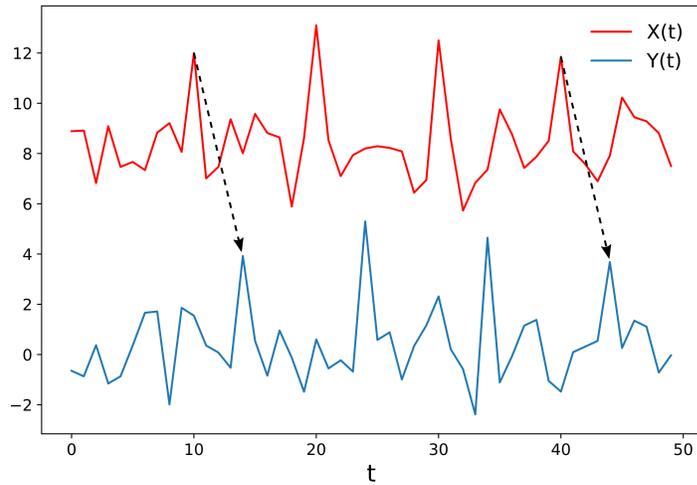


Figura 2.11: En este ejemplo se muestran dos series de tiempo que aparentan tener una relación causal según la definición de causalidad de Granger. Los valores de la serie $Y(t)$ parecen ser dependientes de los valores pasados de $X(t)$; la prueba de causalidad de Granger permite estimar si, en efecto, la información de valores pasados de $X(t)$ permite mejorar predicciones de los valores futuros de $Y(t)$ con respecto a las predicciones que se obtengan con un modelo simple AR.

Para realizar una prueba de causalidad de Granger se hace el ajuste de dos modelos AR, uno reducido y uno completo. El modelo reducido únicamente es autorregresivo (sin valores de $X(t)$ como regresores), y el modelo completo, autorregresivo y con regresiones cruzadas sobre $X(t)$.

Se tiene entonces al siguiente par de modelos AR de orden k , donde k es el número de retrasos temporales:

$$Y_R^k(t) = \beta_0 + \beta_1 Y(t-1) + \dots + \beta_k Y(t-k) + \epsilon(t)$$

$$Y_C^k = \beta_0 + \beta_1 Y(t-1) + \dots + \beta_k Y(t-k) + \gamma_1 X(t-1) + \dots + \gamma_k X(t-k) + \epsilon(t)$$

Ya que ambos modelos son lineales, se puede conducir una prueba F para determinar qué tan buena es la mejora en el modelado de Y que se obtiene al usar el modelo completo en relación con el modelo restringido. Si la prueba de hipótesis rechaza al modelo restringido, se concluye que $X(t)$ Granger causa a $Y(t)$.

Como para la realización de una prueba F una de las características de los datos que se debe cumplir es que la distribución de los residuales de los modelos sea gaussiana y en los datos estudiados en este trabajo esa propiedad no se cumple, realizamos las pruebas utilizando la distribución empírica de estadísticos F^* de versiones aleatorizadas de los datos. Fuera de esta diferencia, la metodología que empleamos es la misma que en una prueba estándar de causalidad de Granger.

Modelos de ruido aditivo

El descubrimiento causal con modelos de ruido aditivo (ANM, por sus siglas en inglés) tiene una formulación parecida a la que propone la causalidad de Granger en el sentido de que las variables causal y explicada se modelan explícitamente a través de una relación funcional entre ellas con la adición de un término de ruido. Las diferencias más importantes son que en la definición de causalidad por ANM no se considera ningún orden cronológico entre las variables involucradas (debe ser introducido por vías alternas) y que permite relaciones funcionales no lineales.

La idea central detrás de este método de descubrimiento causal es suponer que en los datos estudiados, los cuales se modelan como dos variables aleatorias X y Y , exista una relación de causalidad unidireccional en la que X afecte directamente a Y sin que exista al mismo tiempo la relación inversa (que Y afecte a X) y asumiendo que cualquier otro efecto causal que pudiera provenir de variables causales distintas a X se puede representar como ruido aditivo estadísticamente independiente de X . En otras palabras, en el descubrimiento causal por ANM se asume que X afecta directamente a Y y que cualquier otro efecto nacido de otras potenciales variables causales solamente se suma al efecto generado por X , sin que afecten a X . Para esto, se hace una regresión no lineal de Y como función de X y se mide la correlación entre los residuales de la regresión y X ; si no se detecta correlación entonces se concluye que, en efecto, Y es función de X más ruido (los residuales) independiente de X ; si no, se descarta que un modelo de ruido aditivo satisface la relación entre X y Y

Entonces, en un ANM, se propone que [21]:

$$\begin{cases} Y = f_Y(X, U_1, \dots, U_m) \\ X \perp \mathbf{U}, X \sim p_X(x), \mathbf{U} \sim p_{\mathbf{U}}(u_1, \dots, u_m) \end{cases} \quad (2.5)$$

donde \mathbf{U} representa el efecto de potenciales variables con efectos causales sobre Y , además del efecto impartido por X ; $X \perp \mathbf{U}$ es notación para indicar que X y \mathbf{U} son variables aleatorias independientes, $X \sim p_X(x)$ es notación para indicar que X está distribuido de acuerdo a la función de densidad $p_X(x)$; $f : \mathbb{R} \times \mathbb{R}^m \rightarrow \mathbb{R}$ es una función posiblemente no lineal, y $p_{\mathbf{U}}$ es la densidad de \mathbf{U} . En este modelo, es muy importante la hipótesis de independencia entre X y \mathbf{U} , puesto que implica la inexistencia de confusión de causas, sesgos de selección y retroalimentaciones entre X y Y , las cuales son suposiciones que simplifican enormemente la implementación del descubrimiento causal por ANM.

Como las causas latentes no se conocen en realidad, su efecto se puede representar como la influencia de una sola variable efectiva de ruido $E_Y \in \mathbb{R}$, el cual no es un residual de regresión, como es el caso con el término aditivo en las regresiones de la causalidad de Granger (la ϵ de los modelos autoregresivos de Granger); por lo tanto, sí se imponen restricciones a su distribución de probabilidad. De esta manera, el modelo descrito por las relaciones (2.6) se puede sustituir por el siguiente modelo, en el que se incluye la suposición de que el efecto del término de ruido es aditivo:

$$\begin{cases} Y = f_Y(X) + E_Y \\ X \perp E_Y, X \sim p_X, E_Y \sim p_{E_Y} \end{cases} \quad (2.6)$$

Si una densidad conjunta $p(x, y)$ es consistente con el ANM $X \rightarrow Y$ pero no al ANM $Y \rightarrow X$, se dice que el modelo $X \rightarrow Y$ es identificable.

Para poder mostrar si una densidad conjunta de probabilidad obtenida empíricamente satisface un modelo causal como los propuestos, es necesario el siguiente teorema [21]: *sea $p(x, y)$ la densidad conjunta de las variables aleatorias X, Y tal que la esperanza condicional $\mathbb{E}(Y|X = x)$ esté bien definida para toda x . Entonces, $p(x, y)$ satisface un modelo de ruido aditivo $X \rightarrow Y$ si y solo si $E_Y := Y - \mathbb{E}(Y|X = x)$ tiene media finita y es independiente de X .*

Desde luego, uno nunca conoce la densidad conjunta $p(x, y)$ en la práctica, sino que la estima a través de un conjunto finito de observaciones de X y Y ; de la misma manera, uno no sabe si E_Y es independiente de X , sino que se determina por medio de algún estimador de dependencia estadística, preferentemente uno que pueda detectar asociaciones no lineales, como la correlación de distancias o el criterio de independencia de Hilbert-Schmidt (HSIC).

El algoritmo general del procedimiento que decide si se satisface un ANM, propuesto por Hoyer et al. [22], se describe a continuación.

Primero se dividen los datos en dos conjuntos, denominados *conjunto de entrenamiento* $D_N := \{(x_n, y_n)\}_{n=1}^N$ y *conjunto de pruebas* $D'_N := \{(x'_n, y'_n)\}_{n=1}^N$. El conjunto de entrenamiento se utilizará para estimar la esperanza condicional $\mathbb{E}(Y|X)$, y el conjunto de pruebas, para medir la dependencia estadística entre X y E_Y . Los conjuntos de observaciones originales de las variables X y Y ya separadas se representarán como $\mathbf{x} = (x_1, \dots, x_N)$, $\mathbf{y} = (y_1, \dots, y_N)$, $\mathbf{x}' = (x'_1, \dots, x'_N)$ y $\mathbf{y}' = (y'_1, \dots, y'_N)$. Después, se obtienen estimaciones de \hat{f}_Y que es la regresión de $x \mapsto \mathbb{E}(Y|X = x)$ y \hat{f}_X para $y \mapsto \mathbb{E}(X|Y = y)$ mediante algún método de regresión, por ejemplo mediante regresión por procesos gaussianos. Una vez que se tienen las funciones de regresión, se procede a calcular los residuales: $\hat{\mathbf{e}}'_Y := \mathbf{y}' - \hat{f}_Y(\mathbf{x}')$ para la regresión \hat{f}_Y y $\hat{\mathbf{e}}'_X := \mathbf{x}' - \hat{f}_X(\mathbf{y}')$ para \hat{f}_X . Ahora se pueden calcular los puntajes de causalidad en ambas direcciones ($X \rightarrow Y$ y $Y \rightarrow X$) utilizando el estimador de dependencia estadística elegido. Los puntajes se definen como:

$$\hat{C}_{X \rightarrow Y} := \hat{C}(\mathbf{x}', \hat{\mathbf{e}}'_Y)$$

$$\hat{C}_{Y \rightarrow X} := \hat{C}(\mathbf{y}', \hat{\mathbf{e}}'_X)$$

La dirección de la relación causal está finalmente dada por

$$X \rightarrow Y \text{ si } \hat{C}_{X \rightarrow Y} < \hat{C}_{Y \rightarrow X},$$

$$Y \rightarrow X \text{ si } \hat{C}_{X \rightarrow Y} > \hat{C}_{Y \rightarrow X},$$

Si las puntuaciones de causalidad son iguales, la dirección se define como indeterminada.

Se puede considerar como tamaño de efecto al tamaño de la puntuación de causalidad, de manera que además de tener la dirección de la relación se tiene una idea de la fuerza de la relación.

Para la implementación de estas pruebas de causalidad en este trabajo utilizamos la paquetería de software Causal Discovery Toolbox [23].

Mapeos cruzados convergentes

El descubrimiento causal por mapeos cruzados convergentes [24], [25] (CCM, por sus siglas en inglés: *convergent cross mapping*) es una técnica que puede distinguir causalidad a partir de mediciones de sistemas dinámicos. Las técnicas de esta categoría se basan en la idea de que una relación causal se puede establecer si los estados dinámicos de la variable causal pueden recuperarse de la serie de tiempo de la variable afectada. Esto la distingue de los otros dos métodos de descubrimiento causal que basan sus definiciones de causalidad en la predicción puramente estadística de la variable afectada, asumiendo que tanto la variable causal como la causada son aleatorias. En este método, se establece que dos observables (series de tiempo) X y Y están causalmente relacionados si pertenecen al mismo sistema dinámico y comparten al mismo atractor. Si X y Y pertenecen a un mismo sistema dinámico, se puede identificar cada estado de X con un estado correspondiente en Y , ya que ambas variables se asocian a diferentes dimensiones en el espacio fase al que pertenece su atractor, como se muestra en la figura 2.12.

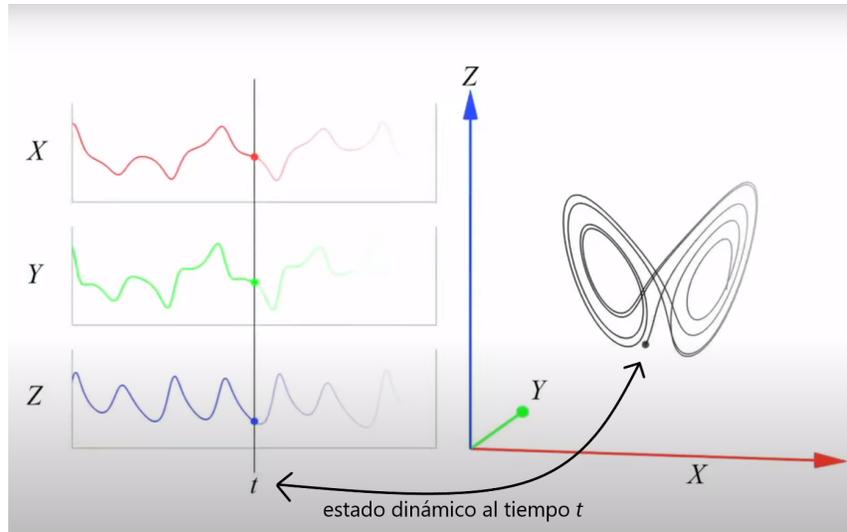


Figura 2.12: Aquí se muestra que cuando dos observables X y Y pertenecen al mismo sistema dinámico, a cada estado de X se le asocia un estado correspondiente de Y , y comparten además al mismo atractor. A la izquierda se presentan tres series de tiempo X , Y y Z , correspondientes a estimaciones del sistema de Lorenz. A la derecha se muestra el atractor del sistema de Lorenz y se indica con el punto gris el elemento en el espacio fase correspondiente a los estados de las tres series al tiempo t . Es evidente que la coordenada X del elemento de espacio fase señalado a la derecha corresponde unívocamente a un valor de la coordenada y , los valores de ambas coordenadas de las series X y Y son señaladas a la izquierda por la línea recta vertical que cruza las series al tiempo t . Imagen tomada y modificada de [25].

Entonces, si se logra establecer que las series X y Y comparten un atractor, podemos concluir que hay una relación causal entre ellas, según esta definición de causalidad dinámica.

Para determinar si las series bajo estudio comparten un mismo atractor, se reconstruye una versión de este a partir de la información contenida en X y otra versión a partir de la información contenida en Y . Si se puede hacer una buena estimación de la reconstrucción del atractor basada en Y usando el atractor reconstruido con X , significa que existe una asociación entre los estados de X y Y , lo que implicaría pertenencia al mismo sistema dinámico. El hecho de que se computa una estimación del reconstruido a partir de Y usando el atractor generado a partir de X da la parte de *mapeo cruzado* al nombre de esta técnica. Y el hecho de que el mapeo cruzado se repite con segmentos de las series X y Y cada vez más largos para medir si se mejora la calidad de la estimación del atractor

basado en Y conforme se toman mayores segmentos, le confiere la parte de *convergente*.

En la figura 2.13 se muestra de manera esquemática la realización de los mapeos cruzados.

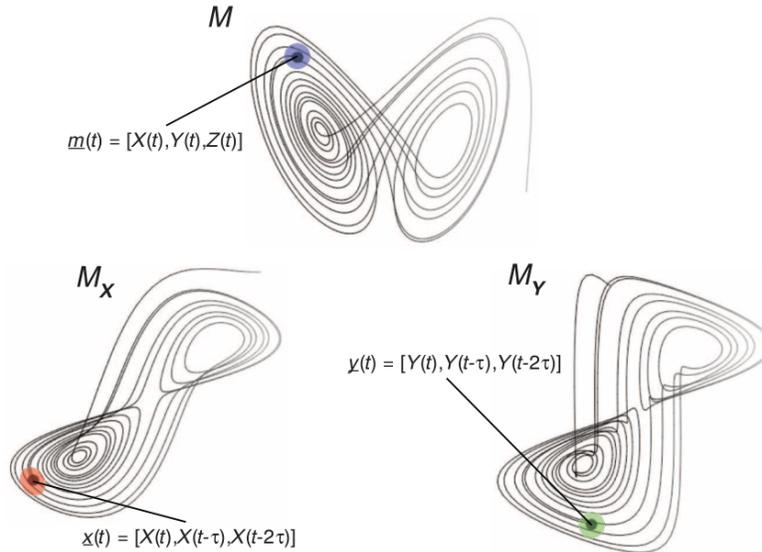


Figura 2.13: Esquema que muestra el funcionamiento de la técnica de mapeos cruzados convergentes basado en el sistema de Lorenz. A la izquierda se observa la sombra del atractor M que se obtiene a través de las observaciones de la variable X ; a la derecha, la sombra obtenida a partir de la variable Y . Como X y Y están acopladas en las ecuaciones dinámicas del sistema de Lorenz, los puntos dentro de una vecindad en M_X corresponderán a puntos dentro de otra vecindad en M_Y . Esta correspondencia de vecindades entre las distintas sombras del atractor permite hacer estimaciones de los valores que Y tomará utilizando valores de X . Conforme se utilizan series de tiempo más largas, las sombras del atractor se volverán más densas, y permitirán predicciones más precisas. Es de la predicción cruzada de las variables y la mejora en las estimaciones que viene con el incremento del tamaño de las series que la técnica toma el nombre de mapeo cruzado convergente. Imagen tomada de [25].

CCM se fundamenta en un teorema demostrado por Takens [26] extendido para soportar estocasticidad por Stark et al. [27], que establece las condiciones bajo las que se puede reconstruir el atractor de un sistema dinámico a partir de una secuencia de observaciones de dicho sistema. Como consecuencia de estos teoremas se tiene de manera general que la información esencial de un sistema dinámico multidimensional está contenida en la serie de tiempo de una sola de sus variables, de manera que, definiendo causalidad como la relación que existe entre dos variables que pertenecen a un mismo sistema, CCM utiliza la idea de

Takens para detectar pertenencia mutua. Este método permite detectar relaciones unidireccionales o bidireccionales, y es capaz de cuantificar la magnitud de las dos direcciones causales si es este el caso, de esta forma, si se tiene un bucle de retroalimentación causal, CCM es capaz de determinar cuál de las dos variables afecta más a la otra.

La implementación de una prueba por CCM consta de manera general de cinco pasos.

Primero, se crea la representación del atractor, a la cual se le suele llamar sombra del atractor, utilizando para esto la serie de tiempo $X(t)$, la cual es considerada considerada como potencial variable causal de otra variable $Y(t)$. La sombra del atractor se crea asociando un vector E -dimensional para cada punto en $X(t)$, construido con observaciones retrasadas entre sí por τ unidades de tiempo, de manera que el vector asociado al tiempo t es $\hat{\mathbf{X}}_t = (X_t, X_{t-\tau}, X_{t-2\tau}, \dots, X_{t-(E-1)\tau})$, asociando así el primer vector al tiempo $t = 1 + (E - 1)\tau$ y el último vector al tiempo L , donde L es el número de observaciones de la serie de tiempo $X(t)$. La dimensión E se puede elegir de acuerdo a dos criterios: se elige el valor de E que maximiza la capacidad predictiva de $X(t)$ sobre $Y(t)$ o se efectúa la prueba de vecinos cercanos falsos, que determina el valor de E a partir del cual no se detecta incremento en las distancias entre vectores vecinos conforme se sigue incrementando a E . Este conjunto de vectores es la reconstrucción del atractor.

En seguida, se encuentran los $E + 1$ vecinos más cercanos del vector $\hat{\mathbf{X}}_t$, y sus distancias al vector se ordenan por cercanía, junto con sus índices temporales correspondientes: $(d_1, d_2, \dots, d_{E+1})$ y $(t_{d_1}, t_{d_2}, \dots, t_{d_{E+1}})$, siendo las distancias a $\hat{\mathbf{X}}_t$ dadas por $d_i = D(\hat{\mathbf{X}}_t, \hat{\mathbf{X}}_{t_{d_i}})$, donde $D(\mathbf{a}, \mathbf{b})$ es la distancia euclidiana entre los vectores \mathbf{a} y \mathbf{b} .

El siguiente paso es calcular un conjunto de pesos asociados a cada uno de los vecinos cercanos. Los pesos se definen como

$$w_i = \frac{u_i}{N}$$

donde $u_i = e^{-d_i/d_1}$ y el factor de normalización N está dado por

$$N = \sum_{i=1}^{E+1} u_i$$

Con los pesos calculados, se procede a estimar cada una de las observaciones de $Y(t)$ a partir de los vecinos cercanos de Y y los pesos obtenidos de X . Cada punto Y_t en Y se

estima como

$$Y_t|\hat{\mathbf{X}} = \sum_{i=1}^{E+1} w_i Y_{t_{d_i}}$$

Finalmente, se calcula la puntuación de relación causal:

$$C_{YX} = \left(\rho(Y, Y|\hat{\mathbf{X}}) \right)^2$$

donde $\rho(A, B)$ es el coeficiente de correlación de Pearson entre A y B .

Este procedimiento se inicia con una submuestra de las series X y Y a las que se denomina librerías, y se repite con librerías cada vez más grandes, hasta eventualmente cubrir la longitud total de las series. Si las puntuaciones de relación causal convergen a valores elevados conforme uno considera librerías mayores, se estima que se ha detectado una relación causal. La convergencia es importante para concluir la presencia de causalidad porque conforme se utilizan librerías mayores, la densidad de las trayectorias en el espacio fase reconstruido es mayor, lo que permite mejores estimaciones de la variable afectada, situación que no ocurre cuando se miden solamente asociaciones estadísticas. El nivel final al que converge C_{YX} se puede considerar como la magnitud de la relación causal entre las variables.

Con respecto a las pruebas de significancia estadística de la puntuación de causalidad, usualmente se utilizan series de tiempo de datos sustitutos, ya sea con copias permutadas de los datos originales o con la técnica de IAFFT [28]. El primer método representa a la hipótesis nula más agresiva, puesto que al permutar los datos originales se destruye no solo la información causal entre las variables, sino también las autocorrelaciones y las correlaciones cruzadas entre ellas, y se conserva solamente la misma densidad de probabilidad. Generando datos sustitutos por IAFFT se mantiene la distribución además de las autocorrelaciones y correlaciones cruzadas lineales. En este trabajo implementamos las pruebas por CCM con la paquetería de software libre Scikit Convergent Cross Mapping, que se encuentra en el repositorio GitHub <https://github.com/nickc1/skccm>.

Capítulo 3

Resultados

3.1. Estadística de gaps

3.1.1. Preprocesamiento de los datos

Todos los análisis se realizaron usando un conjunto de muestreos del estado del libro de ordenes del mercado bitcoin/dólar en la plataforma de bitcoin BTCe, efectuados cada 10 segundos. Los muestreos del libro contienen precios y volúmenes agregados de los primeros 20 niveles para las órdenes de venta y compra; es decir, contiene 20 niveles de precios de venta y 20 de compra, lo cual arroja un total de 40 niveles registrados.

Puesto que los muestreos están ordenados temporalmente, la información de cada nivel se puede considerar como una serie de tiempo individual. Al representar cada nivel de esta manera, se generan dos series de tiempo multivariadas, una asociada a los precios y otra a los volúmenes agregados. Esta representación del libro facilita el empleo de herramientas para el análisis de series de tiempo, tanto para propiedades independientes como dependientes del orden temporal en que se registraron los estados del libro de órdenes.

En la figura 3.1 se muestra una representación gráfica de un libro de órdenes como serie de tiempo multivariada. En esta, cada línea corresponde a un nivel de precios (una serie de tiempo). Se señalan con sombreado las diferencias entre dos niveles contiguos de precio en ambos lados del libro (el primer gap de las ventas y el tercero de las compras), es decir, los gaps. Para ser consistentes con la literatura que mencionamos anteriormente,

utilizaremos los gaps logarítmicos, los cuales se definen como

$$g_i = \ln \left(\frac{p_i}{p_{i-1}} \right)$$

donde g_i es el i -ésimo gap logarítmico, y p_i es el i -ésimo precio de venta (compra). Toda mención de los gaps en el resto de este trabajo se referirá a gaps logarítmicos.

El proceso por el cual se obtienen las poblaciones de percentiles se ilustra en la figura 3.2.

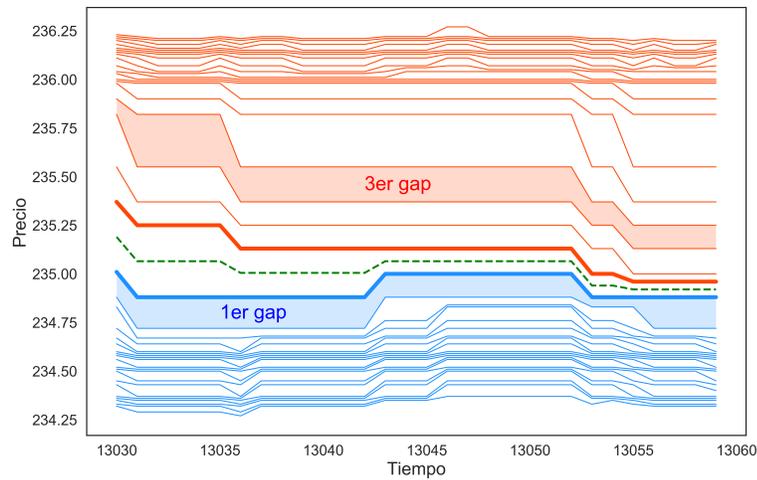


Figura 3.1: Representación de un libro de órdenes como una serie de tiempo multivariada. En esta representación gráfica, las series de tiempo de color rojo corresponden a los precios de las órdenes de venta, mientras que las azules se refieren a los precios de las órdenes de compra. El tercer gap de las órdenes de venta y el primero de las órdenes de compra se señalan con sombreados rojo y azul, respectivamente. De manera similar, las líneas gruesas corresponden a los mejores precios de venta (roja) y de compra (azul). Finalmente, la línea punteada de color verde señala el precio medio.

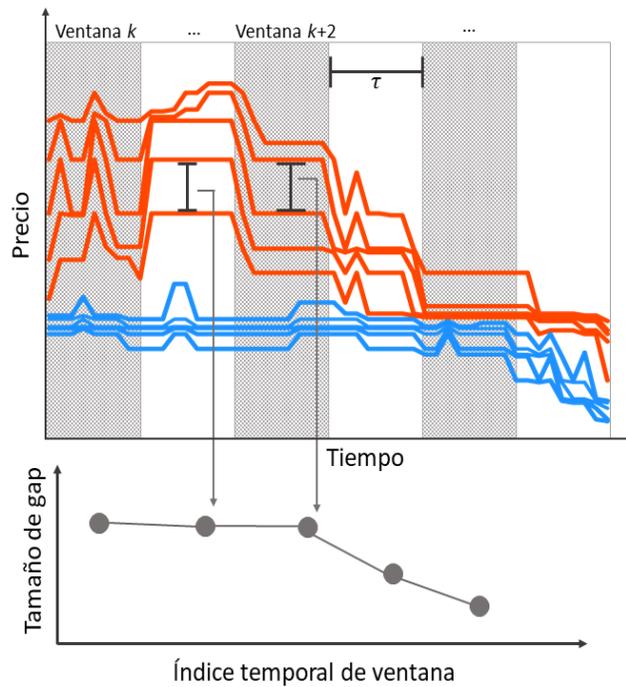


Figura 3.2: Diagrama del proceso para la generación de las series de percentiles de gaps. Los datos correspondientes a un mes de libro de órdenes se dividen en ventanas disjuntas de tamaño τ , señaladas en la figura por las barras alternantes de color gris y blanco. Después, de cada una de estas ventanas, se extrae el gap correspondiente al p-ésimo percentil. En la figura se muestra el ejemplo de la extracción de los máximos gaps (percentil 100) de cada ventana, los cuales se marcan con las barras negras verticales. Asignamos la etiqueta temporal de la orilla izquierda de cada una de las ventanas (k , en el diagrama) a los percentiles extraídos de ellas, formando así la serie de tiempo de los percentiles de gaps que se representa al fondo de la imagen. Entonces, en el ejemplo de la figura, uno termina con la serie de tiempo de los máximos gaps de cada ventana, mostrada en el fondo con los círculos grises. Dicho en otras palabras, se termina con la serie de tiempo del percentil 100 de los gaps.

3.1.2. Distribución de tamaños

En esta sección se presentarán las distribuciones de probabilidad de los gaps logarítmicos del libro de órdenes, con el i -ésimo gap logarítmico definido como $g_i = \ln(p_i/p_{i-1})$, donde p_i es el i -ésimo nivel del libro. El comportamiento de las distribuciones de gaps seccionados por sus percentiles presenta peculiaridades que más tarde servirán para explicar diversas propiedades de las funciones de correlación entre los gaps y los rendimientos.

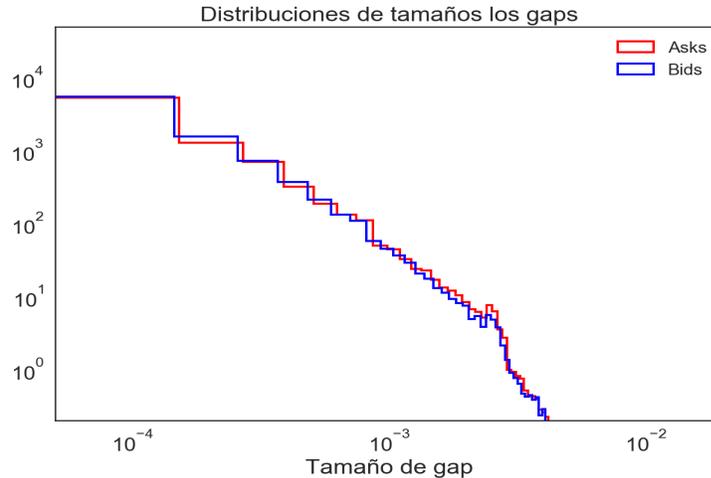


Figura 3.3: Distribución de los tamaños de gap en ambos lados del libro para enero del 2015. En esta distribución están considerados todos los gaps.

La distribución de tamaños de gap para el mes de mayo del 2015 se presenta en la figura 3.3. En esta se aprecia que ambas distribuciones son esencialmente idénticas; no existe asimetría entre los gaps de las ofertas de venta y compra, al menos en lo que respecta a la escala de tiempo con la que se capturaron los datos (un muestreo cada 10 segundos).

En la figura 3.4 se muestran las distribuciones de algunos percentiles del conjunto completo de gaps, los cuales consideran la población de 40 gaps que constituyen el conjunto de datos con el que contamos. Los percentiles se calcularon sobre ventanas de 30 pasos de tiempo, correspondientes a intervalos temporales de 5 min de duración. Elegimos estos tamaños de ventana por dos razones: son suficientemente pequeños para hacer buenas estimaciones de los valores locales y son suficientemente grandes como para permitir que existan diversas posibilidades de encontrar los percentiles que se calculen lejos de las fronteras de las ventanas, evitando así acoplamiento sistemático entre ventanas, lo cual será importante en el cálculo de las correlaciones y relaciones causales que se presentarán más adelante. Cuando utilizamos otros tamaños de ventanas, los resultados siguen siendo muy parecidos en sus propiedades cualitativas.

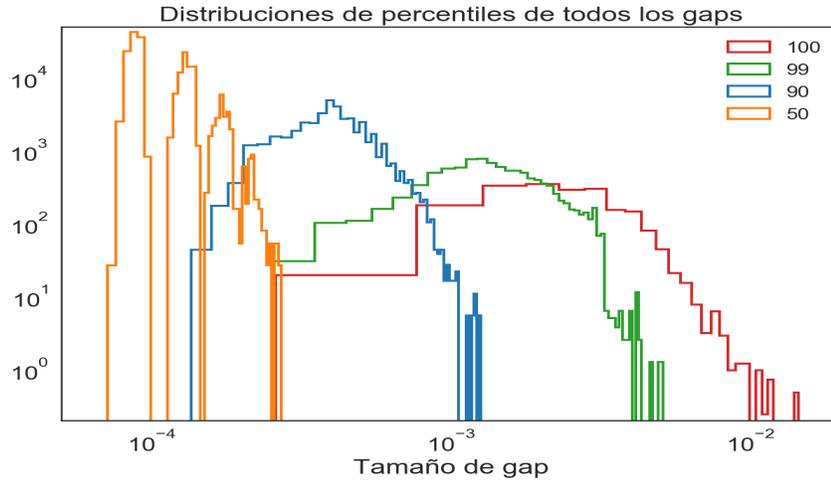


Figura 3.4: Distribución de los percentiles de tamaños de gaps en ambos lados del libro. Los percentiles 100 y 99 tienen distribuciones con colas derechas muy separadas, a pesar del cambio tan pequeño que implica pasar del percentil 100 al 99. La separación es aún mayor cuando se llega al percentil 90. Esto refleja una estratificación de tamaños de los gaps, con separaciones entre las distintas poblaciones de percentiles que no cambian de manera paulatina.

Se ve en esa figura que las distribuciones de los percentiles 100 (el máximo gap) y 99 están considerablemente separadas en las colas, a pesar de que son percentiles muy cercanos. Las separaciones se vuelven drásticas conforme uno se acerca al percentil 50 (la mediana).

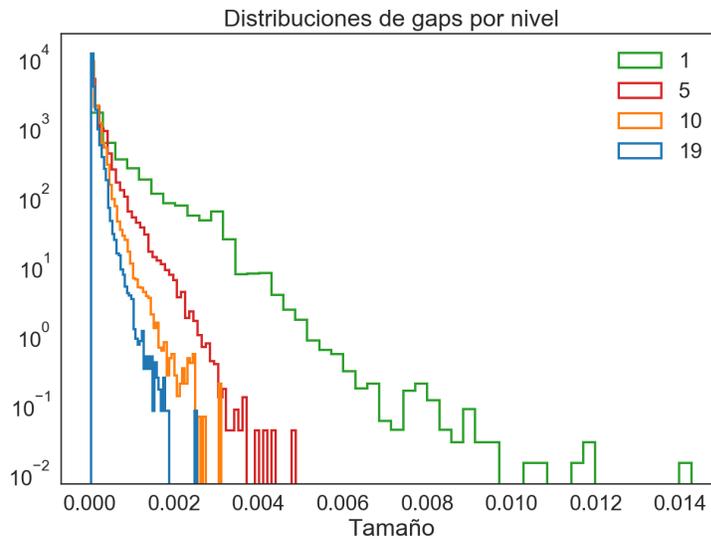


Figura 3.5: Distribución de los tamaños de gap por nivel. Los máximos tamaños posibles por nivel decrecen conforme se avanza en la profundidad del libro.

Además de las fuertes separaciones entre los percentiles, los gaps más grandes se suelen acumular casi exclusivamente en las primeras posiciones, como lo indica la figura 3.6, en la que se muestra la distribución de posiciones del máximo gap, de la cual se extrae que la probabilidad de encontrar al máximo gap en una de las primeras tres posiciones es de 0.76, y decrece rápidamente conforme uno considera profundidades mayores en el libro. Esto genera que, en términos prácticos, el libro se comporte como si existieran dos escalas de gaps: las de las primeras posiciones, cercanas al precio medio, y las que están en las capas superiores. La figura 3.5 muestra las distribuciones de gaps por nivel. A partir del 5to nivel y hasta el último (19avo), los tamaños de gap decrecen como función de la profundidad el libro, pero lo hacen de manera paulatina, formando un grupo de escalas que están considerablemente separados de los primeros gaps, que forman otro grupo.

Esta aparente división de escalas en los gaps será importante para explicar algunas propiedades de las funciones de correlación de los gaps con los rendimientos que se presentarán en secciones ulteriores.

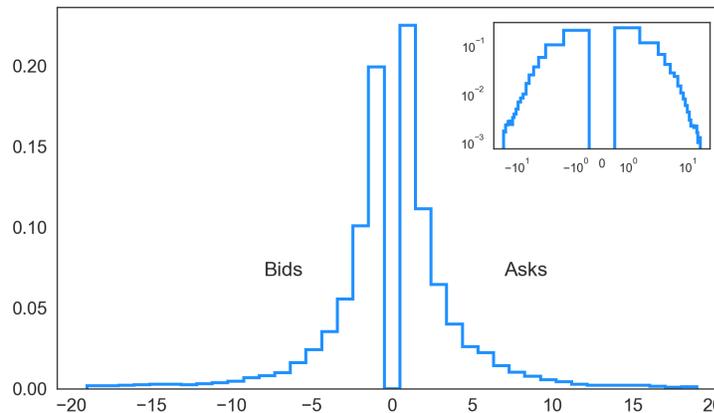


Figura 3.6: Distribución de probabilidad de las posiciones en las que se ubican los máximos gaps. La probabilidad de encontrar al máximo gap se concentra en las primeras posiciones. El inset muestra a la misma gráfica pero en escala doble logarítmica.

3.1.3. Estructura temporal de los gaps

Autocorrelaciones

De manera similar a lo que se observa en las funciones de autocorrelación $C(\tau)$ de los rendimientos absolutos, las autocorrelaciones de series derivadas de los gaps se mantienen

positivas por periodos largos de tiempo, con un decaimiento que se puede aproximar de la forma

$$C(\tau) = \text{corr}(g_t, g_{t+\tau}) \simeq \frac{c}{\tau^\alpha}$$

donde g_t es alguna serie derivada de los gaps. A este tipo de decaimiento se le denomina *dependencia de largo rango* [29], en contraste con el decaimiento más rápido que presenta una función para la cual se cumple que

$$\exists K > 0, c \in (0, 1), \quad |C(\tau)| \leq Kc^\tau$$

comportamiento denominado como *dependencia de corto rango*.

La dependencia de largo rango está fuertemente relacionada con la propiedad de aglomeración, cuya cuantificación más directa se presenta en la siguiente sección.

Los gaps directos (considerados cada paso de tiempo disponible, sin estar separados por ventanas) no presentan el fenómeno de dependencia de largo rango en las primeras posiciones, pero conforme uno ahonda más en el libro y considera gaps de mayores profundidades, el comportamiento en ley de potencias se vuelve una buena descripción de la autocorrelación, como se observa en la figura 3.7. Una posible explicación de la aparición gradual de dependencias de largo rango es el hecho de que la actividad de mercado consume usualmente solo los primeros niveles del libro de órdenes [18], mientras que los niveles profundos suelen persistir por tiempos prolongados sin cambios por la poca exposición que tienen a la actividad mercantil. Como consecuencia, los niveles más superficiales cambian con suficiente frecuencia por unidad de tiempo como para no alcanzar a desarrollar dependencias de largo rango, sino un comportamiento intermedio entre corto y largo. Así mismo, los niveles profundos generan gaps que persisten por largos periodos de tiempo sin experimentar modificaciones, de manera que producen las dependencias de largo rango observadas.

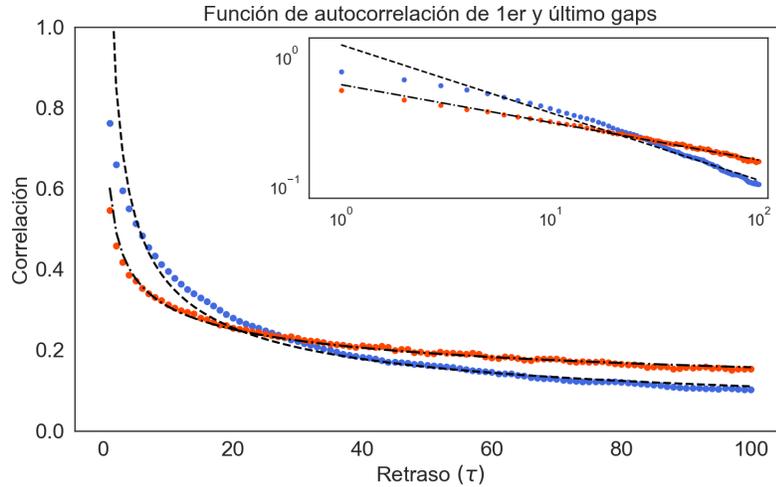


Figura 3.7: Funciones de autocorrelación del primer y último gaps. El primer gap (morado) no presenta dependencia de largo rango, siendo un ajuste en ley de potencias solamente adecuado para describir la última década de la función mientras que el último (rojo) es bien descrito de esta manera. Las dependencias de largo rango se observan de manera gradual conforme uno considera gaps de mayor profundidad; esta aparición gradual puede ser consecuencia de que la actividad de mercado consume casi exclusivamente a los primeros niveles de precios, de manera que las órdenes que se sitúan lejos del spread son más longevas que las órdenes que constituyen los primeros gaps. Como consecuencia, los gaps de la superficie del libro cambian con suficiente rapidez como para no alcanzar a desarrollar dependencias de largo rango, mientras que los gaps profundos, al tardar mucho en ser modificados por la actividad de mercado, dan lugar al efecto contrario, con autocorrelaciones positivas de larga duración.

Habiendo mostrado que los primeros gaps no presentan dependencias de largo rango y que los máximos gaps casi siempre se localizan en las primeras posiciones, se obtiene un resultado que parece contradecir estas dos observaciones cuando se considera la serie de tiempo de los máximos gaps sobre ventanas de tiempo disjuntas, pues las autocorrelaciones de esta serie presentan, con un muy buen acuerdo, dependencias de largo rango, como se puede observar en la figura 3.8. Esto a pesar de que la mayor parte de las observaciones de máximos gaps provienen de los niveles más cercanos al spread, en donde las autocorrelaciones de los gaps no muestran dependencias de largo rango, los máximos sí las presentan. El mismo comportamiento se observa en la serie de medianas de los gaps pero esto es consistente con el hecho de que la mediana usualmente se encuentra en posiciones profundas.

En el siguiente apartado se confirmará de manera directa que consistentemente con la

presencia de correlaciones de larga duración, las series de gaps presentan la propiedad de aglomeración.

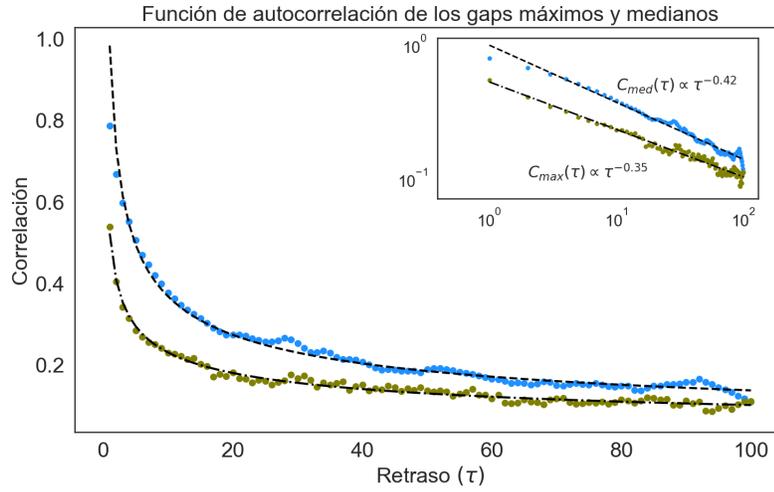


Figura 3.8: Funciones de autocorrelación de los máximos gaps $C_{max}(\tau)$ y los gaps medianos $C_{med}(\tau)$. Ambas funciones son bien descritas como leyes de potencias con exponentes dentro del rango asociado a la definición de dependencia de largo rango.

3.1.4. Correlaciones entre gaps, rendimientos y volatilidades

En esta sección se mostrarán propiedades estadísticas de los gaps que presentan correlaciones dependientes del tiempo con respecto a los rendimientos y volatilidades.

El algoritmo con el que fueron calculadas las funciones de correlación, a las cuales llamamos correlaciones locales, se describe en la figura 3.9.

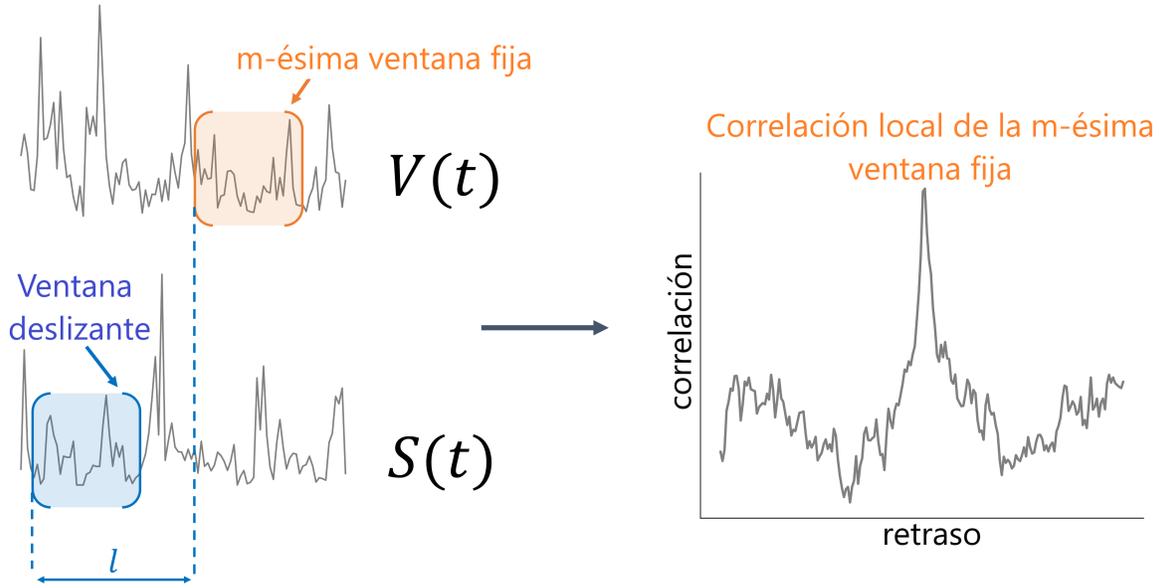


Figura 3.9: Cálculo de las funciones de correlación local. Una función de correlación entre la serie $V(t)$ y la serie $S(t)$ se obtiene al posicionar en $V(t)$ una ventana fija de tiempo de τ pasos de tiempo cuya orilla izquierda comienza en $t = T_m$, de manera que la ventana abarca el intervalo $[T_m, T_m + \tau]$ para después proceder a calcular las correlaciones de esa ventana fija con otra ventana móvil situada en $S(t)$. Conforme se incrementa el retraso que la ventana móvil tiene con respecto a la fija, se obtiene una función de correlación asociada al índice temporal del comienzo de la ventana fija en $V(t)$. Una vez que se tiene la función de correlación de la m -ésima ventana fija, se posiciona una nueva ventana fija que comienza donde la anterior terminó, de manera que las ventanas fijas no presenten intersección entre ellas. Finalmente, se calcula el promedio del ensemble de funciones de correlación local obtenidas de esta manera.

En la figura 3.10 se muestran los promedios de las funciones de correlación local entre los percentiles 100 (el máximo), 99 y 50 (la mediana) de los gaps y de los rendimientos. En dicha figura, los percentiles se obtuvieron sobre ventanas de 30 pasos de tiempo, y las correlaciones, sobre ventanas deslizantes de 100 pasos de tiempo. Las propiedades cualitativas de las funciones de correlación son prácticamente idénticas con distintos tamaños de ventana y a lo largo de todo el tiempo que engloban los datos.

La pareja máximos rendimientos-máximos gaps presenta la mayor correlación para cualquier retraso, y esta es particularmente grande con un retraso nulo. Con solo disminuir un poco el percentil considerado y tomar al percentil 99, se observa una abrupta disminución en el tamaño de las correlaciones. Esto puede parecer contraintuitivo si uno imagina que percentiles cercanos al mayor presentan dinámicas similares entre ellos; pero

como se mostró anteriormente, los máximos gaps suelen estar situados de manera predominante en la primera posición, mientras que los demás miembros de la población de gaps se suelen situar en las profundidades del libro, donde la dinámica responde a escalas de tiempo mayores. Esta diferencia en tiempos de respuesta a la actividad de mercado, que es representada de manera indirecta por los mismos rendimientos, puede provocar el desacople observado entre los percentiles distintos al máximo y los rendimientos, con respecto a la alta correlación que muestran los máximos.

El desacople se vuelve particularmente notorio cuando se consideran los percentiles 50 de ambas series, lo cual tiene la implicación de que los tamaños típicos de rendimiento no están asociados a los tamaños típicos de gap, asociación que podría parecer razonable a priori.

Cuando uno ignora los primeros δ niveles del libro en el cálculo de las funciones de correlación, se obtiene que las funciones decrecen, como es de esperar por la lejanía de gaps profundos con respecto al gap.

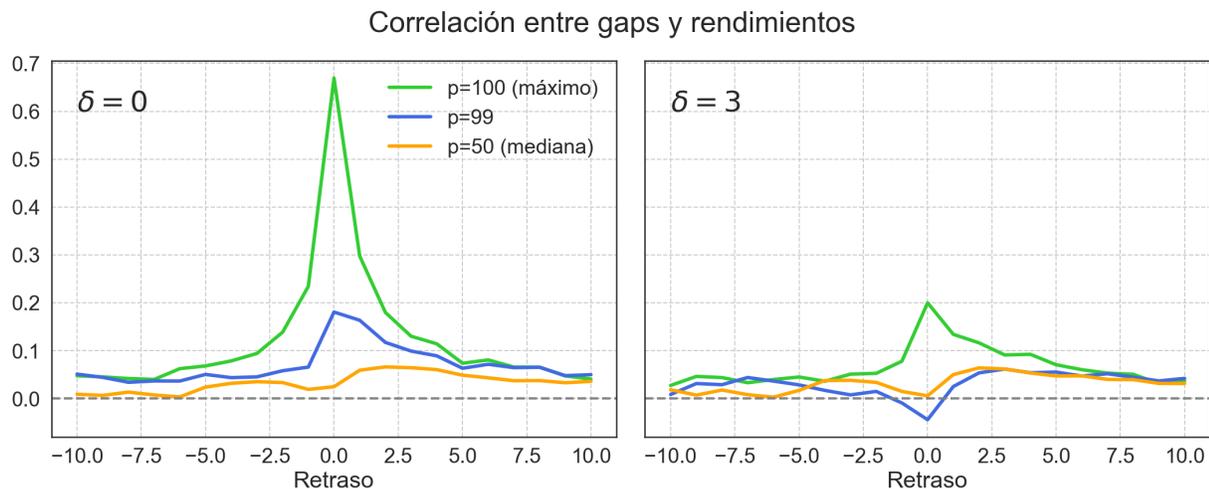


Figura 3.10: Promedio de las funciones de correlación entre distintos percentiles de gaps y rendimientos. A la izquierda las funciones se calcularon sin omitir gaps ($\delta = 0$), a la derecha se calcularon omitiendo los primeros tres gaps ($\delta = 3$). Se puede observar un comportamiento que recuerda a las distribuciones de gaps para ($\delta = 0$) puesto que al pasar del percentil 100 al 99, el pico de la función de correlación es mucho más bajo. El hecho de considerar ($\delta = 3$) tiene un impacto importante en las correlaciones; a pesar de ser un δ pequeño, lo que indica que la mayoría de las asociaciones entre estructura de gaps y precios existe en las primeras capas de gaps del libro, las profundidades del libro se acoplan pobremente a la dinámica de los precios.

Habiendo mostrado que los tamaños típicos de rendimiento no están asociados con los tamaños típicos de gap, en la figura 3.11 se muestra la función de correlación entre los máximos gaps y los rendimientos medianos. En este caso, se presentan mayores magnitudes de las correlaciones, lo que es consistente con el hecho de que los rendimientos consumen siempre el volumen de los primeros gaps y que estos suelen ser los máximos gaps de la población. Entonces no solo el máximo de los gaps está correlacionado con el máximo de los rendimientos; también el tamaño típico de los rendimientos presenta esa asociación.

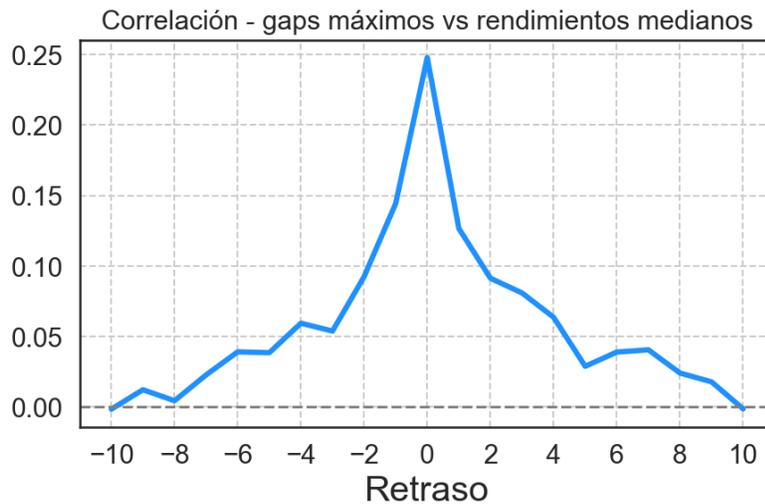


Figura 3.11: Promedio de las funciones de correlación entre los rendimientos medianos y los gaps máximos. El pico de la función alcanza un valor comparable con el de los percentiles 99 en la figura 3.10; de manera que los máximos gaps se correlacionan también con el tamaño típico de los rendimientos. Los máximos gaps parecen ser los que más asociación estadística muestran con la dinámica de los precios, tanto la típica como la extrema.

Finalmente, en la figura 3.12, se muestran los promedios de las funciones de correlación entre la volatilidad y los percentiles 100, 50 y la desviación estándar de los gaps. En este caso, la disminución de las asociaciones con respecto a δ es mucho menor. Se nota además que el pico de la correlación entre el gap mediano y la volatilidad no se sitúa en cero. Esto podría ser consecuencia de que los gaps que representan la mediana en una ventana de tiempo t_0 son consumidos en ventanas de tiempo $t_i, i > 0$, lo que genera asociaciones retrasadas entre ambas señales.

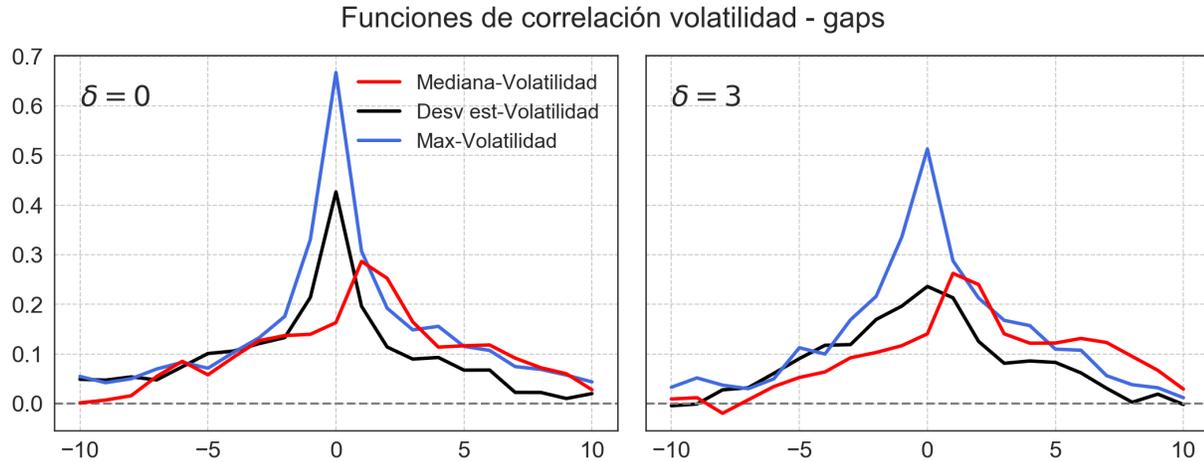


Figura 3.12: Promedios de las funciones de correlación entre las volatilidades y los gaps. El comportamiento es esencialmente idéntico al que muestran las correlaciones entre gaps y rendimientos con la particularidad de que el pico del promedio de correlación entre la mediana de los gaps y la volatilidad se desplaza hacia lags positivos. Este comportamiento se observa con distintos tamaños de ventana y podría ser consecuencia de los diferentes tiempos de respuesta con que la actividad afectan a los primeros niveles del libro con respecto a los niveles más profundos.

Discusión de los resultados en las correlaciones de los gaps

Los resultados obtenidos al analizar la distribución de los gaps son consistentes con la observación que hacen Farmer et al. [18] con respecto a la alta selectividad con que los participantes del mercado introducen órdenes de mercado al libro, las cuales la mayoría de las veces tienen volúmenes idénticos a los volúmenes almacenados en el spread. Esto es importante porque significa que la mayoría de las transacciones ocurren en los primeros niveles del libro, y que las profundidades del libro son de poca relevancia para la dinámica de los precios en escalas cortas de tiempo. El estudio de Farmer et al. se efectuó en el mercado accionario, de manera que las observaciones que obtenemos nosotros extienden la validez de su análisis al mercado de intercambio bitc on/d lar.

Con respecto a las correlaciones de largo rango y la presencia de aglomeraci n en las series de tiempo de gaps, el fen meno de la aglomeraci n de la volatilidad est  ya bien establecido en la literatura de finanzas matem ticas y econof sica [29, 30]. Se han propuesto diversas causas, de las cuales casi todas plantean que la aglomeraci n es consecuencia de las estrategias que siguen los participantes del mercado [31, 32, 33]. El hecho de que

observamos una medición indirecta de aglomeración a través de las correlaciones de largo rango, así como una medición directa con el índice de aglomeración, sugiere indagar en estudios ulteriores si de alguna manera los rendimientos heredan la aglomeración de los gaps; de ser así, la pregunta se movería un paso hacia atrás y la atención se enfocaría en las causas que llevan a los gaps a aglomerarse.

Las funciones de correlación entre los gaps y los rendimientos son consistentes con la literatura que mencionamos anteriormente, en la que se propone que la causa de la aparición de rendimientos de gran magnitud es la existencia de grandes gaps en el libro, además de serlo con el hecho de que la mayoría de la actividad se lleva a cabo solo en los primeros niveles, pues con solo ignorar los primeros tres gaps del libro la magnitud de las correlaciones decrece mucho. Lo que se trae de nuevo a la mesa con nuestro estudio es que el máximo gap usualmente se localiza en la primera posición de los gaps, lo que junto con el hecho de que solo los primeros gaps son relevantes en la dinámica de los precios tiene el efecto de asociar estadísticamente una diversa gama de escalas de rendimientos con el máximo gap. Tanto los tamaños típicos de rendimiento como los máximos se correlacionan con el máximo tamaño de gap; fuera de los máximos gaps las asociaciones son de pequeña magnitud.

3.1.5. Relaciones causales entre gaps y rendimientos

Como consecuencia de que las funciones de correlación entre los gaps y los rendimientos no se desvanecen inmediatamente para retrasos distintos de cero, es natural preguntarse si es posible detectar efectos causales entre los gaps y los rendimientos utilizando las herramientas de descubrimiento causal que se discutieron anteriormente.

En esta sección se presentan los resultados de investigar la existencia de relaciones causales utilizando tres herramientas distintas, que implican tres definiciones de causalidad diferentes: causalidad de Granger, descubrimiento causal con modelos de ruido aditivo y mapeos cruzados convergentes.

Causalidad de Granger

Efectuamos pruebas de causalidad de Granger en ambas direcciones causales: rendimientos \rightarrow gaps y gaps \rightarrow rendimientos. La definición de causalidad en la que se basa esta prueba es una de las menos complejas, y asume solo una dependencia lineal entre el futuro de la variable explicada y el pasado de la variable causal, de manera que tanto su implementación como su interpretación son sencillas.

Con el fin de obtener un ensemble de pruebas suficientemente grande que permita realizar mejor las pruebas de hipótesis no paramétricas, dividimos las series de percentiles de cada mes obtenidas con el procedimiento mostrado en la figura 3.2 en ventanas disjuntas de 500 pasos de tiempo, y sobre cada una de estas ventanas se computó la prueba de causalidad de Granger. Sin esta división por ventanas solo contaríamos con una prueba de causalidad por cada mes, con lo que tendríamos menos de 30 realizaciones de tamaños de efecto, lo que haría poco confiables las pruebas de hipótesis que se realizan al comparar las distribuciones de efectos de los datos originales con las distribuciones de las copias sustitutas aleatorizadas de los datos. Para determinar la cantidad óptima de retrasos utilizada en los modelos de regresión lineal de las pruebas de causalidad de Granger, nos basamos en el criterio de información bayesiano [34] y, con base en él, elegimos un valor de $l = 7$ observaciones pasadas.

Los resultados de las diversas pruebas de Granger se condensan en la figura 3.13, que muestra los tamaños de efecto, definidos como el decremento relativo en la suma de residuales cuadrados que se obtiene en el modelo completo con respecto al modelo restringido. En azul se muestran los resultados de las pruebas sobre el conjunto original de datos, mientras que en rojo se muestran los resultados aplicados a copias aleatorizadas de los datos, las cuales sirven como controles estadísticos. En el caso de las relaciones $r_{max} \rightarrow g_{max}$, $g_{max} \rightarrow r_{max}$ y $g_{med} \rightarrow r_{med}$, las distribuciones y sus soportes son muy parecidas, de manera que una gran mayoría de los tamaños de efecto de la prueba se pueden describir como consistentes con la ausencia de efecto causal genuino. Solamente en el caso de la relación $r_{med} \rightarrow g_{med}$, se obtienen algunas pruebas con resultados incompatibles con los controles, que se presentan en zonas donde la densidad de los controles ya es nula.

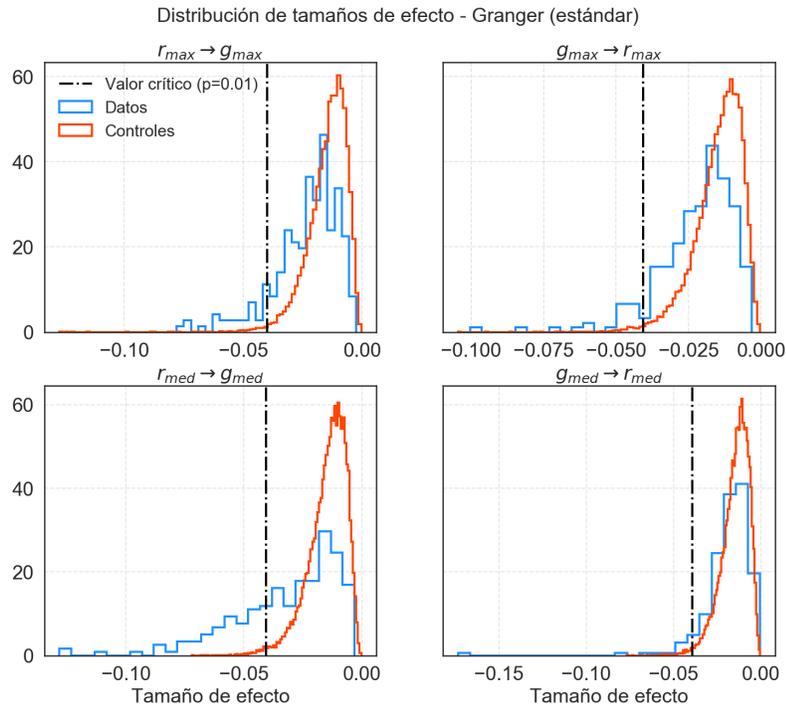


Figura 3.13: Tamaños de efecto de las pruebas de causalidad de Granger estándar. Las líneas negras punteadas señalan el nivel crítico por debajo del cual se concentra una masa de probabilidad de 0.01 en las distribuciones de control. La gran mayoría de las pruebas no superan el nivel crítico, por lo que resultan estadísticamente no significativas. Nótese que a pesar de que los controles fueron generados aleatorizando los datos de cada ventana, lo cual destruye toda información temporal que pudieran tener los datos originales, las distribuciones de control y la de las pruebas sobre los datos son muy similares. Esta similitud aunada al hecho de observar pocas pruebas más allá del nivel crítico hacen muy poco probable la hipótesis de causalidad, al menos en el sentido de Granger.

$g_{max} \rightarrow r_{max}$	$r_{max} \rightarrow g_{max}$	$g_{med} \rightarrow r_{med}$	$r_{med} \rightarrow g_{med}$
0.08	0.10	0.04	0.29

Tabla 3.1: Proporciones de pruebas de causalidad de Granger estadísticamente significativas bajo un nivel de significancia de $p = 0.01$. Muy pocas pruebas resultan estadísticamente significativas, lo que es consistente con la ausencia de efecto causal.

Las fracciones de pruebas estadísticamente significativas bajo un nivel de significancia $p = 0.01$ se enlistan en la tabla 3.1, en la cual se puede observar que muy pocas pruebas resultan significativas.

Cuando consideramos la inclusión de términos instantáneos en el modelo lineal completo, los modelos restringido ($Y_R(t)$) y completo ($Y_F(t)$) adquieren las formas

$$Y_R(t) = c_R + \alpha_1 Y(t-1) + \dots + \alpha_l Y(t-l) + \epsilon(t)$$

$$Y_F(t) = c_F + \alpha_1 Y(t-1) + \dots + \alpha_l Y(t-l) + \beta_0 X(t) + \dots + \beta_l X(t-l) + \eta(t)$$

De esta manera, el término instantáneo $X(t)$ contribuye como variable causal en el modelo de $Y_F(t)$. Esto nos permite comparar qué tanta información cargan las realizaciones simultáneas de los gaps y los rendimientos con respecto a las pruebas de Granger estándar, en relación con el modelo que solo utiliza el pasado de $X(t)$.

La figura 3.14 muestra los resultados de estas pruebas. Saltan a la vista dos características: por un lado, los máximos presentan tamaños de efecto mucho mayores que los controles, de manera que en casi todos los casos no se puede rechazar la hipótesis de causalidad y, por otro lado, las medianas no presentan una diferencia tan marcada como los máximos .

La totalidad de las pruebas entre los máximos gaps y rendimientos resultan estadísticamente significativas, como se puede observar en la tabla 3.2.

$g_{max} \rightarrow r_{max}$	$r_{max} \rightarrow g_{max}$	$g_{med} \rightarrow r_{med}$	$r_{med} \rightarrow g_{med}$
1	1	0.08	0.16

Tabla 3.2: Proporciones de pruebas de causalidad de Granger estadísticamente significativas bajo un nivel de significancia de $p = 0.01$ cuando se incluyen términos instantáneos en los modelos completos. Todas las pruebas entre los máximos resultan ser significativas, mientras que entre las medianas se mantienen muy bajos niveles de significancia.

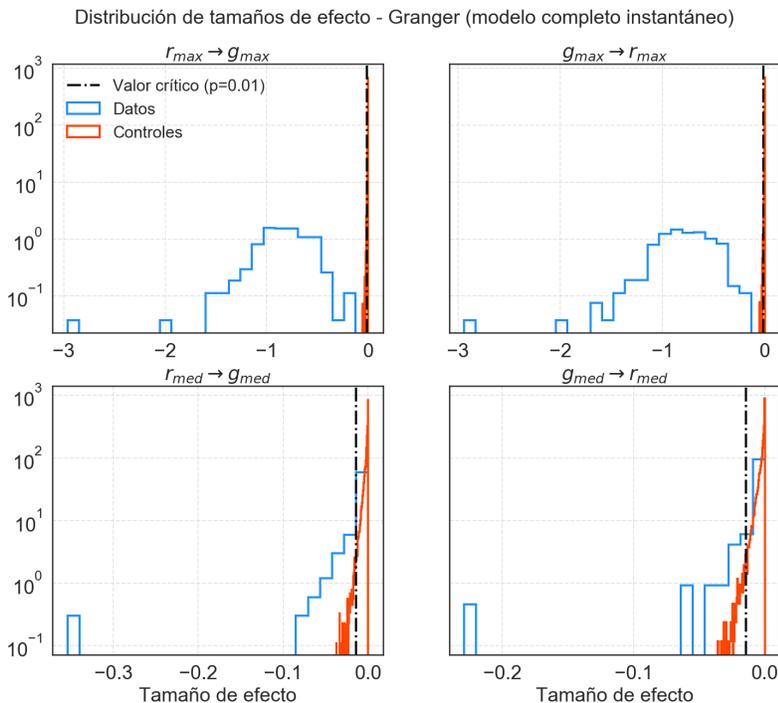


Figura 3.14: Tamaños de efecto de las pruebas de causalidad de Granger con la inclusión de términos instantáneos en los modelos completos. En este escenario, para las relaciones causales entre los máximos, se tiene que un 100 % de las pruebas resultan ser estadísticamente significativas, además de presentar tamaños de efecto muy superiores a los de los controles. Ambas direcciones causales presentan distribuciones muy similares, lo cual, aunado a los tamaños de efecto tan grandes, es compatible con la sospecha de una tercera variable causal que afecta a los rendimientos y los gaps, generando la ilusión de causalidad instantánea entre estas dos señales. Aun con el cambio tan grande que se obtiene entre los máximos al incluir términos instantáneos, los resultados entre las medianas se mantienen muy parecidos a los de los controles.

Modelos de ruido aditivo

Ahora mostraremos los resultados del descubrimiento causal con modelos de ruido aditivo (ANM).

Como las pruebas con ANM no consideran el orden cronológico entre las variables causales y explicadas, condujimos pruebas entre los gaps y los rendimientos sin modificación alguna, de manera que se estudien relaciones causales instantáneas, así como entre versiones con retraso de estas series, tratando de hacer mímica del orden cronológico que toman en cuenta las pruebas de causalidad de Granger.

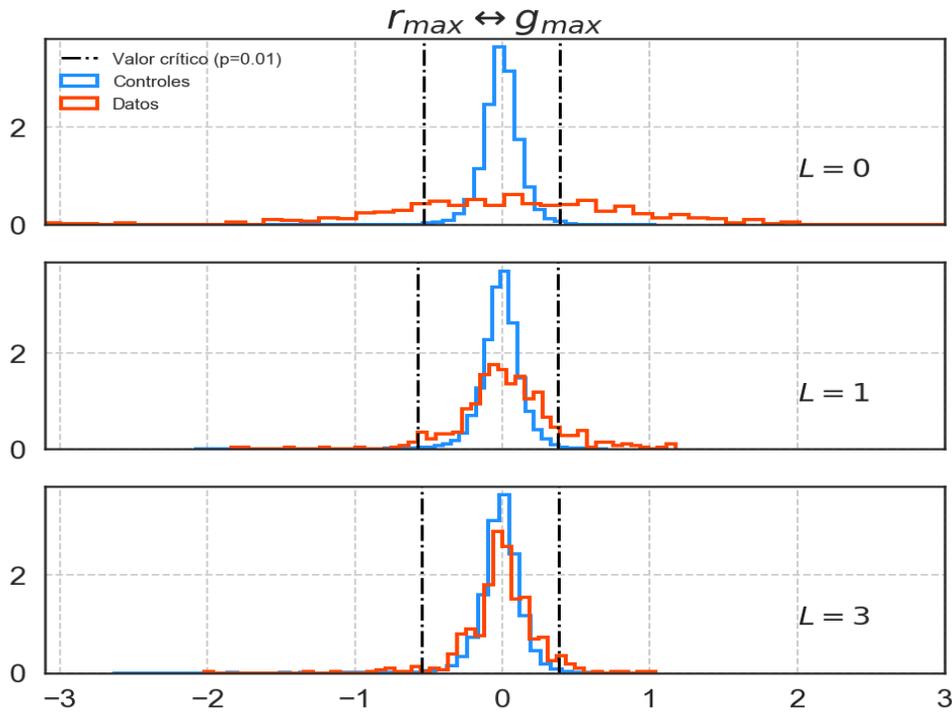


Figura 3.15: Tamaños de efectos para las pruebas de descubrimiento por ANM con retraso cero, uno y tres ($L = 0, 1, 3$) aplicado a los rendimientos y gaps máximos. Con $L = 0$, una proporción significativa de los casos se aleja considerablemente de los valores críticos de la distribución de control, con lo que las pruebas resultan estadísticamente significativas. Sin embargo, este resultado no es sorprendente teniendo en cuenta que existe una relación funcional entre el rendimiento del precio medio y la primera brecha, inducida por la definición misma de rendimiento del precio medio. Con solo un retraso de un paso ($L = 1$), ya se puede observar que las distribuciones de puntuaciones de causalidad recaen abruptamente sobre la distribución de los controles. Esta recaída se vuelve mayor aun con $L = 3$. La similitud entre las distribuciones asociadas a los controles y a los datos originales que resulta al considerar retrasos distintos de cero, y además tan pequeños, sugiere de nuevo la ausencia de un efecto causal genuino.

Los resultados de las pruebas de ANM coinciden con las pruebas de causalidad de Granger en ambos casos: aquel donde no se considera ningún retraso, que muestra la desviación más alta de los puntajes de causalidad con respecto a la distribución de puntajes

de control, y los casos rezagados, donde la distribución de puntajes abruptamente se vuelve similar a la distribución de control, incluso cuando se considera un retraso de un solo paso de tiempo.

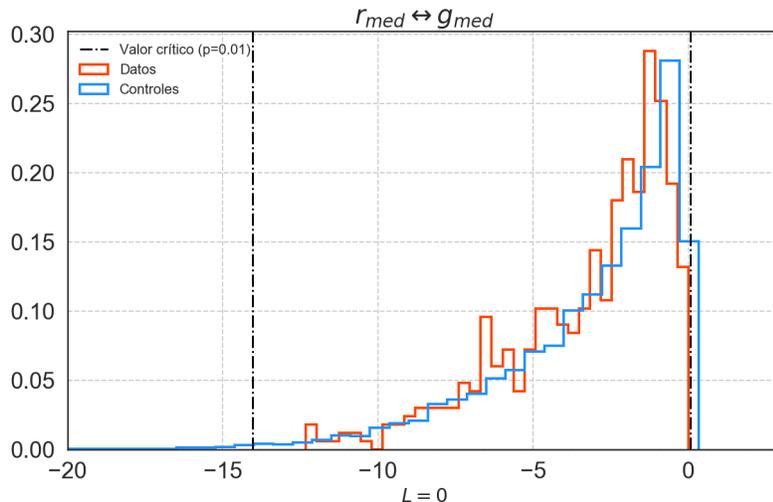


Figura 3.16: Tamaños de efectos para las pruebas de descubrimiento causal por ANM con retraso cero ($L = 0$) aplicado a los rendimientos y gaps medianos.

Dada la relación funcional entre los rendimientos de precio medio y los primeros gaps y el hecho de que una prueba de causalidad por ANM está diseñada para detectar relaciones de la forma $Y = f_Y(X) + E_Y$ para dos variables aleatorias X y Y , no es sorprendente obtener puntuaciones de causalidad ANM estadísticamente significativas para el caso instantáneo ($L = 0$) entre los máximos gaps y rendimientos. Incluso cuando se considera el escenario de ausencia de una relación causal microeconómica / mecanicista real entre los gaps y los rendimientos, se esperaría detectar pruebas positivas como consecuencia de las relaciones inducidas por la definición misma de los rendimientos de precios medios:

$$\begin{aligned} r_\tau(t) &= \ln \left(\frac{p_m(t)}{p_m(t - \tau)} \right) \\ &= \ln \left(\frac{A_0(t) + B_0(t)}{A_0(t - \tau) + B_0(t - \tau)} \right) \end{aligned}$$

En la que, al estar involucrados los mejores asks y bids (A_0 y B_0), se genera indirectamente una dependencia con los primeros gaps. Para mostrar esto, obsérvese que

$$g^A_0(t) = \ln \left(\frac{A_1(t)}{A_0(t)} \right)$$

$$g^B_0(t) = \ln \left(\frac{B_1(t)}{B_0(t)} \right)$$

Siendo $g^A_0(t)$ el primer gap de los asks al tiempo t , $A_0(t)$ es el mejor ask y $A_1(t)$ es el segundo mejor ask. Se ve entonces que tanto los rendimientos de los precios medios como los primeros gaps portan una dependencia de los mejores asks o bids, induciendo así una dependencia funcional que se puede detectar en las pruebas de causalidad por ANM.

Por esta razón, el hecho de que una proporción significativa de las puntuaciones de causalidad están muy lejos de las colas de la distribución de control debe considerarse con extrema precaución. Finalmente, fortaleciendo aun más la hipótesis de ausencia de relación causal entre los gaps y los rendimientos, se tiene la simetría de la distribución de las puntuaciones de causalidad en los datos. Es esencialmente igual de probable detectar cualquiera de las dos direcciones causales entre los gaps y los rendimientos, un resultado que está en desacuerdo con un efecto causal unidireccional en el que los gaps afecten a los rendimientos.

L	$g_{max} \rightarrow r_{max}$	$r_{max} \rightarrow g_{max}$
0	0.34	0.22
1	0.10	0.04
3	0.04	0.03

Tabla 3.3: Proporciones de pruebas de causalidad por ANM bajo un nivel de significancia de $p = 0.01$. Las proporciones de pruebas significativas decrecen rápidamente conforme uno considera retrasos distintos de cero, lo cual es consistente con la ausencia de causalidad, pues una de las características fundamentales de una relación causal es la precedencia de la causa con respecto a su efecto.

Por otro lado, el hecho de que el soporte de la distribución de puntajes de causalidad recaiga abruptamente sobre soporte de la distribución de control cuando se consideran versiones rezagadas de los retornos y los gaps, sugiere que no existe una relación causal subyacente desde el presente de los gaps (rendimientos) al futuro de los rendimientos (gaps).

Para las relaciones causales entre las medianas, no se encuentra ningún efecto estadísticamente significativo independientemente del valor del retraso, al igual que no se encontró ningún efecto a través de las pruebas de causalidad de Granger.

Las proporciones de pruebas estadísticamente significativas para los máximos se enlistan en la tabla 3.3. No incluimos la tabla correspondiente para las medianas porque el 100 % de las pruebas resultan ser no significativas. Es interesante que, además de no ser significativas, resultado que también se obtiene con las pruebas de Granger, la única dirección de causalidad detectada es $r_{med} \rightarrow g_{med}$, que es la relación causal con mayor proporción de significancia con la definición de causalidad de Granger.

Mapeos cruzados convergentes

Las dos medidas de asociación causal ya presentadas, Granger y ANM, definen causalidad como una relación estadística entre las variables causales y causadas. Como mencionamos anteriormente, existe una prueba de causalidad llamada prueba de causalidad por mapeos cruzados convergentes (CCM, por sus siglas en inglés) que utiliza una definición de causalidad no estadística, cimentada en la teoría de sistemas dinámicos.

Incluimos CCM en la batería de pruebas porque las otras dos medidas de causalidad que utilizamos dependen de suponer que la variable causada es una función de la variable causal, ya sea considerando el orden cronológico (Granger) o no (ANM), con un término de ruido aditivo. Como consecuencia de la similitud entre estas pruebas, limitarnos a su aplicación puede resultar en sesgos metodológicos que oculten relaciones causales de difícil detección en un contexto como el que presuponen Granger y ANM.

En la figura 3.17 se muestran los resultados de las pruebas CCM entre los máximos rendimientos y los máximos gaps. Ya que estas pruebas no toman en cuenta el orden temporal, consideramos de nuevo versiones de los datos con retrasos relativos entre ellos.

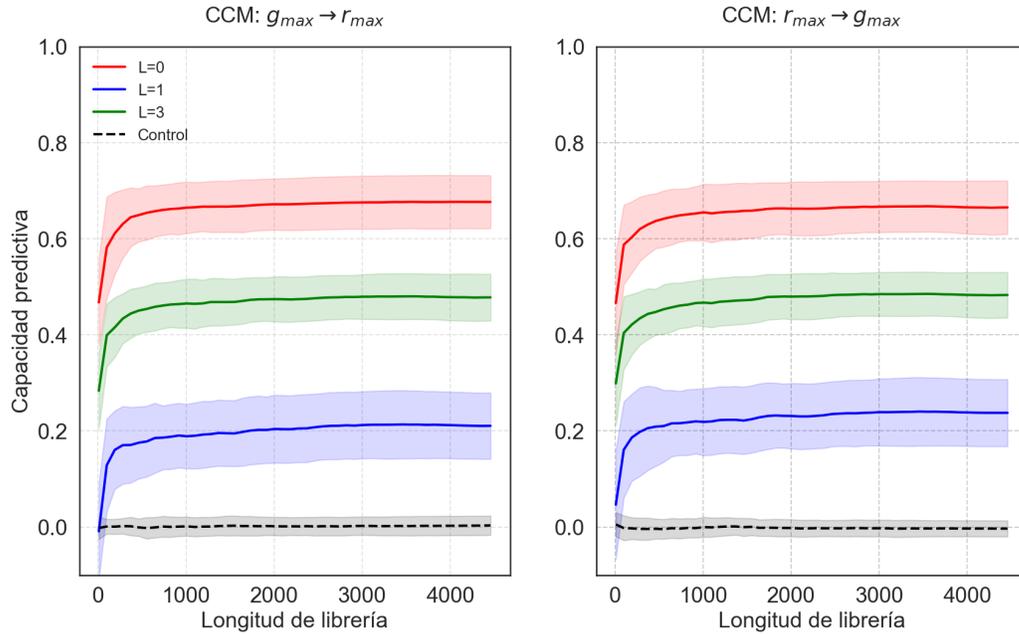


Figura 3.17: Puntuaciones de causalidad en las pruebas de causalidad por mapeos cruzados convergentes para los máximos gaps y rendimientos. Al igual que en las dos técnicas de descubrimiento anteriores, las puntuaciones mayores de causalidad se obtienen en el caso sin retrasos ($L = 0$). Por otro lado, las puntuaciones de causalidad en los casos con retraso son ambiguas: si bien los resultados están muy por encima de los controles aleatorizados, las puntuaciones no convergen a 1 y rápidamente alcanzan los que parecen ser los límites de sus convergencias, es decir, con tamaños pequeños de librería se puede ver el valor final de las puntuaciones de causalidad. Sin embargo, el hecho de que en ambas direcciones causales se obtienen resultados prácticamente iguales nos hace dudar de una relación de causalidad genuina.

Esta vez, no se obtienen resultados compatibles con las pruebas realizadas sobre los controles, los cuales condujimos tanto por aleatorización de los datos originales, como por IAFFT, conservando las correlaciones cruzadas y las autocorrelaciones de los datos originales. En ambos escenarios de control, las pruebas son muy diferentes al aplicarse en los controles y en los datos. Las pruebas sobre los datos originales sugieren una relación causal entre los rendimientos y los gaps, aunque la capacidad predictiva es muy parecida en ambas direcciones causales, lo que hace difícil concluir algo más allá de la mera existencia de una relación causal.

De manera consistente con lo que se obtiene en las pruebas de causalidad por ANM, conforme consideramos mayores retrasos entre los datos, la capacidad predictiva disminuye, lo que significa que el efecto causal es menor y el máximo efecto causal se detecta con cero

retraso. Sin embargo, esta vez la reducción del tamaño del efecto causal no es tan fuerte como con las pruebas ANM.

La definición de causalidad asumida por esta prueba es tan distinta a las otras dos que resulta difícil ponerlas a competir entre sí para alcanzar una conclusión final, unificada. Sin embargo, no se detecta que los mayores efectos causales se alcancen para retrasos distintos de cero, lo que ocurriría si en verdad el pasado de la variable causal afectara al futuro de la variable explicada. Además, el hecho de que los tamaños de efecto sean esencialmente idénticos en ambas direcciones causales sugiere de nuevo que ambas variables podrían estar respondiendo a una tercera variable, no capturada en los datos estudiados.

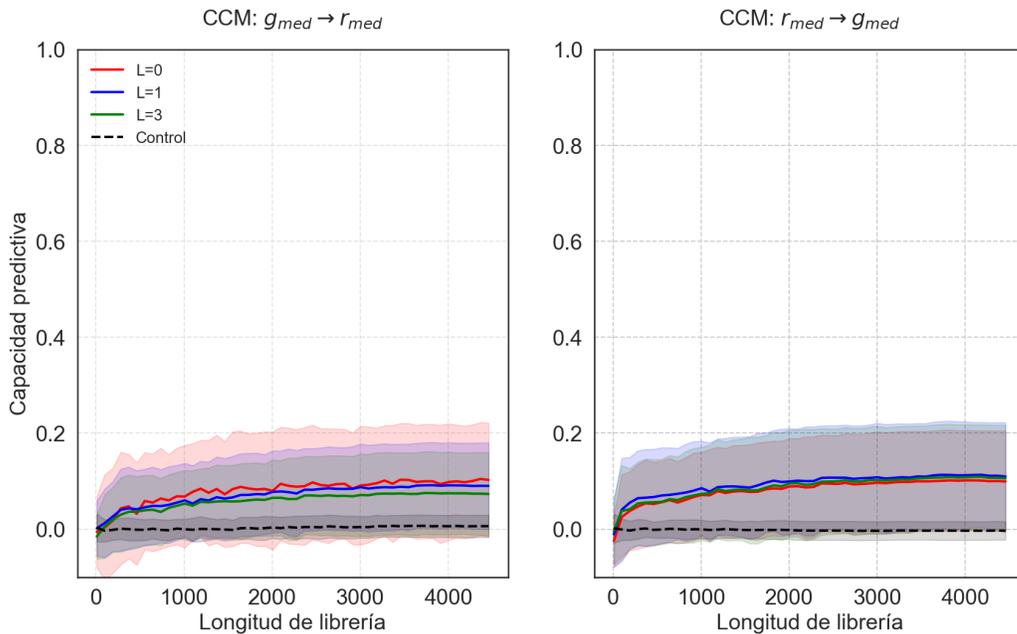


Figura 3.18: Puntuaciones de causalidad en las pruebas de causalidad por mapeos cruzados convergentes para los gaps y rendimientos medianos. En este caso aunque los resultados siguen por encima de los controles, apenas lo hacen así. Además de esto, los rangos de incertidumbre contienen a las puntuaciones de los controles. Estos resultados son muy sugerentes de la ausencia de un efecto causal, como se mide con las otras dos técnicas.

Las pruebas CCM aplicadas sobre las medianas de nuevo presentan los menores tamaños de efecto, como se puede ver en la figura 3.18.

Discusión de los resultados en las pruebas de causalidad

Los resultados de las pruebas de causalidad de Granger y de las pruebas de causalidad por modelos de ruido aditivo sugieren fuertemente que la existencia de grandes gaps en el libro de órdenes en un intervalo de tiempo $[t, t + \tau]$ no causa la generación de rendimientos de gran tamaño en el intervalo futuro $[t + \tau, t + 2\tau]$, según las definiciones de causa asumidas por estos dos métodos, las cuales son más parecidas a la idea de relación causal como se comprende en física. En la prueba de Granger se asume directamente que el futuro de la variable afectada es proporcional al pasado de la variable causal, y en las pruebas ANM, al utilizar retrasos en las series que estudiamos, inducimos como hipótesis que el futuro de la variable afectada se relacione con el pasado de la variable causal a través de un mapeo no necesariamente lineal.

Estos resultados se oponen a los obtenidos en los estudios mencionados previamente, los cuales concluyen que la causa detrás de la aparición de rendimientos de gran magnitud es la presencia de gaps grandes en el libro. Por otro lado, obtenemos tamaños de efecto causal muy grandes cuando incluimos términos instantáneos en los modelos causales. El hecho de que los tamaños de efecto causal en las pruebas estándar son esencialmente irrelevantes, mientras que tamaños de efecto que se tienen con las pruebas de causalidad instantánea superan por mucho los controles estadísticos sugiere que la presencia de gaps de gran tamaño se asocia de manera no causal con los rendimientos extremos. Dicho de otra manera, existe una fuerte correlación que no implica causalidad.

La hipótesis de afectación del estado de los gaps sobre la estadística de los rendimientos es muy natural; después de todo, es de las órdenes que ocupan los niveles del libro de donde se obtienen los precios de las transacciones. Entonces, un libro con órdenes muy separadas debería generar saltos grandes en los precios, mientras que un libro densamente ocupado por órdenes daría lugar a una evolución de los precios con saltos pequeños. Esto hace difícil explicar los resultados negativos de las pruebas de causalidad; si es de las mismas órdenes que se extraen los precios de las transacciones, ¿por qué no sirve conocer el estado de los gaps para mejorar las estimaciones de los tamaños de rendimiento futuros? Creemos que esto es así porque no todos los cambios grandes de precio están relacionados con

la aniquilación de órdenes (y sus consecuentes gaps) en las primeras posiciones. En los datos que estudiamos, detectamos que en algunas ocasiones el arribo de una orden nueva al libro generaba el máximo rendimiento en una ventana de tiempo; en otras ocasiones la desaparición de un gap distinto al máximo generaba el mayor rendimiento, como se muestra en la figura 3.19.

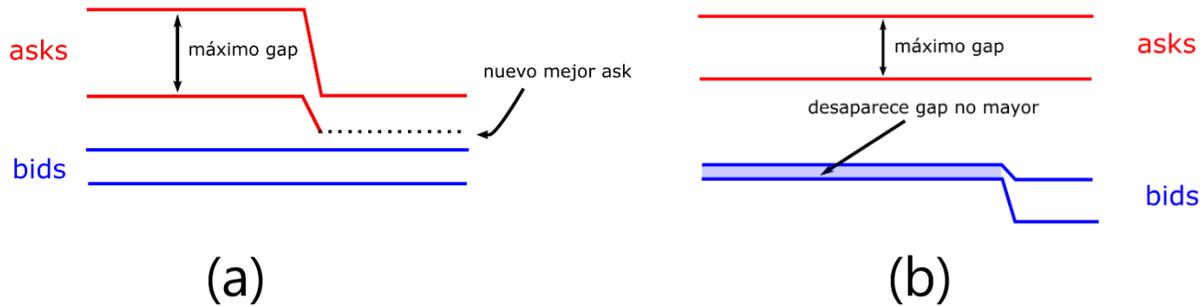


Figura 3.19: Dos escenarios en los que el rendimiento máximo dentro de una ventana de tiempo no se genera por la desaparición del máximo gap. En (a) el rendimiento se generó por el arribo de una orden límite que sustituye a la mejor oferta de venta con un precio mucho más competitivo que el de la orden que sustituye, lo cual genera un cambio grande en el precio medio. En (b) la desaparición del primer gap del lado del libro opuesto al que contiene el mayor gap genera el máximo rendimiento. En ambos casos, los rendimientos se generan por procesos no relacionados causalmente con la existencia del máximo gap, como sí ocurre en la situación mostrada en la figura 2.7.

Pueden existir otros escenarios, además de estos dos, que expliquen la generación de rendimientos. Determinar toda la constelación de situaciones que genera la aparición de rendimientos está fuera del alcance de este trabajo, pero, dado que nos fue relativamente fácil encontrar las situaciones que describimos, resulta razonable pensar que por una diversa gama de mecanismos de generación de cambios de precio las pruebas de causalidad no logran detectar que los tamaños de los gaps aporten suficiente información en la predicción de los rendimientos futuros.

Aun con todo lo mencionado, no podemos descartar los efectos de un observable de mercado no reportado en los datos que estudiamos, que impulsa tanto la desaparición de liquidez de los primeros niveles de precios (que a su vez conduciría a la generación de grandes retornos) como también un aumento en el número de órdenes límite que ingresan al libro (lo que podría contribuir a la creación de grandes gaps), como se muestra de

manera esquemática en la figura 3.20. De una manera muy vaga, el culpable potencial de tal efecto puede ser la intensidad de la actividad de mercado, que puede medirse en el flujo de órdenes entrantes al libro. Cuando los participantes del mercado se vuelven muy activos, se esperaría detectar una relación causal entre la intensidad con que se insertan y consumen órdenes del libro y las propiedades futuras de los primeros gaps y los rendimientos, que daría lugar al efecto causal instantáneo que observamos entre estas dos variables.

Es importante recalcar que utilizamos datos sustitutos permutados aleatoriamente para los casos de control en las pruebas de causalidad. Este tipo de sustitutos mantiene los mismos momentos y distribución que los datos originales, pero cualquier estructura que dependa del orden temporal se destruye [28], lo cual es un escenario de control muy agresivo. Aun así, incluso cuando destruimos toda la estructura cronológica, obtenemos distribuciones de puntajes de causalidad muy similares a las obtenidas en los datos originales, una situación que sugiere que la ausencia de una relación causal entre la presencia de grandes gaps y grandes rendimientos es genuina. Todo esto nos hace creer que solo existen fuertes correlaciones instantáneas pero ningún efecto causal a lo largo del tiempo, y los escenarios que podrían explicar los resultados como consecuencia de otros observables causales son poco plausibles.

La ausencia de relaciones causales entre los gaps y los rendimientos es sugerente de una manifestación más de la hipótesis de mercados eficientes, la cual estipula que los precios de un bien reflejan toda la información disponible que explica su valor; es decir, además de los precios pasados y presentes (que son públicamente accesibles) no existe otra fuente de información a la que un agente pueda acceder para mejorar las valoraciones que este haga sobre el valor futuro de un activo financiero.

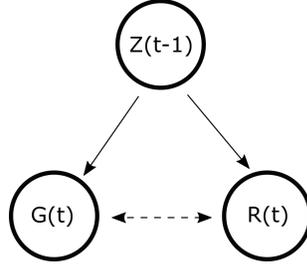


Figura 3.20: Esquema de la relación que podrían tener los gaps y rendimientos con una tercera variable causal que no esté incluida en los datos que estudiamos. Los gaps ($G(t)$) y rendimientos ($R(t)$) podrían estar relacionados causalmente con la actividad de mercado ($Z(t)$) de una manera genuina, dependiente del tiempo, y con ello generar la ilusión de efectos causales instantáneos, como lo detectamos en nuestro estudio.

3.2. Estadística de volúmenes

Con respecto a la estadística de volúmenes, enfocamos las mediciones en el volumen disponible en los primeros niveles de precio del libro, es decir, el menor precio de venta (best ask) y el mayor precio de compra (best bid), ignorando el volumen disponible en otros niveles de precios.

Usaremos la notación $V^{ask}(t)$, $V^{bid}(t)$ para referirnos a los volúmenes disponibles en el best ask y best bid al tiempo t , respectivamente. Aquí, para todo tiempo t se tiene que $V^{ask}(t) > 0$, $V^{bid}(t) < 0$. Esto es así porque las ofertas de venta proveen de moneda al mercado mientras que las de compra consumen moneda.

Replicando la metodología que seguimos en el análisis de los gaps, representamos a los volúmenes de cada lado del libro como una serie de tiempo, generando una serie de tiempo para sus magnitudes y otra para sus cambios relativos logarítmicos, a los que nos referiremos en lo sucesivo como *rendimientos* de volumen, en analogía de los rendimientos tradicionales que están definidos en función de los precios.

Entonces, definimos a los rendimientos del volumen de venta como:

$$r_{\tau}^{ask}(t) = \ln \left(\frac{V^{ask}(t)}{V^{ask}(t - \tau)} \right)$$

Y similarmente para $r_{\tau}^{bid}(t)$, sustituyendo $V^{ask}(t)$ por $V^{bid}(t)$.

El hecho de que los datos que analizamos fueron muestreados con una alta frecuencia

intra-día (10 segundos entre observaciones) es de fundamental importancia. La dinámica de las transacciones en escalas de tiempo muy pequeñas puede mostrar propiedades que no se observarán en escalas más grandes. Muchas de las propiedades estadísticas que presentaremos en este apartado parecen desvanecerse a medida que consideramos escalas de tiempo cada vez mayores, pero debido a la cantidad total de datos disponibles no pudimos confirmar si tal es el caso para escalas de tiempo superiores a unas pocas horas sin disminuir demasiado los tamaños de muestra y afectar negativamente la confiabilidad de las estadísticas.

Por lo tanto, no estudiamos la presencia de estas propiedades estadísticas en el estándar de muestreo diario (un dato por día), pero dada la velocidad a la que estos comportamientos dejaron de ser detectables en nuestras pruebas al incrementar la escala temporal, parece seguro asumir que muestreos diarios no mostrarán las mismas propiedades.

3.2.1. Distribuciones de volumen

A continuación mostramos las funciones de distribución de acumulada ($F_{ask}(v) = \mathbb{P}(V_{ask} \leq v)$) y de manera similar para F_{bid}) de las magnitudes de volumen para asks y bids en la Figura 3.21. Las distribuciones se muestran en escala logarítmica porque los volúmenes toman valores en un amplio rango de órdenes de magnitud (de 10^{-8} hasta 10^3 Bitc oin), as ı que un gr afico en escala lineal ocultar ıa los detalles de la curva de distribuci on.

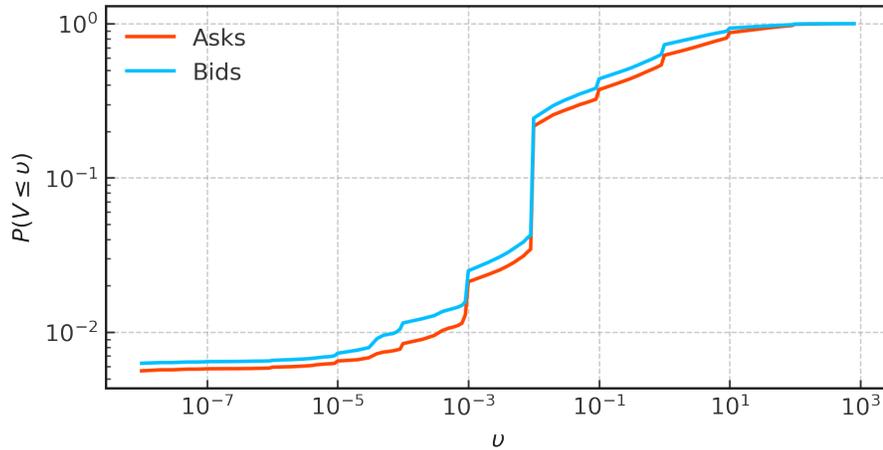


Figura 3.21: Funciones de distribución acumulada de los volúmenes (V) almacenados en el best ask y best bid. Se observa un sesgo en los saltos de la distribución, los cuales son mayores en cada cambio de orden de magnitud, con un salto particularmente grande en 1 centavo de Bitcoin.

Se puede ver un patrón de saltos repentinos en cada cambio de orden de magnitud, lo que sugiere una preferencia de los participantes del mercado para seleccionar volúmenes de pedidos en potencias de diez en lugar de tamaños intermedios. Estando el salto más grande situado en 0.01 Bitcoin.

Además de los saltos en las distribuciones, podemos ver que la distribución del volumen de asks domina estocásticamente la distribución del volumen de bids, es decir, la distribución acumulada de los volúmenes de bids permanece consistentemente por encima de la distribución de asks. Esto implica que $\mathbb{P}(V_{ask} > v) \geq \mathbb{P}(V_{bid} > v)$ para cualquier v dado y como consecuencia, que el valor esperado de los volúmenes de los asks es mayor que el de los bids. Así que se encuentra una asimetría en los volúmenes que eligen los participantes del mercado: por lo general, se almacena más volumen de venta (asks) en lugar de ver el volumen distribuido equitativamente entre compras (bids) y ventas.

3.2.2. Gaussianidad Agregacional

A medida que los rendimientos de volumen se calculan sobre escalas de tiempo cada vez mayores, la función de distribución de éstos tiende a parecerse más a una distribución Gaussiana, como se muestra en la Figura 3.22 en la se presentan los resultados de

pruebas de normalidad de Anderson-Darling realizadas para todas las escalas de tiempo consideradas.

Este comportamiento ha sido observado también en series de cambios de precio y se denomina Gaussianidad agregacional[30].

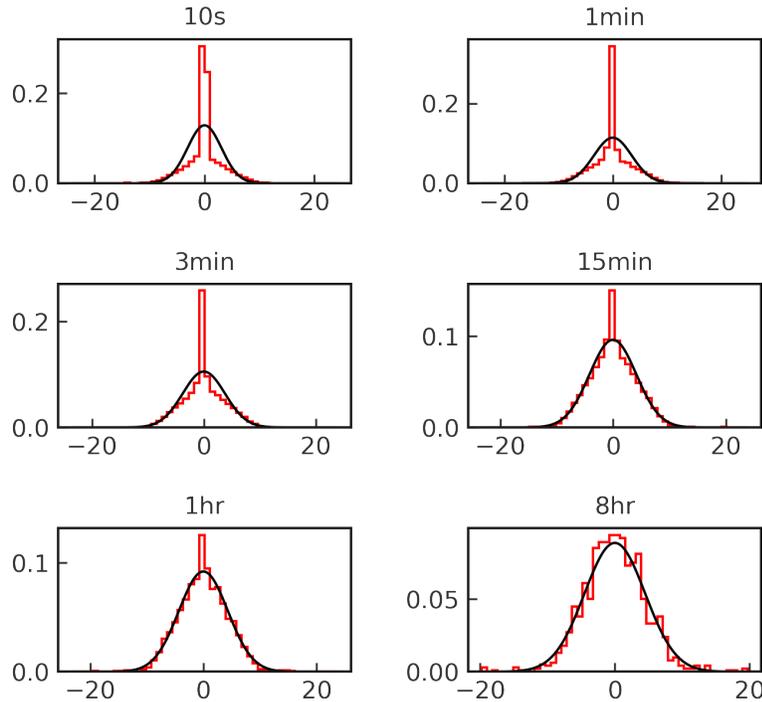


Figura 3.22: Distribuciones de rendimientos de volumen de los bids calculados en diferentes escalas de tiempo. La gaussianidad agregada significa que las distribuciones se parecen cada vez más a una distribución gaussiana conforme se aumentan los intervalos de tiempo en los que se calculan los rendimientos, los cuales van desde 10 segundos hasta 8 horas. La aproximación gaussiana es especialmente buena en el centro en las escalas de tiempo más grandes.

En la Figura 3.22 se puede apreciar que la Gaussianidad se manifiesta con gran rapidez y aunque a una escala de 10 segundos la densidad se aleja mucho de una curva Gaussiana, en particular por la presencia de colas pesadas, al computar los rendimientos cada 8 horas casi se logra indistinguibilidad estadísticamente significativa (a un nivel de confianza del 10%) con respecto a una distribución gaussiana, como se muestra en la figura 3.23.

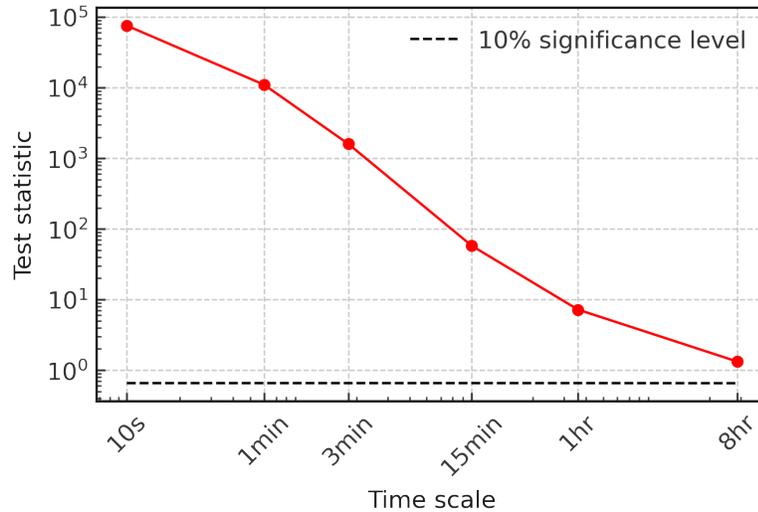


Figura 3.23: Resultados de la prueba de normalidad de Anderson-Darling (rojo) realizada en los rendimientos de volumen calculados cada 10s, 1min, 3min, 5min, 1hr y 8hr. La línea discontinua negra indica el nivel de significancia a 10 %. Aunque ninguna prueba logra ser estadísticamente significativa, es claro que las estadísticas de la prueba tienden rápidamente hacia valores significativos. El conjunto de datos analizado consta de 15 meses, por lo que ir más allá de las 8 horas en el cálculo de los rendimientos nos dejaría con muestras muy pequeñas. Sin embargo, los resultados obtenidos sugieren que, dados los datos suficientes, los retornos de volumen calculados a escalas mayores estarían aún más cerca de mostrar un comportamiento Gaussiano.

3.2.3. ACF y PACF de rendimientos brutos

Cuando calculamos la ACF y PACF de los rendimientos de volumen manteniendo sus signos (rendimientos brutos), observamos que tanto la ACF como la PACF son negativas y se acercan gradualmente a cero, pero la PACF lo hace a un ritmo mucho más lento.

Comparar los comportamientos de la ACF y la PACF, puede dar una mejor idea del tipo de proceso generador de las fluctuaciones de volumen. Una ACF que se reduce a cero lentamente junto con un PACF que se apaga inmediatamente después del retraso p sugiere una serie de tiempo generada por un proceso $AR(p)$. Cuando la ACF se corta después del retraso p pero la PACF decae lentamente a cero, un proceso $MA(p)$ es candidato como proceso generativo. Finalmente, si tanto la ACF como la PACF decaen lentamente a cero, se sospecha de un proceso $ARMA(p)$.

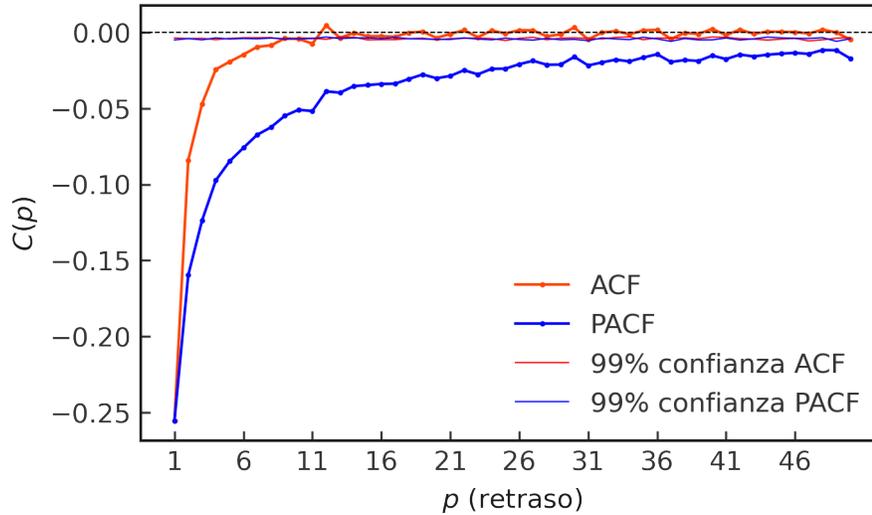


Figura 3.24: Funciones de autocorrelación (rojo) y autocorrelación parcial (azul) de los rendimientos brutos (manteniendo el signo). Los rendimientos se calculan en una escala de 10 segundos. A diferencia de las autocorrelaciones de rendimientos de precios que son insignificantes, los rendimientos de volumen sí presentan correlaciones negativas y altamente significativas que duran aproximadamente 1 minuto (7 lags o 70 segundos) antes de caer a niveles de ruido. Este comportamiento también está presente en la función de autocorrelación parcial pero en la PACF las correlaciones significativas tienen una duración mayor. Una ACF que se apaga rápidamente en relación con la PACF correspondiente implica que los retornos de volumen podrían modelarse como un proceso de promedio móvil.

Como se ve en la Figura 3.24, la PACF se desvanece lentamente hacia 0 mientras que la ACF se corta abruptamente después de 7 retrasos. Por lo tanto, parece que un proceso MA(7) plausible para describir el proceso estocástico que genera a los rendimientos de volumen.

La negatividad de las correlaciones implica que las fluctuaciones de volumen en el presente tienden a oponerse en signo a las fluctuaciones pasadas, es decir, un aumento de volumen tiende a ser seguido por una disminución, y viceversa, como se muestra en la Figura 3.25 en la que se confirma este fenómeno para rendimientos con 1 paso de tiempo de retraso entre ellos.

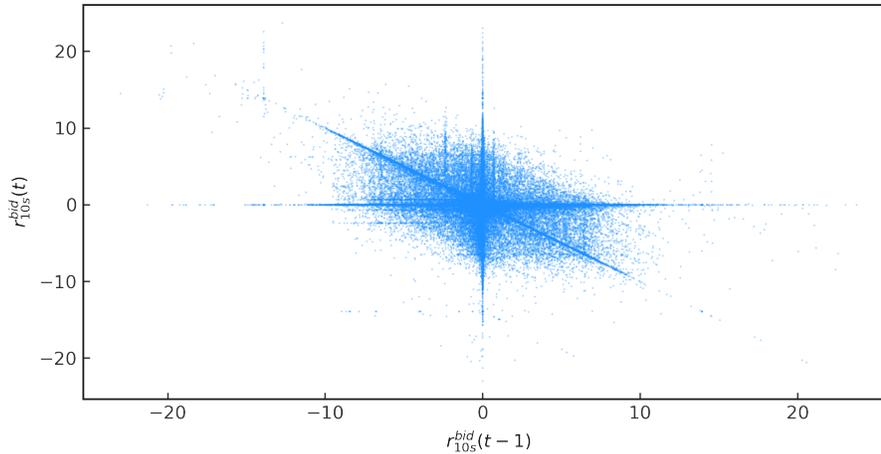


Figura 3.25: Los valores negativos al comienzo de las funciones de correlación indican una dependencia predominantemente lineal y negativa. Esto se confirma en el diagrama de dispersión mostrado. Las correlaciones son genuinamente lineales y no una consecuencia espuria de relaciones no lineales entre $r_{10s}^{bid}(t)$ y $r_{10s}^{bid}(t-1)$ o de la presencia de *outliers* (observaciones mucho mayores que el tamaño típico).

3.2.4. ACF de rendimientos absolutos

Además de medir la ACF de los rendimientos brutos con el objeto de determinar si los aumentos y disminuciones del volumen presentan estructura a lo largo tiempo o si se distribuyen de forma independiente, también medimos la ACF de los rendimientos *absolutos*. Una ACF de retornos absolutos que se mantiene significativamente positiva por tiempos largos mientras decae lentamente hacia niveles de ruido es especialmente interesante porque esto suele suceder cuando es más probable que grandes (pequeños) cambios en el volumen sean seguidos por otros grandes (pequeños) cambios. La tendencia de los cambios de volumen a ser sucedidos por otras fluctuaciones de tamaño similar se denomina *aglomeramiento* y es una ocurrencia muy común en series temporales de cambios de precios sobre una amplia clase de activos financieros. El caso particular de la presencia de aglomeramiento en rendimientos de precios se ha denominado como *aglomeramiento de volatilidad*. Como es el caso con el aglomeramiento de volatilidad, las ACF de rendimientos absoluto de volumen decaen como una ley de potencia del retraso p , como se muestra en la Figura 3.26. En dicha figura mostramos las ACF correspondientes a los únicos tres intervalos de tiempo contiguos que se encuentran en los datos disponibles.

Estos intervalos son de enero a mayo de 2015 (5 meses), de abril a julio de 2016 (4 meses) y de octubre a diciembre de 2015 (3 meses). Elegimos medir las ACF de estos intervalos contiguos para aumentar tanto como sea posible el tamaño de muestra de cada función de autocorrelación y maximizar así la precisión en el cálculo de los exponentes.

Como se ve en la Figura 3.26, el decaimiento en ley de potencia de las funciones de autocorrelación muestra dos cruces ubicados en 60 y 720 retrasos, que corresponden a 10 minutos y 2 horas, respectivamente. El decaimiento de las correlaciones se separa en tres regímenes distintos: un comportamiento de corto plazo hasta los 10 minutos, un comportamiento de rango medio de 10 minutos a 2 horas y finalmente un comportamiento de largo plazo de 2 horas a 10 horas.

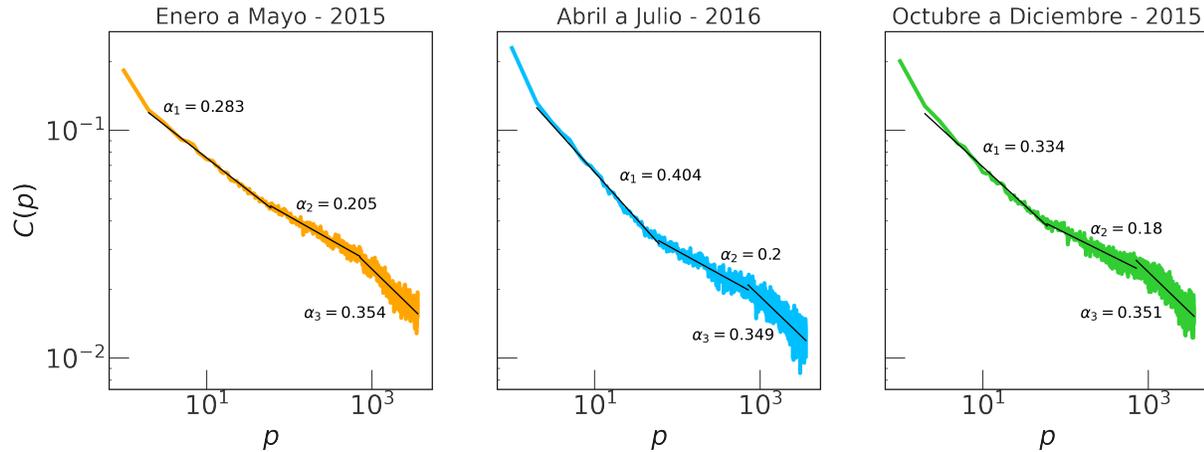


Figura 3.26: Función de autocorrelación de los retornos absolutos. Mostramos solo las correlaciones de los tres intervalos continuos disponibles en los datos. Las líneas negras muestran los ajustes de ley de potencias.

Los valores de los exponentes son en todos los casos consistentes con los valores reportados para el aglomeramiento de volatilidad, que se encuentran en el rango $[0.2, 0.4]$. Cuando las ACF se miden individualmente para cada mes, se obtuvo un resultado similar, aunque algunos meses se pueden describir bien con un solo exponente y ningún mes muestra más de dos leyes de potencia.

Los valores de los exponentes medidos implican la presencia de una dependencia de largo alcance en los tamaños de los rendimientos absolutos. Las variaciones en los valores de los exponentes son más altas en el rango corto, seguidas por las del rango medio y

largo. El comportamiento a largo plazo de las correlaciones es muy estable, mientras que a escalas más cortas, las variaciones locales en el tiempo tienen un efecto más fuerte en la tasa de decaimiento, siendo el efecto más fuerte por debajo de los 10 minutos.

3.2.5. Cuantificación directa de la aglomeración

La función de autocorrelación de rendimientos absolutos sugiere la presencia de aglomeramiento y, como se puede observar en la Figura 3.27, este parece ser el caso.

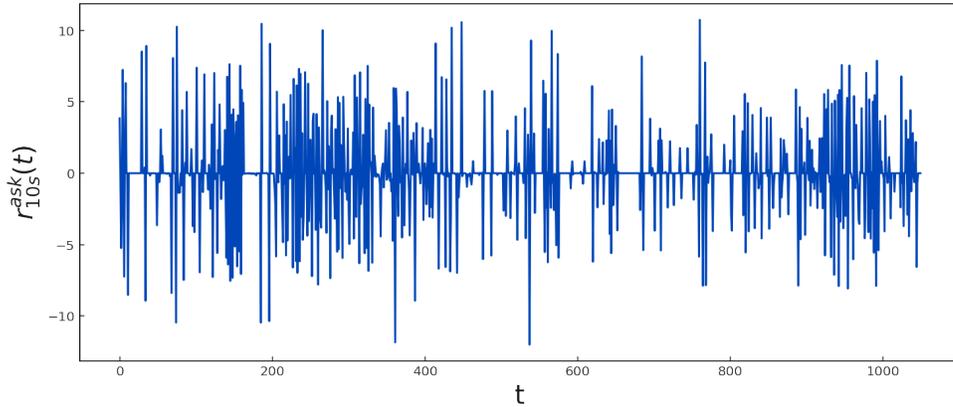


Figura 3.27: Los rendimientos de los volúmenes no se distribuyen de manera independiente a lo largo del tiempo sino que forman aglomeraciones de tamaños similares.

Para cuantificar directamente la tendencia de los rendimientos a agruparse por tamaño, medimos la probabilidad P_τ^Q definida como

$$P_\tau^Q = \mathbb{P}(r(t + \tau) > Q | r(t) > Q)$$

donde Q es algún percentil de la muestra de rendimientos absolutos. Es decir, P_τ^Q es la probabilidad de encontrar dos rendimientos mayores al percentil Q a una distancia de τ pasos de tiempo. Si no existiera aglomeramiento, se esperaría encontrar alrededor de $(100-Q)\%$ de los rendimientos superiores al percentil Q emparejados a una distancia de τ pasos de tiempo para todo valor de τ . Si se encuentra que $P_\tau^Q > 1 - \frac{Q}{100}$ entonces la serie presenta aglomeramiento a lo largo del tiempo.

Los resultados para P_τ^{99} y P_τ^{90} en una escala de 10 segundos, 3 minutos y 15 minutos se muestran en la Figura 3.28.

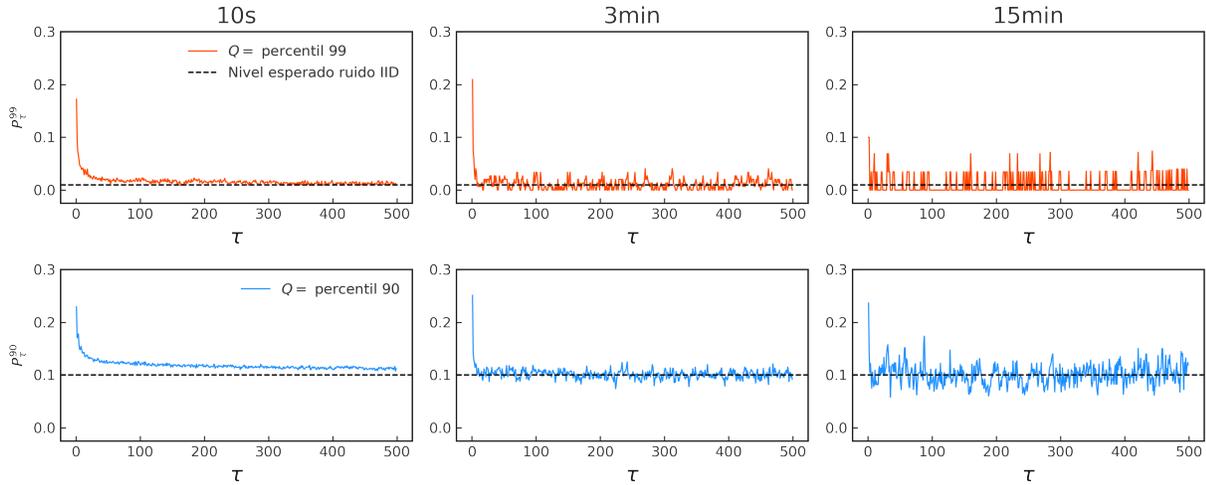


Figura 3.28:

En esa figura, las líneas discontinuas negras corresponden a la frecuencia con la que se espera observar fluctuaciones emparejadas si los rendimientos se distribuyeran aleatoriamente a lo largo del tiempo. Para valores bajos de τ la probabilidad de encontrar pares cercanos de fluctuaciones extremas permanece muy por encima del valor esperado para ruido a escalas de 10s y 3min. Sin embargo, al pasar de medir los rendimientos de volumen cada 10s a medirlos cada 3min, vemos que aunque P_τ^Q es grande para valores bajos de τ , llega al nivel esperado de ruido mucho más rápidamente que con los retornos calculados cada 10s. A los 15min la caída brusca al nivel de ruido es todavía más pronunciada. Entonces, aunque la presencia de aglomeramiento es fuerte en todas las escalas de tiempo consideradas para valores bajos de τ , conforme consideramos escalas de tiempo más grandes sobre las que medimos los rendimientos de volumen, el aglomeramiento de grandes rendimientos tiende a desaparecer. Debido a restricciones en el tamaño de la muestra, no medimos P_τ^Q para una escala de tiempo superior a 15 minutos, pero los resultados obtenidos sugieren que la presencia de aglomeramiento en series temporales de retornos de volumen es una propiedad exclusiva del muestreo de volúmenes en escalas muy cortas de tiempo que no se observará en la típica escala de una medición al día o escalas de mayores.

3.2.6. Tamaños de volumen relativos

El volumen almacenado en los asks constituye una medida directa de la oferta de moneda en el mercado, así como el volumen almacenado en los bids de la demanda. Estas asimetrías entre oferta y demanda son un elemento crucial en la evolución de los precios de transacción, y por ende, es importante cuantificarlas.

Para este propósito definimos la siguiente cantidad, a la que denominamos volumen remanente relativo:

$$V_{rem} = \frac{V_{bid} + V_{ask}}{\max(V_{ask}, |V_{bid}|)} \quad (3.1)$$

Aquí, como mencionamos anteriormente, $V_{bid} < 0$ y $V_{ask} > 0$ y por construcción, tenemos que $-1 < V_{rem} < 1$. Definido de esta manera, V_{rem} es una medida del tamaño relativo del volumen almacenado en el lado (asks o bids) con mayor volumen absoluto con respecto al volumen total disponible entre asks y bids. Si consideramos el escenario en el que el volumen en ambos lados fuera igualado, es decir, que las demandas y ofertas del mercado estuvieran perfectamente equilibradas, tendríamos $V_{rem} = 0$. Por otro lado, si los volúmenes de ambos lados no son exactamente iguales obtendríamos $V_{rem} \neq 0$. Valores de V_{rem} cercanos a sus extremos (-1 y 1), indicarían un gran desequilibrio en las demandas y ofertas del mercado, mientras que valores cercanos a 0 serían consecuencia de equilibrio entre ofertas y demandas.

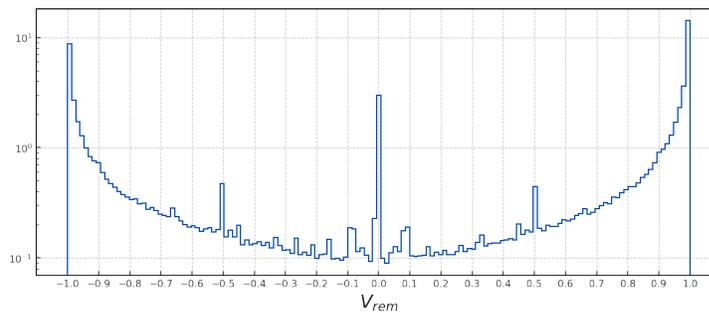


Figura 3.29: Distribution of V_{rem}

La distribución de los volúmenes remanentes se muestra en la Figura 3.29. La primera propiedad de esta distribución que llama la atención es la concentración de masa de pro-

babilidad alrededor del máximo (1) y mínimo (-1). La concentración es tan pronunciada que la probabilidad de observar $|V_{rem}| \geq 0.75$ es 0.7. Esto significa que la ocurrencia más común es tener casi todo el volumen total almacenado ya sea en los asks o en los bids. La otra peculiaridad que podemos observar está en los saltos repentinos en la distribución en $V_{rem} = -0.5, -0.1, 0, 0.1, 0.5$. Dichos saltos implican que, aparte de la situación típica de gran desequilibrio de volúmenes, los siguientes escenarios más comunes son: que ambos lados contengan aproximadamente el mismo volumen igual ($V_{rem}0$), que el lado con más volumen sea aproximadamente 10% mayor que el lado con menor volumen ($V_{rem}0.1$) o que contenga aproximadamente el doble ($V_{rem}0.5$).

3.2.7. Rachas de volumen constante

Hicimos un seguimiento de las ocasiones en que un volumen determinado en un lado del libro permanece constante (rachas) antes de cambiar por la llegada de una nueva que se almacene en el libro volviéndose el nuevo mejor ask o bid, según sea el caso, o por la llegada de una orden que consuma volumen. Este seguimiento tiene el propósito de medir los tiempos típicos que le toma al mercado cambiar el equilibrio de oferta y demanda.

Para este propósito hicimos un seguimiento de las ocasiones en que cada magnitud de volumen en cada lado del libro por separado permaneció constante (es decir, formó una racha de volumen constante) antes de cambiar por alguna de las razones ya mencionadas. Una vez reunidas las duraciones de las rachas de volumen constante, asociamos a cada racha de distinta duración los volúmenes que las constituyeron. De esta manera terminamos con la relación entre los tamaños de volumen y sus duraciones en el libro de órdenes. Relación que no es necesariamente univaluada, ya que se pueden asociar múltiples volúmenes a un tiempo de espera dado, como se muestra en la Figura 3.30

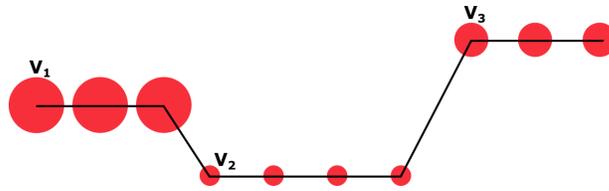


Figura 3.30: Ilustración de la medición de los tiempos de espera y sus volúmenes asociados. Dado un volumen a cada lado del spread (best asks o best bids) hacemos un seguimiento de la duración del intervalo de tiempo t_n en el que dicho volumen, etiquetado como V_{t_n} , se mantiene constante. De manera que el tiempo de espera t_n queda asociado con el volumen V_{t_n} . Después de realizar este procedimiento para cada volumen y tiempo de espera asociado, obtenemos todos los diferentes pares ordenados (t_n, V_{t_n}) que forman la relación entre tiempos de espera y el volúmenes. En el ejemplo del diagrama, tendríamos los pares $(3, V_1)$, $(4, V_2)$ y $(3, V_3)$. La línea negra representa los precios asociados a los volúmenes.

La distribución de rachas de volumen constante se muestra en la Figura 3.31. En ella se observa que la distribución decae rápidamente como función de la duración de la racha, con un 99% de rachas por debajo de los 14 pasos de tiempo, lo que equivale a poco más de 2 minutos. Después de aproximadamente 150 pasos de tiempo, la densidad se comporta aproximadamente como una distribución uniforme.

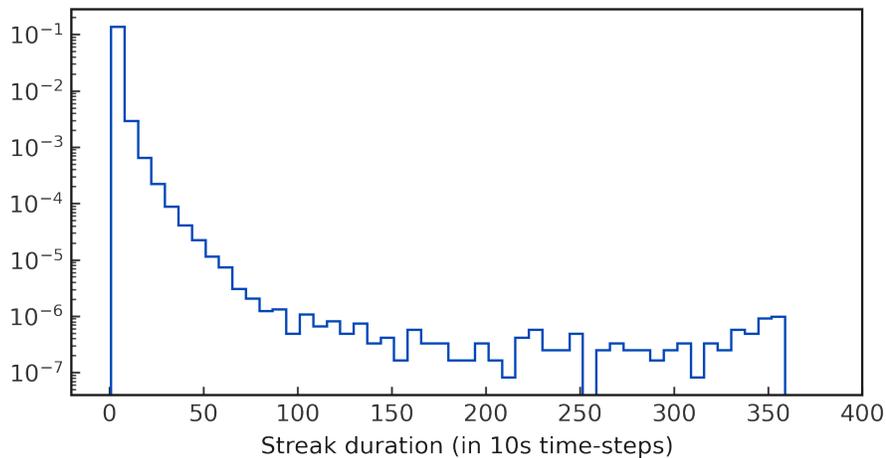
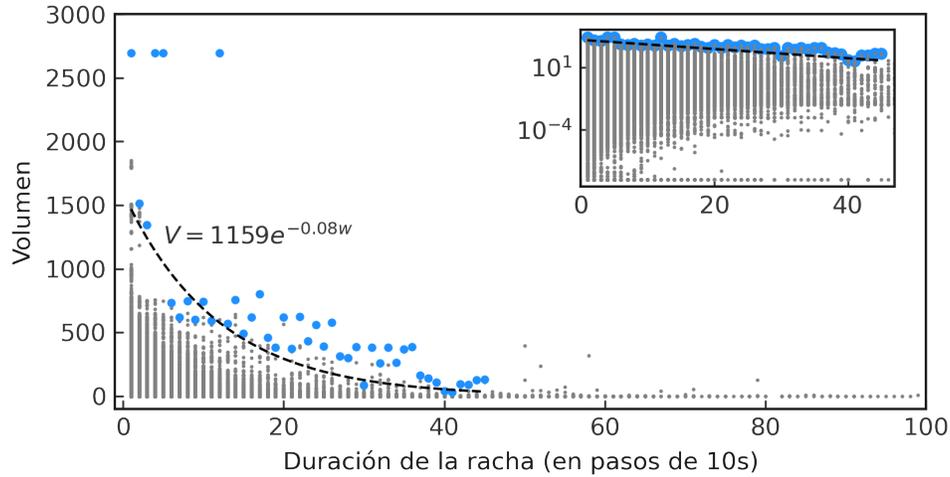


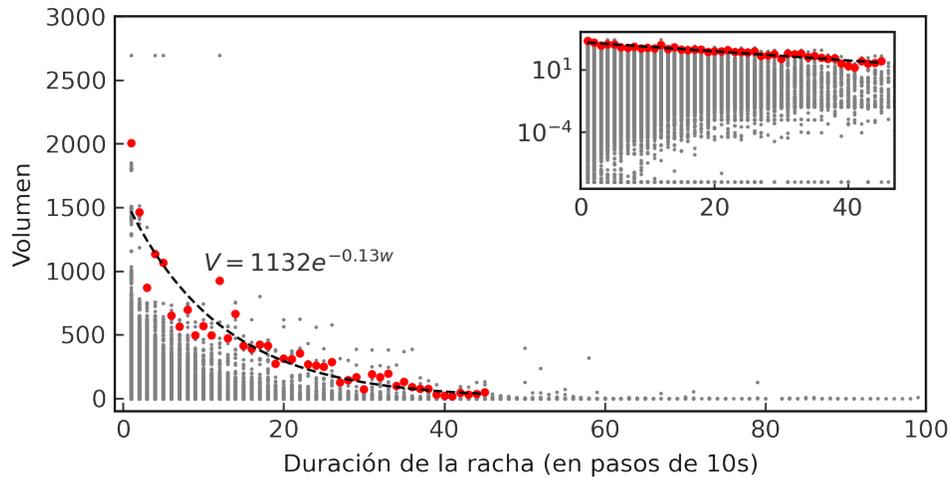
Figura 3.31: Distribución de duraciones de las rachas de volumen constante.

Cuando emparejamos las duraciones de rachas con las magnitudes de volumen que las constituyen, vemos que el volumen máximo asociado a la duración de cada racha también decae rápidamente y un decaimiento exponencial resulta ser una buena descripción de este decaimiento.



0.6

Figura 3.32: Ajuste exponencial (línea negra punteada) de los volúmenes máximos por duración de racha (círculos azules). Este ajuste se ve muy afectado por las grandes fluctuaciones en los volúmenes máximos, las que generan desviaciones sistemáticas en el ajuste.



0.6

Figura 3.33: Ajuste exponencial del promedio de los 10 mayores volúmenes por duración de racha (círculos rojos). Esta vez, el ajuste captura mucho mejor el decaimiento como en función de la longitud de la racha.

Figura 3.34: Gráfico de dispersión de la asociación entre las duraciones de racha y los tamaños de volumen. Aparte de las fluctuaciones estadísticas, los volúmenes máximos asociados a cada racha se reducen conforme los duran más tiempo las rachas. Un ajuste exponencial (línea discontinua negra) describe bien la relación de proporcionalidad inversa. Solo se consideraron para el ajuste las primeras 45 longitudes de racha porque los volúmenes asociados a duraciones mayores tienden a estabilizarse en tamaños muy pequeños.

Discusión de resultados sobre las propiedades de los volúmenes

Las series de rendimientos de volúmenes poseen propiedades que también se encuentran presentes en las series de rendimientos de precio (gaussianidad agregacional y aglomeramiento), al mismo tiempo que se distinguen por la presencia de autocorrelaciones negativas significativas.

La auto-correlación negativa es una propiedad que distingue a los cambios de volumen de los cambios de precio, los cuales en prácticamente todas las series de tiempo financieras presentan correlaciones como las que se obtienen con controles de ruido blanco, a excepción de correlaciones a escalas de tiempo pequeñas (menores a 20 minutos), en donde son también negativas pero duran usualmente un solo retraso. Esto que superficialmente puede parecer una similitud es de hecho una diferencia profunda ya que en el caso de los rendimientos de precio se conoce que la negatividad de las correlaciones es una consecuencia del uso de libros de órdenes y no de la dinámica de los participantes del mercado.

La correlación negativa en los rendimientos de los precios es una consecuencia del fenómeno llamado en inglés *bid-ask bounce*. Este fenómeno ocurre porque los precios de las transacciones se extraen del best ask y el best bid, “rebotando” entre estos dos niveles de precio antes de que se presente un cambio en el best bid o best ask. Este rebote termina generando correlaciones a lo largo en el tiempo en los rendimientos, aún cuando no ingresan órdenes con nuevos precios al libro que cambien el precio del activo. Entonces, la interacción entre *ambos* lados del libro es la causa de estas correlaciones negativas.

Pero las funciones de correlación de los rendimientos de volúmenes que mostramos corresponden a cada uno de los lados del libro *por separado*, de manera que son los cambios en los volúmenes de los best asks por sí solos y sin interacción con los best bids los que presentan autocorrelaciones negativas, ocurriendo lo mismo con las correlaciones de los best bids. Además de esta diferencia, en los rendimientos de volúmenes las correlaciones se mantienen estadísticamente significativas durante un número mucho mayor de retrasos que en los rendimientos de precio. La memoria que tiene la serie de rendimientos de volumen es más relevante porque no se limita a un solo retraso y no es una consecuencia espuria del bid-ask bounce.

En lo que respecta a la presencia de aglomeramiento, si tomamos en cuenta que los cambios de precio también se aglomeran y que usualmente, cuando los precios cambian, estos cambios están asociados con el ingreso de nuevas órdenes al libro, no resulta sorprendente que los volúmenes también estén aglomerados en el tiempo. Para determinar si existe coincidencia temporal entre las aglomeraciones de cambios de precio y las de volumen necesitaríamos una metodología robusta de detección automática de clusters y no contamos con ella en el desarrollo de este trabajo. Pero la pregunta es concreta y se puede investigar en el futuro, con herramientas estadísticas apropiadas. Ahora, si la mera presencia de aglomeramiento no es algo inesperado en las series de rendimientos de volúmenes, el hecho de que es necesario más de un exponente para explicar sí lo es. No es evidente por qué la dinámica de rendimientos de los volúmenes es más compleja que la dinámica de rendimientos de precio, en los que un solo exponente basta para describir el decaimiento de las autocorrelaciones.

Finalmente, las magnitudes de los volúmenes tienen la peculiaridad de tomar ciertos valores discretos con preferencia, en lugar de estar distribuidos de manera suave a lo largo del soporte de su distribución empírica y parece existir un acople entre las magnitudes de los dos lados, como se ve en la distribución de volumen remanente relativo. Debido a la riqueza de propiedades estadísticas que posee la evolución de los volúmenes, no será trivial desarrollar un modelo estocástico que describa simultáneamente todas las propiedades de la dinámica de los volúmenes. Un modelo exitoso deberá reproducir la discretización parcial de los valores en cada señal de volúmenes (bids y asks), generar reversión a la media (autocorrelación negativa) en cada señal y acoplar ambas señales de manera que se acumule casi todo el volumen total en una sola de las señales la mayoría del tiempo.

Capítulo 4

Conclusiones

En este trabajo estudiamos relaciones estadísticas dependientes del tiempo entre los gaps del libro de órdenes y los rendimientos para el mercado de intercambio bitc on/d lar. Para este prop sito, consideramos los estados del libro ordenados con etiquetas temporales como series de tiempo multivariadas, y extrajimos de estas series diversas estadísticas descriptivas de la estructura de los gaps y de la din mica de los precios. Nuestro trabajo se une a estudios previos que establecen una relaci n causal entre los gaps y los rendimientos a trav s de metodologías indirectas que no toman en cuenta el tiempo al utilizar medidas directas de causalidad que consideran expl citamente el tiempo en el que se desarrollan los estados del libro de  rdenes.

Con el empleo de estas medidas de causalidad logramos dar algunas respuestas a las cuestiones de la existencia de alguna relaci n estadística entre los gaps del presente y las fluctuaciones de precio del futuro, y si esta relaci n es de naturaleza causal o solamente una correlaci n.

Para la primera cuesti n, la respuesta es afirmativa, pues detectamos funciones de correlaci n entre los m ximos gaps y los m ximos rendimientos que no decaen de manera abrupta como funci n del tiempo, lo cual hace evidente la presencia de memoria entre la estadística extrema de los rendimientos y los gaps. Similarmente, una funci n que decae de forma no abrupta se encuentra entre las medianas de estas se ales; sin embargo, las magnitudes de estas correlaciones son considerablemente menores que entre los m ximos. De manera consistente con los resultados que se han reportado en la literatura previa,

las magnitudes de las funciones de correlación son altamente dependientes de la profundidad del libro y principalmente se explican como el resultado de las asociaciones entre los primeros gaps y los precios.

Con respecto a la segunda cuestión, la respuesta fue negativa en dos de las tres categorías de pruebas de causalidad que implementamos (causalidad de Granger y causalidad por modelos de ruido aditivo), en las cuales obtuvimos resultados positivos solamente a través del método de mapeos cruzados convergentes. Sin embargo, debido a que este método presupone la existencia de atractores extraños y a que no tenemos evidencia de que esta suposición se cumpla en el mercado que estudiamos, se debe considerar con mucha cautela este resultado positivo. En lo que respecta a las pruebas de Granger y de ANM, los resultados sugieren fuertemente que conociendo los estadísticos que tomamos como variables causales (los gaps máximos y medianos), la capacidad predictiva sobre el futuro de los mismos estadísticos en los rendimientos es tan buena como la que se tiene en los controles, que consisten en ruido sin estructura temporal. Ahora bien, no se puede concluir que por no detectar efectos causales con esas dos técnicas, no existe relación causal alguna, puesto que estos resultados nulos están condicionados al contexto financiero de los datos que estudiamos (mercado bitcoin/dólar), a la naturaleza de los datos (resolución temporal de 10 segundos) y a las herramientas de descubrimiento causal que empleamos. No podemos descartar que exista una relación causal genuina bajo distintas acepciones de relación causal, a distintas escalas de tiempo o en mercados financieros diferentes; sin embargo, este escenario nos parece muy improbable porque, a pesar de la gran agresividad de los controles estadísticos que utilizamos, en estos detectamos resultados muy similares a los que obtuvimos con los datos reales.

Al estudiar los volúmenes disponibles en el spread, encontramos que los volúmenes también presentan el fenómeno de aglomeración, el cual podemos detectar indirectamente a través de la existencia de dependencias de largo rango y directamente a través de la probabilidad de emparejamiento de eventos extremos P_7^Q . La aglomeración de los cambios de volumen se presenta a escalas mucho más pequeñas de las escalas en que es detectable en los cambios de precio y las dependencias de largo rango que generan no pueden ser descritas con un solo exponente. Una característica fundamentalmente diferente que presentan los

cambios de volumen con respecto a los cambios de precio es su función de autocorrelación negativa, lo que sugiere reversión a la media, una propiedad que podría ser de utilidad para la creación de modelos de previsión de volumen de transacción. Finalmente, encontramos que cuando existe un desbalance en el volumen disponible en el spread, lo más probable es que la mayoría del volumen total en magnitud se localice en uno solo de los lados del libro, en palabras más coloquiales, la ocurrencia más frecuente es que cuando llueve, diluvia; en donde por llover nos referimos a la existencia de un desbalance de volumen. Si consideramos que las series de tiempo de volumen son generadas por un proceso estocástico, y lo mismo con las series de rendimientos. La presencia conjunta de las propiedades que observamos en las series de volúmenes sugieren un proceso generador peculiar y más complejo que aquel que está detrás de la generación de los cambios de precio.

Queda como trabajo futuro determinar si la aglomeración de los volúmenes permite predecir la aglomeración de la volatilidad y si las pruebas de causalidad entre gaps y rendimientos resultan también nulas en mercados accionarios.

Bibliografía

- [1] P. G. A. Howells. *Financial Markets and Institutions*. Prentice Hall/Financial Times, 2007.
- [2] Jeff Madura. *Financial Markets and Institutions*. Cengage learning, 2014.
- [3] Frederic S. Mishkin. *The Economics of Money, Banking and Financial Markets*. Addison-Wesley, 2004.
- [4] Kee-Hong Bae, Hasung Jang, and Kyung Suh Park. Traders choice between limit and market orders: evidence from nyse stocks. *Journal of Financial Markets*, 6:517–538, 2003.
- [5] Larry Harris. *Trading and Exchanges: Market Microstructure for Practitioners*. Oxford University Press, 2002.
- [6] Bouchaud2002. Statistical properties of stock order books: empirical results and models. *Quantitative Finance*, 2(4):251–256, 2002.
- [7] J. Bouchaud and M. Potters. *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*. Cambridge University Press, 2003.
- [8] J Doyne Farmer, Paolo Patelli, and Ilija I Zovko. The predictive power of zero intelligence in financial markets. *Proceedings of the National Academy of Sciences*, 102(6):2254–2259, 2005.
- [9] Bruno Biais, Pierre Hillion, and Chester Spatt. An empirical analysis of the limit order book and the order flow in the paris bourse. *The Journal of Finance*, 50(5):1655–1689, 1995.

- [10] Michael A. Goldstein and Kenneth A. Kavajecz. Eighths, sixteenths, and market depth: changes in tick size and liquidity provision on the nyse. *Journal of Financial Economics*, 56:125–149, 2000.
- [11] Gao-Feng Gu, Xiong Xiong, Yong-Jie Zhang, Wei Chen, Wei Zhang, and Wei-Xing Zhou. Stylized facts of price gaps in limit order books. *Chaos, Solitons & Fractals*, 88:48–58, 2016.
- [12] C-K Peng, Sergey V Buldyrev, Shlomo Havlin, Michael Simons, H Eugene Stanley, and Ary L Goldberger. Mosaic organization of dna nucleotides. *Physical review e*, 49(2):1685, 1994.
- [13] Fabrizio Lillo, J Doyne Farmer, and Rosario N Mantegna. Master curve for price-impact function. *Nature*, 421(6919):129–130, 2003.
- [14] Ban Zheng, Eric Moulines, and Frédéric Abergel. Price jump prediction in a limit order book. *Journal of mathematical finance*, 3.
- [15] Thilo A. Schmitt, Rudi Schäfer, Michael C. Münnix, and Thomas Guhr. Microscopic understanding of heavy-tailed return distributions in an agent-based model. *EPL (Europhysics Letters)*, 100, 2012.
- [16] Matthieu Cristelli, V Alfi, L Pietronero, and A Zaccaria. Liquidity crisis, granularity of the order book and price fluctuations. *The European Physical Journal B*, 73(1):41–49, 2010.
- [17] Matthieu Cristelli. *Complexity in Financial Markets: Modeling Psychological Behavior in Agent-Based Models and Order Book Models (Springer Theses)*. Springer, 2013.
- [18] J. Doyne Farmer, László Gillemot, Fabrizio Lillo, Szabolcs Mike, and Anindya Sen. What really causes large price changes? *Quantitative Finance*, 4(4):383–397, 2004.
- [19] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.

- [20] Robert V. Hogg, Joseph W. McKean, and Allen T. Craig. *Introduction to Mathematical Statistics*. Pearson, 2019.
- [21] M Mooij Joris, Peters Jonas, Janzing Dominik, Zscheischler Jakob, and Schölkopf Bernhard. Distinguishing cause from effect using observational data. *Journal of Machine Learning Research*, 2016.
- [22] Hoyer D. Janzing, J.M. Mooij, J. Peters, and B. Schölkopf. Nonlinear causal discovery with additive noise models. *Advances in Neural Information Processing Systems 21*, pages 689–696, 2009.
- [23] Diviyani Kalainathan, Olivier Goudet, and Ritik Dutta. Causal discovery toolbox: Uncovering causal relationships in python. *Journal of Machine Learning Research*, 21(37):1–5, 2020.
- [24] Anastasios A. Tsonis, Ethan R. Deyle, Hao Ye, and George Sugihara. Convergent cross mapping: Theory and an example. In Springer, editor, *Advances in Nonlinear Geosciences*. 2018.
- [25] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *science*, 338(6106):496–500, 2012.
- [26] Floris Takens. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pages 366–381. Springer, 1981.
- [27] Jaroslav Stark, David S Broomhead, ME Davies, and J Huke. Delay embeddings for forced systems. ii. stochastic forcing. *Journal of Nonlinear Science*, 13(6):519–577, 2003.
- [28] Gemma Lancaster, Dmytro Iatsenko, Aleksandra Pidde, Valentina Ticcinelli, and Aneta Stefanovska. Surrogate data for hypothesis testing of physical systems. *Physics Reports*, 748:1–60, 2018.

- [29] Rama Cont. Volatility clustering in financial markets: Empirical facts and agent-based models. In Springer, editor, *Long Memory in Economics*. 2007.
- [30] Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.
- [31] Roberto Mota Navarro and Hernán Larralde. A detailed heterogeneous agent model for a single asset financial market with trading via an order book. *PloS one*, 12(2):e0170766, 2017.
- [32] Lux T. and Marchesi M. Volatility clustering in financial markets: A micro-simulation of interacting agents. *International Journal of Theoretical and Applied Finance*, 3:675, 2000.
- [33] V Alfi, Matthieu Cristelli, L Pietronero, and A Zaccaria. Minimal agent based model for financial markets i. *The European Physical Journal B*, 67(3):385–397, 2009.
- [34] Ernst Wit, Edwin van den Heuvel, and Jan-Willem Romeijn. ‘all models are wrong...’: an introduction to model uncertainty. *Statistica Neerlandica*, 66(3):217–236, 2012.