



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Facultad de Ciencias

PROCESO DE HAWKES PARA FINANZAS DE ALTA
FRECUENCIA

T E S I S

QUE PARA OBTENER EL TÍTULO DE: ACTUARIA

PRESENTA:

PAOLA KAROLINA ESCOBAR SÁNCHEZ

TUTOR:

M. EN C. DANIEL CERVANTES FILOTEO



Cd. Mx. 2023



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Hoja de Datos del Jurado

1. Datos del alumno

Apellido paterno
Apellido materno
Nombre(s)
Teléfono
Universidad Nacional Autónoma de México
Facultad de Ciencias
Carrera
Número de cuenta

Escobar
Sánchez
Paola Karolina
53942705
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
31321127-4

2. Datos del tutor

Grado
Nombre(s)
Apellido paterno
Apellido materno

M. en C.
Daniel
Cervantes
Filoteo

3. Datos del sinodal 1

Grado
Nombre(s)
Apellido paterno
Apellido materno

Dra.
Ruth Selene
Fuentes
García

4. Datos del sinodal 2

Grado
Nombre(s)
Apellido paterno
Apellido materno

Dr.
Oscar
Fontanelli
Espinosa

5. Datos del sinodal 3

Grado
Nombre(s)
Apellido paterno
Apellido materno

M. en C.
Rafael
Miranda
Cordero

6. Datos del sinodal 4

Grado
Nombre(s)
Apellido paterno
Apellido materno

M. en C.
René
Chablé
García

7. Datos del trabajo escrito.

Título
Número de páginas
Año

Proceso de Hawkes para Finanzas de Alta Frecuencia
133
2023

“El hombre no es otra cosa que lo que él se hace.”
- Jean Paul Sartre

Índice general

Introducción	8
1 Mercados de alta frecuencia	11
1.1 Antecedentes	11
1.2 Conceptos básicos	15
1.2.1 Libro de pedidos	16
1.3 Microestructura del mercado de alta frecuencia	17
1.3.1 Formación del precio del mercado	18
1.3.2 Mercados basados en el precio u órdenes	18
1.3.3 Características de los mecanismos comerciales	22
1.3.4 Liquidez del mercado	22
1.3.5 Transparencia del mercado	25
1.3.6 Volatilidad del mercado	25
1.4 Mercados no sincronizados	25
1.4.1 Diferencial oferta-demanda	30
1.5 Hechos estilizados	32
1.5.1 Propiedades de distribución	33
1.5.2 Autocorrelación	35
1.5.3 Estacionalidad	36
1.5.4 Ley de escala	37
1.5.5 Agrupamiento	38
1.6 Base de datos TAQ	39
2 Series de tiempo	41
2.1 Conceptos básicos	42
2.1.1 Series de tiempo no homogéneas	49

2.2	Series de Tiempo para datos de Alta Frecuencia	53
2.3	Procesos ARCH	53
2.4	Procesos GARCH	57
2.4.1	Extensiones del modelo GARCH	58
2.5	Un modelo para la volatilidad intradía	59
3	Procesos de Hawkes	61
3.1	Procesos autoexcitantes	62
3.1.1	Función de intensidad condicional de Hawkes	64
3.2	Estructura Ramificada	67
3.2.1	Factor de Ramificación	67
3.2.2	Propiedades Asintóticas	70
3.3	Estimación de parámetros	72
3.3.1	Función de verosimilitud	72
3.3.2	Simulación	76
3.4	Teorema del cambio de tiempo aleatorio	78
3.5	Covarianza y densidades espectrales de potencia	80
3.6	Generalizaciones del proceso de Hawkes	83
4	Análisis Empírico	85
4.1	SPY 500 ETF	85
4.2	Ajuste del modelo	87
4.2.1	Bondad de ajuste	90
4.3	Medidas de Riesgo	92
4.3.1	Distribución Generalizada de Pareto (GDP)	92
4.3.2	Cálculo del VaR	94
4.4	Conclusiones	97
A	Definiciones básicas	99
A.1	Distribución Exponencial	99
A.2	Distribución Poisson	100
A.3	<i>p-value</i>	101
A.4	QQ-plot	102
A.5	Prueba de Kolmogorov	102

B	Procesos de Punto	104
B.1	Procesos de punto y conteo	104
B.1.1	Función de Intensidad Condicional	109
B.2	Proceso Poisson	110
B.2.1	Notación de Landau	110
B.2.2	Proceso Poisson homogéneo	112
B.2.3	Proceso Poisson no homogéneo	115
C	Medidas de Riesgo	119
C.1	Conceptos básicos	120
C.1.1	Medidas de Riesgo	120
C.2	Valor en Riesgo	121
C.3	<i>Value At Risk</i>	123
C.4	Back-testing para el VaR	124
C.4.1	Prueba Binomial	124
C.4.2	Prueba de Kupiec	125
C.5	Función de Riesgo	126

A mi familia, a mis amigos y a mis mentores.

Introducción

Los últimos años han visto un avance acelerado respecto a la tecnología. El desarrollo de nuevas herramientas tecnológicas ha impulsado diferentes áreas de estudios, como la medicina, la ingeniería, la química, etc. Este avance también ha afectado la forma en la que operan nuestros mercados financieros.

La implementación de sistemas de negociación electrónica ha cambiado la forma en que se manejan los datos en las Bolsas de Valores. Con el comercio o *trading* electrónico, las transacciones se registran automáticamente en tiempo real, lo que permite una mayor precisión y rapidez en el seguimiento de las operaciones del mercado, esto es lo que dio inicio a los mercados de alta frecuencia.

Tener datos intradía implica poseer una gran cantidad de datos, ya que el número de observaciones al día en un mercado de alta frecuencia puede equivaler al número de datos diarios registrados en un período de 30 años. Esto tiene importantes implicaciones estadísticas, ya que el aumento en el número de observaciones independientes mejora la precisión de los estimadores y permite el ajuste de diferentes modelos con mayor exactitud. Además, la cantidad de datos disponibles permite la aplicación de nuevos métodos estadísticos, como el estudio de valores extremos, que permiten examinar con detalle eventos que son “raros” y difíciles de analizar con pocos datos. Con datos de alta frecuencia, se pueden tener muestras con hasta 1,000,000 observaciones independientes, para solo estudiar el percentil de 0.5% y aún así tener 5,000 observaciones con las que trabajar. Esto abre muchas posibilidades ya que nos permite estudiar modelos que vayan acorde a la dinámica de los mercados.

La década de 1990 trajo un auge muy importante en la rama estadística de las finanzas, no sólo porque trajo consigo el comercio electrónico, sino también porque trajo consigo la teoría de la gestión del riesgo de eventos extremos, que se enfoca

en la prevención y mitigación de pérdidas causadas por eventos raros pero catastróficos, como los desastres naturales y los ataques terroristas. Una de las medidas más conocidas de gestión del riesgos financieros es la del Valor en Riesgo (VaR, por sus siglas en inglés).

La estimación precisa del Valor en Riesgo (VaR) es una tarea importante en la gestión de riesgos, ya que proporciona una medida de la pérdida potencial en la que puede incurrir una cartera o inversión en un horizonte de tiempo dado bajo cierto nivel de confianza. En los últimos años ha aumentado el interés por utilizar procesos puntuales, para modelar la llegada de eventos financieros y estimar el VaR.

Uno de los procesos de punto que ha llamado la atención en muchas áreas de estudio en los últimos años, es el proceso de Hawkes. El proceso de Hawkes fue originalmente desarrollado para modelar la ocurrencia de terremotos. En 1971, Alan Hawkes publicó un artículo titulado “Point spectra of some mutually exciting point processes” (1971, [17]), en el que introdujo el modelo de procesos de Poisson auto-regresivos, ahora conocido como el proceso de Hawkes.

Hawkes observó que los terremotos no ocurren de manera independiente, sino que la ocurrencia de un terremoto puede influir en la probabilidad de que ocurran terremotos futuros en la misma área. Esta propiedad, llamada “de autoexcitación” está presente en muchos fenómenos naturales, económicos y sociales y es una característica propia de los datos de alta frecuencia.

En el presente trabajo, proponemos un enfoque para estimar el VaR utilizando un proceso de Hawkes combinado con la Distribución Generalizada de Pareto (GPD), que es ampliamente usada en la Teoría de Valores Extremos. Analizaremos los eventos extremos negativos o excedentes (definidos como pérdidas) de una serie financiera de alta frecuencia, donde el proceso de Hawkes modelará sus tiempos de llegada, y la GPD su tamaño de excedencia. Esperamos obtener un modelo capaz de capturar las características de los datos de alta frecuencia de manera adecuada, que es importante para estimar con precisión el VaR.

El Capítulo 1 busca dar un contexto general de cómo funcionan los mercados financieros actuales, explicando su dinámica, su microestructura y dando una introducción a cómo se comportan y manejan los datos de alta frecuencia.

El Capítulo 2 ofrece una detallada introducción al análisis de Series Tiempo.

Se presentan todos los conceptos básicos necesarios para comprender los modelos clásicos de análisis de series de tiempo, incluyendo los modelos GARCH. Además, se presenta un ejemplo de cómo se utilizan estos modelos para estimar series de alta frecuencia. Es importante destacar que, aunque el enfoque principal de este documento sea el análisis de procesos de punto, el estudio de modelos de series de tiempo es crucial para comprender la evolución de los datos y las tendencias que puedan influir en nuestros procesos. Por lo tanto, el conocimiento y la aplicación de modelos de series de tiempo son fundamentales para cualquier análisis de datos.

El Capítulo 3 presenta al proceso de Hawkes. Principalmente, se muestra una explicación detallada sobre cómo funciona la función de intensidad del proceso de Hawkes y cómo se pueden obtener los parámetros mediante el método de máxima verosimilitud. Además, se aborda cómo escribir este proceso en su forma estructurada y se discuten sus propiedades asintóticas. En general, el capítulo proporciona una base sólida para entender el proceso de Hawkes y su aplicación.

El Capítulo 4 es la culminación de nuestra investigación, en este capítulo se realiza un estudio empírico sobre el precio de un activo de alta frecuencia. En particular, se ajusta el proceso de Hawkes con la distribución generalizada de Pareto (GDP) para poder calcular el Valor en Riesgo (VaR) y se realiza un análisis de la bondad de ajuste del modelo en ambos casos.

Capítulo 1

Mercados de alta frecuencia

En los últimos años los mercados financieros han atravesado importantes cambios, gran parte de ellos debido al avance de la tecnología de la información. Esto ha afectado la forma en que los participantes del mercado interactúan y operan, lo que ha contribuido al desarrollo de un mercado más competente.

Uno de los factores que ha hecho posible este progreso es el uso de datos de alta frecuencia o datos intradía; estos se refieren a las observaciones que son tomadas en intervalos de tiempo sobre una escala refinada. En finanzas, a menudo significan observaciones tomadas por hora, minuto o segundo, lo que nos brinda una gran cantidad de información sobre el comportamiento de precios, tasas, tipos de cambio, etc., y con ello, nuevas perspectivas en el campo de la gestión de riesgos y pronóstico.

Este capítulo es una breve introducción a los conceptos del mercado de alta frecuencia y a los hechos estilizados de los datos intradía. Las definiciones y conceptos se basan principalmente en *An Introduction to High-Frequency Finance* (Dacorona, [1]), *High-frequency financial data modeling using Hawkes processes* (Chavez-Demoulin, V., McGill, J.A., [5]), *Market Microstructure and Modeling of the Trading Flow* (Khalil Antoine al Dayri, [7]) y *Financial econometric analysis at ultra-high frequency: Data handling concerns* (Brownlees C.T., GalloG.M., [10]).

1.1 Antecedentes

Los métodos de negociación de activos financieros han recorrido un largo camino. La evolución del mercado de alta frecuencia ha sido posible gracias al uso de equipos informáticos para la compra-venta de activos financieros en intervalos de

tiempo extremadamente cortos; si bien esto ha cobrado importancia en los últimos años gracias a la digitalización del mercado, desde finales el siglo XVIII ya existía la necesidad de hacer la adquisición de información financiera rápida y accesible para todo tipo de participantes del mercado, con el fin de evitar ventajas entre ellos.

El 28 de enero de 1790, el representante James Jackson¹ del estado de Georgia se dirigió ante la recién formada Cámara de Representantes², reunida en la ciudad de Nueva York. Su propósito, denunciar a los comerciantes de “alta velocidad”.

“Tres barcos, señor; han navegado alrededor de una quincena desde este puerto, cargados para especulaciones; están destinados a comprar el Estado y otros valores en manos de los desinformados...”

En ese tiempo, Jackson y otros congresistas, se debatían la propuesta del secretario del Tesoro, Alexander Hamilton³, de que el entonces nuevo gobierno de los Estados Unidos debía asumir las antiguas deudas en que incurrieron los estados y el Congreso Continental⁴ durante la Revolución, una propuesta que se hizo conocida como el Compromiso de 1790.

Los comerciantes que habían oído hablar de las deliberaciones inmediatamente alquilaron barcos rápidos y comenzaron a comprar la vieja deuda, razonando correctamente que la propuesta aumentaría el valor de mercado de la deuda anterior, que en algunos casos se negociaba al 10% o menos de su valor nominal.

Jackson estaba indignado; exigió al Congreso tomar medidas para “salvar a los habitantes distantes de ser saqueados por estos supuestos especuladores”. Era evidente que se debían tomar moderaciones contra las ventajas informativas que poseían los comerciantes de “alta velocidad”. Así fue como cualquier tecnología que

¹James Jackson fue un político de Georgia nacido en el Reino Unido del Partido Demócrata-Republicano. Fue miembro de la Cámara de Representantes de EE. UU. Desde 1789 hasta 1791.

²La Cámara de Representantes era una de las dos cámaras que formaban el poder Legislativo de Estados Unidos en el año 1787.

³Alexander Hamilton fue el primer secretario del Tesoro de los Estados Unidos. Fue uno de los padres fundadores de los Estados Unidos y el fundador del sistema financiero estadounidense. Como el primer Secretario del Tesoro, Hamilton fue el autor principal de las políticas económicas de la administración de George Washington. Lideró la financiación de las deudas de los estados por el gobierno federal, así como el establecimiento de un banco nacional y un sistema tarifario.

⁴El Congreso Continental fue una asamblea de delegados de las Trece Colonias inglesas del este de Norteamérica, las cuales formarían los Estados Unidos, durante su guerra de Independencia de Inglaterra.

incrementara la velocidad del flujo de información fuera adoptada inmediatamente por la comunidad comercial de Europa y Estados Unidos. Esto llevó a los comerciantes a emplear todos los medios conocidos para comerciar más rápido y más barato, por ejemplo, se incrementó el uso de botes más rápidos, diligencias más rápidas y expresos privados de caballos, de hecho, el comercio de valores fue uno de los primeros y más importantes usos del telégrafo.

Cuando el Compromiso de 1790 fue aprobado, el gobierno de los Estados Unidos asumió la deuda del Congreso Continental y los estados, y el nuevo gobierno emitió bonos. Inmediatamente, se produjo un intercambio rápido de esos bonos entre comerciantes y subastadores. Este negocio resultó ser tan lucrativo que algunos comenzaron a especializarse en el comercio de esos mismos valores. Dos años después, el 17 de mayo de 1792, 24 corredores se reunieron en Wall Street e intentaron formar un monopolio en el comercio de esa deuda. Ese grupo se convirtió en lo que ahora se conoce como la Bolsa de Nueva York, NYSE por sus siglas en inglés *New York Stock Exchange*.

La adopción de técnicas comerciales de “alta velocidad” redujo en gran medida las diferencias de precios entre los mercados y, como era de esperarse, trajo también más quejas sobre ventajas injustas entre participantes. Esto condujo a la constante búsqueda del desarrollo de un sistema capaz de procesar y distribuir información financiera de manera más eficaz. Una vez que los intercambios comenzaron a implementar la comunicación computarizada, las órdenes de compra y venta podrían ejecutarse mucho más rápido; los operadores podrían estar conectados a una plataforma virtual en lugar de estar físicamente presentes en los pisos de negociación.

La década de 1970 vio a los mercados adoptar una tecnología más moderna. En 1971, la Asociación Nacional de Cotizaciones Automatizadas de Distribuidores de Valores (NASDAQ) se convirtió en el primer mercado de valores electrónico del mundo e introdujo un sistema de cotización electrónica para que los creadores y participantes del mercado intercambien valores. Unos años más tarde, en 1976, se introdujo el sistema de cambio de orden designado, en inglés *Designated Order Turnaround (DOT)*, en la Bolsa de Nueva York (NYSE), que permitía la transmisión electrónica de órdenes para comprar y vender valores.

El año de 1980 vio surgir el Programa de Comercio, un tipo computarizado de negociación de valores que abarca una amplia gama de estrategias de negociación de cartera que implican la compra o venta de una cartera de al menos 15

acciones. El mercado electrónico de activos financieros ganó popularidad para los intercambios entre las acciones de S&P 500 y el mercado de futuros, de modo que se podía poner automáticamente una orden en el sistema electrónico de la Bolsa de Nueva York cuando había una diferencia predeterminada entre los dos mercados.

En 1990 el comercio electrónico de instrumentos financieros recibió un gran impulso con la introducción de las redes de comunicaciones electrónicas, *Electronic Communications Networks* (ECN). Un ECN es un sistema computarizado que iguala automáticamente las órdenes de compra y venta de valores en el mercado. Conecta a los principales corredores de bolsa y comerciantes individuales para que puedan negociar directamente entre ellos sin pasar por un intermediario y hace posible que los inversores en diferentes ubicaciones geográficas negocien entre sí de manera rápida y fácil. La Comisión de Bolsa y Valores, o *The Securities and Exchange Commission* (SEC), de Estados Unidos exige que las ECN se registren como corredores de bolsa. Desde 1990, las empresas comerciales han realizado cada vez más inversiones en ECN, disfrutando de todos los beneficios que ofrecen sobre los intercambios tradicionales, incluida una mayor velocidad y eficiencia, menores costos y menos errores manuales, lo que a su vez ha resultado en un mayor uso de este mercado. En 1998, para restringir el monopolio de NYSE y NASDAQ, la Comisión de Bolsa y Valores de Estados Unidos aprobó el Reglamento de Sistemas Alternativos de Comercio que dio lugar a la aparición de una serie de plataformas de comercio electrónico alternativas.

En 2005, la SEC aprobó el Sistema de Regulación del Mercado Nacional, que actualizó la Regla de Comercio, promoviendo la transparencia y la competencia entre mercados y exigiendo que las órdenes comerciales se publiquen a nivel nacional y no en intercambios individuales. Los operadores ahora podrían aprovechar y beneficiarse de cualquier pequeña diferencia de precio de un valor entre dos intercambios diferentes, siempre que pudieran actuar lo suficientemente rápido como para aprovechar el retraso momentáneo entre ellos. Estos cambios, junto con la mejora de la tecnología de alta velocidad, han dando lugar a un nuevo enfoque para el comercio: los mercados de alta frecuencia.

1.2 Conceptos básicos

Hoy en día la mayoría de los mercados financieros operan de forma continua. Consideramos a un **mercado de subasta continua** a aquel donde cualquier participante puede negociar en cualquier momento cuando el mercado esté abierto. Estas negociaciones se llevan a cabo a través de órdenes, que se dividen en **órdenes de mercado**, **órdenes límite** y **órdenes de cancelación**. Una orden de mercado es una orden para comprar o vender un valor de manera inmediata. Este tipo de orden garantiza que la orden se ejecutará, pero no garantiza el precio de ejecución. Una orden de límite es un tipo de orden para comprar o vender un valor a un precio específico o mejor. Para las órdenes de límite de compra, la orden se ejecutará solo al precio límite o inferior, mientras que para las órdenes de límite de venta, la orden se ejecutará solo al precio límite o superior. Una orden de cancelación es una orden enviada previamente para comprar o vender un valor que se cancela antes de que se ejecute en un intercambio. Los inversores pueden cancelar órdenes por cualquier motivo, siempre que la orden no se haya completado todavía.

Como en cualquier otra subasta, los participantes pueden ingresar al mercado presentando sus preferencias por un activo. Si están comprando, informan a la comunidad la cantidad máxima de activos que están dispuestos a comprar por un precio particular o, de manera diferente, el precio que están dispuestos a pagar por cualquier cantidad por debajo de una cantidad máxima, a este precio se llama **precio de demanda** o *bid price*. Del mismo modo, si están vendiendo, informan a la comunidad del precio mínimo que están dispuestos a recibir por una cantidad en particular. Este precio se llama **precio de oferta** o *ask price*. Cuando otro participante del mercado acepta un precio presentado, ambos participan en una transacción.

Si un comerciante tiene un precio de demanda que es más bajo que el precio de oferta de otro comerciante, se producirá una operación a algún precio que ambos acuerden, lo que provocará la desaparición de ambas preferencias del mercado. Una vez que se hayan borrado todas las preferencias acordadas, los que quedan mantienen sus preferencias conocidas y esperan que alguien se relacione con ellos. Esto lleva a la formación en el mercado de un precio de compra máximo y un precio de venta mínimo, llamados respectivamente **la mejor oferta y demanda de precios**. El promedio entre estos precios se denomina **precio medio** o *midpoint price* y a la diferencia entre la mejor oferta y la mejor demanda, que es siempre estrictamente mayor que cero, se le denomina **diferencial de oferta y demanda**.

o *bid-ask spread*. Esta cantidad es muy importante para comprender la dinámica de la oferta y la demanda a nivel de ejecución comercial, igualar compradores y vendedores, y el proceso de negociación de precios en el mercado.

Otro concepto importante es del valor de la **tick**. Una tick mide la fluctuación de precio más pequeña posible para cualquier activo en particular, es decir, es el precio más pequeño que puede ser registrado en el precio. Representa una cantidad específica de dinero, y esta cantidad varía según el activo que se negocie.

1.2.1 Libro de pedidos

Sabemos que un participante se comunica con el mercado mediante órdenes, esencialmente hay 3 tipos: órdenes de mercado, límite y de cancelación. Una orden límite establece un precio límite correspondiente al peor precio permitido para la transacción, por ejemplo, cuando un participante del mercado quiere comprar una cantidad q al precio máximo p , envía una orden de compra límite de q acciones al precio p . Cuando quiere vender, envía una orden límite de venta. El conjunto de preferencias (todas las órdenes límite, sus precios y sus cantidades) de todos los participantes en el mercado se llama el **libro de pedidos u órdenes** (*order book*).

El libro de pedidos contiene los pedidos de compra o venta de cada participante para todos los niveles de precios. Por ejemplo, para cualquier precio dado, el libro de pedidos contiene el volumen total disponible para negociar a ese nivel de precios. En este también se registran las órdenes de mercado y las órdenes de cancelación.

En general, las transacciones son publicadas de forma instantánea en el libro de pedidos con la mayor frecuencia posible. Se pueden ver la cantidad total de acciones disponibles en un número significativo de niveles de precios o tamaño, o valor, de la tick, por debajo de la mejor oferta y por encima de la mejor demanda.

Dependiendo del mercado, el intercambio generalmente publicará el número total de acciones disponibles 5 o 10 niveles de precios (con un volumen distinto de cero) más profundo que la mejor oferta o demanda. En la Figura 1.1 se muestra de manera hipotética, cómo se ve un libro de pedidos para un participante del mercado, donde hay 10 niveles de precios para las órdenes límite de compra y 10 niveles de precio para las órdenes límite de venta para cierto activo con un tamaño de tick de 0.5. Representamos cada orden límite por su propio rectángulo. En cada nivel

de precio, la longitud total de la barra es el número total de acciones disponibles para comprar o vender a nivel de precios.

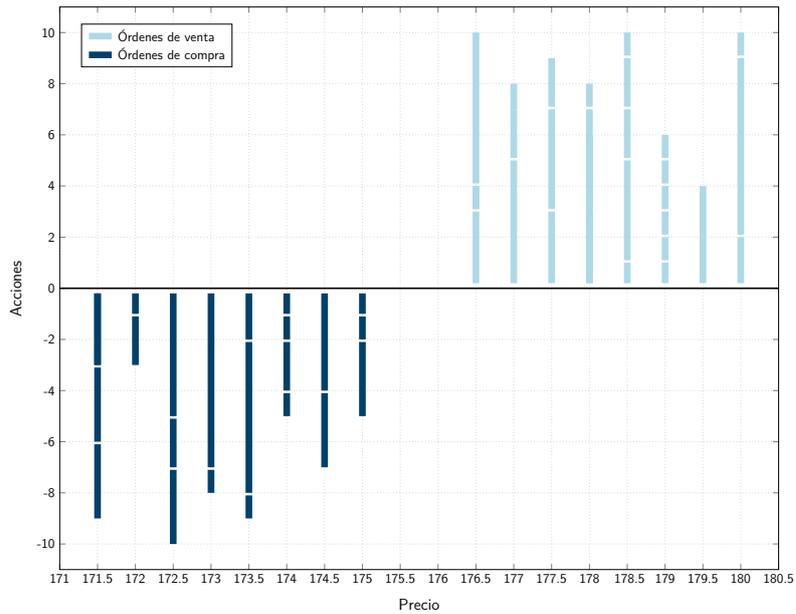


Figura 1.1: La gráfica muestra la cartera de órdenes del contrato futuro DAX de marzo de 2009 el 16 de febrero a las 13 horas con 22 minutos y 48 segundos. El intercambio EUREX solo publica los primeros 10 niveles de cada lado del libro de órdenes en un momento dado, y no publica las diferentes órdenes que componen cada nivel de precios. Mostramos aquí una distribución hipotética de diferentes órdenes límite dentro de cada nivel de precios.

1.3 Microestructura del mercado de alta frecuencia

En esta sección, primero presentamos los tipos más importantes de mecanismos de mercado que se utilizan para negociar acciones en bolsas financieras, como en NYSE, NASDAQ, la Bolsa de París y el FOREX (comercio de divisas). Si bien la lista no se limita a estos cuatro lugares, estos cuatro mercados proporcionan una colección integral de mecanismos comerciales, la mayoría de los cuales también se utilizan en otros centros financieros.

Se introducen los tres mecanismos comerciales teóricos principales, la subasta walrasiana⁵, el mercado basado en el precio y el mercado basado en órdenes. Para cada mercado, proporcionamos una breve revisión de los principales estudios empíricos de microestructura del mercado.

1.3.1 Formación del precio del mercado

Subasta Walrasiana

La subasta walrasiana es un mecanismo comercial antiguo en la literatura económica que consiste en una subasta simultánea donde cada agente calcula su demanda u oferta de un activo a todos los precios posibles y lo envía a un subastador. El precio se establece de modo que la demanda total de todos los agentes sea igual a la cantidad total del activo, lo que se conoce como **precio de compensación del mercado** y que iguala la oferta y la demanda agregadas. Algunos mecanismos de mercado, como el mercado del oro de Londres⁶, están organizados como subastas walrasianas, mientras que otros, como los procedimientos de apertura de la Bolsa de Nueva York y la Bolsa de París, se parecen mucho a la subasta walrasiana, pero son algo diferentes. Algunas bolsas, como la Bolsa de Valores de Arizona, utilizan el proceso de subasta de **órdenes repetidas**, donde los participantes intercambian unidades de un instrumento financiero particular y los precios para comprar o vender se establecen mediante la combinación de los pedidos recopilados en una subasta de llamadas repetidas.

1.3.2 Mercados basados en el precio u órdenes

En la terminología general utilizada en la microestructura del mercado, los mercados generalmente se caracterizan por ser impulsados por los precios (también llamados mercados de **creadores de mercado**) o por órdenes (**mercados de subastas dobles continuas computarizadas** o de libro de pedidos).

Mercados basados en el precio

En un mercado basado en el precio, la existencia de un creador de mercado (o varios creadores de mercado) es fundamental para facilitar la compra y venta de

⁵El principio es llamado así en honor de León Walras, quien lo divulgó en sus cátedras en la Universidad de Lausana.

⁶Como se explica O'Hara en su artículo *Market Microstructure Theory* (1995)

activos negociados. Si un comerciante desea vender un activo a otro comerciante, debe hacerlo a través del creador de mercado designado para el activo, quien compra el activo al primer comerciante y luego lo vende al segundo comerciante.

El creador de mercado es responsable de proporcionar liquidez al mercado publicando cotizaciones de compra y venta del activo en cualquier momento. Estas cotizaciones son precios específicos que el creador de mercado está dispuesto a pagar para comprar el activo o a recibir para vender el activo. El comerciante que desee comprar o vender el activo puede aceptar una de las cotizaciones del creador de mercado.

Es importante tener en cuenta que el creador de mercado está restringido en los precios que puede cotizar debido a que el rango de precios posibles es discreto, y el espacio mínimo entre dos precios posibles sucesivos se llama tick. Por lo tanto, el creador de mercado debe seleccionar sus cotizaciones de compra y venta de una cuadrícula de precios posibles.

El creador de mercado también asume un cierto nivel de riesgo en el proceso de compra y venta de activos. Como mantiene una posición de inventario no deseada durante un período de tiempo, está expuesto a riesgos potenciales. Si, por ejemplo, la compañía emisora del activo anuncia malas noticias justo después de una gran compra por parte del creador de mercado, este puede tener dificultades para vender las acciones a un precio mayor que el precio que pagó para comprarlas, lo que puede resultar en pérdidas.

- **El diferencial:** como el precio de venta es más alto que el precio de oferta, el creador de mercado gana una cantidad positiva de dinero al comprar y vender la acción de inmediato (o casi);
- **Una ventaja informativa:** el creador de mercado generalmente tiene más información sobre las órdenes límite que los comerciantes.

Este es el caso de los especialistas en la NYSE, ya que tienen un conocimiento completo de la cartera de pedidos, mientras que los comerciantes solo conocen la mejor oferta y pueden consultar precios.

Ejemplo:

Suponga que tenemos dos participantes del mercado X e Y , y un creador de mercado Z que hace el mercado de las acciones de la compañía A . X desea vender

100 acciones de A mientras que Y desea comprar 100 acciones. En un sistema de creador de mercado, Z publica cotizaciones de la oferta y demanda que están disponibles para todos los participantes del mercado, como los comerciantes X e Y . Supongamos que el precio de oferta actual es \$10 y que el precio de venta actual es \$11.25. Por lo tanto, el operador X , que desea vender 100 acciones, realiza la operación al precio de oferta de \$10. Una vez que se ha realizado el intercambio, el inventario del creador de mercado ha aumentado en 100 acciones. El operador Y ahora compra las 100 acciones a \$11.25, y el inventario del creador de mercado vuelve a su posición original, lo que resulta en una ganancia neta de \$125 para el creador de mercado. Supongamos ahora que, después de la venta de las acciones al creador de mercado, la compañía A anuncia algunas malas noticias con respecto a uno de sus principales productos, lo que lleva a una revisión a la baja del precio de las acciones de A . Como esta noticia es pública, ningún comerciante estará dispuesto a comprar las 100 acciones al precio de venta vigente de \$11.25. Al mismo tiempo, el creador de mercado Z se queda con un exceso de inventario de acciones que le gustaría vender. No es difícil ver que la única solución que queda es disminuir su precio de venta para atraer órdenes de compra. En este caso y suponiendo que el operador Y todavía desea comprar 100 acciones, Z obtiene una pérdida neta de \$20 en esta operación.

Mercados basados en órdenes

En los mercados basados en órdenes o pedidos, los participantes ingresan directamente las órdenes límite de compra o venta en el libro de pedidos que mantiene el intercambio; actualmente el libro de pedidos es monitoreado por un sistema informático centralizado. Los intercambios ocurren cuando los pedidos pueden ser igualados.

Ejemplo:

Supongamos que el estado actual de el libro de pedidos es el que da la situación en la Tabla 1.1. Podemos ver que se ingresaron cinco órdenes límite en el sistema. Tres comerciantes están vendiendo acciones, con el precio más bajo a \$100 (lo que significa que este comerciante solo venderá las acciones si el precio es igual a \$100 o más) y el precio más alto a \$104. Dos comerciantes están dispuestos a comprar acciones: el primer comerciante acepta comprar 750 acciones siempre que el precio caiga por debajo (o sea igual a) \$98, mientras que el segundo comprará 2000 acciones si el precio cae por debajo (o es igual a) \$96. En esta etapa, un intercambio no es posible ya que la orden de compra más alta es de

\$98, mientras que la orden de venta más baja es de \$100. Supongamos ahora que el comerciante X ingresa una orden de compra por 1000 acciones con un precio límite de \$100. Su pedido coincide inmediatamente con el pedido de venta límite existente a \$100. Al mismo tiempo, las 4000 acciones a \$100 se reducen a 3000 acciones. El estado final del libro de órdenes se encuentra en la Tabla 1.2.

Acciones	Límite compra (\$)	Límite de venta (\$)	Acciones
750	98	100	4000
2000	96	102	800
		104	1000

Tabla 1.1: **Estado de inicio del libro de pedidos.** Se ingresaron cinco órdenes límite en el sistema: tres operadores están vendiendo acciones, con el precio más bajo en \$100 y el precio más alto en \$104; dos operadores están dispuestos a comprar acciones a \$98 y \$96.

Acciones	Límite compra (\$)	Límite de venta (\$)	Acciones
750	98	100	3000
2000	96	102	800
		104	1000

Tabla 1.2: **Estado final del libro de pedidos.** Después del cruce de una orden de compra por 1000 acciones con la orden de venta de límite más bajo.

A pesar de que este ejemplo es muy simple, describe con bastante precisión cómo funciona un sistema de órdenes límite. Aunque no hay creadores de mercado involucrados en este esquema comercial, también existe un diferencial en este caso. En el ejemplo anterior, un comerciante que desee comprar puede hacerlo a un precio de \$100; si decide vender estas acciones inmediatamente después, solo puede venderlas a un precio de \$98: el diferencial es igual a \$2. Como se explicó anteriormente para los mercados basados en los precios, los mercados basados en órdenes también se caracterizan por cuadrículas de precios posibles, y el tamaño de la tick determina el cambio mínimo posible en el precio.

Como se detalla en el ejemplo, la prioridad de precios es una característica muy importante de los sistemas de libro de pedidos, ya que garantiza que las mejores operaciones siempre se ejecuten primero. Por lo tanto, un comerciante que desee comprar (vender) acciones, se asegura que su pedido siempre coincida con los pedidos de venta (compra) de límite más bajo (más alto). Cuando varias órdenes compiten por la ejecución a precios similares, generalmente se aplica la prioridad

de tiempo, lo que significa que las órdenes ingresadas por primera vez en el libro pedidos tendrán prioridad de ejecución. Ejemplos de mercados basados en órdenes incluyen la Bolsa de Toronto, la Bolsa de París, la Bolsa de Bruselas y la mayoría de las Bolsas Europeas.

1.3.3 Características de los mecanismos comerciales

Los mercados se pueden clasificar como **mercados continuos** y **mercados fijos**. Un mercado continuo permite el comercio continuo durante todo el día, mientras que un mercado fijo solo permite el comercio en momentos específicos. La negociación de **subasta por lotes** es un ejemplo de un mecanismo de negociación de mercado fijo que se utiliza para acciones negociadas con poca frecuencia. La negociación continua en tales acciones puede resultar en grandes oscilaciones en el precio ya que pocos participantes del mercado están activos al mismo tiempo. Sin embargo, el comercio en un mercado continuo también puede detenerse debido a eventos “extraordinarios”. Cuando la negociación se reanuda, se utilizan los procedimientos de apertura como si fuera un nuevo día de negociación.

Los mercados también se pueden clasificar como **centralizados** o **descentralizados**. En un mercado centralizado, los participantes del mercado se reúnen en un solo lugar, mientras que en un mercado descentralizado, los comerciantes y creadores de mercado pueden estar en diferentes lugares. Por ejemplo, la Bolsa de Nueva York es un mercado centralizado ya que todo el comercio se realiza a través del especialista en el piso de negociación, mientras que NASDAQ es un mercado descentralizado ya que los creadores de mercado y los comerciantes pueden estar en diferentes ubicaciones.

1.3.4 Liquidez del mercado

Aunque es fácil clasificar los mercados según su tipo de orden/precio, continuidad y centralización, comparar sus características y evaluar su rendimiento relativo es mucho más complicado. Por lo tanto, es esencial contar con un conjunto de criterios que permitan una comparación equitativa de los diferentes tipos de mercados y su desempeño. La literatura existente en microestructura de mercado se centra en la liquidez del mercado para hacer esta evaluación. La **liquidez** se refiere a la capacidad de comprar o vender grandes volúmenes de acciones con un impacto mínimo en los precios y los intercambios que ofrecen mayor liquidez son preferidos

por los inversores, ya que permiten una asignación de capital más eficiente. Por lo tanto, la liquidez es el criterio principal utilizado para evaluar el rendimiento de los mercados. La liquidez del mercado tiene dos componentes distintivos:

- **La dimensión temporal:** ¿Es posible negociar un volumen determinado de acciones en un período corto de tiempo? Esto está relacionado con la posibilidad de comerciar en cualquier momento dado, cuya ausencia introduce costos de inmediatez.
- **La dimensión del precio:** ¿se pagará un precio “justo” por el activo? Más precisamente, ¿es posible negociar un gran volumen de acciones sin interrumpir el proceso de precios, es decir, sin obtener un precio desfavorable?

La liquidez se puede medir en dos dimensiones: **temporal** y **de precio**. La dimensión temporal mide el tiempo entre la presentación de la orden y su ejecución, lo que es importante para los operadores debido al riesgo de movimientos inesperados en el valor subyacente del activo. La dimensión de precio se relaciona estrechamente con el volumen negociado y se puede medir por el diferencial de oferta-demanda, la **profundidad del mercado**⁷, que se refiere a la cantidad de órdenes de compra y venta que se encuentran disponibles en un momento dado para un determinado activo financiero, como una acción, un bono o una divisa, y la **resiliencia**⁸, que se refiere a la capacidad de un mercado para recuperarse rápidamente de una perturbación o evento estresante, como una crisis económica, una caída del mercado, un ciberataque, una pandemia, entre otros. Estas medidas están destinadas a los mercados basados en los precios, pero también se pueden aplicar a los basados en órdenes.

⁷La profundidad del mercado se puede visualizar en un libro de órdenes, que muestra todas las órdenes de compra y venta que están disponibles para un activo financiero en particular en un momento dado. El libro de órdenes también muestra el precio y la cantidad de cada orden. La profundidad del mercado es importante para los operadores financieros porque les permite evaluar la liquidez del mercado, es decir, la facilidad con la que se pueden comprar o vender activos financieros.

⁸La resiliencia en los mercados financieros depende de varios factores, como la fortaleza de los sistemas y procesos de negociación, la eficiencia de los mecanismos de liquidación y compensación, la capacidad de los reguladores para responder y tomar medidas en caso de crisis, la confianza y la credibilidad de los participantes del mercado, entre otros. Los mercados financieros resilientes son capaces de absorber choques y recuperarse rápidamente, sin causar una disrupción significativa en la economía en general.

Liquidez en mercados basados en precios u órdenes

En los mercados basados en precios, los creadores de mercado están siempre disponibles para comprar o vender, lo que elimina el riesgo de ejecución. Sin embargo, los operadores tienen que acordar las cotizaciones con el creador de mercado, lo que tiene un costo. En los mercados basados en libros de órdenes, los operadores enfrentan el riesgo de ejecución al colocar órdenes límite de compra y venta. Si se requiere la negociación inmediata, los operadores pueden optar por una orden de mercado, que se ejecutará contra las órdenes limitadas en el libro de órdenes. En este caso, el riesgo de precio es importante ya que el precio promedio depende de las órdenes límite ingresadas en el libro de órdenes. Con acciones que se comercian con poca frecuencia, puede haber una escasez de pedidos de compra o venta, lo que puede impedir la ejecución del pedido. Las subastas por lotes pueden ser una solución a este problema.

Acciones	Límite compra (\$)	Límite de venta (\$)	Acciones
750	98	100	100
2000	96	104	100
6000	95	110	1000

Tabla 1.3: Ejemplo de libro de pedidos con poca liquidez en el lado de compra (lado derecho de la tabla).

Se da un ejemplo en la Tabla 1.3, que muestra un libro de pedidos con tres órdenes de compra con límite y tres órdenes de venta con límite. El lado de venta es relativamente líquido, ya que se pueden vender 8,750 acciones, pero el lado de compra solo puede manejar 1,200 acciones, de las cuales 1,000 se venden a un precio que está un 10 % alejado del mejor precio de venta. Por lo tanto, un operador que desee comprar 2,000 acciones “en el mejor de los casos” solo podrá comprar 1,200 a un precio promedio de $(100 \times 100 + 100 \times 104 + 1000 \times 110) / (100 + 100 + 1000) = \108.7 . Por lo tanto, este libro de pedidos se caracteriza por una profundidad relativamente buena en el lado de la venta, pero presenta poca profundidad en el lado de la compra. Vale la pena señalar que evaluar la liquidez con solo mirar el diferencial es engañoso, ya que el diferencial de \$2 solo es válido para 100 acciones. Además, la liquidez implica un componente dinámico importante. En el ejemplo anterior, el operador podría optar por enviar una orden de compra de 100 acciones a \$100 y luego esperar. Debido a que la próxima orden de venta límite es de \$104, el margen interno ha aumentado a \$6, lo que podría impulsar a los comerciantes a colocar otras órdenes de venta dentro del rango de \$98 – \$104, trayendo así más liqui-

dez al mercado. Si las órdenes de venta de límite alto con precios de venta bajos se ingresan rápidamente en el libro de pedidos, se restablece la liquidez y la orden de compra de 2,000 acciones se ejecutará a un precio más favorable que \$108.7.

1.3.5 Transparencia del mercado

La transparencia del mercado es la capacidad de los participantes del mercado para observar información sobre el proceso de negociación. Aunque no existe un criterio claro para juzgar la superioridad de un sistema de negociación sobre otro, la mayoría está de acuerdo en que la transparencia del mercado afecta el flujo informativo de pedidos y el proceso de descubrimiento de precios.

1.3.6 Volatilidad del mercado

La volatilidad es una de las métricas de riesgo más importantes. Harris (2003) lo define como la tendencia de los precios a cambiar inesperadamente. Considera que la volatilidad podría considerarse como la reacción del mercado a las noticias reflejadas por los cambios de precios y distingue la **volatilidad fundamental** y la **volatilidad transitoria**. La volatilidad fundamental es causada por variables endógenas que determinan el valor de los instrumentos de negociación. La volatilidad transitoria se debe a la actividad comercial de los participantes desinformados.

La volatilidad no es constante durante la etapa de negociación, sino que cambia con el tiempo. La volatilidad transitoria puede ocurrir en un período de tiempo muy corto antes de que se convierta a su valor fundamental. Los datos de alta frecuencia o intradía reflejan tanto la volatilidad fundamental como la volatilidad transitoria con suficiente significancia estadística.

1.4 Mercados no sincronizados

Las transacciones de acciones, como las de la bolsa de Nueva York, no se producen de forma sincrónica, es decir, diferentes acciones tienen diferentes frecuencias de negociación, e incluso para una sola acción, la intensidad de negociación varía de una hora a otra y de un día a otro. Para datos intradía, los intervalos de tiempo espaciados irregularmente entre observaciones sucesivas son característica sobresaliente en comparación con los datos clásicos de series de tiempo. Sin embargo,

frecuentemente analizamos una serie de rendimientos en intervalos de tiempo fijo, como diario, semanal o mensual.

Para las series diarias, el precio de una acción es su precio de cierre, que es el último precio de transacción de la acción en un día de negociación. El tiempo real de la última transacción de una acción varía de un día a otro. Como tal, asumimos los rendimientos diarios como una serie de tiempo igualmente espaciada dentro de un intervalo de 24 horas. Resulta que tal suposición puede llevar a conclusiones erróneas sobre pronósticos de rendimientos, incluso si las series de rendimiento real son independientes. Para los rendimientos diarios de acciones, el comercio no sincronizado puede introducir:

- I. Correlación cruzada en el retraso 1 entre los rendimientos de activos.
- II. Autocorrelación en el retraso 1 en el rendimiento de la cartera.

Veamos un ejemplo de mercado no sincrónico.

Ejemplo:

Considere las acciones A y B . Suponga que ambas son independientes, y que las acciones A se negocian con más frecuencia que las acciones B . Para noticias especiales que afectan al mercado que llega cerca de la hora de cierre en un día, es más probable que la acción A muestre el efecto de las noticias el mismo día que B simplemente porque A se negocia con más frecuencia. El efecto de las noticias en B finalmente aparecerá, pero puede retrasarse hasta el siguiente día de negociación. Si esta situación realmente sucede, el rendimiento de las acciones A parece liderar las de B . En consecuencia, la serie de rendimiento puede mostrar una correlación cruzada de retraso 1 significativa de A a B , aunque ambas sean independientes.

De una manera más complicada, el comercio no sincronizado también puede inducir autocorrelación negativa errónea para una sola acción. Hay varios modelos disponibles en la literatura para estudiar este fenómeno. Veamos una versión simplificada del modelo propuesto en Lo y MacKinlay (1990, [55]).

Sea r_t el rendimiento compuesto continuo de un activo financiero en el índice de tiempo t . Por simplicidad, suponga que $\{r_t\}$ es una secuencia de variables aleatorias independientes e idénticamente distribuidas con media $\mathbb{E}(r_t) = \mu$ y varianza $\text{Var}(r_t) = \sigma^2$. Para cada período de tiempo, la probabilidad de que el activo no

se negocie es π , que es invariante en el tiempo e independiente de r_t . Sea r_t^o el rendimiento observado. Cuando no hay comercio en el índice de tiempo t , tenemos $r_t^o = 0$ porque no hay información disponible. Sin embargo, cuando hay sí existe comercio en el índice de tiempo t , definimos r_t^o como el rendimiento acumulado del comercio anterior (es decir, $r_t^o = r_t + r_{t-1} + \dots + r_{t-k_t}$, donde k_t es el mayor entero no negativo tal que no se produjo comercio en los periodos $t - k_t, t - k_t + 1, \dots, t - 1$). Matemáticamente, la relación entre r_t y r_t^o es

$$r_t^o = \begin{cases} 0 & \text{con probabilidad } \pi \\ r_t & \text{con probabilidad } (1 - \pi)^2 \\ r_t + r_{t-1} & \text{con probabilidad } (1 - \pi)^2 \pi \\ r_t + r_{t-1} + r_{t-2} & \text{con probabilidad } (1 - \pi)^2 \pi^2 \\ \vdots & \vdots \\ \sum_{i=0}^k r_{t-i} & \text{con probabilidad } (1 - \pi)^2 \pi^k \\ \vdots & \vdots \end{cases} \quad (1.1)$$

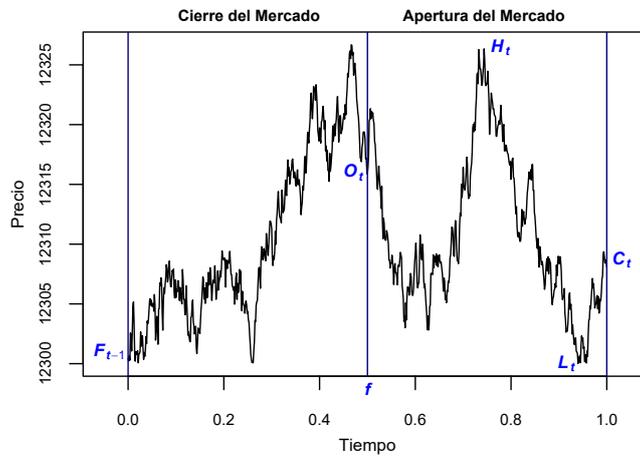


Figura 1.2: Serie de tiempo del precio de un activo financiero en 1 día; la escala del precio es arbitraria. C_t Precio al cierre del t -ésimo día de cotización, O_t es el precio de apertura al t -ésimo día de cotización, f fracción del día (en el intervalo $[0,1]$) que el mercado está cerrado, H_t precio más alto al t -ésimo día de cotización, L_t precio más bajo al t -ésimo día de cotización y F_{t-1} información pública disponible a tiempo $t - 1$. Notemos que si solo tomamos el precio al cierre perdemos mucha información de cómo se comporta el precio.

Que es similar a una distribución geométrica. La probabilidades son fáciles de entender, por ejemplo, $r_t^o = r_t$ si y solo si hay transacciones en t y $t - 1$, $r_t^o = r_t + r_{t-1}$ si y solo si hay transacciones en t y $t - 2$, pero no en $t - 1$ y $r_t^o = r_t + r_{t-1} + r_{t-2}$ si y solo si hay transacciones en t y $t - 3$, pero no hay transacciones en $t - 1$ y $t - 2$, y así sucesivamente. Como se esperaba, la probabilidad total es 1 dada por

$$\pi + (1 - \pi)^2 (1 + \pi + \pi^2 + \dots) = \pi + (1 - \pi)^2 \frac{1}{1 - \pi} = \pi + 1 - \pi = 1$$

Consideremos ahora las ecuaciones de momento de las series de rendimiento observadas $\{r_t^o\}$. Primero, la esperanza de r_t^o es

$$\begin{aligned} \mathbb{E}[r_t^o] &= (1 - \pi)^2 \mathbb{E}(r_t) + (1 - \pi)^2 \pi \mathbb{E}(r_t + r_{t-1}) + \dots \\ &= (1 - \pi)^2 \mu + (1 - \pi)^2 \mu 2\mu + (1 - \pi)^2 \pi^2 3\mu + \dots \\ &= (1 - \pi)^2 \mu (1 + 2\pi + 3\pi^2 + 4\pi^3 + \dots) \\ &= (1 - \pi)^2 \mu \frac{1}{(1 - \pi)^2} = \mu \end{aligned} \quad (1.2)$$

En la derivación anterior, usamos el resultado $1 + 2\pi + 3\pi^2 + 4\pi^3 + \dots = \frac{1}{(1 - \pi)^2}$. Calculemos la varianza de r_t^o , como $\text{Var}(r_t^o) = \mathbb{E}[(r_t^o)^2] \cdot [\mathbb{E}(r_t^o)]^2$ y

$$\begin{aligned} \mathbb{E}(r_t^o)^2 &= (1 - \pi)^2 \mathbb{E}[(r_t)^2] + (1 - \pi)^2 \pi \mathbb{E}[(r_t + r_{t-1})^2] + \dots \\ &= (1 - \pi)^2 [(\sigma^2 + \mu^2) + \pi(2\sigma^2 + 4\mu^2) + \pi^2(3\sigma^2 + 9\mu^2) + \dots] \end{aligned} \quad (1.3)$$

$$= (1 - \pi)^2 [\sigma^2 (1 + 2\pi + 3\pi^2 + \dots) + \mu^2 (1 + 4\pi + 9\pi^2 + \dots)] \quad (1.4)$$

$$= \sigma^2 \mu^2 \left[\frac{2}{1 - \pi} - 1 \right] \quad (1.5)$$

En la Ecuación 1.3 usamos que

$$\mathbb{E} \left[\sum_{i=0}^k r_{t-i} \right]^2 = \text{Var} \left[\sum_{i=0}^k r_{t-i} \right] + \left[\mathbb{E} \left(\sum_{i=0}^k r_{t-i} \right) \right]^2 = (k + 1)\sigma^2 + [(k + 1)\mu]^2$$

bajo el supuesto de independencia de la serie r_t . Empleando técnicas similares en la Ecuación 1.2, podemos mostrar que el primer término de la Ecuación 1.5 se reduce a σ^2 . Para el segundo término de la Ecuación 1.5, usamos la identidad

$$1 + 4\pi + 9\pi^2 + 16\pi^3 + \dots = \frac{2}{(1 - \pi)^3} - \frac{1}{(1 - \pi)^2}$$

que se puede obtener de la siguiente manera:

Sea

$$H = 1 + 4\pi + 9\pi^2 + 16\pi^3 + \dots$$

y

$$G = 1 + 3\pi + 5\pi^2 + 7\pi^3 + \dots$$

entonces $(1 - \pi)H = G$ y

$$\begin{aligned} (1 - \pi)G &= 1 + 2\pi + 2\pi^2 + 2\pi^3 + \dots \\ &= 2(1 + \pi + \pi^2 + \dots) - 1 = \frac{2}{1 - \pi} - 1 \end{aligned}$$

En consecuencia, de las Ecuaciones 1.2 y 1.5, tenemos

$$\text{Var}(r_t^o) = \sigma^2 + \mu^2 \left(\frac{2}{1 - \pi} - 1 \right) - \mu^2 = \sigma^2 + \frac{2\pi\mu^2}{1 - \pi} \quad (1.6)$$

Consideremos a continuación la autocovarianza de $\{r_t^o\}$. Donde empleamos que

$$\text{Cov}(r_t^o, r_{t-1}^o) = \mathbb{E}(r_t^o r_{t-1}^o) - \mathbb{E}(r_t^o) \mathbb{E}(r_{t-1}^o) = \mathbb{E}(r_t^o r_{t-1}^o) - \mu^2$$

La pregunta luego se reduce a $\mathbb{E}(r_t^o r_{t-1}^o)$. Tenga en cuenta que $r_t^o r_{t-1}^o$ es cero si no hay comercio en t y en $t - 1$, o no hay comercio en t y $t - 2$. Por lo tanto, tenemos

$$r_t^o r_{t-1}^o = \begin{cases} 0 & \text{con probabilidad } 2\pi - \pi^2 \\ r_t r_{t-1} & \text{con probabilidad } (1 - \pi)^3 \\ r_t(r_{t-1} + r_{t-2}) & \text{con probabilidad } (1 - \pi)^3 \pi \\ r_t(r_{t-1} + r_{t-2} + r_{t-3}) & \text{con probabilidad } (1 - \pi)^3 \pi^2 \\ \vdots & \vdots \\ r_t \sum_{i=1}^k r_{t-i} & \text{con probabilidad } (1 - \pi)^3 \pi^{k-1} \\ \vdots & \vdots \end{cases} \quad (1.7)$$

Para comprender el resultado anterior, tenga en cuenta que $r_t^o r_{t-1}^o$ si y solo si hay tres transacciones consecutivas en $t - 2$, $t - 1$ y t . Usando la Ecuación 1.7 y el hecho

de que $\mathbb{E}(r_t r_{t-1}) = \mathbb{E}(r_t)E(r_{t-1}) = \mu^2$ para $j > 0$, tendríamos que $\mathbb{E}(r_t^o r_{t-1}^o)$ sería

$$\begin{aligned} &= (1 - \pi)^3 \left[\mathbb{E}(r_t r_{t-1}) + \pi \mathbb{E}[r_t(r_{t-1} + r_{t-2})] + \pi^2 \mathbb{E} \left(r_t \left[\sum_{i=0}^3 r_{t-i} \right] \right) + \dots \right] \\ &= (1 - \pi)^3 \mu^2 (1 + 2\pi + 3\pi^2 + \dots) = (1 - \pi) \mu^2 \end{aligned}$$

La autocovarianza de $\{r_t^o\}$ es entonces

$$\text{Cov}(r_t^o, r_{t-1}^o) = -\pi \mu^2 \quad (1.8)$$

Siempre que μ no sea cero, el comercio no sincronizado induce una autocorrelación de negativa en r_t^o dado por

$$\rho_1(r_t^o) = \frac{-(1 - \pi)\pi \mu^2}{(1 - \pi)\sigma^2 + 2\pi \mu^2}$$

En general, podemos extender el resultado anterior y demostrar que

$$\text{Cov}(r_t^o, r_{t-1}^o) = -\mu^2 \pi^j, j \geq 1$$

La magnitud del ACF depende de las elecciones de μ , π y σ , y puede ser sustancial. Por lo tanto, cuando $\mu = 0$, el comercio no sincronizado induce autocorrelaciones negativas en una serie de rendimientos de un activo financiero. La discusión previa se puede generalizar a la serie de retorno de una cartera que consta de N valores.

1.4.1 Diferencial oferta-demanda

En ciertas bolsas (ej. NYSE), los creadores de mercado son clave para facilitar transacciones, ofreciendo liquidez al estar listos para comprar o vender en cualquier momento. A cambio de esto, se les da el derecho exclusivo por parte del intercambio de publicar diferentes precios de compra y venta para un instrumento. Compran al precio de oferta P_b y venden a un precio de venta más alto P_a . (Para el público, P_b es el precio de venta y P_a es el precio de compra). La diferencia $P_a - P_b$ es lo que conocemos como el **diferencial de oferta-demanda**, que es la fuente principal de compensación por mercado.

Por lo general, el diferencial de oferta-demanda es pequeño, por ejemplo, uno o dos centavos. La existencia de un diferencial de oferta-demanda, aunque de magnitud pequeña, tiene varias consecuencias importantes en las propiedades de las

series de tiempo de los rendimientos de los activos.

El diferencial de oferta-demanda introduce autocorrelación negativa en el retraso 1 en las series de rendimientos de activos. Veamos el modelo simple de Roll (1984). Supongamos que el precio de mercado observado P_t de un activo satisface

$$P_t = P_t^* + I_t \frac{S}{2} \quad (1.9)$$

donde $S = P_a - P_b$ es el diferencial de oferta-demanda, P_t^* es el valor a tiempo t del activo en el mercado, y $\{I_t\}$ es una secuencia de variables aleatorias independientes tal que $I_t = 1$ con probabilidad 0.5 y $I_t = -1$ con probabilidad 0.5. Se puede interpretar como un indicador, que marca 1 cuando una transacción es iniciada por el comprador y -1 si la transacción es iniciada por el vendedor. El modelo también se puede escribir como:

$$P_t = P_t^* + \begin{cases} +\frac{S}{2} & \text{con probabilidad } 0.5 \\ -\frac{S}{2} & \text{con probabilidad } 0.5 \end{cases}$$

Si no hay cambio en P_t^* , entonces

$$\Delta P_t = (I_t - I_{t-1}) \frac{S}{2} \quad (1.10)$$

Bajo el supuesto de que en la ec. 1.9, $\mathbb{E}(I_t) = 0$ y $\text{Var}(I_t) = 1$, y que tenemos $\mathbb{E}(\Delta P_t) = 0$ y

$$\text{Var}(\Delta P_t) = S^2/2 \quad (1.11)$$

$$\text{Cov}(\Delta P_t, \Delta P_{t-1}) = -S^2/4 \quad (1.12)$$

$$\text{Cov}(\Delta P_t, \Delta P_{t-j}) = 0, \quad j > 1 \quad (1.13)$$

Por lo tanto, la función de autocorrelación de P_t es

$$\rho_j(\Delta P_t) \begin{cases} -0.5 & \text{si } j = 1 \\ 0 & \text{si } j > 1 \end{cases} \quad (1.14)$$

El diferencial de oferta-demanda introduce, por lo tanto, autocorrelación negativa en el retraso 1 en la serie de cambios observados en los precios. En la literatura financiera esto se conoce como el **rebote de oferta y demanda**. El rebote se

puede ver de la siguiente manera: Suponga que el precio fundamental P_t^* es igual a

$$\frac{(P_a + P_b)}{2}$$

Entonces P_t asume el valor P_a o P_b . Si el precio observado anteriormente es P_a (el valor más alto), entonces el precio actual observado no cambia o es más bajo en ΔP_t . Por lo tanto, ΔP_t es 0 o $-S$. Sin embargo, si el precio observado anteriormente es P_b (el valor más bajo), entonces ΔP_t es 0 o S . La autocorrelación negativa en el retraso 1 en ΔP_t se hace evidente. Sin embargo, el diferencial oferta-demanda no introduce ninguna autocorrelación más allá del retraso 1. Una formulación más realista es asumir que P_t^* sigue una caminata aleatoria para que

$$\Delta P_t^* = P_t^* - P_{t-1}^* = Z_t$$

que forma una secuencia de variables aleatorias independientes e idénticamente distribuidas con media cero y varianza σ^2 . Además, $\{Z_t\}$ es independiente de $\{I_t\}$. En este caso

$$\text{Var}(\Delta P_t) = \sigma^2 + \frac{S^2}{2}$$

permanece sin cambios. Por lo tanto,

$$\rho_1(\Delta P_t) = \frac{-S^2/4}{S^2/2 + \sigma^2} \leq 0$$

La magnitud de la autocorrelación de P_{t-1} de P_t se reduce, pero el efecto negativo permanece cuando $S = P_a - P_b > 0$.

1.5 Hechos estilizados

Los datos de alta frecuencia, al igual que los datos de frecuencias más bajas, muestran propiedades estadísticas denominadas “hechos estilizados”. La investigación de estos hechos es esencial para crear modelos que expliquen los datos financieros. Los hechos estilizados para datos de alta frecuencia se pueden dividir principalmente en cuatro grupos: propiedades de distribución, autocorrelación de rendimiento, estacionalidad y propiedades de escala.

1.5.1 Propiedades de distribución

En cuanto a la caracterización de las distribuciones de valores de instrumentos financieros, existen varias opciones. Sin embargo, en esta sección, no nos centraremos en proponer un modelo específico para la función de distribución de probabilidad. En su lugar, examinaremos empíricamente cómo se comportan los rendimientos cuando se miden en diferentes frecuencias. En este apartado, analizaremos la distribución en general y nos centraremos en señalar la presencia de colas pesadas.

Desde los años 70's la distribución de las series de rendimientos de instrumentos financieros está definida por

$$F_X(u) = P(r(t, \Delta t) \leq u)$$

Donde $X = r(t, T) = x(t + \Delta t) - x(t)$ es la función de los log-rendimientos, generalmente tomando a $\Delta t = t + 1$ y con función de densidad de probabilidad definida como $f_X = F'_X$.

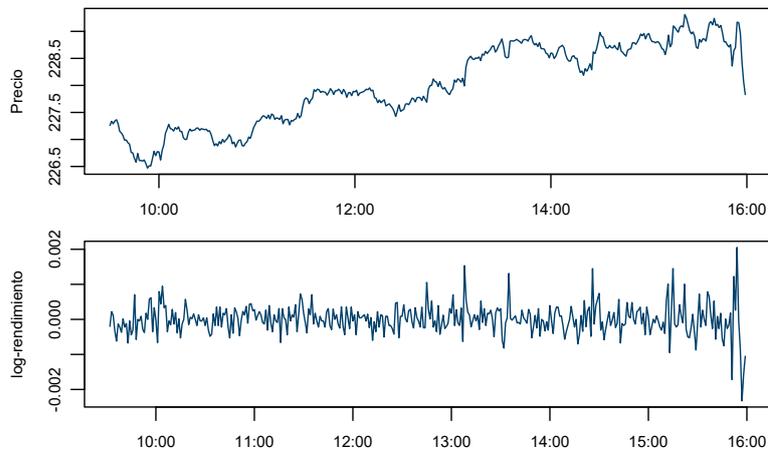


Figura 1.3: Precio de mercado de Apple con su rendimientos y log-rendimiento el 1 de octubre del 2018 entre las 9:00 am y 6:00 pm. Datos obtenidos de la base de datos TAQ.

Ya en los años 60 se destacó que la distribución normal no era adecuada para modelar la distribución de los rendimientos de los activos financieros, debido a la presencia de colas pesadas. Desde entonces, se ha observado de manera recurrente

que la distribución de los cambios de precios no sigue una distribución gaussiana en diferentes datos del mercado. Una forma de cuantificar la desviación de la distribución normal es usando la curtosis de la distribución F_X definida como

$$\kappa = \frac{\mathbb{E}[X_t^4]}{\mathbb{E}[X_t^2]^2} - 3$$

La curtosis se define de manera tal que $\kappa = 0$ para una distribución gaussiana, un valor positivo de κ indicaría una cola pesada.

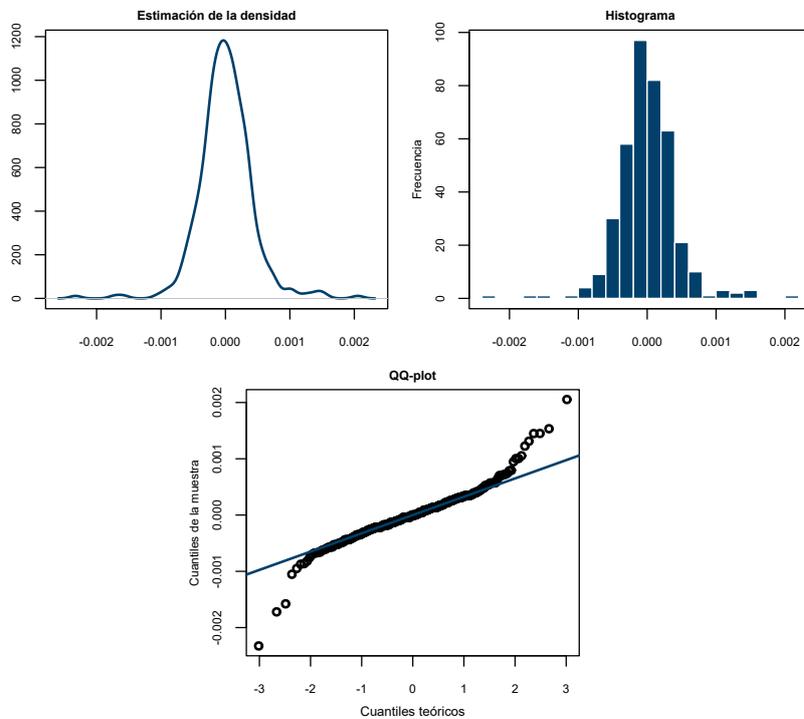


Figura 1.4: Estimación por kernel de la densidad de los log-rendimientos de Apple, donde a pesar de tener cierta forma de campana podemos observar gracias al QQplot que no se sigue una distribución normal ya que tenemos problemas en las colas. Gracias al histograma podemos observar que no son simétricas.

Uno puede resumir los resultados empíricos diciendo que la distribución de $f_{\Delta t}$ tiende a ser no gaussiana, leptocúrtica y de colas pesadas. Estas propiedades son más evidentes para los datos intradía $P(\Delta t < 1 \text{ día})$.

La Figura 1.4 da un ejemplo de la densidad del log-rendimiento de los precios por minuto de Apple durante un día mientras el mercado está abierto ilustrados en la Figura 1.3.

A menudo se argumenta que los rendimientos de los activos financieros son el resultado acumulativo de una gran cantidad de información y decisiones individuales que llegan casi a tiempo. Por lo tanto, en presencia de colas pesadas, es natural suponer que están aproximadamente gobernadas por una distribución no gaussiana.

Wood (1985) (Wood, [56]) muestra que los rendimientos de las acciones tienen distribuciones diferentes en diferentes momentos del día de negociación, siendo diferentes en los primeros 30 minutos del día, al cierre del mercado y durante el resto del día.

1.5.2 Autocorrelación

Goodhart y Figliuoli (1991, [56]) fueron los primeros en publicar la existencia de la autocorrelación negativa de primer orden en series de tiempo de rendimientos, que desaparece una vez que el precio está formado.

En la Figura 1.5 se puede observar la función de autocorrelación de una serie de rendimientos con retrasos de 1 minuto, podemos observar que la autocorrelación es significativa hasta el retraso del minuto 4. Más allá de este, la autocorrelación se encuentra dentro de un intervalo de confianza del 95 % de una distribución gaussiana i.i.d.

Goodhart y Figliuoli (1991, [56]) probaron que la autocorrelación negativa de los rendimientos se debe a lo que se denomina el **rebote de oferta y demanda**. De acuerdo con la explicación del rebote de oferta y demanda, la probabilidad de que una operación se ejecute al precio de oferta y luego sea seguida por una operación que se ejecute al precio de demanda es mayor que una operación al precio de oferta seguido de otra operación al precio de oferta.

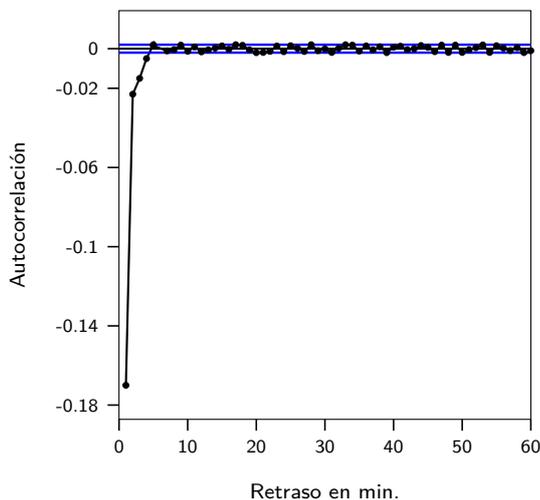


Figura 1.5: La imagen muestra la función de autocorrelación de los rendimientos de una acción para retrasos de tiempo en minutos dentro un intervalo de 1 hora. La autocorrelación es significativamente negativa hasta un retraso de 4 minutos. Las bandas azules representan un intervalo de confianza de 95 %

1.5.3 Estacionalidad

Los datos financieros de alta frecuencia a menudo presentan patrones estacionales fuertes y periódicos, incluyendo la volatilidad, frecuencia de negociación, volumen y diferenciales en diferentes mercados. Uno de los patrones más conocidos es el patrón de “forma de U ” de la negociación diaria, que se caracteriza por una alta volatilidad después de la apertura, seguida de una disminución durante el día y un aumento nuevamente antes del cierre. Como resultado, las series de rendimiento muestran una forma de U con dos picos que corresponden al comienzo y al final del día de negociación, y un fondo entre las 11:30 a.m. y la 1:30 p.m. Podemos ver un ejemplo de esto en la Figura 1.6.

La forma en U de la curva se debe a que al comienzo de un día de negociación, la información es absorbida por el mercado durante la noche y los participantes del mercado ajustan sus posiciones en la apertura, lo que aumenta la volatilidad. Durante el tiempo de descanso del día, hay menos transacciones y la volatilidad disminuye. Finalmente, antes del cierre, los participantes ajustan sus posiciones finales y aumentan la volatilidad nuevamente.

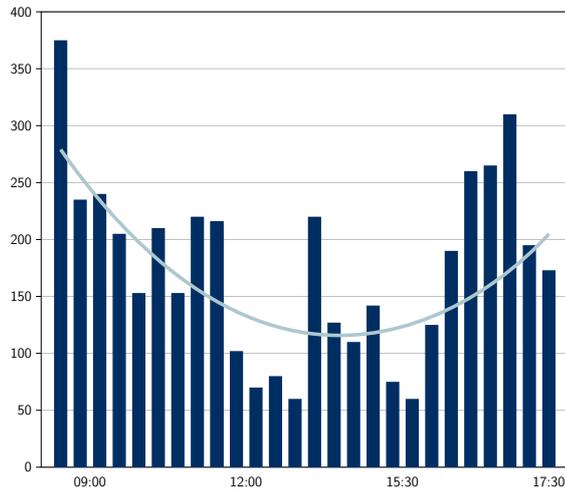


Figura 1.6: Actividad de ticks de las acciones de Deutsche Bank el 8 de mayo de 2006. El eje vertical reporta el número de ticks. Se puede observar una curva en forma de U durante el día de negociación.

1.5.4 Ley de escala

En finanzas se utilizan diferentes intervalos de tiempo para diferentes propósitos de análisis. No existe un intervalo de tiempo mejor en el que los datos deben estudiarse. En la práctica, una forma común de sobrellevarlo es convertir directamente los datos medidos en un horizonte temporal linealmente corto en un horizonte temporal más largo. Este enfoque es cuestionable porque ignora la interdependencia de los datos, lo cual es significativo en el análisis de alta frecuencia. ¿Podemos transferir los resultados de una escala de tiempo a otra sin distorsionarlos? Desde los primeros trabajos de Mandelbort (1967, [54]) sobre los precios del algodón, muchos estudios empíricos han confirmado la existencia de una ley de escala en una amplia gama de datos financieros: una relación directa entre los intervalos de tiempo y la volatilidad.

Aunque es aplicable para una amplia gama de intervalos de tiempo, la ley de escala también tiene su limitación cuando se usa para intervalos de tiempo largos, por lo que la aplicación universal de una ley de escala única también se cuestiona.

1.5.5 Agrupamiento

Muchas series de tiempo financieras muestran agrupaciones de volatilidad. En valores absolutos, se observa que los grandes valores de rendimientos son seguidos por rendimientos más grandes y rendimientos pequeños por rendimientos más pequeños. La agrupación de la volatilidad se manifiesta en los datos intradía.

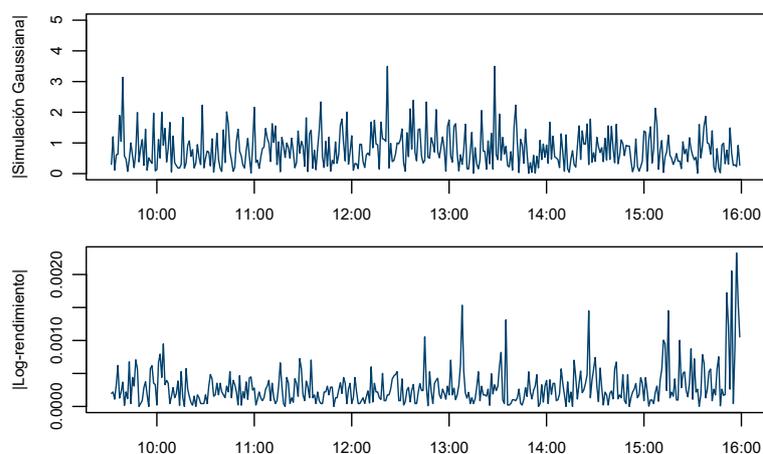


Figura 1.7: Serie de tiempo del valor absoluto log-rendimientos de Apple mostrados en la Figura 1.3 junto con los valores absolutos de una simulación gaussiana.

En la Figura 1.7 podemos apreciar como los log-rendimientos se comportan de forma más errática, teniendo periodos con picos pronunciados y periodos más suaves, mientras que la simulación gaussiana posee varianza constante.

Además de la agrupación de volatilidad, los datos de alta frecuencia muestran agrupación de cotizaciones y agrupación de duración. La agrupación de cotizaciones o precios es la preferencia para algunas cotizaciones/precios sobre otras. El agrupamiento de duración, o de tiempos entre llegadas, significa que las duraciones largas y cortas tienden a ocurrir en grupos.

1.6 Base de datos TAQ

Un ejemplo de una base de datos de una Bolsa de Valores es la base de datos TAQ. En 1990 la bolsa de Nueva York empezó a publicar datos de alta frecuencia y a partir de 1993 comenzó el desarrollo de la base de datos TAQ. La base de datos TAQ (*Trade and Quote*) proporciona información intradía para acciones negociadas en NYSE y NASDAQ-AMEX. Desde 2002, los datos del libro de pedidos también han estado disponibles por separado para fines de investigación.

Aunque a través del tiempo se han agregado muchas mejoras a la calidad de los datos, la NYSE no garantiza el grado de precisión de los datos, por lo que se necesitan más manipulaciones para usarlos en la investigación. La información incluida en esta base de datos consiste en:

- **Información comercial:** todas las operaciones, para todas las acciones negociadas en la Bolsa de Nueva York, NASDAQ-AMEX y bolsas regionales.
- **Información de cotización:** todas las mejores valuaciones de oferta y demanda publicadas por especialistas (en NYSE y AMEX) y por creadores de mercado (en NASDAQ) para todas las existencias.

La Tabla 1.4 contiene un ejemplo de datos comerciales sin procesar para NVDA (NVIDIA Corporación) de la base de datos TAQ.

SYMBOL	DATE	TIME	PRICE	SIZE	G127	CORR	COND
NVDA	20060901	09:30:04	28.67	264,509	0	0	-
NVDA	20060901	09:30:04	28.67	264,509	0	2	-
NVDA	20060901	09:30:05	28.65	100	0	0	-
NVDA	20060901	09:30:05	28.65	200	0	0	-
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Tabla 1.4: Estructura de la base de datos de NVDA. La tabla muestra un extracto de los datos comerciales brutos de la empresa Nvidia registrados en la base TAQ.

- **SYMBOL** o ticker: Denominación del stock que genera los datos.
- **DATE**: Fecha actual de los datos.
- **TIME**: Hora actual de los datos (resolución de 1 s).
- **PRICE**: Precio del comercio actual.
- **SIZE**: Tamaño del pedido
- **G127**: Indicador de comercio de acciones interrumpido.
- **CORR**: Indicador de corrección, puede tomar los valores:

- 0 a 2 Buenas operaciones (0: regular, 1: corregido, 2: símbolo de corrección).
- 7 a 9 Operaciones canceladas debido a errores o correcciones.
- 10 a 12 Instrucciones de corrección (10: cancelar, 11: error, 12: corrección).
- **COND**: Condición de venta, con posibles valores (no exhaustivos):
 - **A** Solo en efectivo
 - **B** precio medio de comercio (comercio agregado de órdenes de mismo precio dentro de 60s).
 - **O** apertura comercial informó más tarde.
 - **G** precio medio de comercio no informado con el intercambio de 90s límite.

Los datos de cotización contienen información sobre las mejores condiciones comerciales disponibles en el intercambio. En los archivos de precios de cierre de TAQ están disponibles cada día de negociación y muestran los precios al cierre, de apertura, el precio más alto, el más bajo, el volumen total, la oferta de cierre y demanda de cierre, etc. para todos los valores de NYSE, NYSE American, NYSE Arca, NYSE National y NYSE Chicago a partir del cierre del mercado.

Capítulo 2

Series de tiempo

Uno de los objetivos que plantea el análisis de datos de alta frecuencia es la creación de modelos que sean capaces de explicar el comportamiento empírico de los datos en cualquier frecuencia con la que se midan, ya sea segundos, minutos, horas, etc. Muchos de los procesos estadísticos propuestos pueden describirse mediante la siguiente fórmula general para series de rendimiento homogéneas r_t :

$$r_t = \sigma_t Z_t \tag{2.1}$$

donde Z_t es una v.a.i.i.d. con media cero y varianza 1 y tal que σ_t representa la volatilidad. Varios de los modelos convencionales se basan en la Ecuación 2.1, pero difieren en gran medida en el modelado de la variable σ_t ya que no se puede observar de manera directa, pero sí se puede estimar si se conoce una serie suficientemente larga de valores de rendimiento pasados. En estos casos, generalmente, se emplean los procesos ARCH (*Autoregressive Conditional Heteroskedasticity*, (Engle, R. [38])) y su versión generalizada GARCH, que son modelos con volatilidad estocástica donde se define la varianza σ_t^2 del rendimiento r_t en función de los rendimientos pasados.

En este capítulo, comenzaremos por dar una breve introducción al análisis de series de tiempo quedándonos dentro del marco del modelado tipo ARCH y explicaremos el concepto de la volatilidad intradía. Todas las definiciones y conceptos están basadas en *Introduction to Time Series and Forecasting* (P.J. Brockwell & R.A. Davis, [36]), *Time Series: Theory and methods* (P.J. Brockwell & R.A. Davis, [37]) y *Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation* (Engle R, [38]).

2.1 Conceptos básicos

En general, a la colección de variables aleatorias $\{X_t\}$ indexadas por $t \in T$ se conoce como un proceso estocástico.

Definición 1

Una **serie de tiempo** es un proceso estocástico indexado al tiempo, es decir, una familia de variables aleatorias, sobre un espacio de probabilidad (Ω, \mathcal{F}, P) :

$$\{X_t; t \in T\}$$

comúnmente tomamos $T = \mathbb{Z}$ o solo $T = \mathbb{N}$.

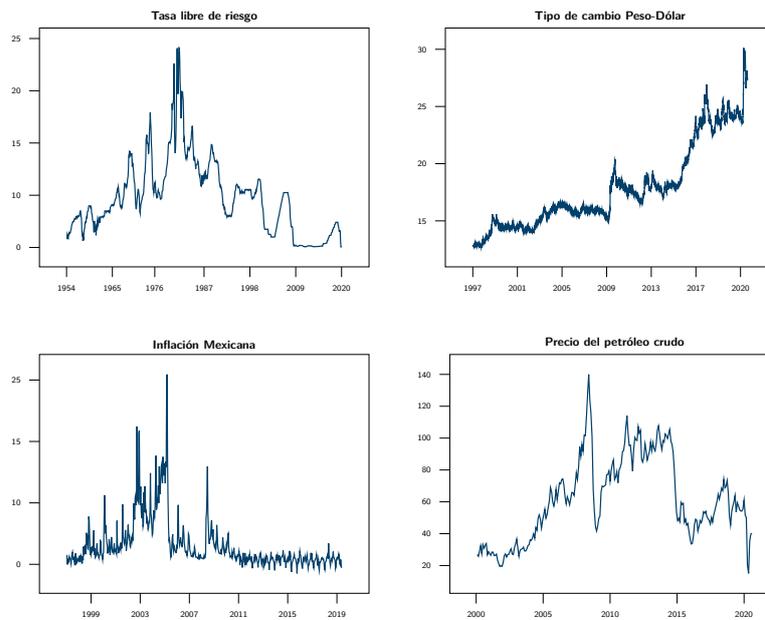


Figura 2.1: Ejemplos de series de tiempo financieras. La primer gráfica muestra la tasa de interés de fondos federales de Estados Unidos con una frecuencia mensual a partir de 1954. La segunda representa el tipo de cambio peso-dólar con una frecuencia diaria entre el 2-01-1997 hasta el 17-07-2020. La tercer serie muestra la tasa de inflación mexicana registrada de forma mensual entre enero de 1970 y mayo del 2019. La cuarta gráfica muestra el precio mensual del petróleo crudo a partir de enero del 2020 y hasta julio del 2020.

Una serie de tiempo es aquella en la que el conjunto T_0 de veces en que se realizan las observaciones es un conjunto discreto, como es el caso, por ejemplo, cuando las observaciones se realizan a intervalos de tiempo fijos como días, meses, años, etc.¹

Definición 2

Un modelo de series de tiempo para los datos observados $\{x_t\}$ es una especificación de las distribuciones conjuntas (o posiblemente sólo las medias y covarianzas) de la secuencia de variables aleatorias $\{X_t\}$ de las cuales $\{x_t\}$ se postula como una realización.

El concepto de estacionariedad es de suma importancia para el análisis de series de tiempo; se dice que una serie de tiempo

$$\{X_t, t \in \mathbb{Z}\}$$

es estacionaria si tiene propiedades estadísticas similares a las de la serie “desplazada en el tiempo”

$$\{X_{t+h}, t \in \mathbb{Z}\}$$

para cada número entero h . Antes de estudiar este concepto, recordemos algunas definiciones.

Definición 3

Sea $\{X_t\}$ una serie de tiempo con $\mathbb{E}[X_t^2] < \infty$ y para todo entero r y s

- La **función media** de $\{X_t\}$ está dada por:

$$\mu_X(t) = \mathbb{E}[X_t]$$
- La **función de autocovarianza** de $\{X_t\}$ es

$$\gamma_X(r, s) = \text{Cov}(X_r, X_s) = \mathbb{E}[(X_r - \mu_X(r))(X_s - \mu_X(s))]$$

Para poder caracterizar un proceso estocástico, casi siempre, es necesario conocer la función de densidad conjunta de las variables aleatorias involucradas; sin embargo en la práctica esto no es muy realista. No obstante, puesto que los primeros momentos de las variables aleatorias resumen en buen medida su distribución, se ha sugerido estudiar al primer y segundo momento, es decir, a las medias, varianzas y covarianzas de las variables involucradas.

¹Un proceso estocástico sobre el tiempo se obtiene cuando las observaciones se registran continuamente durante un intervalo de tiempo, por ejemplo, cuando $T_0 \in [0, 1]$.

Una forma de estacionariedad es cuando pedimos estacionariedad de las características numéricas de la serie.

Definición 4

Diremos que una serie de tiempo, cuyas variables tengan segundo momento finito, es **débilmente estacionaria** (estacionaria de segundo orden) si:

$$\begin{aligned}\mu_X(t) &= \mu \\ \gamma_X(t, s) &= \gamma(t+h, s+h)\end{aligned}$$

Para todos $t, s, h \in \mathbb{Z}$

Si nuestra serie en cuestión es tal que la distribución de cualquier vector $(X_{t_1}, \dots, X_{t_n})$ es invariante bajo traslaciones, se dice que es **estrictamente estacionaria**.

Definición 5

Diremos que una serie de tiempo, $(X_t)_{t \in \mathbb{Z}}$ es **estrictamente estacionaria** si

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+k}, \dots, X_{t_n+k})$$

para cualesquiera valores $t_1, \dots, t_n \in \mathbb{Z}$ con $n \in \mathbb{N}$, donde $\stackrel{d}{=}$ significa igualdad en distribución.

Podemos ver, directamente de las definiciones, que si una serie de tiempo es estrictamente estacionaria y con varianza finita, entonces es débilmente estacionaria. De ahora en adelante, si decimos que un proceso cumple la propiedad de estacionariedad, nos referiremos a que es débilmente estacionario. Para las series débilmente estacionarias se tiene:

$$\begin{aligned}\gamma(t-s, 0) &= \gamma(t, s) \\ &= \gamma(s, t) \\ &= \gamma(s-t, 0)\end{aligned}$$

Por lo que concluimos que la autocovarianza sólo va a depender de la separación de las variables (el *lag* o rezago). Para este tipo de series definiremos la función de autocovarianza como:

$$\gamma_X(h) = \gamma_X(h, 0)$$

que es el valor de la función en el retraso h .

La autocovarianza mide la dependencia lineal entre dos puntos en la mismas series pero observadas en diferentes momentos de esta. Está claro que, para $s = t$, la autocovarianza se reduce a la **varianza** (supuesta finita), porque

$$\gamma_t(s, s) = \mathbb{E} [(X_t - \mu_t)^2] = \text{var}(X_t)$$

Quizás el modelo más simple para una serie de tiempo es uno tal que las observaciones son simplemente variables aleatorias independientes e idénticamente distribuidas con media cero.

Definición 6

Diremos que $\{Z_t\}_{t \in T}$ es un **ruido blanco** si es un proceso débilmente estacionario y su función de covarianza es:

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2 & \text{si } h = 0 \\ 0 & \text{si } h \neq 0 \end{cases}$$

Notación: $\{Z_t\}$ es un $WN(\mu, \sigma^2)$ donde μ es la media del proceso y σ^2 su varianza.

Las gráficas de las series de ruido blanco exhiben comportamiento errático e impredecible, como se puede ver en la Figura 2.3. En las series de tiempo financieras, a menudo se considera que el ruido blanco representa **innovaciones**. Un ruido blanco centrado es aquel que tiene media 0, es decir, $Z_t \sim WN(0, \sigma^2)$.

Definición 7

Un proceso estocástico $(Z_t)_{t \in \mathbb{Z}}$ es un **ruido blanco estricto** si es una sucesión de variables aleatorias independientes e idénticamente distribuidas con varianza finita. Notación: Z_t es un $SWN(\mu, \sigma^2)$, μ es la media y σ^2 la varianza.

Es decir, un proceso estocástico es un ruido blanco estricto si las variables aleatorias $Z(t_1)$, $Z(t_2)$ son independientes y no hay ninguna relación, ni lineal, ni no lineal, entre ellas.

Ejemplo:

La Figura 2.2 muestra un ejemplo de una **caminata aleatoria** $\{S_t, t = 0, 1, 2, \dots\}$, que es obtenida a través de la suma acumulativa de variables aleatorias independientes e idénticamente distribuidas. Por lo tanto, una caminata aleatoria,

donde $\{Z_t\}$ es un ruido blanco, se define como

$$S_n = S_{n-1} + Z_n$$

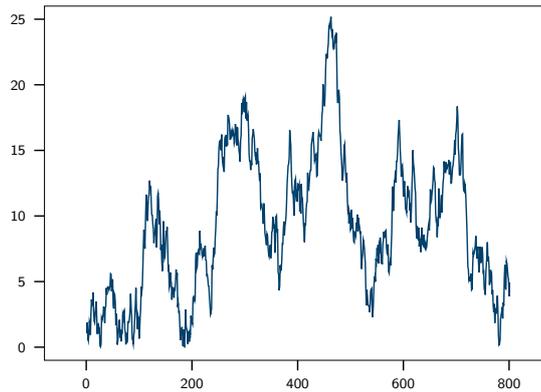


Figura 2.2: Simulación de una caminata aleatoria.

Para evaluar el grado de dependencia de los datos y seleccionar un modelo que refleje esto, una de las herramientas importantes que utilizamos es la muestra de la función de autocorrelación.

Definición 8

Definiremos a la **función de autocorrelación** (ACF) de $\{X_t\}$, $t = 0, 1, 2, 3, \dots, T$ en el rezago h como

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} \text{ para toda } h \in \mathbb{Z}$$

La función de autocorrelación parcial (PACF) mide la correlación entre los valores de la serie distanciados un retraso de tiempo h .

Definición 9

La **función de autocorrelación parcial** (Partial Autocorrelation Function) o PACF de un proceso $\{X_t\}$ está definido por la función $\alpha(\cdot)$ tal que

- $\alpha(0) = 1$
 - $\alpha(h) = \phi_{hh}, h \geq 1$
- donde ϕ_{hh} es el último componente de $\phi_h = \Gamma_h^{-1}\gamma_h$ tal que

$$\Gamma_h = [\gamma(i-j)]_{i,j=1}^h \text{ y } \gamma_h = [\gamma(1), \gamma(2), \dots, \gamma(h)]'$$

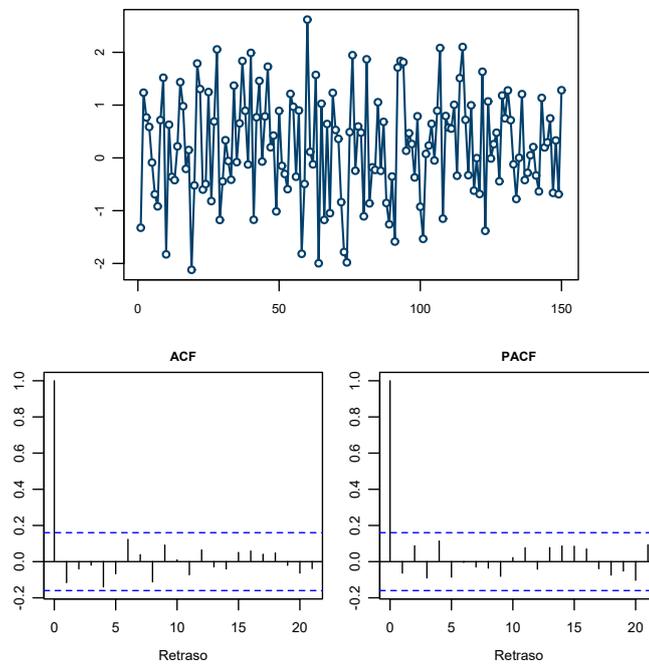


Figura 2.3: Simulación de un ruido blanco con $n = 150$. Podemos observar del ACF y el PACF que no existen retrasos importantes.

Un concepto importante que vamos a introducir es el *operador de retraso*. Los modelos de series de tiempo suelen definirse en términos de éste.

Definición 10

Sea $(X_t)_{t \in \mathbb{Z}}$ una serie de tiempo, entonces, el **operador de retraso** B es tal

que

$$BX_t = X_{t-1}, \quad B^2X_t = B(BX_t) = X_{t-2}, \dots$$

para toda $t \in \mathbb{Z}$

Otro concepto que usaremos después es la diferencia-martingala. Para estudiarlo debemos suponer que la serie de tiempo $(X_t)_{t \in \mathbb{Z}}$ es adaptada a cierta filtración² $(\mathcal{F}_t)_{t \in \mathbb{Z}}$ que representa la información obtenida hasta el momento t .

Definición 11

Diremos que un proceso $(X_t)_{t \in \mathbb{Z}}$ es una **diferencia martingala**, respecto a una filtración $(\mathcal{F}_t)_{t \in \mathbb{Z}}$ si:

- Es integrable ($\mathbb{E}[|X_t|] < \infty$),
- Adaptada a la filtración (X_t es v.a. respecto a \mathcal{F}_t) y
- $\mathbb{E}[X_t | \mathcal{F}_{t-1}] = 0$

para toda $t \in \mathbb{Z}$.

Lo más común es tomar la filtración natural del proceso, esto es,

$$\mathcal{F}_t = \sigma(X_s : s \leq t)$$

Esta filtración nos da la información generada por lo que ya ha ocurrido en el proceso que estamos estudiando y es la mínima σ -álgebra que contiene esa información.

Podemos ver directamente que:

$$\mathbb{E}[X_t] = 0, \quad \forall t \in \mathbb{Z}.$$

Además si $\mathbb{E}[X_t^2] < \infty$, podemos dar información sobre la función de autocova-

²Una filtración es una familia de σ -álgebras, $(\mathcal{F}_t)_{t \in \mathbb{Z}}$, tales que $\mathcal{F}_n \subseteq \mathcal{F}_m$ cuando $n \leq m$.

rianza (sin pérdida de generalidad supongamos que $s < t$):

$$\begin{aligned}\gamma_{(t,s)} &= \mathbb{E}[X_t X_s] \\ &= \mathbb{E}[\mathbb{E}[X_t X_s | \mathcal{F}_{s-1}]] \\ &= \mathbb{E}[X_t \mathbb{E}[X_s | \mathcal{F}_{s-1}]] \\ &= 0.\end{aligned}$$

Por lo que si una diferencia-martingala tiene varianza constante en el tiempo, entonces será un ruido blanco.

2.1.1 Series de tiempo no homogéneas

Un análisis adecuado de datos de alta frecuencia se basa en definiciones explícitas de las variables en estudio. Para la mayoría de los métodos, estas series sin procesar no son adecuadas para trabajar, porque las ticks³ del mercado llegan en momentos aleatorios, para solucionar esto, se utilizan métodos de interpolación o diferentes escalas de tiempo.

VARIABLES BÁSICAS como el rendimiento y la volatilidad realizada se definen a continuación. También hablaremos brevemente de la frecuencia de la tick, ya que esta sirve para conocer la dinámica del mercado intradía.

Interpolación

Antes de definir diferentes variables, se debe explicar el concepto de una serie de tiempo homogénea. Una serie de tiempo homogénea, es un conjunto de datos numéricos que se obtienen en periodos regulares a través del tiempo bajo cualquier horizonte temporal que sea de nuestro interés. Como ya hemos mencionado, esto no sucede con datos de frecuencias altas. Es por eso que se deben emplear métodos para obtener una serie homogénea z_j a partir de la original (no homogénea) t_j , es decir:

$$z_j = z(t_j)$$

El índice j se refiere a la secuencia espaciada por irregularidades de la serie original. Al utilizar un método de interpolación⁴, construimos una serie de tiempo homogénea con valores en los tiempos $t_0 + i\Delta t$, espaciados regularmente por Δt ,

³Una tick es una medida del movimiento mínimo hacia arriba o abajo del precio del valor.

⁴Otra manera de enfrentar este problema, es por medio de modelos de duración.

arraigados en el tiempo t_0 . El índice i se refiere a la serie homogénea.

El tiempo $t_0 + i\Delta t$ está entre corchetes por dos veces t_j de la serie bruta

$$j' = \max(j | t_j \leq t_0 + i\Delta t), \quad t_{j'} \leq t_0 + i\Delta t < t_{j'+1}$$

Interpolamos entre $t_{j'}$ y $t_{j'+1}$. Los dos métodos de interpolación más importantes son la interpolación lineal

$$z(t_0 + i\Delta t) = z_{j'} + \frac{t_0 + i\Delta t - t_{j'}}{t_{j'+1} - t_{j'}}(z_{j'+1} - z_{j'})$$

y la extrapolación de la tick previa (tomando el valor más reciente),

$$x(t_0 + i\Delta t) = z_{j'}$$

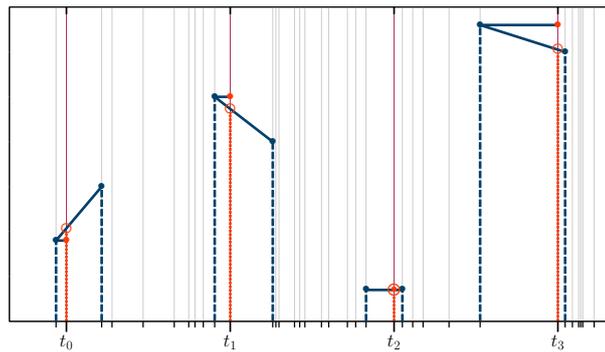


Figura 2.4: Métodos de interpolación para obtener una serie de tiempo homogénea: seleccionar valores en puntos igualmente espaciados en el tiempo t_i , indicados por líneas rojas verticales punteadas. La secuencia de tiempo no homogénea de las observaciones sin procesar se indican por las marcas debajo del eje x y con líneas verticales grises. Dos métodos de interpolación importantes se ilustran: interpolación lineal (círculos rojos grandes y vacíos) y la extrapolación de ticks previas (círculos rojos y pequeños).

Ahora ya podemos introducir algunas de las variables usadas en el análisis de datos de alta frecuencia.

Precio

Los precios de los activos son las variables más importantes exploradas en finanzas. Dependiendo de la estructura del mercado y el proveedor de datos, los

precios están disponibles como cotizaciones en diferentes formas:

- Precios de oferta-demanda en pares: p_{oferta} y p_{demanda} .
- Precios de transacción (que pueden ser o no cotizaciones anteriores de oferta o demanda).
- Oferta, demanda, precios de transacción en secuencia irregular (no en pares).
- Precios medios.

Una observación individual en el momento t_j , también en el caso de precios de oferta y demanda, se denomina tick.

Rendimiento

El rendimiento o retorno en el momento t_i , se define como

$$r(t_i) = r(\Delta t; t_i) = \ln(P_t) - \ln(P_{\Delta t}) \quad (2.2)$$

donde Δt es un intervalo de tiempo de tamaño fijo. En el caso normal, Δt es el intervalo de la serie homogénea y $r(t_i)$ es la serie de las primeras diferencias de $\ln(P_{t_i})$.

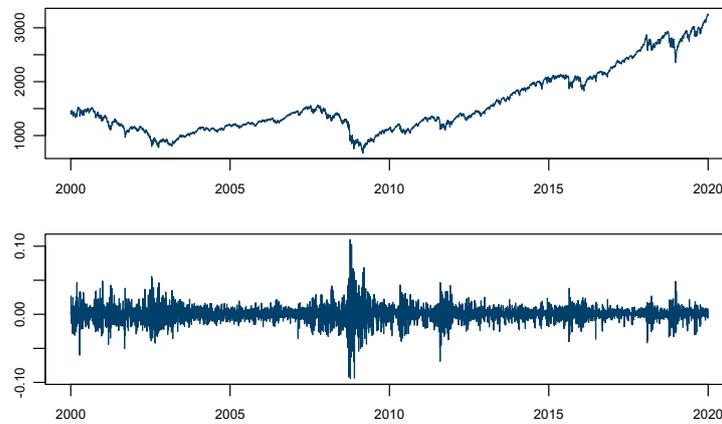


Figura 2.5: Las gráficas muestra el índice y el rendimiento del Standard & Poor's 500, también conocido como S&P 500, desde el 2000-01-03 con frecuencia diaria respectivamente. El índice se basa en la capitalización bursátil de 500 grandes empresas que poseen acciones que cotizan en las bolsas NYSE o NASDAQ, y captura aproximadamente el 80 % de toda la capitalización de mercado en Estados Unidos.

El rendimiento suele ser una variable de análisis más adecuada que el precio, por varias razones. Es la variable de interés comúnmente utilizada como medida directa del éxito de una inversión. Además, la distribución de rendimientos es más simétrica y estable en el tiempo que la distribución de precios. Los precios de los rendimientos son casi estacionarios, mientras que el proceso de precios no lo es. Este hecho puede verse en la Figura 2.5.

Volatilidad Realizada

La volatilidad realizada $v(t_i)$ en el momento t_i se calcula a partir de datos históricos y también se denomina volatilidad histórica. Se define como

$$v(t_i) = v(\Delta t, n, p; t_i) = \left[\frac{1}{n} \sum_{j=1}^n |r(\Delta t : t_{i-n+j})|^p \right]^{\frac{1}{p}}$$

donde los rendimientos regularmente espaciados r fueron definidos en la Ecuación 2.2 y n es el número de observaciones de retorno. En muchos casos, se prefiere un valor de $p = 1$, aunque p también puede ser una fracción, $p > 0$

Frecuencia *tick*

Como ya habíamos discutido, el mercado fija una cuadrícula en la que los comerciantes pueden colocar sus precios. El paso de la cuadrícula es el intervalo más pequeño entre dos precios y se llama valor de la *tick*⁵. Para un valor dado, es seguro considerar que esta cuadrícula está espaciada uniformemente aunque el mercado pueda cambiarla a veces.

La frecuencia *tick* en el momento t_i , $f(t_i)$ se define como

$$f(t_i) = f(\Delta t; t_i) = \frac{1}{\Delta t} N\{x(t_j) | t_i - \Delta t < t_j \leq t_i\}$$

Donde $N\{x(t_j)\}$ es la función de conteo y Δt es el tamaño del intervalo de tiempo en el que se cuentan las *ticks*.

De manera más simple, el **tamaño tick** es el incremento más pequeño por el cual el precio puede moverse, la **frecuencia tick** es el parámetro que define la

⁵Se mide en la moneda del activo.

cantidad de nuevos precios que ingresan a una plataforma de negociación. Cuanto mayor sea el número, mejores condiciones para los comerciantes. Una frecuencia tick alta reduce la devaluación y aumenta el cumplimiento de pedidos a precios objetivo, lo cual es vital para todos los participantes del mercado.

2.2 Series de Tiempo para datos de Alta Frecuencia

A continuación presentaremos a los modelos tipo ARCH y GARCH y su uso en series de tiempo de alta frecuencia con datos regularmente espaciados en el tiempo construidos a partir de datos intradía. Esto es lo que nos permite caracterizar la volatilidad intradiaria de los activos financieros. Destacamos la necesidad de tener en cuenta la estacionalidad intradiaria de la volatilidad antes de estimar los modelos GARCH.

D

2.3 Procesos ARCH

El modelo autorregresivo de heterocedasticidad condicional (ARCH) de Engle (1982) y su versión generalizada (GARCH) de Bollerslev (1986) son ampliamente utilizados, no solo en la literatura cambiaria (FX) sino también como el marco básico de estudios empíricos de la economía, microestructura del mercado, como el impacto de las noticias y las intervenciones del banco central, o las relaciones entre mercados.

En la subsección anterior, se ha argumentado que un modelo adecuado para los rendimientos de los activos debe incluir un término de error no correlacionado, pero aún debe permitir la dependencia. El modelo ARCH tiene como objetivo abordar este problema mediante la introducción de dependencia temporal entre los cuadrados de los términos de error. Dicho proceso se ha convertido en una de las herramientas más importantes en el análisis de datos de series de tiempo. Demos una breve introducción del proceso.

Definición 12

Sea Z_t un $SWN(0, 1)$. Decimos que X_t es **un proceso ARCH**(p) si es estrictamente estacionario y satisface (para algún proceso σ_t):

$$X_t = \sigma_t Z_t \tag{2.3}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 \quad (2.4)$$

para $t \in \mathbb{Z}$, $\alpha_0 > 0$ y $\alpha_i \geq 0$, $i = 1, 2, \dots, p$.

Es importante que el proceso σ_t sea un proceso positivo ya que podemos notar fácilmente, de la Definición 7 de $SWN(0, 1)$ y tomando nuevamente a $\mathcal{F}_t = \sigma(X_s : s \leq t)$, que

$$\mathbb{E}[X_t | \mathcal{F}_{t-1}] = \sigma_t \mathbb{E}[Z_t | \mathcal{F}_{t-1}] = \sigma_t \mathbb{E}[Z_t] = 0$$

$$\text{Var}[X_t | \mathcal{F}_{t-1}] = \sigma_t^2 \mathbb{E}[Z_t^2 | \mathcal{F}_{t-1}] = \sigma_t^2$$

Tenga en cuenta que la independencia de Z_t y \mathcal{F}_{t-1} que hemos asumido anteriormente se deriva del hecho de que un proceso ARCH debe ser causal, es decir, las Ecuaciones 2.3 y 2.4 deben tener una solución de la forma $X_t = f(Z_t, Z_{t-1}, \dots)$ para alguna f , de modo que Z_t sea independiente de los valores previos del proceso⁶.

Los coeficientes X_t deben satisfacer algunas condiciones de regularidad para garantizar que la varianza incondicional de X_t sea finita. En la práctica, a menudo se supone que Z_t sigue la distribución normal estándar o una distribución t -Student estandarizada o alguna distribución de error generalizada. A partir de la estructura del modelo, se ve que los grandes valores de X_t^2 del pasado implican una gran varianza condicional σ_t^2 para la innovación en X_t . En consecuencia, X_t tiende a asumir un valor grande. Esto significa que, en el marco del ARCH, los grandes valores de X_t , o choques, tienden a ir seguidos de otro gran choque. Aquí uso la palabra “*tender*” porque una gran variación no necesariamente produce una gran realización. Solo dice que la probabilidad de obtener una variante grande es mayor que la de una varianza menor. Esta característica es similar a las agrupaciones de volatilidad observadas en el rendimiento de los activos.

ARCH(1)

El proceso ARCH(1) es la forma más simple de la familia ARCH. Usando la Definición 12 tenemos que el proceso ARCH(1) puede ser escrito como:

$$X_t^2 = \alpha_0 Z_t^2 + \alpha_1 Z_t^2 X_{t-1}^2$$

⁶Algunos autores usan h_t para denotar la varianza condicional en la Ecuación 2.4, en este caso $X_t = \sqrt{h_t} Z_t$.

En la Figura 2.6 podemos ver una simulación de un proceso ARCH(1), con $\alpha_1 = 0.4$ y $\alpha_0 = 0$. Para garantizar la positividad de σ_t^2 , asumimos que $\alpha_0 > 0$ y $\alpha_1 \geq 0$. Primero, la media incondicional de X_t permanece cero ya que

$$\mathbb{E}[X_t] = \mathbb{E}[\mathbb{E}(X_t|\mathcal{F}_{t-1})] = [\sigma_t \mathbb{E}(Z_t)] = 0$$

Segundo, la varianza incondicional de X_t puede ser obtenida de la siguiente manera:

$$\begin{aligned} \text{Var}(X_t) &= \mathbb{E}[X_t^2] = \mathbb{E}[\mathbb{E}(X_t^2|\mathcal{F}_{t-1})] \\ &= \mathbb{E}[\alpha_0 + \alpha_1 X_{t-1}^2] \\ &= \alpha_0 + \alpha_1 \mathbb{E}[X_{t-1}^2] \end{aligned}$$

Como X_t es un proceso estacionario con $\mathbb{E}[X_t] = 0$,

$$\text{Var}(X_t) = \text{Var}(X_{t-1}) = \mathbb{E}(X_{t-1}^2)$$

Por lo tanto, tenemos que

$$\text{Var}(X_t) = \alpha_0 + \alpha_1 \text{Var}(X_t)$$

y

$$\text{Var}(X_t) = \frac{\alpha_0}{1 - \alpha_1}$$

Dado que la varianza de X_t debe ser positiva, requerimos $0 \leq \alpha_i < 1$. En tercer lugar, en algunas aplicaciones, necesitamos momentos de orden superior de X_t y, por lo tanto, α_1 también debe satisfacer algunas restricciones adicionales. Por ejemplo, para estudiar el comportamiento de su cola, necesitamos que el cuarto momento de X_t sea finito. Bajo el supuesto de normalidad de Z_t en la Ecuación 2.3, tenemos

$$\mathbb{E}[X_t^4|\mathcal{F}_{t-1}] = 3[\mathbb{E}(X_t^2|\mathcal{F}_{t-1})]^2 = 3(\alpha_0 + \alpha_1 X_{t-1}^2)^2$$

Entonces,

$$\mathbb{E}[X_t^4] = \mathbb{E}[(X_t^4|\mathcal{F}_{t-1})] = 3\mathbb{E}(\alpha_0 + \alpha_1 X_{t-1}^2)^2 = 3\mathbb{E}(\alpha_0^2 + 2\alpha_0\alpha_1 X_{t-1}^2 + \alpha_1^2 X_{t-1}^4)$$

Si X_t es estacionario de cuarto orden con $m_4 = \mathbb{E}[X_t^4]$, entonces tendremos

$$\begin{aligned} m_4 &= 3[\alpha_0^2 + 2\alpha_0\alpha_1 \text{Var}(X_t) + \alpha_1^2 m_4] \\ &= 3\alpha_0^2 \left(1 + 2\frac{\alpha_1}{1 - \alpha_1}\right) + 3\alpha_1^2 m_4 \end{aligned}$$

Por consiguiente

$$m_4 = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}$$

Este resultado tiene dos implicaciones importantes: (a) dado que el cuarto momento de X_t es positivo, vemos que α_1 también debe satisfacer la condición $1 - 3\alpha_1^2 > 0$; es decir, $0 \leq \alpha_1^2 < \frac{1}{3}$; y (b) la curtosis incondicional de X_t es

$$\frac{[X_t^4]}{[\text{Var}(X_t)]^2} = 3 \frac{\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)} \times \frac{(1 - \alpha_1^2)^2}{\alpha_0^2} = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3$$

Por tanto, el exceso de curtosis de X_t es positivo y la distribución de la cola de X_t es más pesada que la de una distribución normal.

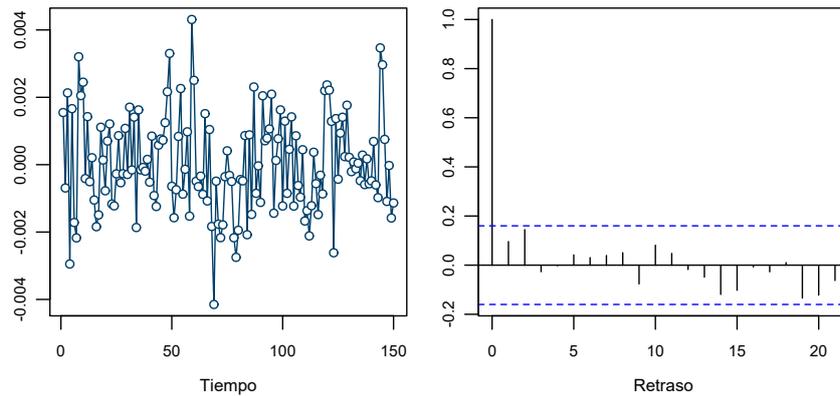


Figura 2.6: Simulación de un proceso ARCH(1) con su ACF.

```

1 library(fGarch)
2 spec = garchSpec(model=list(alpha=c(0.2,0.4), beta=0))
3 plot(garchSim(spec, n=150))
4 acf(garchSim(spec, n=150))

```

Código 2.1: La librería `fGarch` proporciona la función `garchspec` que nos permite especificar un modelo GARCH univariado y la función `garchSim` que simula un modelo de serie de tiempo GARCH univariado.

En otras palabras, es más probable que el impacto de un modelo ARCH(1) gaussiano condicional produzca “valores atípicos” que una serie de ruido blanco gaussiano. Esto está de acuerdo con el hallazgo empírico de que los “valores atípicos” aparecen con más frecuencia en los rendimientos de los activos que los que implica una secuencia iid de variables aleatorias normales. Estas propiedades continúan siendo válidas para los modelos ARCH generales, pero las fórmulas se vuelven más complicadas para los modelos ARCH de orden superior.

2.4 Procesos GARCH

Tim Peter Bollerslev en 1986 hace una de las principales aportaciones a esta nueva familia de modelos al hacer una generalización del modelo de Engle.

Definición 13

Sea Z_t un $SWN(0, 1)$. Decimos que X_t es **un proceso GARCH**(p, q) si es estrictamente estacionario y satisface (para algún proceso σ_t):

$$X_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

para $t \in \mathbb{Z}$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, 2, \dots, p$ y $\beta_j \geq 0$, $j = 1, 2, \dots, q$

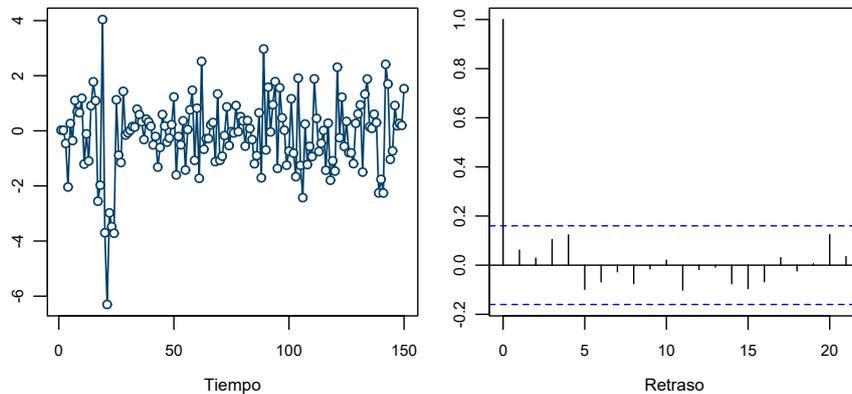


Figura 2.7: Simulación de un proceso GARCH(1,1) con $\alpha_1 = 0.4$, $\beta_1 = 0.2$ y $\alpha = 0.5$ y su ACF.

```

1 spec = garchSpec(model=list(alpha=c(0.5,0.4), beta=0.2))
2 plot(garchSim(spec, n=150))
3 acf(garchSim(spec, n=150))

```

Código 2.2: Podemos observar que a diferencia del código para simular un ARCH(1), aquí asignamos un valor a beta.

Los procesos GARCH son procesos ARCH generalizados; ahora, la volatilidad en algún tiempo va a depender de las volatilidades anteriores, así como de los valores anteriores del proceso (como el caso del ARCH).

Las fortalezas y debilidades de los modelos GARCH se pueden ver fácilmente centrándose en el modelo GARCH(1,1):

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad 0 \leq \alpha_1, \beta_1 \leq 1, (\alpha_1 + \beta_1) < 1$$

El modelo GARCH(p, q) fue una de las primeras generalizaciones del modelo ARCH después de la publicación de Engle (1982). En los años siguientes, siguieron muchos otros, por lo general involucrando especificaciones más complejas de la varianza condicional.

2.4.1 Extensiones del modelo GARCH

En el contexto de datos financieros de alta frecuencia, los modelos GARCH se utilizan a menudo para modelar la volatilidad intradía de los rendimientos financieros, que pueden ser muy difíciles de predecir. Los datos financieros de alta frecuencia pueden contener mucho ruido y variaciones aleatorias, y los modelos GARCH pueden ayudar a filtrar parte de este ruido e identificar patrones significativos en los datos.

Una de las ventajas de los modelos GARCH es que pueden capturar el fenómeno de agrupamiento de volatilidad, que se refiere a la tendencia de los rendimientos financieros a exhibir períodos de alta volatilidad seguidos de períodos de baja volatilidad. Esto es particularmente importante en los datos financieros de alta frecuencia, que, como ya hemos discutido, pueden mostrar cambios rápidos en la volatilidad durante períodos cortos de tiempo.

Hay varios tipos diferentes de modelos GARCH que se pueden usar para datos financieros de alta frecuencia, incluido el modelo GARCH tradicional, el modelo GARCH exponencial (EGARCH), el modelo GARCH de potencia (PGARCH o

Power GARCH) o el modelo GARCH integrado (IGARCH). Cada uno de estos modelos tiene sus propias fortalezas y debilidades, y la elección del modelo dependerá de las características específicas de los datos que se analizan.

Veamos el caso del modelo EGARCH.

Definición 14

Sea Z_t un SWN(0,1), el modelo GARCH Exponencial general (EGARCH) tiene como varianza condicional la siguiente expresión:

$$\log(\sigma_t^2) = \alpha_t + \sum_{k=1}^{\infty} \beta_k g(Z_{t-k})$$

donde α_t son coeficientes deterministas y

$$g(Z_t) = \theta Z_t + \gamma (|Z_t| - \mathbb{E}[|Z_t|])$$

De donde vemos que $\mathbb{E}(g(Z_t)) = 0$.

Este modelo permite una respuesta asimétrica de σ_t^2 a los choques de volatilidad en el término de innovación Z_{t-1} cuando θ difiere de cero. La derivada de $\ln \sigma_t^2$ con respecto a Z_{t-1} es igual a $\gamma - \theta$ cuando Z_{t-1} es negativo y es igual $\gamma + \theta$ cuando Z_{t-1} es positivo. Estudios empíricos realizados con datos diarios utilizando especificaciones EGARCH, generalmente encuentran que $\gamma + \theta$ es positivo mientras que θ es negativo, por lo que los choques negativos tienen un mayor impacto pronunciado sobre la volatilidad que los choques positivos. Esto conduce a una curva asimétrica de impacto de noticias (el gráfico de $\ln h_t$ sobre ϵ_{t-1}), con una pendiente mayor cuando la innovación es negativa que cuando es positiva (Engle y Ng, 1993).

2.5 Un modelo para la volatilidad intradía

Como ya habíamos discutido la volatilidad de rendimientos intradía no es constante durante el día de negociación. Más precisamente, se ha encontrado un patrón intradía, donde los rendimientos son mucho más variables al comienzo y al final del día que alrededor de la hora del almuerzo. Por tanto, la varianza de los rendimientos sigue una función en forma de U . Si se supone que el patrón de volatilidad

intradía sigue una función determinista del tiempo, se puede medir calculando promedios (por ejemplo, en intervalos de 30 minutos) de los rendimientos al cuadrado registrados durante estos intervalos. Una segunda posibilidad es suponer que la volatilidad intradía se compone de dos componentes distintos: en primer lugar, un patrón determinista intradía como el definido anteriormente; en segundo lugar, una función estocástica que modela la parte restante de la volatilidad, es decir, la parte no capturada por la función determinista. Este es el método utilizado en Andersen y Bollerslev (1997, [54]), donde la función estocástica pertenece a la clase GARCH. Se necesitan tres pasos para implementar este procedimiento:

- (1) Se calcula un patrón intradía determinista, por ejemplo, utilizando el método descrito anteriormente. Por lo tanto, para todos los intervalos de un minuto, calculamos los rendimientos cuadrados promedio (previamente definidos en la cuadrícula de muestreo) teniendo en cuenta un efecto de tiempo de la semana, es decir, que la volatilidad promedio depende de la hora del día y del día de la semana.
- (2) Rendimientos desestacionalizados rd_i ; se calculan dividiendo los rendimientos por la raíz cuadrada de la función determinista de volatilidad intradía:

$$rd_i = r_i / \sqrt{\phi(t_i)}$$

- (3) Los modelos GARCH se especifican para capturar la posible dependencia en los rendimientos desestacionalizados.

En resumen, los rendimientos desestacionalizados al cuadrado están significativamente autocorrelacionados, lo que indica un agrupamiento de volatilidad que no se explica por un efecto horario. Los fuertes ciclos en el ACF se han eliminado por completo, lo que es un buen augurio para la aplicación de modelos de volatilidad estándar en estos datos. La curtosis de estos retornos es menor que la curtosis de los retornos originales.

En general, los modelos GARCH pueden ser una herramienta poderosa para analizar datos financieros de alta frecuencia, particularmente cuando se combinan con otras técnicas como la descrita. Sin embargo, es importante usar estos modelos con precaución, ya que pueden ser sensibles a los cambios en los supuestos subyacentes y es posible que no siempre proporcionen predicciones precisas de la volatilidad futura.

Capítulo 3

Procesos de Hawkes

Es común encontrar que eventos, cuya realización es estudiada a lo largo del tiempo, tiendan a agruparse mostrando algún tipo de dependencia. Por ejemplo, luego de un gran temblor de tierra tenemos réplicas que lo siguen, o determinados fenómenos sociales que disparan o disminuyen las ventas de ciertos artículos, es decir, procesos donde se sabe que la tasa de llegada de eventos futuros depende explícitamente de los eventos pasados. Este tipo de procesos son conocidos como de “autoexcitación” y, particularmente, en este capítulo hablaremos sobre el más conocido, el proceso de Hawkes.

Este proceso fue introducido por el matemático Alan Hawkes en 1971 en su trabajo “*Spectra of some self-exciting and mutually exciting point processes*” [17] y “*Point spectra of some mutually exciting point processes*” [22]. En los últimos años, el proceso de Hawkes se ha vuelto muy popular en muchos campos, como las finanzas (Bacry, [18]), la sismología (Ogata, [20]) y la biología (Reynaud-Bouret & Schbath, [21]). Se trata de un caso particular de un proceso puntual¹ cuya intensidad condicional depende de la historia de los eventos. Cada realización de un evento excita el proceso, en el sentido de que los cambios en una secuencia de ocurrencias sucesivas son afectados durante algún periodo de tiempo por la ocurrencia del evento que las desencadenó.

La característica de auto-excitación hace a los procesos de Hawkes casos especiales de procesos Poisson no homogéneos, en los que la intensidad del proceso es descrita mediante una función específica. Esta característica resulta muy útil para representar diversas situaciones financieras como, por ejemplo, relaciones de

¹En el Apéndice B “Procesos de Punto” se presenta la teoría básica necesaria para entender estos procesos y se introduce el más conocido, el proceso de Poisson.

dependencia temporal entre variables, presencia de ciclos de negociación, microestructura de los mercados, entre muchas otras.

Este capítulo ilustrará el marco analítico del proceso de Hawkes, describiremos específicamente el proceso y sus propiedades. El contenido está basado en *Hawkes Processes*, (P. J. Laub & T. Taimre & P. K. Pollett [28]), *Analysis of Order Clustering Using High Frequency Data: A Point Process Approach* (F. Lorenzen, [34]), *Hawkes Processes: Simulation, Estimation, and Validation* (Patrick Laub, [29]), *A Tutorial on Hawkes Processes for events in Social Media.* (Rizoiu, M.A., & Lee, Y. & Swapnil, M. & Xie, L., [44]).

3.1 Procesos autoexcitantes

Los procesos de punto ganaron mucha atención en el campo de las estadística durante las décadas de 1950 y 1960. En 1955 Cox introdujo la noción de un doble proceso estocástico de Poisson² y Bartlett en 1963 investigó métodos estadísticos para procesos puntuales basados en sus densidades espectrales de potencia. En 1964 Lewis formuló un modelo de proceso puntual³ que fue un paso en la dirección del proceso de Hawkes. La actividad culminó en la importante monografía de Cox y Lewis de 1996 en análisis de series de tiempo [24].

Fue en este contexto que Hawkes se propuso llevar el análisis espectral de Bartlett para un nuevo proceso de punto autoexcitante. En una dimensión, el proceso de Hawkes es un proceso de conteo $N(t)$ con una intensidad $\lambda^*(t)$, que representa el número esperado de eventos por unidad de tiempo, dado por un término constante y un término “autoexcitante”, que es una función del historial de eventos.

$$\lambda^*(t) = \lambda(t|\mathcal{F}_t) = \lambda_0 + \int_{-\infty}^t \phi(t-s)dN(s) \quad (3.1)$$

El término autoexcitante forma agrupaciones de una forma retroactiva: Los eventos pasados contribuyen a la tasa de eventos futuros. En la Ecuación 3.1 la función $\phi(\tau) \geq 0$ representa una función kernel que expresa la influencia positiva de eventos pasados T_i en el valor actual de la función de intensidad $\lambda^*(t)$, es decir, es la función que decide el peso a atribuir a los eventos que ocurrieron con un retraso

²Ahora llamado proceso de Cox.

³Para patrones de fallas de computadoras.

en el pasado⁴. La función $\lambda_0 > 0$ es la **intensidad base**⁵ o **intensidad de fondo** que describe la llegada de eventos exógenos, es posiblemente no estacionaria y representa la parte esperada, predecible o determinista de la intensidad.

Definición 15

Considere $\{N(t); t \in \mathbb{R}\}$ un proceso de conteo asociado a una filtración \mathcal{F}_t , $t \in \mathbb{R}$, tal que:

$$P[N(t+h) - N(t) = m | \mathcal{F}_t] = \begin{cases} \lambda^*(t)h + o(h), & \text{si } m = 1 \\ o(h), & \text{si } m > 1 \\ 1 - \lambda^*(t)h + o(h), & \text{si } m = 0 \end{cases}$$

Suponga que la función de intensidad condicional del proceso tiene la forma

$$\lambda^*(t) = \lambda_0 + \int_{-\infty}^t \phi(t-u) dN(u)$$

para alguna $\lambda_0 > 0$ denominada como **intensidad de fondo** o **intensidad base**, y una función $\phi : (0, \infty) \rightarrow [0, \infty)$ denominada **función de auto-excitación**. Entonces tal proceso $N(t)$ es un **Proceso de Hawkes**.

En pocas palabras, el proceso de Hawkes es un proceso puntual que posee una función de respuesta, o kernel, $\phi(\tau)$ que tiene en cuenta la influencia de los eventos pasados en la intensidad condicional actual. Es un modelo específico de proceso de Poisson, que tiene el potencial de explicar algunos de los hechos estilizados relacionados con las dinámicas de cotización y comercio. Como explicaron Daley y Vere-Jones (2003, p.183), el proceso de Hawkes:

“Se acerca más al cumplimiento, para procesos puntuales, al tipo de papel que juega el modelo autorregresivo para las series de tiempo convencionales”.

⁴Cada salto $dN(\tau) \neq 0$ aumenta la probabilidad de eventos futuros a través del kernel ϕ

⁵Realmente λ_0 debería escribirse como $\lambda_0(t)$ pero observando que la suma de la Ecuación 3.2 es sobre los tiempos de eventos (discretos) que son menores o iguales a t_i , podemos asumir que $\lambda_0(t) = \lambda_0$.

3.1.1 Función de intensidad condicional de Hawkes

La definición de intensidad condicional dada en la Definición 3.1 es simplemente una versión generalizada de la que es más común en la literatura. La forma de la función de intensidad condicional del proceso Hawkes puede especificarse si se consideran los instantes de tiempo $\{t_1, t_2, \dots, t_k\}$ que conforman el proceso punto asociado a $N(t)$ hasta el instante t , de manera que:

$$\lambda^*(t) = \lambda_0 + \int_0^t \phi(t-s)dN(s) = \lambda_0 + \sum_{t_i < t} \phi(t-t_i) \quad (3.2)$$

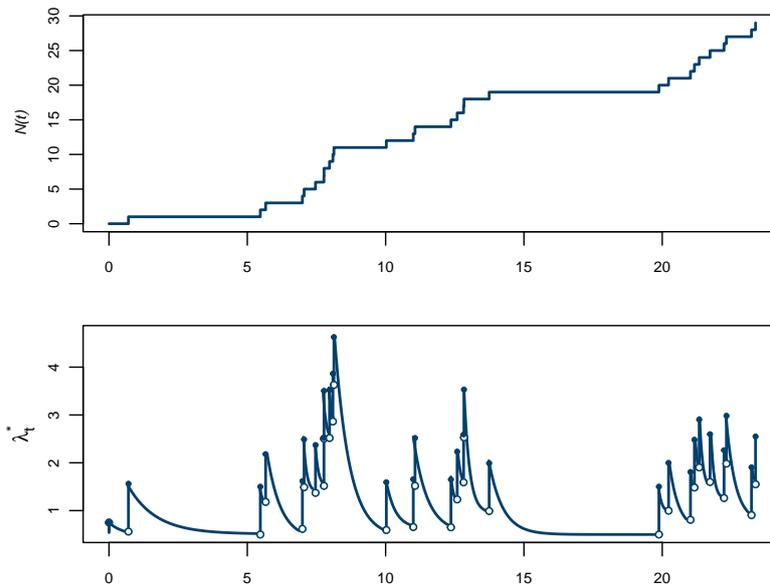


Figura 3.1: Simulación de un proceso de Hawkes unidimensional de tamaño $n = 30$ con kernel que decae exponencialmente especificado por los parámetros $(\lambda, \alpha, \beta) = (0.5, 1, 2)$.

La estructura de la función de intensidad condicional $\lambda^*(t)$ es bastante flexible y solo requiere la especificación de la función de intensidad de fondo λ_0 y la función de auto-excitación $\phi(\tau)$. En el artículo original, Hawkes utilizó a ϕ cuando decae

exponencialmente, es decir, $\phi(t) = \alpha e^{-\beta t}$ con constantes $\alpha, \beta > 0$.

$$\lambda^*(t) = \lambda_0 + \int_{-\infty}^t \alpha e^{-\beta(t-s)} dN(s) = \lambda_0 + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}$$

donde: λ_0 indica el nivel inicial del proceso, α el incremento en la intensidad de ocurrencias por cada ocurrencia presentada y β es la tasa de decaimiento en el tiempo de la influencia de estas ocurrencias, es decir, cada llegada al sistema instantáneamente aumenta la intensidad de la llegada en α , luego, con el tiempo, la influencia de esta llegada decae a un ritmo β .

Un problema práctico a considerar es que no es posible observar en la naturaleza procesos desde el tiempo menos infinito. Si el proceso de Hawkes está restringido a \mathbb{R}^+ con alguna condición inicial $\lambda^*(0) = \lambda$, entonces el proceso de intensidad condicional satisface la ecuación diferencial estocástica:

$$d\lambda^*(t) = \beta(\lambda_0 - \lambda^*(t))dt + \alpha dN(t), \quad t \geq 0$$

con $\lambda^*(0) = \lambda$, cuya solución se tiene que al aplicar el factor integrante $e^{\beta t}$:

$$\begin{aligned} d\lambda^*(t) &= \beta(\lambda_0 - \lambda^*(t))dt + \alpha dN(t) \\ d\lambda^*(t) + \beta\lambda^*(t)dt &= \beta\lambda_0 dt + \alpha dN(t) \\ e^{\beta t} d\lambda^*(t) + \beta\lambda^*(t)e^{\beta t} &= \beta\lambda_0 e^{\beta t} dt + \alpha e^{\beta t} dN(t) \\ d[e^{\beta t}\lambda^*(t)] &= \beta\lambda_0 e^{\beta t} dt + \alpha e^{\beta t} dN(t) \\ e^{\beta t}\lambda^*(t) &= \lambda^*(0) + \int_0^t \beta\lambda_0 e^{\beta s} ds + \int_0^t \alpha e^{\beta s} dN(s) \\ e^{\beta t}\lambda^*(t) &= \lambda^*(0) + \lambda_0(e^{\beta t} - 1) + \int_0^t \alpha e^{\beta s} dN(s) \\ \lambda^*(t) &= \lambda^*(0)e^{-\beta t} + \lambda_0 e^{-\beta t}(e^{\beta t} - 1) + e^{-\beta t} \int_0^t \alpha e^{\beta s} dN(s) \\ \lambda^*(t) &= e^{-\beta t}(\lambda - \lambda_0) + \lambda_0 + \int_0^t \alpha e^{-\beta(t-s)} dN(s) \end{aligned}$$

Se puede ver que

$$\lambda^*(t) = e^{-\beta t}(\lambda - \lambda_0) + \lambda_0 + \int_0^t \alpha e^{-\beta(t-s)} dN(s), \quad t \geq 0$$

Que es una extensión natural de

$$\begin{aligned}\lambda^*(t) &= \lambda_0 + \int_{-\infty}^t \alpha e^{-\beta(t-s)} dN(s) \\ &= \lambda_0 + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}\end{aligned}$$

Otra opción frecuente es una función de ley de potencias, definida como

$$\begin{aligned}\lambda^*(t) &= \lambda_0 + \int_{-\infty}^t \frac{k}{(c + (t - t_i))^p} dN(s) \\ &= \lambda_0 + \sum_{t_i < t} \frac{k}{(c + (t - t_i))^p}\end{aligned}$$

para algunos escalares positivas c , k y p . Esta forma fue popularizada por el modelo geológico llamado ley de Omori, que se utiliza para predecir la tasa de réplicas causadas por un terremoto (Ogata, 1999 [25]).

Relación con modelos AR

Una característica interesante de la función de intensidad condicional es su similitud con un modelos Autorregresivo (AR). Consideremos una función de intensidad condicional con un kernel que decae exponencialmente para algún punto del tiempo en el pasado t_i , entonces la función de intensidad se escribiría como:

$$\lambda^*(t_i) - \lambda_0 = \sum_{t_k < t_i} \alpha e^{-\beta(t_i - t_k)}$$

donde t_k representa todos los eventos que ocurrieron antes de t_i . Si multiplicamos ambas partes de la ecuación por $e^{-\beta(t-t_i)}$ obtendríamos:

$$[\lambda^*(t_i) - \lambda_0] e^{-\beta(t-t_i)} = \sum_{t_k < t_i} \alpha e^{-\beta(t-t_k)}$$

La función $\sum_{t_i < t} \alpha e^{-\beta(t-t_i)}$ se puede descomponer como:

$$\sum_{t_i < t} \alpha e^{-\beta(t-t_i)} = \sum_{t_k < t_i} \alpha e^{-\beta(t_i - t_k)} + \sum_{t_k > t_i < t} \alpha e^{-\beta(t-t_k)}$$

Combinando las últimas 2 ecuaciones podemos escribir a $\lambda^*(t_i) - \lambda_0$ como:

$$\lambda^*(t) - \lambda_0 = [\lambda^*(t) - \lambda_0] e^{-\beta(t-t_i)} + \sum_{t_k > t_i} \alpha e^{-\beta(t-t_k)}$$

Notemos que está ultima ecuación es parecida a un modelo autorregresivo

$$X_t - \lambda_0 = e^{-\beta(t-s)}(X_s - \lambda_0) + \epsilon$$

“Dónde $[\lambda^*(t) - \lambda_0] e^{-\beta(t-t_i)}$ es el término autorregresivo y $\sum_{t_k > t_i} \alpha e^{-\beta(t-t_k)}$ representa la suma de innovaciones de un proceso AR”

-Lorenzen, 2012.

3.2 Estructura Ramificada

Otra visión equivalente del proceso de Hawkes se refiere a la interpretación del proceso de agrupación de Poisson (Hawkes y Oakes, 1974 [32]), que separa los eventos en un proceso de Hawkes en dos categorías: **inmigrantes** y **descendencia**. Los eventos de la descendencia son activados por eventos previos en el proceso, mientras que los inmigrantes llegan de forma independiente y, por lo tanto, no tienen un evento principal existente. Se dice que la descendencia está estructurada en grupos, asociados con cada evento inmigrante. A esto se le llama **estructura ramificada**⁶.

En el resto de esta sección, detallaremos dicha representación y calculamos dos cantidades: el **factor de ramificación**, que es el número esperado de eventos desencadenados directamente por un evento dado en un proceso de Hawkes, y el **número total estimado de eventos en un grupo de descendientes**.

3.2.1 Factor de Ramificación

Consideramos el caso en el que los eventos inmigrantes siguen un proceso Poisson homogéneo con intensidad base λ_0 . Para una persona que ingresa al sistema en el tiempo $t_i \in \mathbb{R}$, la tasa a la que produce descendencia en los tiempos futuros $t > t_i$ es $\phi(t - t_i)$. Digamos que la descendencia directa de esta persona son la primera generación, y sus hijos son la segunda generación, y así sucesivamente. El evento que produce una descendencia se describe como el **antepasado inmediato** o **raíz**

⁶También se e conoce como estructura de Inmigración-Nacimiento.

de la descendencia. Los eventos que están directa o indirectamente relacionados con un inmigrante forman el grupo de descendientes asociados con ese inmigrante.

Ejemplo:

La Figura 3.2(d) ilustra en los primeros 7 tiempos la estructura de ramificación de un proceso de Hawkes con un kernel que decae exponencialmente. Los tiempos de los eventos T_i se indican mediante círculos y las relaciones “origen-descendencia” entre los eventos se muestran mediante flechas. Introducimos las variables aleatorias Z_{ij} , donde $Z_{i0} = 1$ si el evento i es un inmigrante, y $Z_{ij} = 1$ si el evento i es descendiente del evento j . Cada círculo tiene la generación a la que pertenece el evento, es decir, G_k denota la k -ésima generación. Los inmigrantes son etiquetado como G_0 , mientras que las generaciones $G_k; k > 0$ denotan su descendencia.

El **factor de ramificación** es una cantidad clave que describe los procesos de Hawkes, definido como el número esperado de descendientes directos engendrados por un solo evento. El factor de ramificación n^* describe intuitivamente la cantidad de eventos que aparecen en el proceso. Además, da una indicación de si el grupo de descendientes asociado con un inmigrante es un conjunto finito ($n^* < 1$) o infinito ($n^* > 1$).

Para $n^* < 1$, se dice que el proceso se encuentra en **régimen subcrítico**: el total el número de eventos en cualquier grupo está limitado. El evento inmigrante ocurre de acuerdo con la intensidad base λ_0 , pero cada uno de ellos tiene asociado un grupo finito de descendencia, tanto en número como en extensión temporal. Cuando $n^* > 1$, el proceso está en un régimen llamado **super crítico** con $\lambda^*(t)$ creciente y el número total de eventos en cada grupo es ilimitado. Calculamos el factor de ramificación integrando $\phi(\tau)$ (la contribución de cada evento) durante el tiempo del evento t como:

$$n^* = \int_0^{\infty} \phi(\tau) d\tau$$

En el caso de una intensidad con un kernel que decae exponencialmente es

$$n^* := \int_0^{\infty} \alpha e^{-\beta s} ds = \frac{\alpha}{\beta}$$

Cuando $n^* < 1$ se puede obtener una estimación más precisa del tamaño de cada grupo.

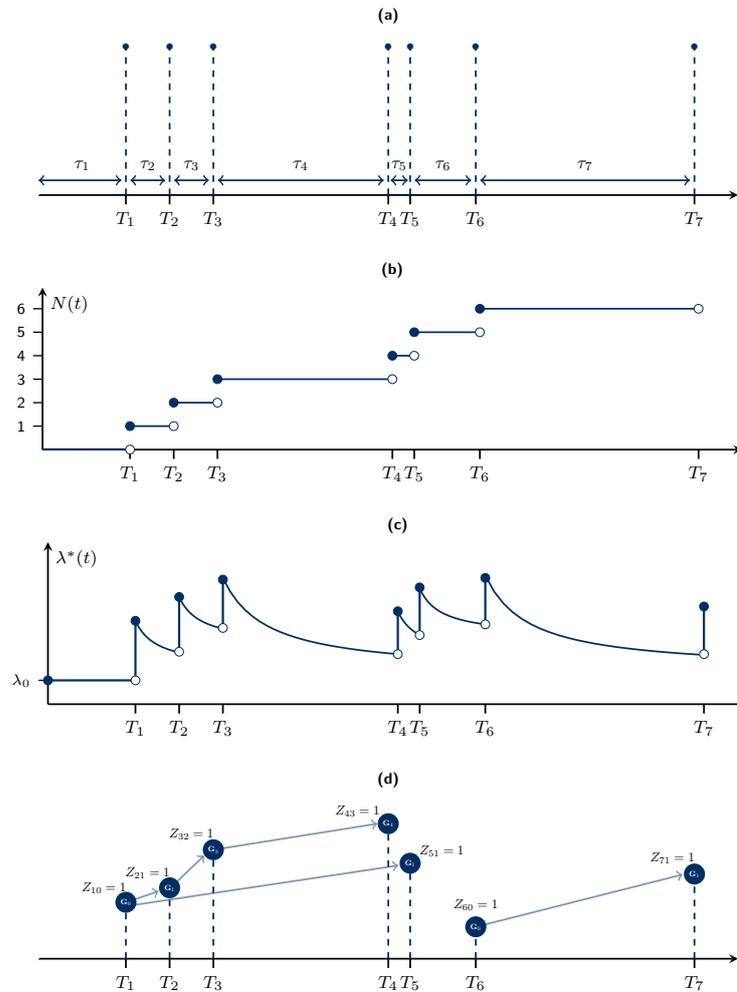


Figura 3.2: Proceso de Hawkes con un kernel que decae exponencialmente. (a) Muestra los primeros siete tiempos del evento. T_i representa los tiempos de eventos, mientras que τ_i representa los tiempos entre llegadas. (b) Es el proceso de conteo N_i que aumenta en una unidad en cada momento del evento T_i . (c) Representa la función de intensidad a lo largo del tiempo. Observe cómo cada evento provoca un salto, seguido por una disminución exponencial. (d) La estructura ramificada latente o no observada del proceso de Hawkes. Cada círculo representa un evento que ha ocurrido en T_i , las flechas representan la relación raíz-descendencia. G_i especifica la generación del evento, con $i = 0$ para inmigrantes o $i > 0$ para la descendencia. Z_{ij} son variables aleatorias, de modo que $Z_{i0} = 1$ si el evento i es un inmigrante, y $Z_{ij} = 1$ si el evento i es un descendiente del evento j .

3.2.2 Propiedades Asintóticas

Las propiedades de estabilidad del proceso de Hawkes son a menudo más simples de adivinar si se considera como un proceso de ramificación⁷. En terminología de la teoría de procesos de ramificación, esta representación de inmigración-descendencia describe un proceso de Galton Watson⁸ con una dimensión de tiempo modificada.

En 1974 los matemáticos Hawkes y Oakes utilizaron la representación de estructura ramificada para derivar características asintóticas del proceso.

Teorema 1

Si

$$0 < n^* := \int_0^\infty \phi(\tau) d\tau < 1$$

y

$$\int_0^\infty \tau \phi(\tau) d\tau < \infty$$

entonces, el número de eventos entre $(0, t]$ cuando $t \rightarrow \infty$ tiene una distribución normal. Usando la siguiente notación $N(0, t] = N(t) - N(0)$ entonces

$$P \left(\frac{N(0, t] - \lambda t / (1 - n)}{\sqrt{\lambda t / (1 - n^*)^3}} \right) \rightarrow \Phi(y)$$

El valor de n^* determina cuándo el proceso Hawkes explota⁹. Para ver esto se puede considerar la función $g(t) = \mathbb{E}[\lambda^*(t)]$ y determinar su comportamiento en el límite, lo que permite caracterizar el comportamiento asintótico del proceso.

$$g(t) = \mathbb{E}[\lambda^*(t)] = \mathbb{E} \left[\lambda_0 + \int_0^t \phi(t-s) dN(s) \right] = \lambda_0 + \int_0^t \phi(t-s) \mathbb{E}[dN(s)]$$

Para calcular el valor esperado en la expresión anterior, se realiza lo siguiente:

⁷Esta representación es útil para modelos de Black S., caso multivariado en: "Hawkes Processes and Their Applications to High-Frequency Data Modeling".

⁸El proceso de Galton-Watson, nombrado así en honor del naturalista británico Francis Galton y su compatriota el matemático Henry William Watson, es un proceso estocástico utilizado para modelizar el desarrollo de una población de individuos auto replicantes.

⁹Usando la definición estándar como el evento en el que $N(t) - N(s) = \infty$ para $t - s < \infty$.

$$\begin{aligned}
 g(s) &= \mathbb{E}[\lambda^*(t)] = \mathbb{E}\left[\lim_{h \rightarrow 0^+} \mathbb{E}\left[\frac{N_{s+h} - N_s}{h} \middle| \mathcal{F}_s\right]\right] \\
 &= \mathbb{E}\left[\mathbb{E}\left[\frac{dN(s)}{ds} \middle| \mathcal{F}_s\right]\right] \\
 &= \frac{\mathbb{E}[\mathbb{E}[dN(s) | \mathcal{F}_s]]}{ds} \\
 &= \frac{\mathbb{E}[dN(s)]}{ds}
 \end{aligned}$$

de donde se sigue que $\mathbb{E}[dN(s)] = g(s)ds$, luego:

$$g(t) = \lambda_0 \int_0^t \phi(t-s) \mathbb{E}[dN(s)] = \lambda_0 + \underbrace{\int_0^t \phi(t-s)g(s)ds}_{\text{Convolución}} = \lambda_0 + \phi * g$$

Esta ecuación tiene diferentes soluciones de acuerdo con el valor de n^* . En general se consideran tres casos: i) el caso por defecto ($n^* < 1$); ii) el caso propio ($n^* = 1$), y iii) el caso super crítico ($n^* > 1$). Se puede demostrar que para el caso por defecto se tiene que:

$$g(t) = \mathbb{E}[\lambda^*(t)] \rightarrow \frac{\lambda_0}{1 - n^*}, \text{ cuando } t \rightarrow \infty \quad (3.3)$$

y que en el caso $n^* > 1$ el proceso eventualmente explota al presentar crecimiento exponencial no acotado. La explosión para $n^* > 1$ se apoya al ver las llegadas como un proceso de ramificación.

Sea $\mathbb{E}[Z_i]$ el número esperado de eventos en la Generación $_i$, con $\mathbb{E}[Z_0] = 1$ (puesto cada grupo tiene un solo inmigrante). Sabemos que $\mathbb{E}[Z_i] = (n^*)^i$ ya que

$$\mathbb{E}[Z_i] = \mathbb{E}[Z_{i-1}]n^* = \mathbb{E}[Z_{i-2}](n^*)^2 = \dots = \mathbb{E}[Z_0](n^*)^i = 1 \cdot (n^*)^i, \quad i \geq 1$$

el número esperado de descendientes para un grupo es:

$$\mathbb{E}\left[\sum_{i=0}^{\infty} Z_i\right] = \sum_{i=0}^{\infty} \mathbb{E}[Z_i] = \sum_{i=0}^{\infty} n^i = \begin{cases} \frac{n^*}{1-n^*} & \text{si } n^* < 1 \\ \infty & \text{si } n^* \geq 1 \end{cases}$$

Por tanto, $n^* \geq 1$ significa que un inmigrante generaría una cantidad infinita de descendientes en promedio.

Cuando $n^* \in (0, 1)$ la razón de ramificación es una probabilidad y se entiende intuitivamente como la relación entre el número total de descendientes y el tamaño de toda la familia (en otras palabras, es la descendencia total más el inmigrante original); es decir:

$$\frac{\mathbb{E}[\sum_{i=0}^{\infty} Z_i]}{1 + \mathbb{E}[\sum_{i=0}^{\infty} Z_i]} = \frac{\frac{n^*}{1-n^*}}{1 + \frac{n^*}{1-n^*}} = \frac{\frac{n^*}{1-n^*}}{\frac{1}{1-n^*}} = n^*$$

Por lo tanto, cualquier llegada de proceso de Hawkes seleccionada al azar se generó de forma endógena (un descendiente) con probabilidad n^* o exógenamente (un inmigrante) $1 - n^*$. La mayoría de las propiedades del proceso de Hawkes se basan en que el proceso sea estacionario, que es otra forma de insistir en que $n^* \in (0, 1)$, por lo que se asume de ahora en adelante.

3.3 Estimación de parámetros

Ahora, veamos el problema de generar estimaciones para los parámetros

$$\hat{\theta} = (\hat{\lambda}, \hat{\alpha}, \hat{\beta})$$

dado un conjunto finito de tiempos de llegada $T = \{t_1, t_2, \dots, t_k\}$ que se suponen provenientes de un proceso de Hawkes.

El método considerado es la estimación por máxima verosimilitud, que comienza por encontrar la función de verosimilitud y estima los parámetros del modelo como las entradas que maximizan esta función.

3.3.1 Función de verosimilitud

Los matemáticos Daley y Vere-Jones (2003, Proposición 7.2.III [23]) dan el siguiente resultado,

Teorema 2

Sea $N(t)$ un proceso de conteo en $[0, T]$ para algún positivo T , y sea t_1, t_2, \dots, t_k una realización de $N(t)$ sobre $[0, T]$. Entonces, la verosimilitud L de $N(t)$ se puede expresar de la siguiente forma:

$$L = \left[\prod_{i=1}^k \lambda^*(t_i) \right] \exp \left(- \int_0^T \lambda^*(u) du \right)$$

Demostración. La función de densidad condicional se puede escribir como:

$$L = L(\hat{\theta}; t) = f(t_1, t_2, \dots, t_k) = \prod_{i=1}^k f^*(t_i)$$

Esta función puede escribirse en términos de la función de intensidad condicional:

$$\begin{aligned} \lambda^*(t) &= \frac{f^*(t)}{1 - F^*(t)} \\ &= \frac{\frac{d}{dt} F^*(t)}{1 - F^*(t)} \\ &= -\frac{d \log(1 - F^*(t))}{dt} \end{aligned}$$

Integrando ambos lados sobre el intervalo (t_k, t) :

$$-\int_{t_k}^t \lambda^*(u) du = \log(1 - F^*(t)) - \log(1 - F^*(t_k))$$

El proceso de Hawkes es un proceso de punto simple, lo que implica que múltiples llegadas no pueden ocurrir al mismo tiempo. Por tanto $F^*(t_k) = 0$ con $T_{k+1} > t_k$, así

$$-\int_{t_k}^t \lambda^*(u) du = \log(1 - F^*(t))$$

Reorganizando

$$\begin{aligned} F^*(t) &= 1 - \exp\left(-\int_{t_k}^t \lambda^*(u) du\right) \\ f^*(t) &= \lambda^*(t) \exp\left(-\int_{t_k}^t \lambda^*(u) du\right) \end{aligned}$$

Así la verosimilitud resulta

$$\begin{aligned} L &= \prod_{i=1}^k f^*(t_i) \\ &= \prod_{i=1}^k \lambda^*(t_i) \exp\left(-\int_{t_{i-1}}^{t_i} \lambda^*(u) du\right) \\ &= \left[\prod_{i=1}^k \lambda^*(t_i) \right] \exp\left(-\int_0^{t_k} \lambda^*(u) du\right) \end{aligned} \tag{3.4}$$

Esta verosimilitud se define para observar un proceso hasta el momento de la n -ésima llegada. Cuando el proceso se observa durante un período de tiempo $[0; t_k] \subset [0, T]$ la verosimilitud debe incluir la probabilidad de no ver llegadas en el intervalo $(t_k, T]$:

$$\therefore L = \left[\prod_{i=1}^k f^*(t_i) \right] (1 - F^*(T))$$

Usando la misma construcción de $F^*(t)$

$$L = \left[\prod_{i=1}^k \lambda^*(t_i) \right] \exp \left(- \int_0^T \lambda^*(u) du \right)$$

■

Simplificación para el caso que decae exponencialmente

Con la función de verosimilitud de la Ecuación 3.4, la log-verosimilitud del intervalo $[0, t_k]$ puede obtenerse como¹⁰:

$$l = l(\hat{\theta}, t) = \sum_{i=1}^k \log(\lambda^*(t_i)) - \int_0^{t_k} \lambda^*(u) du = \sum_{i=1}^k \log(\lambda^*(t_i)) - \Lambda(t_k) \quad (3.5)$$

Notemos que la integral sobre $[0, t_k]$ puede ser separada sobre $[0, t_1], [t_1, t_2], \dots, [t_{k-1}, t_k]$ por tanto:

$$\Lambda(t_k) = \int_0^{t_k} \lambda^*(u) du = \int_0^{t_1} \lambda^*(u) du + \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \lambda^*(u) du$$

Esto se puede simplificar en el caso donde $\lambda^*(t_i)$ decae exponencialmente:

$$\Lambda(t_k) = \int_0^{t_1} \lambda_0 du + \sum_{i=1}^{k-1} \left[\int_{t_i}^{t_{i+1}} \lambda_0 + \sum_{t_j < u} \alpha e^{-\beta(u-t_j)} du \right]$$

¹⁰**Nota:** La notación omitirá los argumentos t y $\hat{\theta}$ de las funciones, es decir, $L = L(\hat{\theta}; t)$, $l = l(\hat{\theta}; t)$, $\lambda^*(t) = \lambda^*(t; \hat{\theta})$ y $\Lambda(t; \hat{\theta})$

Que se puede reescribir como:

$$\begin{aligned}
 \Lambda(t_k) &= \int_0^{t_1} \lambda_0 du + \sum_{i=1}^{k-1} \left[\int_{t_i}^{t_{i+1}} \lambda_0 + \sum_{t_j < u} \alpha e^{-\beta(u-t_j)} du \right] \\
 &= \lambda_0 t_k + \alpha \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \sum_{j=1}^i e^{-\beta(u-t_j)} du \\
 &= \lambda_0 t_k + \alpha \sum_{i=1}^{k-1} \sum_{j=1}^i \int_{t_i}^{t_{i+1}} e^{-\beta(u-t_j)} du \\
 &= \lambda_0 t_k - \frac{\alpha}{\beta} \sum_{i=1}^{k-1} \sum_{j=1}^i \left[e^{-\beta(t_{i+1}-t_j)} - e^{-\beta(t_i-t_j)} \right]
 \end{aligned}$$

Muchos términos de la doble suma se cancelaran:

$$\begin{aligned}
 \Lambda(t_k) &= \lambda_0 t_k - \frac{\alpha}{\beta} \sum_{i=1}^{k-1} \left[e^{-\beta(t_k-t_i)} - e^{-\beta(t_i-t_i)} \right] \\
 &= \lambda_0 t_k - \frac{\alpha}{\beta} \sum_{i=1}^k \left[e^{-\beta(t_k-t_i)} - 1 \right]
 \end{aligned}$$

Sustituyendo $\lambda^*(\cdot)$ y $\Lambda(\cdot)$ en la Ecuación 3.5:

$$l = \sum_{i=1}^k \log \left[\lambda_0 + \alpha \sum_{j=1}^{i-1} e^{-\beta(t_i-t_j)} \right] - \lambda_0 t_k + \frac{\alpha}{\beta} \sum_{i=1}^k \left[e^{-\beta(t_k-t_i)} - 1 \right]$$

Este enfoque directo es computacionalmente inviable ya que la suma doble del primer término implica $\mathcal{O}(k^2)$ complejidad. Afortunadamente, la estructura similar de las sumas anteriores permiten que l se calcule con complejidad $\mathcal{O}(k)$. Denote la suma interna, para algún valor $i \in \{2, \dots, k\}$, como

$$A(i) = \sum_{j=1}^{i-1} e^{-\beta(t_i-t_j)}$$

Esto se puede definir de manera recursiva en términos de $A(i-1)$:

$$\begin{aligned}
 A(i) &= \sum_{j=1}^{i-1} e^{-\beta t_i + \beta t_j} \\
 &= e^{-\beta t_i + \beta t_{i-1}} e^{\beta t_i - \beta t_{i-1}} \sum_{j=1}^{i-1} e^{-\beta t_i + \beta t_j} \\
 &= e^{-\beta t_i + \beta t_{i-1}} \sum_{j=1}^{i-1} e^{-\beta t_i + \beta t_j} \\
 &= e^{-\beta(t_i - t_{i-1})} \left(1 + \sum_{j=1}^{i-2} e^{-\beta(t_{i-1} - t_j)} \right) \\
 &= e^{-\beta(t_i - t_{i-1})} (1 + A(i-1))
 \end{aligned}$$

Por lo tanto

$$A(i) = e^{-\beta(t_i - t_{i-1})} (1 + A(i-1)) \quad (3.6)$$

Añadiendo el caso base $A(1) = 0$, l puede reescribirse como:

$$l = \sum_{i=1}^k \log(\lambda_0 + \alpha A(i)) - \lambda_0 t_k + \frac{\alpha}{\beta} \sum_{i=1}^k \left[e^{-\beta(t_k - t_i)} - 1 \right]$$

3.3.2 Simulación

Esta sección contiene un método detallado para simular datos de un proceso de Hawkes descrito en *MLE of Hawkes' Self-Exciting Point Process* (RK, [15]). Este se hace principalmente para verificar los resultados de las estimaciones de máxima verosimilitud. La idea básica detrás de la simulación se deriva de la función condicional de Hazar

$$\frac{f(t|t_1, \dots, t_n, \theta)}{1 - F(t|t_1, \dots, t_k, \theta)} = \Lambda(t|t_1, t_2, \dots, t_k, \theta)$$

Entonces

$$\log(1 - F(t|t_1, \dots, t_k, \theta)) = - \int_{t_k}^u \Lambda(t|t_1, \dots, t_k, \theta) dt = - \int_{t_k}^u \lambda_0 + \sum_{i=1}^k g(t - t_i | \theta) dt$$

En la ecuación anterior, el tiempo del $(k+1)$ -evento u satisface la ecuación anterior, se puede generar una variable aleatoria uniforme U y resolver para u en la siguiente ecuación para generar el tiempo del próximo evento

$$\log U + \int_{t_k}^u \left(\lambda_0 + \sum_{i=1}^k g(t - t_i | \theta) \right) dt = 0$$

Para la función kernel que decae exponencialmente, la ecuación anterior se reduce a:

$$\log U + \lambda_0(u - t_k) - \frac{\alpha}{\beta} \left[\sum_{i=1}^k e^{-\beta(u-t_i)} - \sum_{i=1}^k e^{-\beta(t_k-t_i)} \right] = 0$$

Considere la expresión:

$$Y(k) = \sum_{i=1}^k e^{-\beta(t_k-t_i)} - \sum_{i=1}^k e^{-\beta(u-t_i)}$$

Esto puede ser escrito como

$$\sum_{i=1}^k e^{-\beta(t_k-t_i)} - \sum_{i=1}^k e^{-\beta(t_k-t_i)} e^{-\beta(u-t_k)}$$

Así,

$$\left(1 - e^{-\beta(u-t_k)} \right) \left(1 + e^{-\beta(t_k-t_{k-1})} \sum_{i=1}^{k-1} e^{-\beta(t_{k-1}-t_i)} \right)$$

Entonces,

$$Y(k) = S(k) \left(1 - e^{-\beta(u-t_k)} \right), S(k) = e^{-\beta(t_k-t_{k-1})} S(k-1) + 1$$

Por tanto, se puede resolver la siguiente ecuación

$$\log U + \lambda_0(u - t_k) + \frac{\alpha}{\beta} S(k) \left(1 - e^{-\beta(u-t_k)} \right) = 0$$

Donde $S(1) = 1$ y $S(k) = e^{-\beta(t_k-t_{k-1})} S(k-1) + 1$.

Para resolver la ecuación numéricamente, se podría requerir la primera derivada de la ecuación anterior.

Veamos un ejemplo de algoritmo. Tomemos a T como $[0, T]$, el último tiempo de llegada del proceso que vamos a simular y a X como el tamaño del salto en λ^* .

Algoritmo de Simulación
1. Inicio: Definimos $\lambda^* \leftarrow \lambda_0(0)$, $n \leftarrow 1$
2. Primer Evento: Generamos $U \sim Unif(0, 1)$, y fijamos $s \leftarrow -\frac{1}{\lambda^*} \ln U$ Si $s \geq T$ Entonces $T_1 \leftarrow s$ De otro modo ir al último paso.
3. Generar Proceso. Sea $n \leftarrow n + 1$ (a) Actualizar el máximo de la intensidad: Sea $\lambda^* \leftarrow \lambda_{n-1}^- + Y$ (b) Nuevo Evento. Generar $U \sim Unif(0, 1)$ y fije $s \leftarrow s - \frac{1}{\lambda^*} \ln U$. Si $s \geq T$ Entonces Regresar al último paso (c) Prueba de Rechazo. Generar $D \sim Unif(0, 1)$ Si $D \geq \frac{\lambda(s)}{\lambda^*}$ Entonces $T_n \leftarrow s$ y volvemos a generar el proceso. De otro modo actualizar $\lambda^* \leftarrow \lambda(s)$ y tomar otro evento del paso (b) del proceso.
4. Resultado. Recuperar el proceso simulado $\{T_n\}$ en $[0, T]$.

Tabla 3.1: Algoritmo tomado de *Simulation, Estimation and Applications of Hawkes Processes* (Kartazyna, [19]).

3.4 Teorema del cambio de tiempo aleatorio

Sabemos que la función media de un proceso estocástico se puede escribir como la integral de la función de intensidad, es decir

$$A(t) = \int_0^t \lambda(s) ds$$

Para un proceso Poisson homogéneo de intensidad λ la función media se puede escribir como $A(t) = \lambda t$, ya que en este caso λ es una constante.

Para un proceso puntual general, la función media lleva el proceso puntual con intensidad $\lambda(s)$ a un proceso de Poisson de parámetro $\lambda = 1$. Por lo tanto, los tiempos entre llegadas se distribuyen exponencialmente

$$A(t_i, t_{i+1}) = \int_{t_i}^{t_{i+1}} \lambda(s) ds$$

y se puede usar para probar el proceso simulado mediante un QQplot. La última ecuación realiza un cambio aleatorio en la escala de tiempo del proceso. El proceso resultante de este cambio de tiempo aleatorio se denomina **proceso residual**. En

el caso de la función de intensidad de un proceso de Hawkes que decae exponencialmente, tenemos:

$$\begin{aligned} \Lambda(t_i, t_{i+1}) &= \int_{t_i}^{t_{i+1}} \lambda_0 ds + \int_{t_i}^{t_{i+1}} \sum_{t_k < s} \alpha \exp(-\beta(s - t_k)) ds \\ \Lambda(t_i, t_{i+1}) &= \lambda_0(t_{i+1} - t_i) + \sum_{k=1}^i \int_{t_i}^{t_{i+1}} \alpha \exp(-\beta(s - t_k)) ds \\ \Lambda(t_i, t_{i+1}) &= \lambda_0(t_{i+1} - t_i) - \sum_{k=1}^i \frac{\alpha}{\beta} [\exp(-\beta(t_{i+1} - t_k)) - \exp(-\beta(t_i - t_k))], \end{aligned}$$

Una forma común de medir la bondad de ajuste del modelo de Hawkes es hacer uso del proceso residual derivado del modelo. La propiedad de cambio de tiempo de los procesos puntuales asegura que el proceso de Hawkes integrado es un proceso Poisson de parámetro $\lambda = 1$. Por lo tanto, los tiempos entre llegadas del proceso integrado se distribuyen exponencialmente de parámetro $\lambda = 1$. Siguiendo a Ogata (1988), es posible hacer una transformación uno a uno del proceso puntual descrito por los eventos $\{t_i\}$ al conjunto aleatorio cambiado en el tiempo definido por $\{\zeta_i\} = \Lambda(t_i, t_{i+1})$. El conjunto de tiempos $\{\zeta_i\}$ es el proceso residual y, usando la propiedad de cambio de tiempo, sigue un proceso de Poisson de parámetro $\lambda = 1$.

Si además tomamos $U_k = 1 - \exp\{-\Lambda(t_{i-1}, t_i)\}$, entonces U_k se distribuye como una variable aleatoria *Unif*(0, 1). Por tanto, una forma muy sencilla de evaluar la bondad de ajuste del modelo de Hawkes es calcular $\hat{\Lambda}(t_i)$ haciendo uso del vector de parámetros $\hat{\theta} = (\hat{\mu}, \hat{\alpha}, \hat{\beta})$, para obtener el proceso residual $\{\hat{\zeta}_i\}$.

Entonces se pueden calcular U_k 's y compararlos con variables uniformes aleatorias en el rango $[0, 1]$. La idea general es que si el modelo de Hawkes es una buena descripción de nuestros datos, entonces podríamos esperar que el proceso residual estimado siga un proceso de Poisson de parámetro $\lambda = 1$, o de manera equivalente, que las duraciones, o tiempos entre llegada, del proceso residual estimado tengan una distribución exponencial de parámetro $\lambda = 1$.

Para probar la hipótesis de que el proceso residual estimado proviene de un proceso de Poisson(1), podemos utilizar la estadística de Kolmogorov-Smirnov (KS) para dibujar límites de confianza para el proceso¹¹.

¹¹Daley y Vere-Jones (2003, p.262) describen un algoritmo que se puede usar para evaluar la

3.5 Covarianza y densidades espectrales de potencia

Los procesos de Hawkes se originaron del análisis espectral de los procesos de punto estacionarios. Encontrar la potencia de la densidad espectral del proceso de Hawkes da acceso a muchas técnicas del campo; por ejemplo, el ajuste del modelo se puede lograr usando el periodograma observado de la realización. La densidad espectral de potencia se define en términos de la densidad de covarianza

$$dN(t) = \lim_{h \rightarrow 0} N(t+h) - N(t)$$

Desafortunadamente, el término “estacionario” tiene muchos significados diferentes en probabilidad. En este contexto, el proceso de Hawkes es estacionario cuando el proceso de salto ($dN(t) : t \in \mathbb{R}$) que toma valores en $\{0, 1\}$ es débilmente estacionario. Esto significa que $\mathbb{E}[dN(t)]$ y $\text{Cov}(dN(t), dN(t+s))$ no dependen de t . Estacionariedad en este el sentido no implica la estacionariedad de $N(t)$ o estacionariedad en los tiempos entre llegadas. Una consecuencia de la estacionariedad es que $\lambda^*(t)$ tendrá una media (como está dado en la ecuación 3.3) a largo plazo:

$$\bar{\lambda}^* = \mathbb{E}[\lambda^*(t)] = \frac{\mathbb{E}[dN(t)]}{dt} = \frac{\lambda_0}{1 - n^*} \quad (3.7)$$

La densidad para autocovarianza estará definida, para $\tau > 0$ como:

$$R(\tau) = \text{Cov} \left(\frac{dN(t)}{dt}, \frac{dN(t+\tau)}{d\tau} \right)$$

Debido a la simetría de covarianza, sucede que $R(-\tau) = R(\tau)$, sin embargo, $R(\tau)$ no puede extenderse a todo \mathbb{R} debido a un átomo en el origen. Para procesos de punto simples $\mathbb{E}[(dN(t))^2] = \mathbb{E}[dN(t)]$ (como $N(t) \in \{0, 1\}$) por lo tanto para $\tau = 0$

$$\mathbb{E}[(dN(t))^2] = \mathbb{E}[dN(t)] = \bar{\lambda}^* dt$$

Por lo tanto, la densidad de la covarianza completa (completa en el sentido que su dominio es todo \mathbb{R}) se define como

$$R^{(c)}(\tau) = \bar{\lambda}^* \delta(\tau) + R(\tau)$$

bondad de ajuste de un proceso puntual usando la estadística KS.

donde $\delta(\cdot)$ es la función delta de Dirac¹². La función de densidad espectral de potencia sería:

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\tau\omega} R^{(c)}(\tau) d\tau = \frac{1}{2\pi} \left[\bar{\lambda}^* + \int_{-\infty}^{\infty} e^{-i\tau\omega} R(\tau) d\tau \right] \quad (3.8)$$

hasta este punto, excluyendo al valor final en la ecuación 3.7, solo se han considerado procesos generales de puntos estacionarios. Para aplicar la teoría específicamente a los procesos de Hawkes entonces debemos considerar:

Teorema 3

Considere un proceso de Hawkes con una intensidad que decae exponencialmente tal que $\alpha < \beta$. El proceso de intensidad tiene entonces densidad de covarianza, para $\tau > 0$,

$$R(\tau) = \frac{\alpha\beta\lambda_0(2\beta - \alpha)}{2(\beta - \alpha)^2} e^{-(\beta - \alpha)\tau}$$

Por tanto, la densidad espectral de la potencia es $\forall \omega \in \mathbb{R}$

$$S(\omega) = \frac{\lambda_0\beta}{2\pi(\beta - \alpha)} \left(1 + \frac{\alpha(2\beta - \alpha)}{(\beta - \alpha)^2 + \omega^2} \right)$$

Demostración. Considere la densidad de la covarianza para $\tau \in \mathbb{R} \setminus \{0\}$:

$$R(\tau) = \mathbb{E} \left[\frac{dN(t)}{dt} \frac{dN(t + \tau)}{d\tau} \right] - \bar{\lambda}^{*2}$$

Tenemos que:

$$\begin{aligned} \mathbb{E} \left[\frac{dN(t)}{dt} \frac{dN(t + \tau)}{d\tau} \right] &= \mathbb{E} \left[\mathbb{E} \left[\frac{dN(t)}{dt} \frac{dN(t + \tau)}{d\tau} \middle| \mathcal{F}(t + \tau) \right] \right] \\ &= \mathbb{E} \left[\frac{dN(t)}{dt} \mathbb{E} \left[\frac{dN(t + \tau)}{d\tau} \middle| \mathcal{F}(t + \tau) \right] \right] \\ &= \mathbb{E} \left[\frac{dN(t)}{dt} \lambda^*(t + \tau) \right] \end{aligned}$$

¹²Normalmente, $R(0)$ se define de manera que R^C es continuo en todas partes. Lewis (1970, p. 357) afirma que estrictamente hablando R^C “no tiene un ‘valor’ en $\tau = 0$ ”. Véase Bartlett (1963b), Cox y Lewis(1966) y Hawkes (1971a) para más detalles.

Así

$$R(\tau) = \mathbb{E} \left[\frac{dN(t)}{dt} \left(\lambda_0 + \int_{-\infty}^{t+\tau} \phi(t + \tau - s) dN(s) \right) \right] - \bar{\lambda}^{*2},$$

Lo que nos da:

$$\begin{aligned} R(\tau) &= \bar{\lambda}^* \phi(\tau) + \int_{-\infty}^{\tau} \phi(\tau - v) R(v) dv \\ &= \bar{\lambda}^* \phi(\tau) + \int_0^{\infty} \phi(\tau + v) R(v) dv + \int_0^{\tau} \phi(\tau - u) R(u) dv \end{aligned}$$

Tomando la transformada de LaPlace:

$$\mathcal{L}\{R(\tau)\}(s) = \frac{\alpha \bar{\lambda}^* (2\beta - \alpha)}{2(\beta - \alpha)(s + \beta - \alpha)} \quad (3.9)$$

Ahora

$$\bar{\lambda}^* = \frac{\beta \lambda_0}{\beta - \alpha} \Rightarrow \mathcal{L}\{R(\tau)\}(s) = \frac{\alpha \beta \lambda_0 (2\beta - \alpha)}{2(\beta - \alpha)^2 (s + \beta - \alpha)} \quad (3.10)$$

Por tanto

$$R(\tau) = \mathcal{L}^{-1} \left\{ \frac{\alpha \beta \lambda_0 (2\beta - \alpha)}{2(\beta - \alpha)^2 (s + \beta - \alpha)} \right\} = \frac{\alpha \beta \lambda_0 (2\beta - \alpha)}{2(\beta - \alpha)^2} e^{-(\beta - \alpha)\tau}$$

El valor de $\bar{\lambda}^*$ de las Ecuaciones 3.9 y 3.10 puede ser sustituida en la función de densidad espectral de potencia dada en la ecuación 3.8:

$$\begin{aligned} S(\omega) &= \frac{1}{2\pi} \left[\bar{\lambda}^* + \int_{-\infty}^{\infty} e^{-i\tau\omega} R(\tau) d\tau \right] \\ &= \frac{1}{2\pi} \left[\bar{\lambda}^* + \int_0^{\infty} e^{-i\tau\omega} R(\tau) d\tau + \int_0^{\infty} e^{i\tau\omega} R(\tau) d\tau \right] \\ &= \frac{1}{2\pi} \left[\bar{\lambda}^* + \frac{\alpha \bar{\lambda}^* (2\beta - \alpha)}{2(\beta - \alpha)(i\omega + \beta - \alpha)} + \frac{\alpha \bar{\lambda}^* (2\beta - \alpha)}{2(\beta - \alpha)(-i\omega + \beta - \alpha)} \right] \\ &= \frac{\lambda_0 \beta}{2\pi(\beta - \alpha)} \left[1 + \frac{\alpha(2\beta - \alpha)}{(\beta - \alpha)^2 + \omega^2} \right] \end{aligned}$$

■

Como $R(w)$ es una función simétrica de valor real, su transformada de Fourier $S(w)$ también es real y simétrica, es decir,

$$S(\omega) = \frac{1}{2\pi} \left[\overline{\lambda^*} + \int_{-\infty}^{\infty} e^{-i\tau\omega} R(\tau) d\tau \right] = \frac{1}{2\pi} \left[\overline{\lambda^*} + \int_{-\infty}^{\infty} \cos(\tau\omega) R(\tau) d\tau \right], \text{ y}$$

$$S_+(\omega) = S(-\omega) + S(\omega) = 2S(\omega)$$

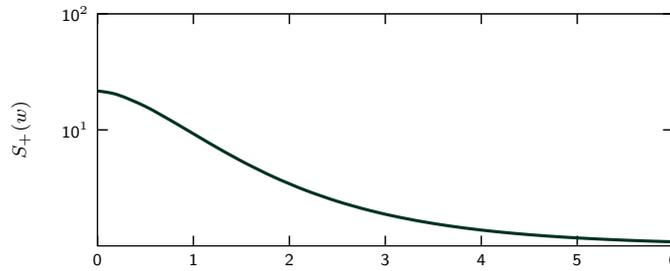


Figura 3.3: Densidad espectral teórica de potencia para un proceso Hawkes especificado por $(\lambda_0, \alpha, \beta) = (0.5, 4, 5)$

Es común que se grafique $S_+(w)$ en lugar de $S(w)$; esto es equivalente a envolver las frecuencias negativas en la media línea positiva (Cox y Lewis 1966, Sección 4.5). La Figura 3.3 muestra un ejemplo de densidad espectral.

3.6 Generalizaciones del proceso de Hawkes

La representación de inmigración-descendencia es útil tanto teórica como prácticamente. Sin embargo, solo se puede utilizar para describir procesos lineales de Hawkes. Brémaud y Massouli en 1996 generalizaron el proceso de Hawkes a su forma no lineal:

Definición 16

Considere a $N(t)$ un proceso de conteo con una función de intensidad condicional:

$$\lambda^*(t) = \Psi \left(\int_0^t \phi(t-s) dN(s) \right)$$

donde $\Psi : \mathbb{R} \rightarrow [0, \infty)$ y $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}$, entonces $N(t)$ es un **Proceso de Hawkes no lineal**. Tenga en cuenta que si tomamos $\Psi(x) = \lambda_0 + x$ entonces $N(t)$ se reduce al proceso lineal de Hawkes de la Definición 15.

Volvamos ahora a la extensión mencionada anteriormente, la de una colección de procesos de Hawkes auto-excitantes y mutuamente auto-excitantes. Los procesos que se examinan son colecciones de procesos de Hawkes unidimensionales que se “auto excitan” a sí mismos y entre sí. Existen modelos para procesos de Hawkes donde los puntos en sí mismos son multidimensionales, por ejemplo, procesos espaciales de Hawkes o procesos temporo-espaciales de Hawkes, que no se examinan en esta tesis.

Definición 17

Considere una colección de m procesos de conteo $\{N_1(t), \dots, N_m(t)\}$ denotado como \mathbf{N} . Sean $\{T_{i,j} : i \in \{1, \dots, m\}, j \in \mathbb{N}\}$ tiempos aleatorios de llegada para cada proceso de conteo (y $t_{i,j}$ para tiempos de llegada observados). Si para cada $i = 1, \dots, m$ $N_i(t)$ tiene una función de intensidad condicional

$$\lambda^*(t) = \lambda_i + \sum_{j=1}^m \int_{-\infty}^t \phi_j(t-u) dN_j(u) \tag{3.11}$$

para alguna $\lambda_i \in \mathbb{R}^+$ y $\phi_i : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \cup \{0\}, \phi_i(\tau) \neq 0$, entonces \mathbf{N} es un **Proceso de Hawkes mutuamente autoexcitante**.

Cuando la función $\phi(\tau)$ tiene la forma que decae exponencialmente, entonces 3.11 puede ser escrita como

$$\begin{aligned} \lambda_i^*(t) &= \lambda_i + \sum_{j=1}^m \int_{-\infty}^t \alpha_{i,j} e^{-\beta_{i,j}(t-s)} dN_j(s) \\ &= \lambda_i + \sum_{j=1}^m \sum_{t_{j,k} < t} \alpha_{i,j} e^{-\beta_{i,j}(t-t_{j,k})} \end{aligned}$$

Para constantes no negativas $\{\alpha_{i,j}, \beta_{i,j} : i, j = 1, \dots, m\}$.

Capítulo 4

Análisis Empírico

En este análisis empírico, se utilizó un enfoque basado en procesos de Hawkes para modelar la dinámica del precio del activo financiero SPY 500 ETF y con esto, calcular el VaR.

Para ello, se recopilaron datos de alta frecuencia y se ajustó un proceso de Hawkes utilizando un algoritmo de máxima verosimilitud. En este enfoque, se utilizan los procesos de Hawkes para modelar la dinámica de las tasas de eventos en los datos financieros de alta frecuencia sobre la cola izquierda de la distribución. Luego, se utiliza la Distribución Generalizada de Pareto (GDP) para modelar el tamaño de estas pérdidas para poder calcular el VaR.

Los resultados obtenidos mostraron que el enfoque basado en procesos de Hawkes puede ser una herramienta útil para calcular el VaR en datos de alta frecuencia, proporcionando una estimación precisa y confiable del riesgo.

4.1 SPY 500 ETF

El SPY 500 ETF, también conocido como “SPDR S&P 500 ETF Trust”, es un fondo cotizado en bolsa que sigue el índice S&P 500, que es un índice de referencia para el rendimiento de las acciones de 500 grandes empresas cotizadas en bolsa en los Estados Unidos, es decir, se forma mediante la compra de acciones de las empresas que componen el índice S&P 500. Este ETF¹ se lanzó en 1993 y es uno de los fondos cotizados más populares y líquidos en el mercado, con un volumen

¹Un ETF Trust (*Exchange-Traded Fund Trust*) es una estructura legal utilizada para crear un fondo de inversión cotizado (ETF). Un ETF es un tipo de fondo de inversión que cotiza en bolsa.

de negociación diario promedio de varios millones de acciones, lo que significa que se negocia con alta frecuencia.

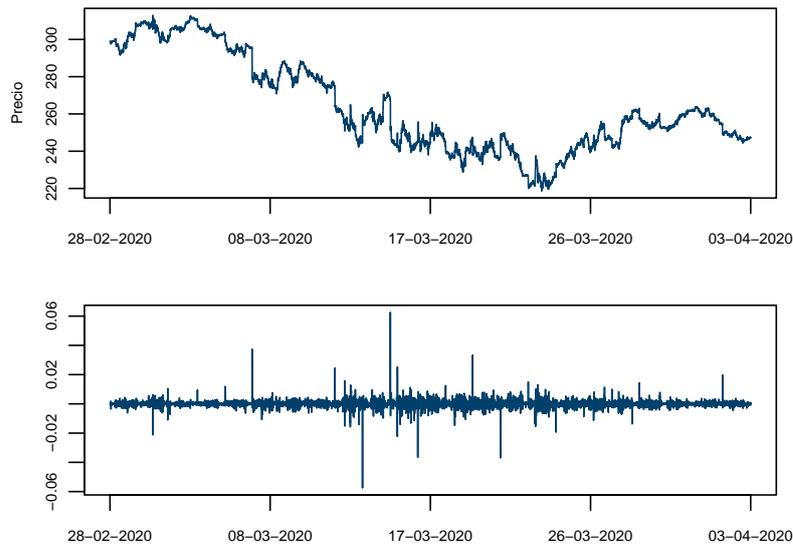


Figura 4.1: La primer gráfica nos muestra la serie de tiempo del precio del SPY 500 ETF entre las 9:30 del 28 de febrero de 2020 y las 23:53 del 3 de abril de 2020. Los datos tienen un frecuencia de 7 minutos. La segunda gráfica nos muestra las log-pérdidas o los log-rendimientos negativos de dichos precios.

Así como ya mencionado anteriormente, nos interesa estudiar la serie de rendimientos de los precios de nuestro activo, y ya que nuestra meta es calcular el VaR, nos enfocaremos en solo una cola de la distribución. Sabemos que la cola derecha representa los eventos extremos positivos (por ejemplo, grandes aumentos en el precio de un activo o grandes ganancias). Sin embargo, al estudiar la cola izquierda de la distribución, que representa los eventos extremos negativos (por ejemplo, grandes caídas en el precio de un activo o pérdidas), podemos entender la probabilidad y la severidad de estos eventos para gestionar adecuadamente el riesgo. Es por eso que el ajuste lo haremos sobre las pérdidas o log rendimientos negativos, como se muestra en la segunda gráfica de la Figura 4.1, donde podemos apreciar agrupaciones de volatilidad características de los datos financieros intradía.

4.2 Ajuste del modelo

Antes de proceder a ajustar los parámetros del modelo propuesto, vamos a discutir 2 puntos muy importantes que afectan la estructura de este, la elección del umbral, dónde consideramos a un valor como uno extremo, y la elección del kernel de nuestra función de intensidad.

La elección de un umbral no es algo intuitivo o fácil de hacer, de hecho la teoría de valores extremos nos dice que encontrar el valor de umbral adecuado es un compromiso entre reducir el sesgo y minimizar la varianza de los estimadores de los parámetros de la distribución de valores extremos. Un umbral alto reduce el sesgo pero aumenta la varianza, mientras que un umbral bajo hace lo contrario. Por lo tanto, se debe buscar un equilibrio adecuado para obtener una estimación precisa de los parámetros de la distribución de los valores extremos.

Existen diferentes métodos para seleccionar un umbral, aunque ninguno proporciona una solución definitiva. Por ejemplo, Chávez-Demoulin y McGill (2012, [5]) recomiendan seleccionar umbrales entre el 92 % y el 95 % utilizando la gráfica de exceso medio en el caso de datos no independientes ni idénticamente distribuidos. En el caso de datos no estacionarios, Chávez-Demoulin (2014, [58]) eligieron el 10 % de los eventos semanales más extremos. Por su parte, Herrera y Schipp (2013, [59]) proponen una metodología basada en un análisis de sensibilidad del VaR, eligiendo un umbral donde las estimaciones de riesgo VaR sean más estables. Concluyen que un umbral entre los cuantiles 0.90 y 0.92 puede ser óptimo².

Por simplicidad, para seleccionar nuestro umbral, tomamos valores diferentes de u entre el cuantil .90 y el .95. Seleccionamos el que nos permitió un mejor ajuste, que resultó el cuantil 0.95.

Sabemos que los kernels son funciones que describen la influencia de los eventos anteriores en la probabilidad de que ocurra un evento en el futuro. En muchos estudios de aplicación, la simple observación de la distribución de eventos proporciona suficiente evidencia para decidir cómo debería verse el núcleo/kernel o qué propiedades debería tener.

²Un artículo muy famoso para elegir umbrales en la Teoría de Valores Extremos es “*Estimating the Tail-Expectation of an Empirical Distribution*” de Hill (1975).

- **Kernel exponencial:** es el kernel más comúnmente utilizado en los procesos de Hawkes. La función del kernel exponencial decae exponencialmente en el tiempo desde el evento anterior.
- **Kernel de ley de potencias:** es un kernel que decae en función de una ley de potencia. Es útil para modelar eventos que tienen memoria larga.
- **Kernel de Gauss:** es un kernel que tiene una forma de campana de Gauss. Este kernel se utiliza a menudo cuando se desea modelar eventos que tienen un patrón estacional o diurno.
- **Kernel de Cauchy:** es un kernel que tiene una forma de distribución de Cauchy. Este kernel es útil para modelar eventos extremadamente raros, como terremotos o eventos de mercado muy extremos.
- **Kernel logístico:** es un kernel que tiene una forma de curva en S . Este kernel se utiliza a menudo en aplicaciones de aprendizaje automático y se puede utilizar para modelar eventos que tienen una tasa de crecimiento exponencial.

El Capítulo 3 desarrolla toda la teoría de procesos de Hawkes usando un kernel exponencial debido a su simplicidad y que requiere menos esfuerzo computacional para ser aproximado. En la literatura es común que al tratarse de datos financieros de alta frecuencia se emplee un kernel exponencial o de ley de potencias. En datos de alta frecuencia, el uso del kernel de ley de potencias puede resultar en un modelo que es demasiado flexible y sobreajustado, lo que puede dar lugar a una mala generalización y problemas de pronóstico. En cambio, el kernel exponencial puede ser una mejor opción porque es más simple y tiene menos parámetros para ajustar.

En resumen, elegimos el kernel exponencial debido a que es una buena opción para modelar eventos financieros de alta frecuencia por su simplicidad, flexibilidad, interpretación y eficiencia computacional. Sin embargo, debemos tener en cuenta que esto no es una generalización, la elección del kernel adecuado depende del contexto específico y requiere una detallada exploración empírica de los datos.

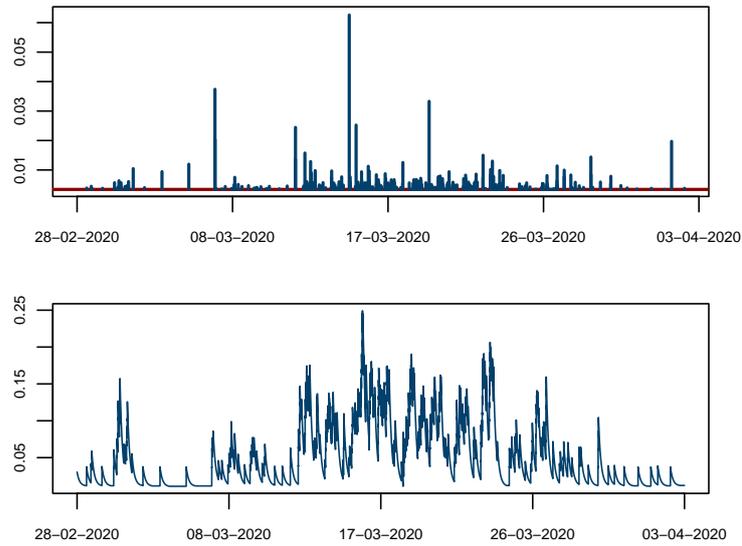


Figura 4.2: La primera gráfica nos muestra todos los excedentes que pasan el cuantil 0.95 de los log-rendimientos negativos de la Figura 4.1. La segunda gráfica muestra el cálculo de la función de intensidad con un kernel que decae exponencialmente con parámetros $(\hat{\lambda}_0, \hat{\alpha}, \hat{\beta}) = (0.0115, 0.0257, 0.0335)$

Utilizando un algoritmo de máxima verosimilitud, obtuvimos los siguientes parámetros de nuestra función de intensidad con un kernel que decae exponencialmente:

Estimador	Valor Estimado	<i>p</i> - value
$\hat{\lambda}_0$	0.0115	0.092
$\hat{\alpha}$	0.0257	0.056
$\hat{\beta}$	0.0335	0.061

Tabla 4.1: Parámetros estimados por máxima verosimilitud. Podemos observar que todos los parámetros son significantes, siendo λ_0 el que tiene la significancia más alta, lo que era de esperarse siendo la parte determinista de la función.

Usando los valores estimados de nuestra Tabla 4.1 nuestra función de intensidad

condicional estaría definida por la ecuación:

$$\lambda^*(t) = 0.0115 + \int_{-\infty}^t 0.0257e^{-0.0335(t-s)} dN(s) = 0.0115 + \sum_{t_i < t} 0.0257e^{0.0335(t-t_i)}$$

Evaluemos qué tan bien un modelo ajustado se ajusta a los datos observados.

4.2.1 Bondad de ajuste

Para evaluar la bondad de ajuste de nuestro modelo, haremos 3 pruebas usando el teorema de cambio de tiempo aleatorio definido en la Sección 3.4:

1. Si tomamos a $U_k = 1 - \exp\{\Lambda(t_{i-1}, t_i)\}$, entonces $U_k \sim Unif(0, 1)$
2. Usando $\Lambda(t)$ y el teorema de cambio de tiempo aleatorio, comprobaremos que el proceso de Hawkes integrado es un proceso Poisson de parámetro 1, Por lo tanto los tiempos entre llegadas se distribuyen $Exp(1)$ Una manera de corroborarlo es creando un QQplot.
3. Si hay razones para creer que nuestros tiempos entre llegadas están distribuidos exponencialmente, debemos comprobar que son independientes, esto se puede hacer buscando la autocorrelación de τ_i . Obviamente, la autocorrelación cero no implica independencia, pero una cantidad distinta de cero ciertamente implica un modelo no Poisson. Se puede realizar un examen visual trazando los puntos del proceso residual (U_{i+1}, U_i) . Si encontramos patrones notables, los τ_i están autocorrelacionados.

En general para los procesos de punto, usamos el teorema de cambio de tiempo aleatorio para medir la bondad de ajuste. Para los procesos de Hawkes, el algoritmo fue propuesto por Ozaki (1979, [50]) y se basa en la siguiente ecuación

$$\int_{t_k}^t \lambda^*(u) du = -\log(1 - F^*(t)),$$

que relaciona la función de distribución acumulativa de la siguiente llegada al historial de llegadas $\{t_1, \dots, t_k\}$. Esta relación significa que la siguiente llegada a tiempo T_{k+1} puede generarse usando el método de transformación inversa, es decir, si $U \sim Unif(0, 1)$ entonces t_{k+1} es encontrado resolviendo

$$\int_{t_k}^{t_{k+1}} \lambda^*(u) du = -\log(U)$$

En el caso que decae exponencialmente

$$\log(U) + \lambda(t_{k+1} - t_k) - \frac{\alpha}{\beta} \left(\sum_{i=1}^k e^{\beta(t_{k-1}-t_i)} - \sum_{i=1}^k e^{-\beta(t_k-t_i)} \right) = 0$$

Para resolver dicha ecuación para t_{k+1} , podemos usar la recursión de la Ecuación 3.6:

$$A(i) = e^{-\beta(t_i-t_{i-1})}(1 + A(i-1))$$

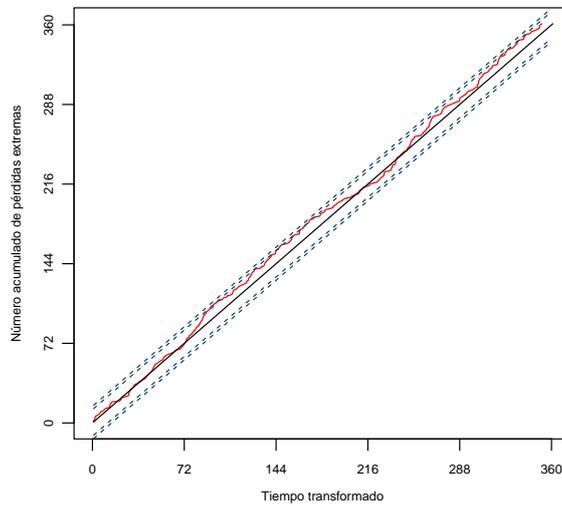


Figura 4.3: Las líneas punteadas son los límites de confianza bilaterales al 95 % y el 99 % basados en la estadística de Kolmogorov-Smirnov. Ya que nuestro proceso U_k no sale de los límites de confianza, podemos decir que nuestras variables siguen una distribución $unif(0,1)$.

Podemos ver de la Figura 4.3 que nuestro proceso definido como $U_k = 1 - \exp\{A(t_{i-1}, t_i)\} \sim Unif(0, 1)$ ya que se mantiene dentro de los límites de confianza de la prueba de KS al 95 % y 99 %.

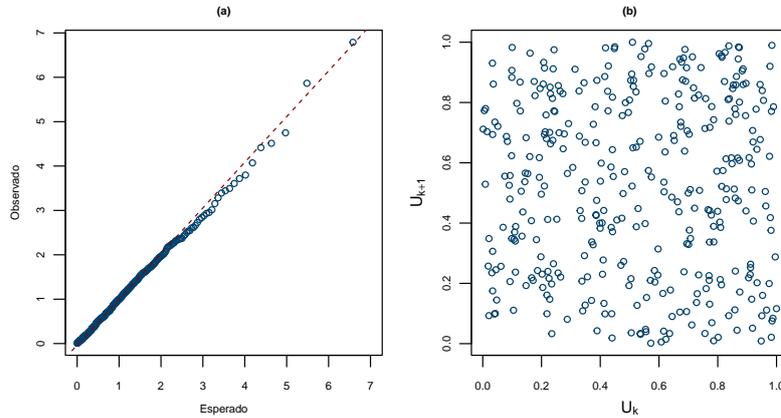


Figura 4.4: (a) Nos muestra el QQplot para v.a.i.d. Exp(1) para los tiempos entre llegadas, donde podemos apreciar que los datos sí se ajustan. (b) **Prueba de Autocorrelación**. Los valores de U_k se definen como $U_k = F(t_k - t_{k-1}) = 1 - e^{-(t_k - t_{k-1})}$. No podemos apreciar una autocorrelación evidente.

En la Figura 4.4 podemos corroborar que los tiempos entre llegadas se distribuyen Exp(1) y que no existe una correlación visible entre ellas.

4.3 Medidas de Riesgo

Antes de proceder a nuestro cálculo del VaR, es necesario definir la distribución que usaremos para el tamaño de los excedentes. Para este fin, usaremos la Distribución Generalizada de Pareto, que es ampliamente utilizada en el estudio de valores extremos.

4.3.1 Distribución Generalizada de Pareto (GDP)

La función de distribución de la distribución Generalizada de Pareto de dos parámetros está definida como

$$G_X(x) = 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi}; \xi \neq 0$$

donde $x \geq 0$ para $\xi > 0$ y $0 \leq x \leq \frac{-\sigma}{\xi}$ para $\xi < 0$. Los parámetros ξ y σ son los de forma y escala respectivamente.

para el caso donde $\xi = 0$

$$G_X(x) = 1 - e^{-x/\sigma}; \xi = 0$$

Que es una distribución exponencial definida para $x \geq 0$ con parámetro de escala σ . La función de densidad es entonces

$$g_X(x) = \begin{cases} \frac{1}{\sigma} \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi-1} & \text{si } \xi \neq 0 \\ \frac{1}{\sigma} e^{-x/\sigma} & \text{si } \xi = 0 \end{cases}$$

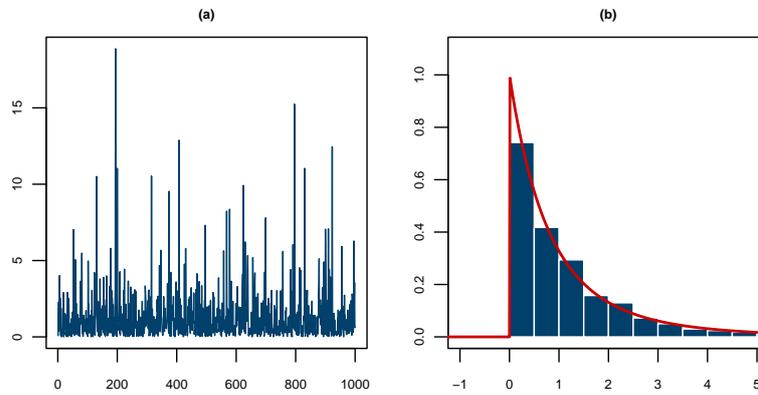


Figura 4.5: (a) Simulación de una variable aleatoria con Distribución Generalizada de Pareto con parámetros $\xi = .25, \sigma = 1$ (b) Densidad empírica y densidad real.

La función de Distribución Generalizada de Pareto (GPD) tiene tres distribuciones. Cuando $\xi > 0$, la GPD es una distribución de Pareto del segundo tipo³ con una cola pesada que decae a una velocidad polinomial; cuando $\xi = 0$ la GPD es una distribución exponencial con un cola mediana que se descompone exponencialmente; para $\xi < 0$, la GPD se convierte en una distribución de cola ligera, el límite superior del soporte de la distribución es finito.

El k -ésimo momento de la GPD es existente para $\xi < 1/k$; por ejemplo, la media y la varianza son finitas solo cuando $\xi < 1$ y $\xi < 0.5$, respectivamente.

³O distribución Lomax en la literatura de seguros

La función de log-verosimilitud para la función de distribución generalizada de Pareto está dada por

$$\ln L(\xi, \sigma; x) = -n \ln \sigma + \left(\frac{1}{\xi} - 1 \right) \sum_{i=1}^n \ln \left(1 - \frac{\xi x_i}{\sigma} \right)$$

Al hacer el ajuste usando la ecuación anterior, obtuvimos las siguientes estimaciones:

Estimador	Valor Estimado	<i>p</i> - value
$\hat{\xi}$	0.3634	0.081
$\hat{\sigma}$	0.0014	0.100

Tabla 4.2: Parámetros estimados por máxima verosimilitud. Podemos observar que todos los parámetros son significativos.

4.3.2 Cálculo del VaR

En la práctica, uno de los principales propósitos de ajustar un modelo a nuestros datos, es el de calcular alguna medida de riesgo para predecir pérdidas futuras. Para nuestro análisis, nuestro proceso de Hawkes nos dice cómo se modelan los tiempos de llegada de eventos de pérdida extrema que vamos a combinar con la distribución Generalizada de Pareto para el tamaño de dichos eventos. La GPD es una distribución de probabilidad comúnmente utilizada en la Teoría de Valores Extremos para modelar la distribución de los valores extremos.

Para estimar la medida de Valor en Riesgo (VaR) tenemos que encontrar

$$z_q^t = \inf\{z \in \mathbb{R} : F_{Z_{t+1}|\mathcal{F}_t}(z) \geq q\}$$

Usando los datos observados hasta tiempo t . Es común que en la práctica se tome $q = 0.95$. Para nuestro modelo, vamos a utilizar la siguiente expresión:

$$P(Z_{t+1} > z | \mathcal{F}_t) = P(Z_{t+1} - u > z - u | Z_{t+1} > u, \mathcal{F}_t) \times P(Z_{t+1} > u | \mathcal{F}_t),$$

Donde el segundo término de la expresión representa la probabilidad condicional de un evento en $(t, t + 1]$, esto es,

$$1 - P\{N(t, t + 1) = 0 | \mathcal{F}_t\} = 1 - \exp\left(-\int_0^{t+1} \lambda_{\mathcal{F}}(u) du\right)$$

Y el primer término puede ser calculado usando un modelo de Distribución Generalizada de Pareto:

$$Z_{t+1} - u | \mathcal{F}_t; Z_{t+1} \geq u \sim GPD_{\xi, \sigma}$$

Buscamos el cuantil z_q^t que resuelva la ecuación

$$P(Z_{t+1} > z_q^t | \mathcal{F}_t) = 1 - q$$

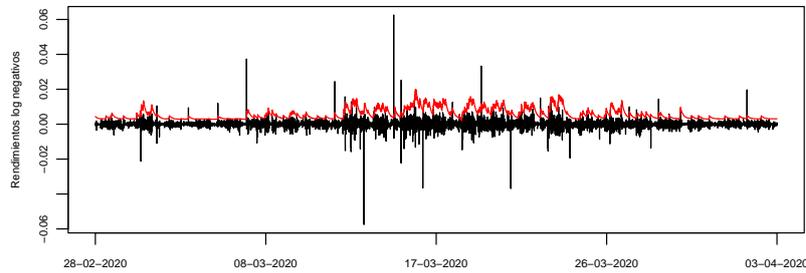


Figura 4.6: Log rendimientos negativos del precio del SPY 500 ETD entre el 28 de febrero de 2020 y el 03 de abril del 2020 interpolados cada 7 minutos. La curva roja es el 95 %VaR condicional estimado cuando el proceso de autoexcitación se ajusta a los tiempos de llegada, combinado con el modelo GPD para el tamaño de las marcas.

De la Figura 4.6 podemos observar la capacidad de la función de intensidad de los procesos de Hawkes de capturar las agrupaciones de eventos y los elevaciones bruscas y sostenidas.

Back-testing del VaR

El back-testing del cálculo del VaR (Value at Risk) es un proceso fundamental para evaluar la eficacia de cualquier modelo del VaR, especialmente en datos de alta frecuencia. El back-testing se refiere a la prueba del modelo para evaluar la precisión de las predicciones de éste.

Hay muchas pruebas para realizar un back-testing, sin embargo debemos tener en cuenta que no todas ellas toman en cuenta la volatilidad condicional, es decir, no podemos asegurar que los fallos sean independientes, debido a la tendencia de formar grupos de los datos de alta frecuencia.

Christoffersen (1998, [51]) propone un método para evaluar las estimaciones del VaR, introduciendo una prueba condicional, esta prueba parte de observar que si una aproximación para la estimación del VaR captura de manera precisa la distribución condicional de los rendimientos y sus propiedades dinámicas, entonces, las excepciones o fallos deben ser impredecibles.

Consideremos \hat{q} como la proporción de excepciones del VaR en los periodos $\{1, 2, \dots, T\}$. Supongamos que K es el número de estas excepciones, entonces la variable K es una suma de las variables aleatorias Bernoulli definidas como:

$$1_{t+1} = \begin{cases} 0, & \text{si } L_{t+1} < VaR(q) \\ 1, & \text{si } L_{t+1} \geq VaR(q) \end{cases}$$

Donde L_{t+1} es la pérdida observada a tiempo $t+1$. Por lo tanto, K tiene distribución binomial dada por

$$P(K) = \binom{T}{K} q^K (1-q)^{T-K}$$

Por una parte se quiere probar la hipótesis $H_0 : \hat{q} = 0$, para ello se considera como estadística de prueba la función de verosimilitud siguiente:

$$LR_{uc} = 2 \log \left[\left(1 - \frac{K}{T}\right)^{T-K} \left(\frac{K}{T}\right)^K \right] - 2 \log[(1-\alpha)^{T-K} \alpha^K]$$

Esta estadística se distribuye asintóticamente como una distribución $\chi_2^{(1)}$. Esta parte es equivalente a la prueba de Kupiec descrita en el Apéndice C, sin embargo la prueba de Christoffersen considera además la independencia del modelo por medio del siguiente estadístico

$$LR_{ind} = -2 \log[(1 - \pi_0)^{n_{00} + n_{11}}] + 2 \log[(1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1 - \pi_{11})^{n_{10}} \pi_{11}^{n_{11}}]$$

donde n_{ij} es el número de días en que el estado i -ésimo es seguido del estado j -ésimo, con $i, j \in \{0, 1\}$. Consideramos cero el estado en el cual la pérdida es menor que el VaR estimado, y 1 cuando el rendimiento actual es mayor que el VaR estimado. Este estadístico sigue una distribución $\chi_2^{(1)}$. Por lo tanto, en la prueba de Christoffersen se considera tanto la condición de cobertura como la de independencia, el estadístico de esta prueba es

$$\begin{aligned} LC_{cc} &= LR_{uc} + LR_{ind} \\ &= 2 \log[(1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1 - \pi_{11})^{n_{10}} \pi_{11}^{n_{11}}] - 2 \log[(1 - \alpha)^{T-K} \alpha^K] \end{aligned}$$

donde π_i denota la probabilidad que la excepción ocurra al tiempo $t + 1$ condicionando sobre el estado i al tiempo t , entonces

$$\pi_0 = \frac{n_{01}}{n_{00} + n - 01}$$

y

$$\pi_1 = \frac{n_{11}}{n_{10} + n - 11}$$

Por su definición, la estadística LC_{cc} tiene una distribución $\chi_2^{(2)}$.

	Número de violaciones	Prueba incondicional	Prueba de independencia	Condicional
95%	96	0.72	0.68	0.14
99%	42	0.65	0.34	0.48

Tabla 4.3: Los resultados son el número de violaciones observadas sobre la muestra y con diferentes valores p . Se tomaron muestras de tamaño $T = 2000$ para el nivel del 95% y $T = 4000$ para el nivel 99%.

De la Tabla 4.3 observamos que para distintos valores de p con diferentes tamaños de muestra T nuestros excedentes pasaron las 3 pruebas de Christoffersen.

4.4 Conclusiones

En los últimos años, la recopilación de información ha experimentado una evolución significativa gracias a la tecnología y a la creciente cantidad de datos generados por las empresas y los consumidores. Esto ha impactado directamente en cómo funcionan las bolsas de valores del mundo. Hoy en día el comercio de alta frecuencia representa más del 70% del comercio a nivel mundial.

Si bien es cierto que hay una variedad bastante considerable de modelos para predecir series de tiempo financieras, los datos de alta frecuencia han puesto nivel de dificultad extra, gracias a las propiedades que poseen, que difieren de las series de tiempo financieras convencionales o de frecuencias más bajas.

Como ya hemos visto en la Sección 2.4.1, algunos modelos de series de tiempo, como algunas extensiones de modelos GARCH, son usados para modelar datos intradía. Los procesos de Hawkes y los modelos de series de tiempo son dos enfoques diferentes para modelar la volatilidad en los datos financieros de alta frecuencia. Aunque ambos modelos son útiles para estimar el riesgo financiero, los procesos de

Hawkes presentan los siguientes beneficios:

1. **Modelado de eventos extremos:** Son más flexibles para modelar eventos extremos, como las colas pesadas en la distribución de los rendimientos. Esto se debe a que permiten modelar la autocorrelación de los eventos extremos, sin hacer modificaciones a las escalas de tiempo que, como ya se mencionó en el capítulo podrían alterar las propiedades de los datos.
2. **Capacidad para modelar la dinámica del mercado:** Resultan valiosos para capturar la dinámica del mercado, incluidas las interacciones entre los agentes del mercado, detalladas en el Capítulo 1, ya sea por sus propiedades estacionales o porque no ponen condiciones de distribución de los datos, como algunos modelos de series de tiempo.
3. **Mejor ajuste a los datos de alta frecuencia:** Son capaces de modelar la estructura de los datos de alta frecuencia. Esto se debe a que pueden capturar la propiedad de agrupación de la volatilidad gracias a la estructura de su función de intensidad.

Otro enfoque diferente al de series de tiempo tipo GARCH, es el de modelos POT (*Peaks over threshold*) empleados en la Teoría de Valores Extremos, dichos modelos al igual que los procesos de Hawkes, pueden poseer una parte autoexcitante. Ya en artículos como *High-frequency financial data modeling using Hawkes processes* (Chavez-Demoulin, [5]) se prueban estos modelos contra los procesos de Hawkes, demostrando que en ocasiones, los modelos POT subestiman el cálculo del VaR.

En conclusión, el uso de procesos de Hawkes para modelar los datos financieros de alta frecuencia puede ser beneficioso para el cálculo del VaR debido a su capacidad para capturar las propiedades de agrupación o dependencia de eventos y de la frecuencia de las series de tiempo intradía. Tienen la ventaja de ser muy flexibles respecto a la elección de las especificaciones del modelo, que van de la mano con las características particulares de los datos estudiados. Esto permite desarrollar modelos más precisos y efectivos para el cálculo del VaR, que son particularmente útiles en entornos de alta frecuencia y alta volatilidad, donde los riesgos cambian rápidamente y deben ser evaluados con precisión.

Apéndice A

Definiciones básicas

En el siguiente apéndice podremos revisar algunos de los conceptos de estadística que fueron empleados a lo largo del presente trabajo.

El apéndice está basado en *Introducción a la Probabilidad* (Luis Rincón, [45]), *Exponentiated Generalized Pareto Distribution: Properties and applications towards Extreme Value Theory* (Lee, Se Yoon & H.T. Kim, J., [46]) y *Statistical Inference* (Silvey, S.D., [47]).

A.1 Distribución Exponencial

Decimos que una variable aleatoria continua X tiene distribución exponencial con parámetro $\lambda > 0$, y escribimos $X \sim \text{exp}(\lambda)$, cuando su función de densidad es

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

La gráfica de esta función está en la Figura A.1 para varios valores de λ .

Esta distribución se usa para modelar tiempos de espera para la ocurrencia de un cierto evento. Integrando la función de densidad desde menos infinito hasta un valor arbitrario x se encuentra que la función de distribución tiene la expresión:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

Aplicando el método de integración por partes, es sencillo corroborar que

$$\mathbb{E}(X) = \frac{1}{\lambda}$$

$$\text{Var}(X) = \frac{1}{\lambda^2}$$

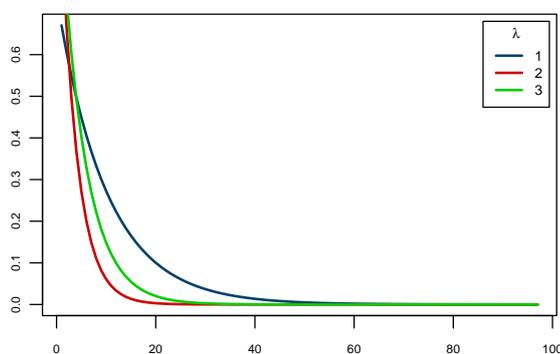


Figura A.1: Función de densidad de una variable aleatoria exponencial con diferente valor de parámetro λ .

A.2 Distribución Poisson

Decimos que X tiene una distribución de Poisson con parámetro $\lambda > 0$, y escribimos a $X \sim \text{Poisson}(\lambda)$ cuando

$$P[X = x] = \begin{cases} e^{-\lambda} \frac{\lambda^x}{x!} & \text{si } x = 0, 1, 2, \dots \\ 0 & \text{si en otro caso} \end{cases}$$

La forma de esta función de probabilidad se muestra en la Figura A.2.

La función de distribución $F(x)$ para esta distribución, como suma de los valores de $f(x)$, no tiene una expresión reducida y no la escribiremos, sin embargo el cálculo computacional puede realizarse de manera muy sencilla.

Se puede comprobar que

$$\begin{aligned}\mathbb{E}(X) &= \lambda \\ \text{Var}(X) &= \lambda\end{aligned}$$

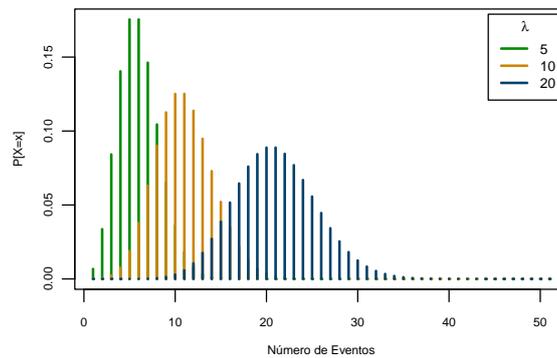


Figura A.2: Función de densidad de una variable aleatoria con distribución Poisson con diferente valor de parámetro λ .

A.3 *p-value*

Supongamos que tenemos una muestra de datos x_1, x_2, \dots, x_n y que contamos con la estimación de un parámetro de la distribución de la muestra, digamos β , con la estadística $\hat{\beta}$, hagamos la siguiente hipótesis:

$$H_0 : \beta = \hat{\beta}$$

La prueba la realizaremos con una estadística que provenga de la muestra, digamos

$$T = T(x_1, \dots, x_n)$$

que, en teoría, tiene la misma distribución que cierta variable X . Entonces, definimos al *p-value* como:

$$P[X > T]$$

Entonces, el *p-value* es el nivel de significancia de la prueba, el nivel de confianza con el que no se rechaza la hipótesis.

A.4 QQ-plot

Una de las pruebas más simples para evaluar el ajuste de un modelo teórico a datos es el gráfico QQ-plot.

Supongamos una realización x_1, \dots, x_n , y una distribución teórica, F , que suponemos es la distribución de los datos, el QQplot se construye de la siguiente manera:

1. La muestra se ordena de menor a mayor: $x_{(1)}, \dots, x_{(n)}$. Entonces, $x_{(i)}$ debe ser el cuantil $\frac{i}{n}$. Usualmente se considera el cuantil $\frac{i-0.5}{n}$.
2. Se obtienen los cuantiles teóricos de la distribución a ajustar a los datos. Los cuantiles a obtener serían: q_1, \dots, q_n tal que:

$$F(q_i) = \frac{i - 0.5}{n}$$

3. Se grafican los puntos $(q_i, x_{(i)})$. Si el ajuste es correcto, entonces el gráfico se parecerá a la identidad.

A.5 Prueba de Kolmogorov

Otra prueba muy común para corroborar que una muestra empírica tiene una distribución F es la prueba de Kolmogorov Smirnov.

Suponga una muestra ordenada

$$x_{(1)}, x_{(2)}, \dots, x_{(n)}$$

y sea $F_n(z)$ la distribución empírica de la función definida por

$$F_n(z) = \begin{cases} 0 & \text{si } z < x_{(1)}, \\ \frac{j}{n} & \text{si } x_{(j)} \leq x_{(j+1)} (j = 1, 2, \dots, n-1) \\ 1 & \text{si } x_{(n)} \end{cases}$$

Si la función de distribución de nuestra muestra es realmente F , entonces la función empírica F_n debe de aproximarse a F . Por lo tanto, parece sensato basar una región crítica de una prueba de F en la distancia de F_n a F . Esto, por supuesto,

deja el problema de elegir una medida de la distancia entre dos funciones. La prueba de Kolmogorov-Smirnov utiliza la estadística:

$$D_n(x) = \sup |F_n(z) - F(z)|$$

donde la región crítica tiene la forma

$$\{x : D_n(x) > k\}$$

siendo elegida la constante k para dar el tamaño de prueba deseado. Por lo tanto, para una prueba de tamaño α , k es elegida tal que

$$Pr\{D_n(x) > k|F\} = \alpha$$

Apéndice B

Procesos de Punto

Algunos procesos que debemos definir para entender los procesos de Hawkes son los procesos de punto y conteo. Definirlos es importante para poder explicar de manera formal al proceso Poisson, la discusión de dicho proceso es necesaria ya que el proceso de Hawkes puede verse como una generalización del proceso de Poisson no homogéneo. Este apartado asumirá que los procesos considerados están definidos sobre un espacio de probabilidad filtrado (Ω, \mathcal{F}, P) , para alguna medida de probabilidad P .

El apéndice está basado en *Introducción a los procesos estocásticos* (Luis Rincón, [39]), *Simulation, Estimation and Applications of Hawkes Processes* (Katarzyna Obral, [40]) y *Point Processes in Time and Space* (Snyder, D., Miller, M., Random, [41]).

B.1 Procesos de punto y conteo

Un proceso de punto es un proceso estocástico que se utiliza para modelar eventos que ocurren en intervalos aleatorios en el tiempo. Si $\{T_i, i \in \mathbb{N}\}$ es una secuencia de variables aleatorias no negativas tales que

$$\forall i \in \mathbb{N} : T_i < T_{i+1}$$

y asumimos que $T_0 = 0$ entonces llamaremos a $\{T_i, i \in \mathbb{N}\}$ un proceso de punto en \mathbb{R}^+ .

Definición 18

Sea \mathbf{T} una secuencia de variables aleatorias $\mathbf{T} = \{T_1, T_2, \dots\}$ que toma valores

en \mathbb{R}^+ , tal que

$$P[0 < T_1 \leq T_2 \leq T_3 \leq \dots] = 1 \text{ y}$$

$$P[T_i < T_{i+1}, T_i < \infty] = P[T_i < \infty]$$

para $i \geq 1$, y el número de puntos en una región acotada es finita (c.s.), entonces \mathbf{T} es un **proceso de punto**.

Como $T_i < t_i \forall i$, el proceso se denomina **proceso puntual simple**, ya que excluye la posibilidad de la ocurrencia simultánea de eventos. Notemos que n representa el número de eventos observados y T_n el último punto observado del proceso.

Si $\{W_i\}_{i \in \{1, \dots, n\}}$ es una secuencia de marcas con $W_i \in \{1, \dots, K\}$ que denota el tipo del i -ésimo tipo de evento, entonces, la secuencia doble $\{T_i, W_i\}_{i \in \{1, 2, 3, \dots, n\}}$ se llama un **proceso de punto marcado simple**. Las “marcas” específicamente indican diferentes tipos de eventos, como, por ejemplo, la llegada de compras, ventas o ciertas órdenes límite en procesos de negociación financiera. Los tiempos de llegada específicos del evento se indican como $\{T_i^k\}_{i \in \{1, \dots, n^k\}}$, $k = 1, \dots, K$, donde n^k es el número de eventos tipo k observados.

Un proceso estocástico $\{N(t); t \geq 0\}$ se dice es un proceso de conteo si $N(t)$ representa el número total de eventos que ocurren en el tiempo t . Un proceso de conteo se puede definir formalmente, con la ayuda de la definición de procesos de punto,

Definición 19

Sea $\{T_i, i \in \mathbb{N}\}$ un proceso de punto, entonces

$$N(t) = \sum_{i \in \mathbb{N}} 1_{\{T_i < t\}}$$

Es llamado el **proceso de conteo** asociado a $\{T_i, i \in \mathbb{N}\}$.

Un proceso de conteo se puede ver como un recuento acumulativo del número de “llegadas” a un sistema hasta un tiempo dado u . Otra forma de caracterizar dicho proceso es considerar la secuencia de tiempos de llegada aleatorios $\mathbf{T} = \{T_1, T_2, \dots\}$ en el cual el proceso de conteo $N(t)$ ha saltado.

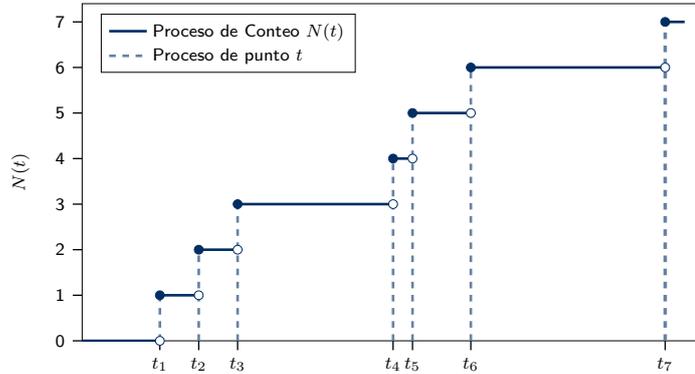


Figura B.1: La gráfica muestra ejemplo de realización de un proceso puntual $\{t_1, t_2, \dots\}$ (eventos) y el proceso de conteo correspondiente $N(t)$.

Definición 20

Sea x_i como el tiempo de espera entre dos puntos sucesivos, definido como

$$x_i = \begin{cases} t_i - t_{i-1} & \text{si } i = 2, \dots, n \\ t_i & \text{si } i = 1 \end{cases}$$

con $t_0 = 0$. Entonces, $\{x_i\}_{i \in \{1, \dots, n\}}$ se denomina como un **proceso de duración asociado** a $\{t_i\}$.

Una forma de caracterizar un proceso de punto particular es especificar la función de distribución del próximo tiempo de llegada con condicional en el pasado. Digamos que $\{\mathcal{F}(u) : u \geq 0\}$ es la historia de las llegadas hasta el momento u^1 .

Dado el historial hasta la última llegada u , $\mathcal{F}(u)$ define la función de distribución condicional (y de densidad) del próximo tiempo de llegada T_{k+1} como:

$$F^*(t|\mathcal{F}(u)) = \int_u^t P [T_{k+1} \in [s, s + ds] | \mathcal{F}(u)] = \int_u^t f^*(s|\mathcal{F}(u)) ds$$

¹Estrictamente hablando $\mathcal{F}(u)$ es una filtración, es decir, una secuencia creciente de álgebras; Usualmente se toma la filtración natural.

La función de densidad conjunta para la realización $\{t_1, t_2, \dots, t_k\}$ es entonces:

$$f(t_1, t_2, \dots, t_k) = \prod_{i=1}^k f^*(t_i | \mathcal{F}(t_{i-1}))$$

En la literatura, la notación raramente condiona $\mathcal{F}(u)$ explícitamente, en su lugar coloca una estrella superíndice sobre la función. Como tal, estas funciones se escriben como $F^*(t)$ y $f^*(t)$. Esta caracterización proporciona una definición para varias clases de procesos puntuales. Si un proceso de punto tiene una distribución $f^*(t)$ que es independiente de $\mathcal{F}(\cdot)$, entonces el proceso es llamado un **proceso de renovación**.

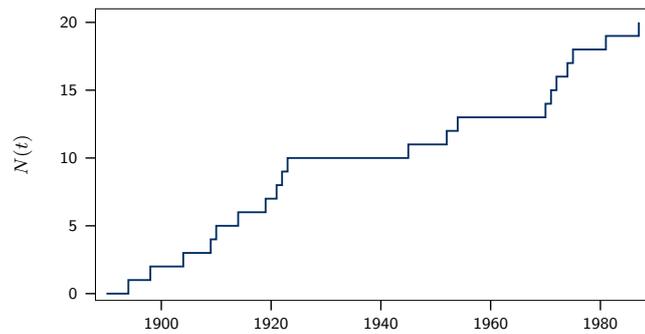


Figura B.2: La figura muestra las inundaciones en el río Amazonas cerca de Manaus, Brasil, durante los años 1892–1992.

Veamos a un par de ejemplos de procesos de conteo de la vida real. Si dejamos que $N(t)$ sea igual al número de personas que ingresan a un super mercado o banco en particular en el momento t o antes, entonces $\{N(t), t \geq 0\}$ es un proceso de conteo en el que un evento corresponde a una persona que ingresa a algunos de estos establecimientos. Si decimos que un evento ocurre siempre que alguien compra una casa, entonces $\{N(t), t \geq 0\}$ es un proceso de conteo cuando $N(t)$ es igual al número total de personas que compraron una casa en el tiempo t . La Figura B.2 muestra un ejemplo gráfico de un proceso $N(t)$.

Corolario 1

Un de proceso de conteo $N(t)$ satisface:

- i) $N(t) \geq 0$,
- ii) $N(t) \in \mathbb{Z}$,
- iii) Si $s < t$, entonces $N(s) \leq N(t)$,
- iv) Para $s < t$, $N(t) - N(s)$ es igual al número de eventos que ocurren en el intervalo $(s, t]$.

Si consideramos el intervalo $(0, \mathcal{T}]$, sea

$$0 < T_1 < T_2 < \dots < T_{N(t)} \leq t \leq \mathcal{T}$$

un conjunto de tiempo de eventos de un proceso puntual. Para $t \in (0, \mathcal{T}]$, $N(t)$ es la **trayectoria** del proceso de conteo asociado. La trayectoria es una función continua por la derecha con límite por la izquierda, con saltos de tamaño 1 en los momentos del evento y, casi siempre, es constante. De esta manera, $N(t)$ cuenta el número y la ubicación de los eventos en el intervalo $(0, t]$. Ahora veamos tres propiedades importantes de los procesos de conteo.

Definición 21

Se dice que un proceso de conteo posee incrementos independientes si el número de eventos que ocurren en intervalos de tiempo separados son independientes.

Definición 22

Se dice que un proceso de conteo posee incrementos estacionarios si la distribución del número de eventos que ocurren en cualquier intervalo de tiempo depende sólo de la duración del intervalo de tiempo.

En otras palabras, el proceso tiene incrementos estacionarios si el número de eventos en el intervalo $(s, s + t]$ tiene la misma distribución para todos los s .

Definición 23

Se dice que un proceso estocástico es homogéneo en el espacio si la probabilidad de transición entre dos valores de estado cualesquiera en dos momentos determinados depende únicamente de la diferencia entre esos valores de estado.

El proceso es **homogéneo** en el tiempo si la probabilidad de transición entre dos valores de estado dados en dos momentos cualesquiera depende únicamente de la diferencia entre esos momentos.

B.1.1 Función de Intensidad Condicional

A menudo es difícil trabajar con la distribución de llegada condicional $f^*(t)$, por eso, se utiliza otra caracterización de procesos puntuales: **la función de intensidad condicional**. De hecho si el condicional de la función de intensidad existe, caracteriza de manera única las distribuciones de dimensiones finitas del proceso de punto. Se define como

$$\lambda^*(t) = \frac{f^*(t)}{1 - F^*(t)}$$

Aunque esta definición sigue siendo válida, la siguiente definición ofrece una representación más intuitiva de la función de intensidad condicional como la tasa esperada de llegadas condicionada en $\mathcal{F}(t)$.

Definición 24

Considere un proceso de conteo $N(t)$ asociando a una filtración \mathcal{F}_t , si $\lambda^*(t)$ existe, entonces

$$\lambda^*(t) = \lambda(t|\mathcal{F}_t) = \lim_{h \rightarrow 0^+} \mathbb{E} \left[\frac{N_{t+h} - N_t}{h} \middle| \mathcal{F}_t \right]$$

$\lambda^*(t)$ es la **función de intensidad condicional** de $N(t)$

De manera equivalente, podemos escribir lo anterior como:

$$\lambda^*(t) = \lim_{h \rightarrow 0^+} \frac{1}{h} P[N_{t+h} - N_t > 0 | \mathcal{F}_t]$$

Se puede ver que la intensidad del proceso depende de la filtración seleccionada (que por lo general se asume como la filtración natural para el proceso $N(t)$).

Definición 25

Para un proceso de conteo $N(t)$ la función no decreciente

$$\Lambda(t) = \int_0^t \lambda(s) ds$$

es llamada la **función media** del proceso de punto.

B.2 Proceso Poisson

Definamos a uno de los tipos de procesos de punto más importantes, el proceso Poisson.

Suponga que un mismo evento ocurre repetidas veces de manera aleatoria a lo largo del tiempo. Tal evento puede ser, por ejemplo, la llegada de una reclamación a un banco o la entrada de una llamada a un *call center*, la llegada de un cliente a una taquilla de cine o los momentos en que una cierta maquinaria requiere reparación, etcétera.

Suponga que las variables aleatorias T_1, T_2, \dots representan los tiempos que transcurren entre una ocurrencia del evento y la siguiente ocurrencia. También que estos tiempos son independientes uno del otro y que cada uno tiene distribución $\exp(\lambda)$. Se define al proceso Poisson a tiempo t como el número de ocurrencias del evento que se han observado hasta ese instante t .

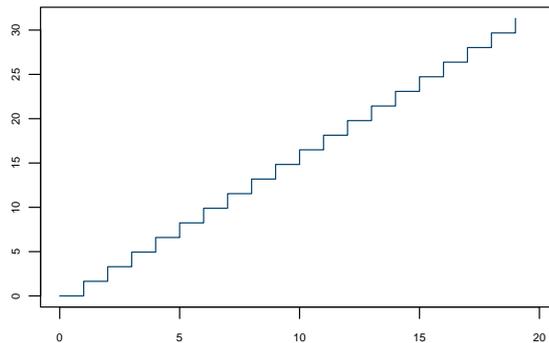


Figura B.3: Simulación de un Proceso Poisson homogéneo con $\lambda = 1$. Podemos observar que tenemos 1 evento por cada unidad de tiempo transcurrido. Es decir, las llegadas son constantes en el tiempo.

B.2.1 Notación de Landau

Introduciremos la notación \mathcal{O} *mayúscula* y o *minúscula*, a veces también llamados **notación de Landau** gracias al matemático alemán Edmund Landau².

²El símbolo $\mathcal{O}(\xi)$ aparece por primera vez en el segundo volumen del *Tratado sobre Teoría de Números*, escrito en 1894 por P. G. H. Bachmann (1837 - 1920) y E. G. H. Landau (1877 - 1938)

Definición 26

i) Sean $f = f(x)$ y $g = g(x)$, si x_0 es finito, la expresión

$$f = \mathcal{O}(g), \quad \text{cuando } x \rightarrow x_0 \quad (\text{B.1})$$

significa que existen constantes k y x_1 (independientes de x) tales que

$$|f(x)| \leq k|g(x)| \quad \text{para } |x - x_0| < x_1$$

Si x_0 es infinito, la expresión B.1 significa que existen constantes k y $M > 0$ tales que

$$|f(x)| \leq k|g(x)| \quad \text{para } |x| > M$$

En este caso se dice que “ f es \mathcal{O} mayúscula de g cuando $x \rightarrow x_0$ ”, o que “ f y g son del mismo orden cuando $x \rightarrow x_0$ ”.

ii) Sean $f = f(x)$ y $g = g(x)$, si x_0 es finito, la expresión:

$$f = o(g) \quad \text{cuando } x \rightarrow x_0 \quad (\text{B.2})$$

significa que para cualquier k existe un x_1 (independiente de x) se satisface que

$$|f(x)| \leq k|g(x)| \quad \text{para } |x - x_0| < x_1$$

Si x_0 es infinito, la expresión B.2 significa que para cualquier k existe M (independiente de x) tal que

$$|f(x)| \leq k|g(x)| \quad \text{para } |x| > M$$

En este caso se dice que “ f es o minúscula de g cuando $x \rightarrow x_0$ ”.

Los símbolos de orden \mathcal{O} y o , pueden ser definidos también de la siguiente manera.

Supongamos que $g(x) \neq 0$ para $|x - x_0| < x_r$ para algún $x_r > 0$ y consideremos el límite

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} \quad (\text{B.3})$$

Se dice que “ f es \mathcal{O} de g cuando $x \rightarrow x_0$ ” (o que “ f es del orden de g ” cuando

adquiere esta notación del libro de Bachmann. Sin embargo, el símbolo $o(x)$ fue inventado por Landau para sustituir la notación anterior, $\{x\}$).

$x \rightarrow x_0$) si el límite B.3 existe y es diferente de cero.

A continuación se establece la relación que existe entre estas dos maneras de definir a \mathcal{O} y o .

Teorema 4

- i) Si el límite B.3 existe y es diferente de cero, entonces $f = \mathcal{O}(g)$ cuando $x \rightarrow x_0$.
- ii) Si $\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = 0$ entonces $f = o(g)$ cuando $x \rightarrow x_0$.

B.2.2 Proceso Poisson homogéneo

Una de las clases más simples de un proceso puntual es el proceso de Poisson.

Definición 27

Sea $(\tau_i)_{i \geq 1}$ una secuencia de variables aleatorias independientes e idénticamente distribuidas con distribución exponencial de parámetro λ y tiempos de llegada

$$T_n = \sum_{i=1}^n \tau_i$$

entonces al proceso $\{N(t), t \geq 0\}$ definido como

$$N(t) := \sum_{i \geq 1} I_{t \geq T_i}$$

es un **Proceso Poisson** con intensidad λ .

A la secuencia τ_j se le denomina **tiempos entre llegadas** o **duraciones**, es decir, el primer evento ocurre en el momento τ_1 , el segundo en τ_2 después del primero, etc. Los tiempos entre llegadas τ_i son independientes, y cada uno de ellos siguen una distribución exponencial con parámetro λ . Aquí, la notación $f_\tau(t)$ denota la función de densidad de probabilidad de la variable aleatoria τ tomando valores denotados por t .

$$f_{\tau}(t) = \begin{cases} \lambda e^{-\lambda t} & \text{si } t \geq 0 \\ 0 & \text{si } t < 0 \end{cases}$$

Aquí $\lambda > 0$ es una constante positiva. La esperanza de τ_i se puede calcular de la siguiente manera:

$$\begin{aligned} \mathbb{E}[\tau] &= \int_0^{\infty} t f_{\tau}(t) dt = \lambda \int_0^{\infty} t e^{-\lambda t} dt = [-t e^{-\lambda t}]_{t=0}^{t=\infty} + \int_0^{\infty} e^{-\lambda t} dt \\ &= 0 - \left[\frac{1}{\lambda} e^{-\lambda t} \right]_{t=0}^{t=\infty} = \frac{1}{\lambda} \end{aligned}$$

En general, la intensidad del evento no necesita ser constante, sino que es una función del tiempo, escrito como $\lambda(t)$, este caso general se le conoce como **proceso poisson no homogéneo**.

Definición 28

Los *tiempos de llegada* se definen como:

$$T_n = \sum_{j=1}^n \tau_j$$

donde T_n es el tiempo de la n -ésima llegada. Notemos que $N(t)$ tiene incrementos de tamaño 1 por cada T_i , es decir:

$$N(t) = \begin{cases} 0 & \text{si } 0 \leq t < T_1 \\ 1 & \text{si } T_1 \leq t < T_2 \\ 2 & \text{si } T_2 \leq t < T_3 \\ \vdots & \\ n & \text{si } T_n \leq t < T_{n+1} \end{cases}$$

Observamos que $N(t)$ se define de modo que sea continuo por la derecha con límites por la izquierda.

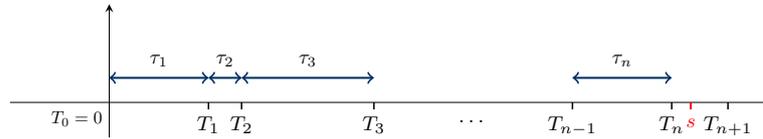


Figura B.4: Las variables τ_n representan los intervalos de tiempo entre eventos sucesivos o tiempos entre llegadas (llegadas de clientes a una fila, de llamadas a una central telefónica, de pacientes a la emergencia de un hospital, etc.). $T_n = \tau_1 + \dots + \tau_n$ es el instante en el ocurre el n -ésimo evento y $N(s)$ es el número de eventos que han ocurrido hasta el instante s .

Falta de memoria del proceso

No tener memoria en un proceso de punto significa que la distribución de los tiempos futuros entre llegadas depende solo sobre información relevante sobre el tiempo actual, pero no sobre información de más allá del pasado. Este es el caso de los procesos de Poisson. Calculamos la probabilidad de observar un tiempo entre llegadas τ más largo que un período de tiempo predefinido t . Si F_τ es la función de distribución de τ , que se define como

$$F_\tau(t) := P[\tau \leq t] = \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t}, t \geq 0$$

y, por tanto, la probabilidad de observar un evento en el tiempo $\tau > t$ está dada por

$$P[\tau > t] = e^{-\lambda t}, t \geq 0$$

Supongamos que estábamos esperando la llegada de un evento, cuyos tiempos entre llegadas siguen una distribución exponencial con parámetro λ . Suponga que han transcurrido m unidades de tiempo y durante este período no ha llegado ningún evento, es decir, no hay eventos durante el intervalo de tiempo $[0, m]$. La probabilidad de que tengamos que esperar otras t unidades de tiempo dada por

$$\begin{aligned} P[\tau < t + m | \tau > m] &= \frac{P[\tau > t + m, \tau > m]}{P[\tau > m]} \\ &= \frac{P[\tau > t + m]}{P[\tau > m]} = \frac{e^{-\lambda(t+m)}}{e^{-\lambda m}} = e^{-\lambda t} = P[\tau < t] \end{aligned}$$

Esta última ecuación denota la propiedad de falta de memoria de los procesos de Poisson. Es decir, la probabilidad de tener esperar t unidades de tiempo adicionales después de haber esperado m unidades de tiempo es la misma que la probabilidad de tener que esperar t unidades de tiempo al comenzar en el tiempo 0. Dicho de otra manera, si se interpreta a τ como el tiempo de llegada de un evento donde τ sigue una distribución exponencial, la distribución de $\tau - m$ dado $\tau > m$ es igual que la distribución de τ en sí.

B.2.3 Proceso Poisson no homogéneo

En los procesos Poisson, los eventos llegan aleatoriamente con la intensidad constante $\lambda > 0$. Este modelo inicial es suficiente para describir procesos simples, digamos la llegada de automóviles a una calle en un corto período de tiempo. Sin embargo, necesitamos poder variar la intensidad del evento con el tiempo para describir procesos más complejos, como simular las llegadas de automóviles durante las horas concurridas y fuera de éstas. En un proceso de Poisson no homogéneo, la tasa de llegadas de eventos es una función del tiempo, es decir, $\lambda = \lambda(t)$.

Definición 29

El proceso de conteo $\{N(t), t \geq 0\}$ es un **proceso de Poisson no homogéneo** con función de intensidad $\lambda(t)$, $t \geq 0$ si:

- i) $N(0) = 0$,
- ii) $\{N(t), t \geq 0\}$ tiene incrementos independientes,
- iii) $P[N(t+h) = n+m | N(t) = n] = \lambda(t)h + o(h)$, si $m = 1$,
- iv) $P[N(t+h) = n+m | N(t) = n] = o(h)$, si $m > 1$,
- v) $P[N(t+h) = n+m | N(t) = n] = 1 - \lambda(t)h + o(h)$, si $m = 0$.

Donde $o(h)$ es una función tal que $\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0$. En otras palabras, la probabilidad de observar un evento durante el intervalo infinitesimal de tiempo t y $t+h$ cuando $h \rightarrow 0$ es $\lambda(t)h$. La probabilidad de observar más de un evento durante el mismo intervalo es insignificante.

Veamos otra definición más común de encontrar en libros de texto.

Definición 30

Un Proceso de Poisson de parámetro $\lambda > 0$ es un proceso continuo $\{X(t), t \geq 0\}$

con espacio de estados $\lambda = \{0, 1, \dots\}$, con trayectorias no decrecientes y que cumple con las siguientes propiedades:

- i) $X(0) = 0$,
- ii) $\{X(t), t \geq 0\}$ tiene incrementos independientes,
- iii) $X_{t+s} - X_s \sim Poi(\lambda t)$, para cualesquiera $s \geq 0, t > 0$.

Esta es posiblemente la definición del proceso de Poisson que con más frecuencia se puede encontrar en las referencias. A partir de ella inmediatamente sabemos que la variable X_t tiene distribución Poisson(λt). La independencia de los incrementos es explícita, y la estacionariedad de los mismos aparece de manera implícita en el tercer postulado.

Definición 31

A la función $A(t)$ definida como

$$A(t) = \int_0^t \lambda(y) dy$$

se le conoce como la **función media** de un proceso Poisson no homogéneo.

El proceso de Poisson no homogéneo posee mucha similitud al proceso Poisson homogéneo excepto por dos aspectos: en lugar de la constante λ se escribe ahora la función $\lambda(t)$, y la hipótesis de incrementos estacionarios ya no aparece. Esto es consecuencia de que el parámetro varía con el tiempo, generalmente de manera decreciente. Es decir, la distribución de probabilidad de la variable incremento $N_{t+s} - N_t$ depende de los valores de la función λ en el intervalo $(s, s+t)$. Sin embargo, y en completa analogía con el caso homogéneo, la variable N_t continua teniendo distribución Poisson.

Proposición 1

Si $\{N(t), t \geq 0\}$ es un proceso Poisson no estacionario con función de intensidad $\lambda(t)$, $t \geq 0$, entonces $N(t+s) - N(s)$ tiene distribución Poisson con media

$$A(t+s) - A(s) = \int_s^{s+t} \lambda(y) dy$$

es decir, para $n = 0, 1, 2, \dots$

$$P[X_t = n] = e^{-\Lambda(t)} \frac{[\Lambda(t)]^n}{n!}$$

Las trayectorias de un proceso de Poisson no homogéneo son semejantes a las trayectorias de un proceso de Poisson, es decir, son trayectorias no decrecientes y con saltos unitarios hacia arriba, pero la frecuencia promedio con la que aparecen los saltos cambia a lo largo del tiempo. De manera análoga al caso homogéneo, los incrementos de este proceso también tienen distribución Poisson.

Ahora, si $\Lambda(t) = \int_0^t \lambda(s) ds$, entonces $N(t) \sim Poi(\Lambda(t))$ y

$$P[T_1 < t] = P[N(t) = 0] = e^{-\Lambda(t)}$$

Derivando obtenemos la densidad

$$f_{T_1}(t) = -\frac{d}{dt} P[T_1 > t] = \lambda(t) e^{-\int_0^t \lambda(s) ds} = \lambda(t) e^{-\Lambda(t)}$$

para $t \geq 0$. De manera similar podemos obtener la densidad conjunta

$$f_{\tau_1, \tau_2}(u, v) = \lambda(u) e^{-\Lambda(u)} \lambda(v) e^{-(\Lambda(v) - \Lambda(u))} = \lambda(u) \lambda(v) e^{-\Lambda(v)}$$

y haciendo ahora un cambio de variables

$$f_{T_1, T_2}(s, t) = \lambda(s) e^{-\Lambda(s)} \lambda(s+t) e^{-(\Lambda(s+t) - \Lambda(s))} = \lambda(s) \lambda(s+t) e^{-\Lambda(s+t)}$$

de modo que T_1 y T_2 no son independientes si $\lambda(s)$ no es constante. Las relaciones anteriores se pueden generalizar de la siguiente manera

$$f_{\tau_1, \dots, \tau_n}(v_1, \dots, v_n) = \lambda(v_1) \cdots \lambda(v_n) e^{-\Lambda(s+t)}$$

y

$$f_{T_1, \dots, T_n}(t_1, \dots, t_n) = \lambda(t_1) \cdots \lambda(t_1 + \dots + t_n) e^{-\Lambda(t_1 + \dots + t_n)}$$

Proposición 2

Sea $\{X_t : t \geq 0\}$ un Proceso Poisson no homogéneo de parámetro $\lambda(t)$ y $\Lambda(t) =$

$\int_0^t \lambda(s) ds$. Defina la función

$$\Lambda^{-1}(t) = \inf\{u \geq 0 : \Lambda(u) = t\}$$

Entonces el proceso $\{X_{\Lambda^{-1}(t)} : t \geq 0\}$ es un proceso Poisson de parámetro 1.

Demostración. Usando la Definición 30. Sea $\{X_{\Lambda^{-1}(t)} : t \geq 0\}$ un Proceso Poisson que empieza en 0 y tiene incrementos independientes, si se consideran los tiempos $0 \leq t_1 < t_2 < \dots < t_n$. bajo la función creciente $\Lambda^{-1}(t)$, estos tiempos se transforman en una nueva colección monótona de tiempos

$$0 \leq \Lambda^{-1}(t_1) < \Lambda^{-1}(t_2) < \dots < \Lambda^{-1}(t_n)$$

Por lo tanto las siguientes variables son independientes

$$X_{\Lambda^{-1}(t_1)}, X_{\Lambda^{-1}(t_2)} - X_{\Lambda^{-1}(t_1)}, \dots, X_{\Lambda^{-1}(t_n)} - X_{\Lambda^{-1}(t_{n-1})}$$

Finalmente, para cualesquiera tiempos $s, t \geq 0$, el incremento $X_{\Lambda^{-1}(t+s)} - X_{\Lambda^{-1}(t)}$ tiene distribución Poisson de parámetro

$$\Lambda(\Lambda^{-1}(t+s)) - \Lambda(\Lambda^{-1}(t)) = (t+s) - t = s$$

■

Apéndice C

Medidas de Riesgo

El riesgo se define normalmente como “la posibilidad de malas consecuencias, pérdida o exposición a un accidente”. En el caso del riesgo financiero, podríamos definirlo como “cualquier evento o acción que pueda afectar adversamente la capacidad de una organización para lograr sus objetivos y ejecutar sus estrategias” o, alternativamente, “la probabilidad de pérdida o rendimientos inferiores a los esperados”.

El riesgo se relaciona fuertemente con la incertidumbre y por tanto con la probabilidad. Considere los siguientes ejemplos: un inversor que tiene acciones en una empresa en particular; una compañía de seguros que ha vendido una póliza de seguro; una persona que decide convertir una hipoteca de tasa fija en una variable. Todas estas situaciones tienen algo importante en común: el inversor posee hoy un activo con un valor futuro incierto. Esto es muy claro en el caso de las acciones. Para la compañía de seguros, la póliza vendida puede o no ser activada por el evento subyacente cubierto. En el caso de una hipoteca, nuestra decisión de hoy de celebrar este acuerdo de refinanciamiento cambiará (para bien o para mal) los reembolsos futuros.

Por tanto, la aleatoriedad juega un papel crucial en la valoración de los productos actuales que posee el inversor, la compañía de seguros o el propietario de la vivienda. Para modelar estas situaciones, un matemático ahora definiría una posición de riesgo de un período (o simplemente riesgo) X como una función en el espacio de probabilidad (Ω, \mathcal{F}, P) .

El siguiente apéndice pretende dar una breve introducción a la función *Value at Risk* (VaR), que es probablemente la medida de riesgo más usada en la rama de

administración de riesgos. También hablaremos de algunas pruebas de back-testing que usualmente se emplean cuando calculamos el VaR.

Las definiciones fueron tomadas principalmente de *Quantitative Risk Management: Concepts, Techniques and Tools* (MCNEIL A.J., FREY R. Y EMBRECHETS P. [48]) y *Internal Models Validation in Brazil: Analysis of VaR Backtesting Methodologies* (RODRIGUEZ, A., & SILVERIA, C. & ARAÚJO, G. & EIRAS DAS NEVES, M., [35]).

C.1 Conceptos básicos

Consideraremos a un portafolio de instrumentos financieros arbitrarios, cuyo valor en el tiempo es una variable aleatoria. Denotaremos el valor de este portafolio a tiempo, t , como $V(t)$.

Definición 32

Supongamos que tenemos un portafolio financiero, entonces, para un horizonte de tiempo dado, digamos Δ , y para un tiempo s , definimos a la **pérdida del portafolio** como:

$$L_{[s,s+\Delta]} = -(V(s + \Delta) - V(s))$$

$L_{[s,s+\Delta]}$ es una variable aleatoria cuya distribución suele llamarse distribución de pérdidas.

C.1.1 Medidas de Riesgo

Para lo siguiente supondremos que la pérdida $L_{[s,s+\Delta]}$ es una variable aleatoria definida sobre cierto espacio de probabilidad, $(\Omega, \mathcal{F}, \mathbb{P})$, para cierto horizonte de tiempo Δ determinado.

Denotaremos por $L^0(\Omega, \mathcal{F}, \mathbb{P})$ al espacio de todas las variables definidas en (Ω, \mathcal{F}) que son casi seguramente finitas, en riesgo financiero se suele usar el conjunto $\mathcal{M} \subset L^0(\Omega, \mathcal{F}, \mathbb{P})$ de las variables que interpretamos como la pérdida de un portafolios en cierto horizonte de tiempo.

Asumiremos que \mathcal{M} es un cono convexo, es decir, que se cumple que:

1. Si $L_1 \in \mathcal{M}$ y $L_2 \in \mathcal{M}$ entonces: $L_1 + L_2 \in \mathcal{M}$.

2. Si $L_1 \in \mathcal{M}$ y $\lambda \in \mathbb{R}$ entonces: $\lambda L_1 \in \mathcal{M}$.

Con estas ideas, daremos los axiomas de coherencia para las medidas de riesgo. Una medida de riesgo es una función $\varrho : \mathcal{M} \rightarrow \mathbb{R}$.

Decimos que ϱ es una medida de riesgo coherente, si cumple los siguientes axiomas.

Axioma 1. Invarianza bajo traslación.

Para todas $L \in \mathcal{M}$ y $\lambda \in \mathbb{R}$ se tiene que: $\varrho(L + \lambda) = \varrho(L) + \lambda$.

Axioma 2. Subaditividad.

Para todas $L_1, L_2 \in \mathcal{M}$ se tiene que: $\varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2)$.

Axioma 3. Homogeneidad positiva.

Para todas $L \in \mathcal{M}$ y $0 < \lambda \in \mathbb{R}$ se tiene que: $\varrho(\lambda L) = \lambda \varrho(L)$.

Axioma 4. Monotonicidad.

Para todas $L_1, L_2 \in \mathcal{M}$ con $L_1 \leq L_2$ se tiene que: $\varrho(L_1) \leq \varrho(L_2)$.

C.2 Valor en Riesgo

En términos generales, el VaR es una herramienta cuantitativa que tiene como objetivo evaluar la posible pérdida en la que se puede incurrir en un período de tiempo determinado y para una cartera de activos determinada. Durante los últimos diez años, esta técnica ha sido utilizada cada vez más por bancos y reguladores de todo el mundo como una forma de estimar las posibles pérdidas relacionadas con la tenencia y negociación de activos financieros.

Como se mencionó anteriormente, el principal objetivo del VaR es evaluar la posible pérdida en la que puede incurrir un comerciante o una institución financiera durante un período de tiempo determinado. Debido a razones regulatorias, el horizonte temporal suele ser un período de diez días y los modelos se evalúan en

función de los rendimientos diarios. Sin embargo, para los participantes activos del mercado, como los creadores de mercado, el horizonte temporal de sus rendimientos es mucho más corto y el riesgo comercial correspondiente por tanto, debe evaluarse en intervalos de tiempo igual de cortos. Veamos la definición matemática del VaR.

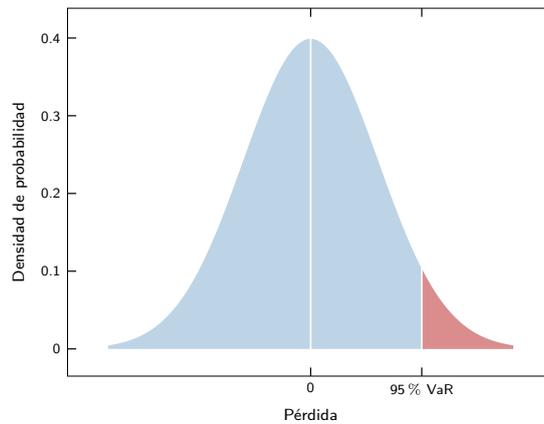


Figura C.1: Ejemplo de VaR al 95 %

Considere una cartera de activos de riesgo y un horizonte de tiempo fijo, y denote por

$$F_L(l) = P(L \leq l)$$

Queremos definir una estadística basado en F_L que mide la severidad del riesgo de mantener nuestra cartera durante el período de tiempo. Un candidato obvio es la pérdida máxima posible, dada por $\inf\{l \in \mathbb{R} : F_L(l) = 1\}$, una medida de riesgo importante en el reaseguro. Sin embargo, en la mayoría de los modelos de interés, el soporte de F_L es no acotado, por lo que la pérdida máxima es simplemente infinita.

El valor en riesgo es una extensión directa de la pérdida máxima, la idea es simplemente reemplazar “pérdida máxima” por “pérdida máxima que no se excede con una alta probabilidad dada”, a esta probabilidad se lo conoce como nivel de confianza.

Definición 33

Dado algún nivel de confianza $\alpha \in (0, 1)$. El VaR de nuestra cartera en el nivel de

confianza α está dado por el número más pequeño l , de modo que la probabilidad de que la pérdida L exceda l no sea mayor que $(1 - \alpha)$. Formalmente,

$$\text{VaR}_\alpha(L) = \inf\{l \in \mathbb{R} : P[L > l] \leq 1 - \alpha\}$$

En términos probabilísticos, VaR es, por tanto, simplemente un cuantil de la distribución de pérdidas. Los valores típicos de α son $\alpha = 0.95$ o $\alpha = 0.99$; en la gestión del riesgo de mercado el horizonte temporal Δ suele ser de 1 o 10 días. La Figura C.1 ilustra la noción de VaR. Artzner y col. (1999) critican al VaR alegando que no cumple el Axioma 2 de subaditividad de una medida de riesgo, es decir, que el VaR de una cartera puede ser superior a la suma de los VaR de los activos individuales de la cartera. En otras palabras, el VaR no es una medida de riesgo “coherente”. Este problema se debe al hecho de que el VaR es un cuantil de la distribución de pérdidas y ganancias y no una expectativa, por lo que la forma de la cola antes y después de la probabilidad de VaR no tiene por qué tener relación con el número de VaR real.

C.3 Value At Risk

En aplicaciones, el propósito de ajustar un modelo a nuestros datos, sería estimar el cuantil condicional de la distribución predictiva para el rendimiento del día siguiente:

$$z_q^t = \inf\{z \in \mathbb{R} : F_{Z_{t+1}|\mathcal{F}_t}(z) \geq q\}$$

Usando los datos observados hasta tiempo t . En la práctica tomaremos $q = 0.95$ o $q = 0.99$, para definir nuestro VaR Condicional, que se puede escribir como:

$$P(Z_{t+1} > z | \mathcal{F}_t) = P(Z_{t+1} - u > z - u | Z_{t+1} > u, \mathcal{F}_t) \times P(Z_{t+1} > u | \mathcal{F}_t),$$

Donde el segundo término de la expresión puede ser aproximado por la probabilidad condicional de un evento en $(t, t + 1]$, esto es,

$$1 - P\{N(t, t + 1) = 0 | \mathcal{F}_t\} = 1 - \exp\left(-\int_0^{t+1} \lambda_{\mathcal{F}}(u) du\right)$$

El primer término puede ser calculado usando un modelo de Distribución Generalizada de Pareto:

$$Z_{t+1} - u | \mathcal{F}_t; Z_{t+1} \geq u \sim GPD_{\xi, \sigma}$$

Buscamos el cuantil z_q^t que resuelva la ecuación

$$P(Z_{t+1} > z_q^t | \mathcal{F}_t) = 1 - q,$$

Con los parámetros resultantes de la estimación por máxima verosimilitud. Esto tiene solución cuando $1 - q \leq P(Z_{t+1} > u | \mathcal{F}_t)$.

C.4 Back-testing para el VaR

Podemos clasificar las pruebas para la validación de modelos VaR en tres grupos: (i) las que validan si el número de excepciones observadas es consistente con el número de excepciones esperadas; (ii) las que analizan el tamaño de las excepciones ocurridas y; (iii) las que verifiquen la consistencia entre los modelos de riesgo y sus supuestos de construcción. El primer grupo está compuesto por el Procedimiento del Comité de Basilea, la frecuencia básica de la prueba de pérdidas de cola (o Kupiec) y el enfoque condicional de Christoffersen (1998), que se basa no sólo en la frecuencia de las excepciones observadas sino también de la independencia entre ellas.

C.4.1 Prueba Binomial

Este método es una extensión del backtesting de Christoffersen (1998) que se analiza más adelante. Esta prueba establece la siguiente estadística

$$T = \sum_{i=1}^k I_{\{x_{n-k+i} > VaR_\alpha(n-k+i)\}}$$

donde $VaR_\alpha(n-k+i) - x_{n-k+i}$ son los excedentes del modelo o excepciones.

De estar bien construido el modelo, el porcentaje de datos que exceden su VaR estimado debe de ser $1 - \alpha \approx \frac{T}{k}$.

Si el modelo es correcto entonces la variable T debe comportarse como una Binomial con parámetros $(k, 1 - \alpha)$. Para probar que verifica este último hecho es la prueba binomial.

C.4.2 Prueba de Kupiec

El test más utilizado es el de frecuencia de pérdidas en los extremos propuesto por Kupiec (1995), que se basa en la frecuencia de pérdidas que superan el VaR.

Sea x el número de “fallos” o excepciones (el número de casos en los que la pérdida supera el VaR previsto) en una muestra de tamaño n . Si el modelo de VaR es correcto, sabemos que x sigue la distribución binomial con parámetro (n, p) . Bajo la hipótesis nula, el modelo de pronóstico es correcto y la frecuencia observada de pérdidas en los extremos es consistente con la frecuencia de excepciones predichas por el modelo. La prueba se basa en la razón de verosimilitud (LR) para la hipótesis nula, que viene dada por

$$LR = -2 \log \left[(1 - p^*)^{n-x} (p^*)^x \right] + 2 \log \left[\left(1 - \frac{x}{n} \right)^{n-x} \left(\frac{x}{n} \right)^x \right]$$

donde p^* es la probabilidad de excepciones bajo la hipótesis nula, n es el tamaño de la muestra, x es el número de excepciones y $\frac{x}{n}$ es la tasa de excepciones. Kupiec llama a la prueba LR_{UC} la prueba de proporción de fallas (PF). Bajo la hipótesis nula, la probabilidad de una excepción (p) es igual al nivel de significancia (p^*) del VaR y PF tiene una distribución chi-cuadrada con un grado de libertad.

La región, donde la hipótesis nula no pueden ser rechazada, está determinada por la intersección entre el PF y la función chi-cuadrada. Para un tamaño de muestra y un nivel de significación dados, la prueba determina límites inferior y superior donde la hipótesis nula no puede ser rechazado al nivel de significación de la prueba, es decir:

- Si $\chi_1^2 > LR_{UC}$ no se rechaza $H_0 : p = p^*$
- Si $\chi_1^2 < LR_{UC}$ se rechaza $H_0 : p = p^*$

Sin embargo, la prueba presenta baja potencia en muestras pequeñas, por lo que existe una probabilidad significativa de no rechazar la hipótesis nula cuando es falsa.

Al igual que el criterio de Basilea, esta prueba se basa exclusivamente en la frecuencia de pérdidas extremas. También descuida información potencialmente importante como el tamaño de las pérdidas extremas y la dependencia temporal en el comportamiento de las excepciones. Este es un tema importante dado las evidencias de volatilidades condicionales en series financieras (Fierli, 2002).

C.5 Función de Riesgo

Un concepto útil para modelar la probabilidad de que a un individuo le ocurra cierto suceso a lo largo del tiempo es la función de riesgo, o *hazard function* implícita en una función de distribución. Para una variable aleatoria X , la función de supervivencia se define como:

$$S(x) = P(X > x) = 1 - P(X \leq x), \quad x > 0$$

que representa la probabilidad de que un individuo, que sigue la distribución de X , sobreviva en el momento x . La función de riesgo de X es entonces definida por

$$h(x) = \frac{f(x)}{S(x)}$$

donde f y S son la función de densidad de probabilidad y la función de supervivencia de X , respectivamente.

La función de riesgo no es una densidad ni una probabilidad. Sin embargo, podemos pensar en ella como la probabilidad de riesgo en un período de tiempo infinitesimalmente pequeño entre $y + \partial y$ dado que el sujeto ha sobrevivido hasta el momento y . En particular, ya que por definición

$$f_Y(y) = \lim_{\partial y \rightarrow 0} \frac{F_Y(y + \partial y) - F_Y(y)}{\partial y}$$

podemos escribir a la función de riesgo como:

$$\begin{aligned} h_Y(y) &= \lim_{\partial y \rightarrow 0} \frac{F_Y(y + \partial y) - F_Y(y)}{\partial y \cdot S_Y(y)} \\ &= \lim_{\partial y \rightarrow 0} \frac{P(y < Y \leq y + \partial y)}{\partial y \cdot S_Y(y)} \\ &= \lim_{\partial y \rightarrow 0} \frac{P(y < Y \leq y + \partial y | Y > y)}{\partial y} \end{aligned}$$

Entonces:

$$\begin{aligned}h_Y(y) &= \frac{f_Y(y)}{S_Y(y)} \\ &= \frac{f_y(y)}{1 - F_Y(y)} \\ &= -\frac{\partial}{\partial y} \log[1 - F_Y(y)] \\ &= -\frac{\partial}{\partial y} \log[S_Y(y)]\end{aligned}$$

por lo tanto

$$S_Y(y) = \exp[H_Y(y)]$$

donde

$$H_Y(y) = \int_0^y h_Y(t) dt$$

La función $H_Y(y)$ se denomina función de riesgo acumulativo. Como la función de riesgo, la función de riesgo acumulativo no es una probabilidad. Sin embargo, también es una medida de riesgo: cuanto mayor es el valor de $H_Y(y)$, mayor es el riesgo de falla en el tiempo y .

Bibliografía

- [1] DACOROGNA, M. M., GENÇAY, R., MÜLLER, U. A., OLSEN, R. B. AND PICTET, O. V. (2001), “*An Introduction to High-Frequency Finance*”; Academic Press.
- [2] TSAY R. S. (2010), “*Analysis of financial time series*”;(3ra ed.) John Wiley & Sons.
- [3] BOB PISANI (2014), “*Plundered by harpies: An Early History of High-Speed Trading*”; Financial History.
- [4] BROWNLEES,C.T., GALLO,G.M., (2006), “*Financial econometric analysis at ultra-high frequency: Data handling concerns*”; Computational Statistics & Data Analysis, Elsevier, vol. 51(4), pp. 2232-2245.
- [5] CHAVEZ-DEMOULIN, V., MCGILL, J.A., (2012), “*High-frequency financial data modeling using Hawkes processes*”; Journal of Banking & Finance, Elsevier, vol. 36(12), pp. 3415-3426.
- [6] MICHAEL J. MCGOWAN (2010), “*The rise of computerized high frequency trading: use and controversy*”; Duke Law & technology review, No. 016.
- [7] KHALIL ANTOINE AL DAYRI (2011), “*Market Microstructure and Modeling of the Trading Flow*” (tesis de doctorado); L'École Polytechnique, Francia.
- [8] ANUJ AGARWAL (2017), “*High Frequency Trading: Evolution and the Future*”; Capgemini.
- [9] A. PLATANIA & ROGERS, L., (2004), “*Particle Filtering in High Frequency Data*”.

- [10] BROWNLEES, C.T., GALLO, G.M., (2006), “*Financial econometric analysis at ultra-high frequency: Data handling concerns*”; Computational Statistics & Data Analysis, Elsevier, vol. 51(4), pp. 2232-2245.
- [11] BAUWENS, LUC AND GIOT, PIERRE, (2001), “*Econometric Modelling of Stock Market Intraday Activity*”; Springer.
- [12] WEI SUN (2007), “*Quantitative Methods in High-Frequency Financial Econometrics: Modeling Univariate and Multivariate Time Series*” (tesis de doctorado); Universität Karlsruhe, Alemania.
- [13] CONT, RAMA (2002), “*Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues*”; vol. 1 Quantitative Finance.
- [14] WANG, FENGZHONG AND YAMASAKI, KAZUKO AND HAVLIN, SHLOMO AND STANLEY, H. (2006), “*Scaling and memory of intraday volatility return intervals in stock market*”; vol. 3, Physical review. E, Statistical, nonlinear, and soft matter physics.
- [15] RK (2015), “*MLE of Hawkes’ Self-Exciting Point Process*”.
- [16] MCNEIL A.J., FREY R. Y EMBRECHETS P. (2005), “*Quantitative Risk Management: Concepts, Techniques and Tools*”; primera edición, USA, Princeton University Press.
- [17] ALAN G HAWKES (1971), “*Point spectra of some mutually exciting point processes*”; In: Journal of the Royal Statistical Society. Series B (Methodological), pp. 438-443.
- [18] EMMANUEL BACRY (2013), “*Modelling microstructure noise with mutually exciting point processes*”; In: Quantitative Finance 13.1, pp. 65- 77.
- [19] KATARZYNA OBRAL (2016), “*Simulation, Estimation and Applications of Hawkes Processes*”; University of Minnesota.
- [20] YOSHIKO OGATA (1988), “*Statistical models for earthquake occurrences and residual analysis for point processes*”; In: Journal of the American Statistical association 83.401, pp. 9-27.
- [21] PATRICIA REYNAUD-BOURET & SOPHIE SCHBATH (2010), “*Adaptive estimation for Hawkes processes; application to genome analysis*”; In: The Annals of Statistics 38.5, pp. 2781-2822.

- [22] ALAN G HAWKES (1971), “*Spectra of some self-exciting and mutually exciting point processes*”; In: *Biometrika* 58.1, pp. 83–90.
- [23] DJ DALEY & D VERE-JONES (2002), “*An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*”; Springer Science & Business Media.
- [24] COX, DAVID AND LEWIS, P. (1966), “*The statistical analysis of series of events*”; isbn = 978-94-011-7803-7, journal = SERBIULA (sistema Librum 2.0), doi = 10.1007/978-94-011-7801-3.
- [25] Y. OGATA (1981), “*On Lewis’ simulation method for point processes*”; in *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 23-31, doi: 10.1109/TIT.1981.1056305.
- [26] E. BACRY & S. DELATTRE, M. HOFFMANN & J.F. MUZY (2015), “*Hawkes Processes in Finance*”; Springer.
- [27] E. BACRY & S. DELATTRE, M. HOFFMANN & J.F. MUZY (2012), “*Scaling limits for Hawkes processes and application to financial statistics.*”; arXiv:1202.0842.
- [28] P. J. LAUB & T. TAIMRE & P. K. POLLETT (2015), “*Hawkes Processes.*”; arXiv:1507.02822v1
- [29] PATRICK LAUB, (2014), “*Hawkes Processes: Simulation, Estimation, and Validation.*” (tesis de posgrado); Universidad de Queensland, Australia.
- [30] MORENO TRUJILLO, J. F. (2018). “*Modelo estocástico para el precio de activos riesgosos utilizando procesos Hawkes*”. *odeon*, 15, 161-172. doi: <https://doi.org/10.18601/17941113.n15.06>
- [31] LAW & BARON & VIENS, FREDER, (2016), “*Hawkes Processes and Their Applications to High-Frequency Data Modeling.*”. 183-219, doi = 10.1002/9781118593486.ch7
- [32] A. G. Hawkes & D. Oakes, (1974), “*A cluster process representation of a self-exciting process.*”. *Journal of Applied Probability*, pp. 493–503.
- [33] CHAVEZ-DEMOULIN, V. AND DAVISON, A. AND MCNEIL, ALEXANDER, (2005), “*Estimating value-at-risk: A point process approach.*”. Vol. 5 pp. 227-234, doi = 10.1080/14697680500039613.

- [34] F. LORENZEN, (2012), “*Analysis of Order Clustering Using High Frequency Data: A Point Process Approach.*”. Tilburg School of Economics and Management, Finance Department.
- [35] RODRIGUEZ, A., & SILVERIA, C. & ARAÚJO, G. & EIRAS DAS NEVES, M. (2017), “*Internal Models Validation in Brazil: Analysis of VaR Backtesting Methodologies.*”. Revista Brasileira de Finanças, vol. 4, núm. 1, 2006, pp. 363-384.
- [36] P.J. BROCKWELL & R.A. DAVIS (2002), “*Introduction to Time Series and Forecasting*”; Springer.
- [37] P.J. BROCKWELL & R.A. DAVIS (1991), “*Time Series: Theory and methods.* Springer.”; Springer.
- [38] ENGLE, R. (1982), *Autoregressive Conditional Heterosedasticity with Estimates of the Variance of United Kingdom Inflation*, Econometrica, vol. 50, issue 4.
- [39] RINCÓN, LUIS (2012), “*Introducción a los procesos estocásticos*”; Facultad de Ciencias, UNAM.
- [40] KATARZYNA OBRAL (2016), “*Simulation, Estimation and Applications of Hawkes Processes*”; Universidad de Minesota.
- [41] SNYDER, D., MILLER, M., RANDOM (1991), “*Point Processes in Time and Space*”; 2nd Edition, New York: Springer Verlag.
- [42] NIKOLAUS HAUTSCH (2011), “*Econometrics of Financial High-Frequency Data*”; 1st Edition, New York: Springer Verlag.
- [43] MORENO TRUJILLO, J. F. (2018). “*Modelo estocástico para el precio de activos riesgosos utilizando procesos Hawkes*”. odeon, 15, 161-172. doi: <https://doi.org/10.18601/17941113.n15.06>
- [44] RIZOIU, MARIAN-ANDREI AND LEE, YOUNG AND MISHRA, SWAPNIL (2017). “*A Tutorial on Hawkes Processes for Events in Social Media*”. pp. 191-218. <https://doi.org/10.48550/arXiv.1708.06401>
- [45] RINCÓN, LUIS, (2017), “*Introducción a la Probabilidad.*”. (2014) Facultad de Ciencias, UNAM, México, pp 217-232.

- [46] LEE, SE YOON & H.T. KIM, J., (2017), “*Exponentiated Generalized Pareto Distribution: Properties and applications towards Extreme Value Theory.*”. (2017), doi=1708.01686v1.
- [47] SILVEY, S.D., (2017), “*Statistical Inference.*”. (1975) Second Edition, Chapman and Hall, Londres, pp 140-141.
- [48] MCNEIL A.J., FREY R. Y EMBRECHETS P. (2005), “*Quantitative Risk Management: Concepts, Techniques and Tools*”; primera edición, USA, Princeton University Press.
- [49] ALARCÓN, VERÓNICA (2019), “*Medición de riesgo en portafolios aplicando un modelo GARCH-EVT-COPULA*”; CIMAT.
- [50] T. OZAKI (1979), “*Maximum likelihood estimation of Hawkes’ self-exciting point processes.*”; Ann Inst Stat Math 31, 145–155 (1979). <https://doi.org/10.1007/BF02480272>
- [51] CHRISTOFFERSEN, P. (1998), “*Evaluating Interval Forecasts.*”; Evaluating Interval Forecasts. International Economic Review, 39, 841-862.
- [52] MANDELBROT, BENOIT (1967), “*The Variation of Some Other Speculative Prices.*”; The Journal of Business 40, no. 4 (1967): 393–413.
- [53] MANDELBROT, BENOIT (1967), “*The Variation of Some Other Speculative Prices.*”; The Journal of Business 40, no. 4 (1967): 393–413.
- [54] ANDERSEN, TORBEN G. & BOLLERSLEV, TIM, (1997), “*Intraday periodicity and volatility persistence in financial markets*”; Journal of Empirical Finance, Elsevier, vol. 4(2-3), pages 115-158.
- [55] ANDREW W. LO , A. CRAIG MACKINLAY, (1990), “*When Are Contrarian Profits Due to Stock Market Overreaction?*”; The Review of Financial Studies, Volume 3, Issue 2, Pages 175–205.
- [56] C.A.E. GOODHART, L. FIGLIUOLI, (1991), “*Every minute counts in financial markets*”; Journal of International Money and Finance, Volume 10, Issue 1, Pages 23-52.
- [57] ROBERT A. WOOD AND THOMAS H. MCINISH AND J. KEITH ORD, (1991), “*An Investigation of Transactions Data for NYSE Stocks*”; Journal of Finance, Volume 40, Pages 723-739.

-
- [58] CHAVEZ-DEMOULIN, V. & EMBRECHTS, PAUL & HOFERT, MARIUS (2014), ‘*An Extreme Value Approach for Modeling Operational Risk Losses Depending on Covariates*’; *Journal of Risk and Insurance*. 83. 10.1111/jori.12059.
- [59] HERRERA, RODRIGO & SCHIPP, BERNHARD (2014), ‘*Value at risk forecasts by extreme value models in a conditional duration framework*’; *Journal of Empirical Finance*, Elsevier, vol. 23(C), pages 33-47.