[Rama Cont]

Statistical Modeling of High-Frequency Financial Data

[Facts, models, and challenges]

The availability of high-frequency data on transactions, quotes, and order flow in electronic order-driven markets has revolutionized data processing and statistical modeling techniques in finance and brought up new theoretical and computational challenges. Market dynamics at the transaction level cannot be characterized solely in terms the dynamics of a single price, and one must also take into account the interaction between buy and sell orders of different types by modeling the order flow at the bid price, ask price, and possibly other levels of the limit order book. We outline the empirical characteristics of high-frequency financial time series and provide an overview of stochastic models for the continuous-time dynamics of a limit order book, focusing in particular on models that describe the limit order book as a queuing system. We describe some applications of such models and point to some open problems.

© iSTOCKPHOTO.COM/NIKADA

ELECTRONIC ORDER-DRIVEN MARKETS

In recent years, automated trading and dealing have largely replaced floor-based trading in equity markets. Electronic Crossing Networks (ECNs) such as Archipelago, Instinet, Brut, and Tradebook providing order-driven trading systems have captured a large share of the market. In contrast to markets where a market maker or specialist centralizes buy and sell orders and provides liquidity by setting bid and ask quotes, these electronic platforms aggregate all outstanding limit orders in a limit order book that is available to market participants, and market orders are executed against the best available prices, in a mechanical manner. As a result of the ECNs' popularity, established exchanges such as the NYSE, Nasdaq, the Tokyo Stock Exchange, Toronto Stock Exchange, Vancouver Stock Exchange, Euronext (Paris, Amsterdam, Brussels), and the London Stock Exchange have fully or partially adopted electronic order-driven platforms.

At the same time, the frequency of submission of orders has increased and the time to execution of market orders on these electronic markets has dropped from more than 25 ms in 2000 to less than 1 ms in 2010. As shown in Table 1, thousands of orders are submitted in a 10-s interval, resulting in frequent updates in bid and ask quotes, up to 100,000 times a day, as well as frequent changes in transaction prices.

As a result, the evolution of supply, demand, and price behavior in equity markets is being increasingly recorded: this data is available to market participants in real time and for researchers in the form of high-frequency databases. The analysis of such high-frequency data constitutes a challenge, not the least because of their sheer volume and complexity. These data provide us with a detailed view of the complex dynamic process through which the market "digests" the inflow of supply and demand to generate the price [1], [2].

The large volume of data available, the presence of statistical regularities in the data, and the mechanical nature of execution of orders makes order-driven markets interesting candidates for statistical analysis and stochastic modeling.

At a fundamental level, statistical analysis and modeling of high-frequency data can provide insight into the interplay between order flow, liquidity, and price dynamics [3]–[5] and might help bridge the gap between market microstructure theory [6]–[11], which has provided useful insights by focusing on models of price formation mechanism in stylized equilibrium settings, and "black box" stochastic models used in financial risk management, which represent the price as an exogenous random process.

At the level of applications, models of high-frequency data provide a quantitative framework for market making [12] and optimal execution of trades [13]–[16]. Another obvious application is the development of statistical models in view of predicting short term behavior of market variables such as price, trading volume and order flow.

The study of high-frequency market dynamics is also important for risk management and regulation. Even though the horizons traditionally considered by risk managers and regulators have been longer ones (typically, daily or longer), trading strategies at different frequencies may interact in a complex manner, leading to ripples across time scales that propagate from high frequency to low frequency and even leading to possible market disruptions, as shown by the Flash Crash of May 2010 [17], [18].

TRADES, QUOTES, AND ORDER FLOW

High-frequency data in finance involve time series of prices and quantities associated to these prices. However, when one looks at transaction-level frequencies, a security is not characterized by a single price: one must distinguish between transaction data, which record quantities that were actually traded and the prices at which they were traded, and quotes, which are proposals to buy or sell a given quantity of an asset at a given price.

[**THE STUDY OF HIGH-FREQUENCY MARKET DYNAMICS IS IMPORTANT FOR RISK MANAGEMENT AND REGULATION.**]

TRANSACTION DATA
Transaction data record trades that occur on an exchange, each record being associated with a time stamp, a price, and quantity (trade volume, in number of shares). Transaction data present several particularities that differentiate them from financial time series sampled, say, at daily frequency [2], [19].

[TABLE 1] AVERAGE NUMBER OF ORDERS IN 10 S AND NUMBER OF PRICE CHANGES (26 JUNE 2008).		
	AVERAGE NUMBER OF ORDERS IN 10 S	**PRICE CHANGES IN ONE DAY**
CITIGROUP	4,469	12,499
GENERAL ELECTRIC	2,356	7,862
GENERAL MOTORS	1,275	9,016

Price changes are discrete: they are multiples of the minimal price change δ—the "tick"—and more than often equal to one, two, or a few ticks. These price changes are not independent: the autocorrelation function of transaction price returns is significantly negative at the first lag and then it rapidly decreases to zero [20]. This is a typical microstructure effect that disappears when one considers returns at longer time scales.

Trades occur at irregular intervals: the durations between successive trades reflect the intensity of trading activity [19]: they are random and endogenous i.e., linked to the behavior of the price and possibly to previous trading history. With respect to traditional time-series models, this poses the additional problem of modeling the durations. These durations are neither independent—they show significant autocorrelation—nor exponentially distributed: Figure 1 displays a
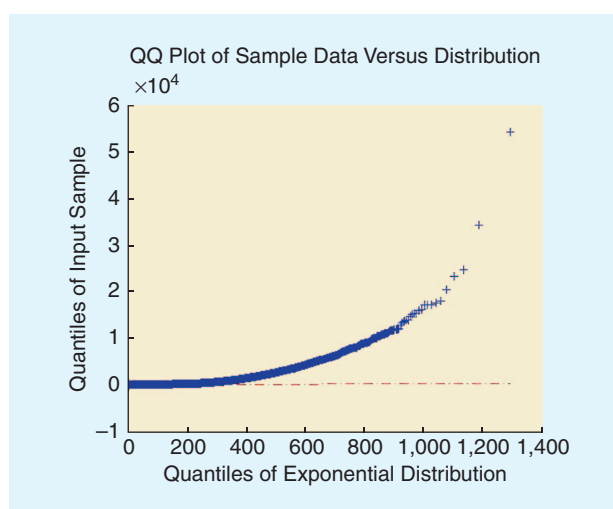
[FIG1] Quantiles of interevent durations compared with quantiles of an exponential distribution with the same mean (Citigroup, June 2008). The dotted line represents the quantiles of an exponential distribution fitted to the observed durations, which are found to be quite different from the empirical quantiles.

quantile-quantile plot comparing the distribution of duration for Citigroup to an exponential distribution. The irregularity of observations makes it much easier to work with continuous-time model, as opposed to time-series models based on a fixed time step. The sizes of these transactions (trading volume) are both heterogeneous and strongly autocorrelated [21].

Another issue with high-frequency data is the presence of seasonality effects. Trading activity is far from uniform during the trading day and exhibits a U-shaped pattern: activity is typically highest at the market open and close, and lowest around lunch time. Quantities related to price volatility and trading volume also follow strong intraday seasonality patterns. As a result, intraday data cannot be treated as stationary and should be first deseasonalized before further analysis. This is done in practice by averaging across many days to obtain a seasonality profile or by using Fourier-based filtering methods. Statistical models of transaction data need to jointly model the durations and the price changes. It is thus natural to model such data as a marked point process [22]. One approach is to consider a stochastic model for durations and, conditional on the duration, model the price changes using a classical autoregressive moving average (ARMA) or generalized autoregressive conditional heteroscedastic (GARCH)

model [23], [24]. The duration τ_i between transactions $i-1$ and i may be represented as

$$\tau_i = \psi_i \epsilon_i,$$

where $(\epsilon_i)_{i \geq 1}$ is a sequence of independent positive random variables with common distribution and $\mathbb{E}[\epsilon_i] = 1$ and the conditional duration $\psi_i = \mathbb{E}[\tau_i | \psi_{i-j}, \tau_{i-j}, j \geq 1]$ is modeled as a function of past history of the process

$$\psi_i = G(\psi_{i-1}, \psi_{i-2}, \ldots, \ldots ; \tau_{i-1}, \tau_{i-2}, \ldots, \ldots).$$

Engle and Russell's autoregressive conditional duration (ACD) model [23] proposes an ARMA (p, q) representation for G

$$\psi_i = a_0 + \sum_{i=1}^{p} a_k \psi_{i-k} + \sum_{i=1}^{q} b_q \tau_{j-k},$$

where $(a_0, \ldots, a_p)$ and $(b_1, \ldots, b_q)$ are positive constants. The ACD-GARCH model of Ghysels and Jasiak [24] combines this model with a GARCH model for the returns. Engle [21] proposes a GARCH-type model with random durations where the volatility of a price change may depend on the previous durations. Variants and extensions are discussed in [25] and [19]. Such models, like ARMA or GARCH models defined on fixed time intervals, have easily computable likelihood functions so maximum likelihood estimation may be done numerically.

QUOTES AND ORDER FLOW

Market participants can post two types of buy/sell orders. A limit order is an order to trade a certain amount of a security at a given price. Limit orders are posted to an electronic trading system and the state of outstanding limit orders posted by market participants can be summarized by stating the quantities posted at each price level: this is known as the limit order book. An example of a limit order book is shown in Figure 2. The lowest price for which there is an outstanding limit sell order is called the ask price and the highest buy price is called the bid price.

A market order is an order to buy/sell a certain quantity of the asset at the best available price in the limit order book. When a market order arrives, it is matched with the best available price in the limit order book and a trade occurs. The quantities available at the bid/ask in the limit order book are decreased by x when a market order of size x is executed. Figure 3 shows the evolution of a limit order book when a market order is executed.

A limit order sits in the order book until it is either executed against a market order or it is canceled. A limit order may be executed very quickly if it corresponds to a price near the bid and the ask but may take a long time if the market price moves away from the requested price or if the requested price is too far from the bid/ask. Alternatively, a limit order can be canceled at any time. On most electronic exchanges, most limit orders are canceled shortly after being posted: for many liquid stocks on the NYSE and NASDAQ, up to 80% of limit orders are canceled within less than a second of posting.

[FIG2] A limit buy order: Buy two at 69,200.

[FIG3] A market sell order of ten.

Quote data contain the evolution of the bid and ask prices and the quantities (number of shares) quoted at these prices. These quantities vary every time a limit order is posted, executed against a market order or canceled. Figure 4 shows the intraday evolution of the ask price for Citigroup on 26 June 2008.

Trades and quotes data are available for a wide range of stock exchanges and markets worldwide. The most well-known trade and quote database is the Trade and Quote (TAQ) database, maintained by New York Stock Exchange (NYSE) since 1993. Trade and quote data allow to reconstruct the sequence of limit and market orders and infer whether the transaction was buyer or seller initiated. The idea is to compare the trade price with the bid and ask price of the prevailing quote, which is the most recent past quote. This is the basis of the Lee and Ready algorithm [26] for assigning signs to trades (see also [27] for a discussion).

Most stocks are simultaneously quoted on many exchanges and, at each point in time, there is a reference National Best Bid and Offer (NBBO) quote but the uniqueness of the NBBO should not hide the multiplicity of order books, with different depths and bid-ask spreads, which is the reality of this market.

Quote data are sometimes known as the "level-1 order book," since they correspond to the first (i.e., best price) level of the order book. More generally high-frequency time series of order book data have become available in the recent years: level-2 order book data give the prices and quantities for the first five nonempty levels of the book on each side, while complete order book data correspond to quantities at all price levels [28]. Order books for many stocks, including liquid ones, have typically less than five active queues on each side of the order book, the rest being empty or sparsely filled [1].

DYNAMICS OF LIMIT ORDER BOOKS

In a limit order market, outstanding limit orders at any given time are represented by the limit order book, which summarizes the state of supply and demand. Not surprisingly, empirical studies [28], [29] show that the state of the order book contains information on short-term price movements, so it is of interest to provide forecasts of various quantities conditional on the state of the order book. Order books presents a rich palette of statistical features [1], [5], [30], [31], [4] that are challenging to incorporate in a statistical model, partly because of the high dimensionality of the order book and the complexity of its evolution. Providing analytically tractable models that enable to compute and/or reproduce conditional quantities that are relevant for trading and intraday risk management has proven to be challenging, given the complex relation between order book dynamics and price behavior.

THE LIMIT ORDER BOOK AS A QUEUEING SYSTEM

The state of the order book is modified by order book events: limit orders (at the bid or ask), market orders and cancellations [28],

> **ORDER BOOKS PRESENTS A RICH PALETTE OF STATISTICAL FEATURES THAT ARE CHALLENGING TO INCORPORATE IN A STATISTICAL MODEL.**

[27], [4]. A limit buy (respectively sell) order of size x increases the size of the bid (respectively ask) queue by x.

Market orders are executed against limit orders at the best available price: a market order decreases of size x the corresponding queue size by x. Limit orders placed at the best available price are automatically executed against market orders according to a priority rule. In many electronic equity markets, priority is based on the time of arrival (first come, first served). However, other priority rules exist: size-based priority grants priority to larger orders, while pro-rata matching, used in some fixed income futures markets such as EUREX, allocates shares to limit orders proportionally to their size [32]. Cancellation of x orders in a given queue reduces the queue size by x.

It is thus natural to represent a limit order book as a multi-class queueing system [28], [4] i.e., a system of (interacting) queues of buy and sell limit orders, executed against market orders ("the server") or cancelled before execution. Point processes and queueing analogies have been used early on for modeling order flows in securities markets [33] and were considered by Smith et al. [4] for limit order book modeling. Queueing models allow for analytical computation of many quantities of interest [28], [34], [35].

A queueing model of limit order book dynamics consists in specifying the arrival rates of different types of order book events (limit buy, limit sell, market buy, market sell, buy/sell cancel) and the rules of execution of these orders. The evolution of the order book is then defined by the order flow. In such models, the dynamics of durations, prices, and bid-ask spreads are endogenous. The arrival rates may be taken constant—leading to a Poisson order flow [28]—or be stochastic processes affected by the state of the order book and by recent order flow [35].

Cont et al. [28] study a Markovian queueing model in which arrivals of market orders and limit orders at each price level are
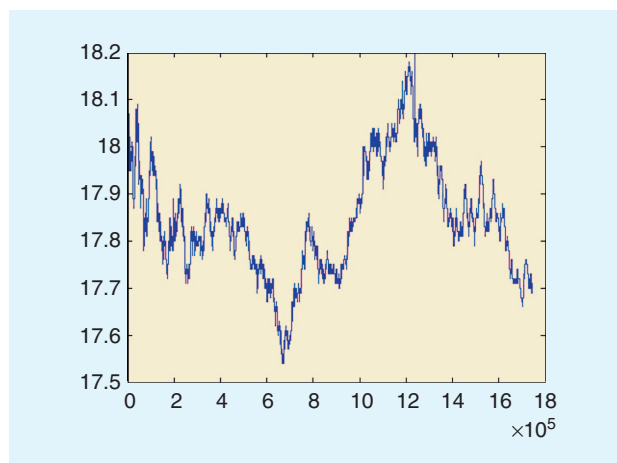
[FIG4] Intraday evolution of the ask price: Citigroup, 26 June 2008. Horizontal axis: number of events.

independent Poisson processes and compute analytically, using simple matrix computations and Laplace transform methods, the distribution of time to fill (execution time), the probability of "making the spread" and the conditional distribution of price moves given the state of the order book [28]. Though the Poisson assumption is not very realistic—durations are neither exponentially distributed nor independent—the model is shown to capture effectively the short-term dynamics of a limit order book.

In reality, the order flow is clustered in time: as shown in Figure 5, we observe periods with a lot of market buy orders and periods with a lot of market sell orders [35], [36]. This is partly due to the fact that many observed orders are in fact components of a larger "parent" order that is executed in small blocks to minimize market impact [13]. This results in a

[FIG5] Changes in the size of the ask queue: each data point represents one order or cancellation. Positive values represent sizes of limit orders and negative values represent market orders or cancellations. Data from Citigroup, 26 June 2008.

[FIG6] The level-1 representation of a limit order book represents the quantities of shares available at the bid (best buy price) and the ask (best sell price).

significant autocorrelation in durations, as well as a positive cross-correlation of arrival rates across event types (market buy, limit sell). These features, which are not captured in models based on Poisson processes, may be adequately represented by a multidimensional self-exciting point process [36], [35]: in such models, the arrival rate $\lambda_i(t)$ of an order of type i is represented as a stochastic process whose value depends on the recent history of the order flow: each new order increases the rate of arrival for subsequent orders of the same type (self-exciting property) and may also affect the rate of arrival of other order types (mutually exciting property)

$$\lambda_i(t) = \theta_i + \sum_{j=1}^{J} \delta_{ij} \sum_{T_n^j \le t} e^{-\kappa_i(t - T_n^j)},$$

where $T_1^j, \ldots, T_n^j$ are the occurrence times of events of type j. Here δ_{ij} measures the impact of events of type j on the rate of arrival of subsequent events of type i: as each event of type j occurs, λ_i increases by δ_{ij}. In between events, $\lambda_i(t)$ decays exponentially at rate κ_i. This baseline model, known as the Hawkes process, does not account for the fact that order flow reacts to changes in the bid-ask spread. Andersen et al. [35] enhance this basic model by incorporating dependence of parameters δ, κ on the bid-ask spread. Maximum likelihood estimation of this model on TAQ data [35] shows evidence of self-exciting and mutually exciting features in order flow: the coefficients δ_{ij} are all significantly different from zero and positive, with $\delta_{ii} > \delta_{ij}$ for $j \ne i$ and depend significantly on the bid-ask spread.
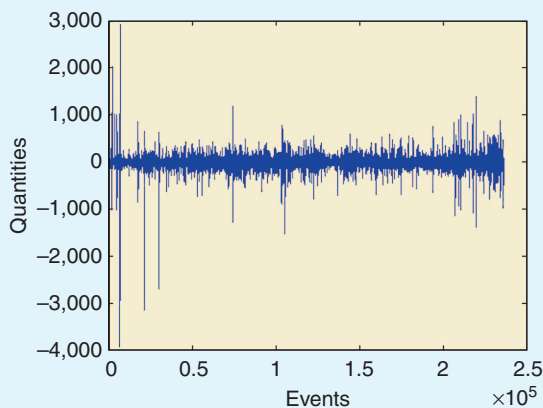
Like models based on Poisson order flow, these models allow for analytical computations of many quantities, without recourse to simulation. The availability of such analytical formulae is a key feature when using such models in real-time applications.

QUEUEING MODELS FOR LEVEL-1 ORDER BOOKS

Empirical studies of limit order markets suggest that the major component of the order flow occurs at the (best) bid and ask price levels (see, e.g., [37]). Furthermore, studies on the price impact of order book events show that the net effect of orders on the bid and ask queue sizes is the main factor driving price variations [27]. These observations motivate a reduced-form modeling approach in which, instead of modeling the order queues at all price levels, one focuses on queue sizes at the best bid and ask [34]. In this approach one focuses on the dynamics of the state variables $(S_t^b, S_t^a, q_t^b, q_t^a)$ where
- the bid price S_t^b and the ask price S_t^a
- the size of the bid queue q_t^b representing the outstanding limit buy orders at the bid
- the size of the ask queue q_t^a representing the outstanding limit sell orders at the ask.

Figure 6 summarizes this representation. In principle, the evolution of the size of the queues at the best quotes depends on the deeper levels of the order book: indeed, once the bid (respectively the ask) queue is depleted, the price will move to
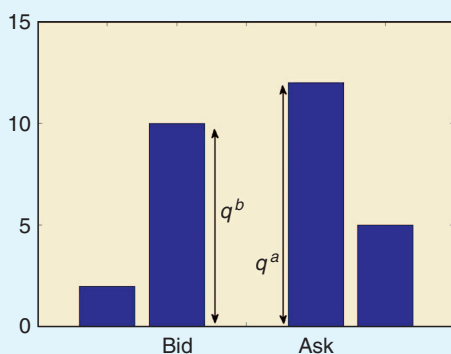
the queue at the next level, which is queue at the next level below (respectively above) the current bid (ask). The new queue size thus corresponds to what was previously the number of orders sitting at the second-best price. Instead of keeping track of these queues (and the corresponding order flow) at all price levels (as in [28] and [4]), Cont and Larrard [34] propose a model in which the bid/ask queue sizes (q_t^b, q_t^a) are renewed at each price change, i.e., every time one of the queues is depleted. More precisely,

- if the ask queue is depleted at t then (q_t^b, q_t^a) is a variable with distribution f, where $f(x, y)$ represents the probability of observing $(q_t^b, q_t^a) = (x, y)$ right after a price increase.
- if the bid queue is depleted at t then (q_t^b, q_t^a) is a variable with distribution $\tilde{f}$, where $\tilde{f}(x, y)$ represents the probability of observing $(q_t^b, q_t^a) = (x, y)$ right after a price decrease.

The distributions $f, \tilde{f}$ may be estimated from empirical data: an example is shown in Figure 7. They summarize the interaction of the queues at the best bid/ask levels with the rest of the order book, viewed here as a reservoir of limit orders. Thus (q_t^b, q_t^a) is a random process in the positive orthant, renewed every time it hits the axes. This approach considerably simplifies the model, enhances tractability, and makes it easier to estimate parameters.

Between two hitting times, one can specify various dynamics for the queues. Cont and Larrard [34] consider the simplest specification, namely a Markovian model with Poisson arrivals of orders. This model allows to obtain analytical expressions for various quantities of interest such as the distribution of the duration between price changes, the distribution and autocorrelation of price changes, and the probability of an upward move in the price, conditional on the state of the order book. Interestingly, these results provide reasonable fits to empirical data [34], showing that such low-dimensional models may be suitable for many applications. Figure 8 shows an example of such a result, comparing theoretical transition probabilities for the price computed in this reduced-form model with empirical transition probabilities estimated from high-frequency data.

## HEAVY-TRAFFIC LIMITS AND DIFFUSION APPROXIMATIONS

Analytically tractable models of order book dynamics may contain assumptions that are not statistically realistic: durations between orders are neither independent nor exponentially distributed, and various order types exhibit strong dependence [35]. In addition, the order flow exhibits seasonality as well as dependence with respect to price and spread behavior [1].

While a model incorporating all these features may be too complex to study analytically, two remarks show that the situation is not as hopeless as it may seem at first glance.

First, as shown in Table 2, most applications involve the behavior of prices over time scales an order of magnitude larger than the typical interevent duration: for example, in optimal trade execution the benchmark is the volume weighted average price (VWAP) computed over a day or an hour: over
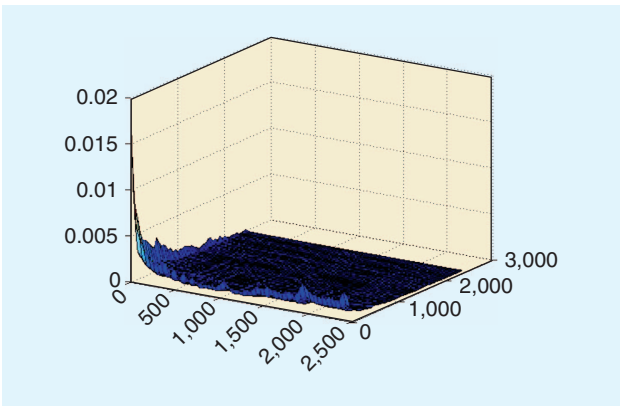
[FIG7] Empirical (joint) distribution of bid and ask queue sizes after a price move (Citigroup, June 2008) [34]. Unit: blocks of 100 shares.
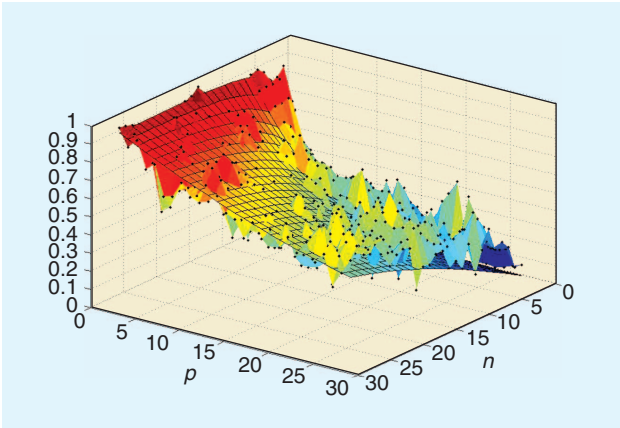
[FIG8] Conditional probability of a price increase, as a function of the bid and ask queue size, computed using a Markovian queuing model for the level-1 limit order book [34] (solid curve) compared with transition frequencies for Citigroup tick-by-tick data on 26 June 2008 (points).

such time scales much of the microstructural details of the market are averaged out. Second, as noted in Table 1, in liquid equity markets, the number of events affecting the state of the order book over such time scales is quite large; of the order of hundreds or thousands. The typical duration $\tau_L = 1/\lambda$ (respectively $\tau_M = 1/\mu$) between limit orders (respectively, market orders and cancelations) is typically $0.001 - 0.01 \ll 1$ (in seconds). These observations show that it is relevant to consider asymptotic methods in which the rate of arrival of orders is large, similar to those used in the study of heavy-traffic limits in queueing theory [38]. In this limit, the possibly complex

[TABLE 2] A HIERARCHY OF TIME SCALES.		
REGIME	TIME SCALE	ISSUES
ULTRAHIGH FREQUENCY (UHF)	$\sim 10^{-3} - 0.1$	MICROSTRUCTURE, LATENCY
HIGH FREQUENCY (HF)	$\sim 1 - 100$ S	TRADE EXECUTION
"DAILY"	$\sim 10^3 - 10^4$	TRADING STRATEGIES, OPTION HEDGING

discrete dynamics of the queue-ing system is approximated by a simpler system with a continuous state space, which can be either described by a system of ordinary differential equations (in the "fluid limit," where random fluctuations in queue size vanish) or a system of stochastic differential equations (in the "diffusion limit," where random fluctuations dominate). Intuitively, the fluid limit corresponds to the regime of the law of large numbers, where random fluctuations average out and the limit is described by average queue size, whereas the diffusion limit corresponds to the regime of the central limit theorem, where fluctuations in queue size are asymptotically Gaussian. The relevance of each of these asymptotic regimes is, of course, not a matter of taste but an empirical question that depends on the behavior of order flow in the system.

As argued in [39], for most liquid stocks, while the rate of arrival of market orders and limit orders is large, the imbalance between limit orders, which increase queue size, and market orders and cancels, which decrease queue size, is an order of magnitude smaller: over, say, a 30-min interval, one observes an imbalance ranging from 1 to 5% of order flow. In other words, over a time scale $\tau_1 \gg \tau_L, \tau_M$ (say, 1 min) a large number $N \sim \lambda \tau_1$ of events occur, but the bid/ask imbalance accumulating over the same interval is of order $\sqrt{N} \ll N$ i.e., $(\lambda - \mu)$ is small, of the order $1/\sqrt{N}$. This corresponds to a situation where limit orders accumulate and disappear roughly at the same rate. In this regime, the queues follow a diffusive behavior and it is relevant to consider the diffusion limit of the limit order book. When the sequences of order sizes at the bid and the ask $(V_i^a, V_i^b, i \geq 1)$ and interevent durations $(\tau_i^a, \tau_i^b, i \geq 1)$ are weakly dependent covariance-stationary sequences, the rescaled order book process converges weakly [39]

$$\left(\frac{q^a(nt)}{\sqrt{n}}, \frac{q^b(nt)}{\sqrt{n}} \right)_{t \geq 0} \overset{d}{\Rightarrow} (Q_t)_{t \geq 0}$$

to a Markov process $(Q_t)_{t \geq 0}$ diffusing in the quarter plane

$$\{(x, y), \quad x \geq 0, y \geq 0\},$$

which is "renewed" every time it hits one of the axes. In the case of a balanced order book where the arrival rate of limit orders matches on average their depletion rate by market orders and cancels and $E[V_a^i] = E[V_b^i] = 0$, the process Q_t

- behaves like a planar Brownian motion with drift b and covariance matrix

$$\begin{pmatrix} \lambda_a v_a^2 & \rho v_a v_b \sqrt{\lambda_a \lambda_b} \\ \rho v_a v_b \sqrt{\lambda_a \lambda_b} & \lambda_b v_b^2 \end{pmatrix} \qquad (1)$$

on the interior of the orthant $\{(x, y), \quad x > 0, y > 0\}$
- is reinitialized according to F each time it hits the x-axis
- is reinitialized according to $\widetilde{F}$ each time it hits the y-axis.

Q_t is an example of "regulated Brownian motion" [40], a notion introduced in the context of heavy-traffic limits of multiclass

A PARTICULARLY IMPORTANT ISSUE FOR APPLICATIONS IS THE IMPACT OF ORDERS ON PRICES.

queueing systems. In this case, Q is in fact a process with jumps: it jumps into the interior of the orthant every time it hits the axes. Here δ is the tick size,

- $1/\lambda_a = \lim_{n \to \infty} (\tau_1^a + \ldots + \tau_n^a)/n$ is the average duration between events at the ask
- $v_a^2 = \mathbb{E}[(V_1^a)^2] + 2 \sum_{i=2}^{\infty} \text{Cov}(V_1^a, V_i^a)$ represents the variance of event sizes at the ask
- $1/\lambda_b = \lim_{n \to \infty} (\tau_1^b + \ldots + \tau_n^b)/n$ is the average duration between events at the bid
- $v_b^2 = \mathbb{E}[(V_1^b)^2] + 2 \sum_{i=2}^{\infty} \text{Cov}(V_1^b, V_i^b)$ represents the variance of event sizes at the ask
- F (respectively $\widetilde{F}$) is the (rescaled) distribution of (q^a, q^b) after a price increase (respectively decrease)
- $\rho = (\text{cov}[V_1^a V_1^b] + 2 \sum_{i=1}^{\infty} (\text{cov}[V_1^a V_i^b] + \text{cov}[V_1^b V_i^a]))/(v_a v_b)$ measures the correlation between order sizes at the bid and at the ask. If order sizes at the bid and ask are symmetric and uncorrelated then $\rho = 0$. Empirically one finds that $\rho < 0$ for equity order books [39].

This diffusion approximation is analytically tractable and allows to compute analytically [39] many quantities such as the duration until the next price change, the probability of an increase in the price, and the distribution of the time to execution conditionally on the state of the order book. For example, in the symmetric case $\lambda_a = \lambda_b, v_a = v_b$, the probability $p_{\text{up}}(x, y)$ that the next price move is an increase, given a queue of x shares on the bid side and y shares on the ask side, has a simple expression [39]

$$p_{\text{up}}(x, y) = \frac{1}{2} - \frac{\arctan\left(\sqrt{\frac{1 + \rho}{1 - \rho}} \frac{y - x}{y + x} \right)}{2 \arctan\left(\sqrt{\frac{1 + \rho}{1 - \rho}} \right)}, \qquad (2)$$

which only depends on the size x of the bid queue, the size y of the ask queue, and the correlation ρ between event sizes at the bid and the ask.

Beyond their computational aspect, these analytical results shed light on the relation between order flow and price dynamics. A nice application is given by Avellaneda, Reed, and Stoikov [41]: using formula (2) in the case $\rho = -1$, they invert the effective size (x, y) of the queues from the empirical transition probabilities of the price and compare it to the visible queue size: the difference, which they call *hidden liquidity*, is shown to account for a significant portion of market depth for some U.S. stocks.

PRICE IMPACT MODELING

A particularly important issue for applications is the impact of orders on prices: the optimal liquidation of a large block of shares, given a fixed time horizon, crucially involves assumptions on price impact (see [13]–[16]). Of course, any model for the dynamics of the (full) order book implies a model of price impact. However, even in simple models for order book dynamics, the price impact may be difficult to characterize. Hence, the typical approach is to

investigate the relation between price changes and order flow directly using high-frequency data.

Various models for price impact have been proposed in the literature but there is little agreement on how to model it: price impact has been described by various authors as linear, nonlinear, square root, temporary, instantaneous, permanent, or transient. The only consensus seems to be the intuitive notion that imbalance between supply and demand moves prices.

The empirical literature on price impact has primarily focused on trades. Starting with Hasbrouck [42], empirical studies on public data [42]–[48] have investigated the relation between the direction and sizes of trades and price changes and typically conclude that the price impact of trades is an increasing, concave ("square root") function of their size.

Studies based on proprietary data [49]–[51] are able to measure impact of "parent orders" gradually executed over time and give a different picture. Using large-cap U.S. equity orders executed by Citigroup Equity Trading, Almgren et al. [50] find evidence for a nonlinear dependence of temporary price impact ΔS as function of trade size V, of the form $\Delta S \sim V^{3/5}$ across the range of trade sizes considered, but reject the "square root" law. Obizhaeva [51] studies the short run price-volume relation using a proprietary data set of portfolio transitions using a methodology that corrects for the endogeneity of trading decisions and finds that, in the short run, purchases of stocks induce permanent price increases, the magnitude of which is proportional to the size of the trade. Moreover, if volume is expressed as a percentage of average daily volume and returns are expressed as a percentage of daily volatility, the proportionality constant is relatively stable across stocks.

This focus on trades leaves out the information in quotes, which provide a more detailed picture of price formation [52], [53], [27]. In fact, knowledge of the limit order book allows to estimate the short-term price impact of incoming orders. This can be understood by considering the following example of an order book with constant number of shares D per level (see Figure 9). If, during a short time interval $[t, t + \Delta]$, the volume of incoming limit orders (respectively market orders and cancellations) at the bid are L (respectively M, C) shares, then the net change in the number of limit orders at the bid is $L - M - C$. If the depth of the order book is D shares (on average) at each level then the change in the bid price will be

$$S(t + \Delta) - S(t) = \delta \left[\frac{L - M - C}{D} \right], \quad (3)$$

where δ is the tick size and $[\cdot]$ denotes the integer part. This example, illustrated in Figure 9, shows that the change in the bid (respectively ask) price is determined by the order flow imbalance $L - M - C$ at the bid (respectively the ask), defined as the sum of

> **WITH THE ADVENT OF INTRADAY AND HIGH-FREQUENCY TRADING, THERE IS AN INCREASING NEED FOR DEVELOPING REALISTIC STATISTICAL MODELS FOR INTRADAY VOLATILITY PREDICTION AND RISK MANAGEMENT.**

limit order minus the market orders and cancellations during the interval $[t, t + \Delta]$ [27]

$$OFI(t, t + \Delta) = L - M - C. \quad (4)$$

Motivated by this example, Cont, Kukanov, and Stoikov [27] investigate a model in which high-frequency price changes are driven by order flow imbalance

$$\frac{S(t + \Delta) - S(t)}{\delta} = c \frac{OFI(t, t + \Delta)}{D(t)} + \epsilon(t), \quad (5)$$

where δ is the tick size, $\epsilon(t)$ is a (white) noise term, and $D(t)$ is a measure of order book depth (number of limit orders at the bid/ask). Empirical analysis of 1 s and 10 s midprice changes using the TAQ database for U.S. stocks show empirical evidence for the linear price impact model [27]: c is estimated to be typically between 0.1 and 1.

The linear model (5) is quite different from models of price impact that consider only the size of trades [44], [42], [46]–[48], [54]. Instead of looking at price impact of trades as a (nonlinear) function of trade size, it shows that all order book events (including trades) have a linear price impact, equal to $c/D(t)$ on average, and this impact can be summarized by aggregating all events into a single variable: the order flow imbalance. This representation also implies that price impact has strong intraday seasonality, which follows the seasonality patterns in order book depth; this is indeed shown to be the case empirically [27].

CONNECTING VOLATILITY ACROSS TIME SCALES
While most econometric models of price volatility were initially developed with a focus on daily returns, with the advent of intraday and high-frequency trading there is an increasing need for developing realistic statistical models for intraday volatility prediction and risk management. High-frequency data are also of interest for the estimation of volatility at daily level: as shown by Andersen and Bollerslev [55]–[57], realized variance computed from high-frequency data yields a better estimator of daily volatility than estimates computed from daily returns.

One approach is to combine time-series models, such as GARCH models, developed for daily returns, with a model for
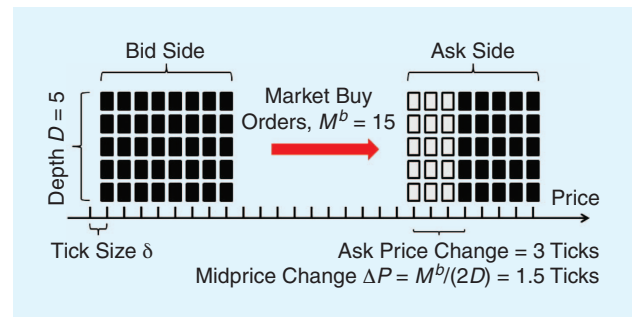
[FIG9] Price impact of a market order.

durations between trades, such as the ACD model [23]: this allows to incorporate the information on variable rates of activity, peaks in trading activity being represented by shorter durations [21], [24]. This approach has been mostly applied to transaction data but a similar approach may also be considered for quotes [52].

Another approach for establishing a link between volatility and order flow is to start from a model for order book dynamics and compute the diffusion limit of the price process that describes the diffusive behavior of the price at longer time scales. This idea is implemented in [34] and [39] by proving a functional central limit theorem for the order book models described above. This result allows in particular to compute the variance of price changes over a time scale $T \gg \tau_0$ much larger than the typical interevent duration. For example, in a Markovian model for a symmetric Level 1 order book, Cont and Larrard [34] derive the following relation between the standard deviation σ of price increments $S(t + T) - S(t)$ and the statistics of order flow:

$$\sigma = \frac{v\delta}{D(F)}\sqrt{\frac{\pi\lambda T}{\tau_0}} = v\delta\lambda\frac{\sqrt{\pi T}}{D(F)}, \tag{6}$$

where $\lambda = 1/\tau_0$ is the order flow in shares per unit time, F is the distribution of queue sizes after a price change, and $D(F) = \sqrt{\int xy\,dF(x, y)}$ is a measure of order book depth: it is the geometric mean of the queue size at the bid and the ask after a price change (and π is the familiar ratio of a circle's perimeter to its diameter). In particular, this relation suggests that the high-frequency properties of order flow contribute to price volatility mainly through λ and $D(F)$: for stocks that trade at the same frequency, volatility increases with the ratio $\lambda/D(F)$. This prediction is empirically verified for major U.S. stocks [34]. Similar relations between volatility and order flow statistics may be obtained in more general settings with non-Markovian and asymmetric order flows [58].

PERSPECTIVES AND CHALLENGES
Existing studies have only scratched the surface of high-frequency financial data: much remains to be done towards a better analysis and understanding of intraday dynamics of financial markets and the development of operational models.

These studies have mostly focused on equity markets and, to a lesser extent, futures markets and foreign exchange markets. Further studies are necessary on other markets, in particular, fixed-income markets and markets for listed derivatives, which have their own features.

Also, almost all existing models focus on single-asset markets, whereas, in practice, most applications—in trading, risk management, or regulation—involve multiple assets and portfolios. Extending univariate models to a multidimensional setting may not be straightforward: for example, in models where durations are random, keeping track of synchronicity across different assets

> **MORE REMAINS TO BE DONE TO BRIDGE THE GAP BETWEEN MICROSTRUCTURE MODELS AND STOCHASTIC MODELS.**

is not trivial to implement. For this reason, whereas single-asset models go to a great length to model randomness in durations, multiasset modeling has mostly focused on modeling of fixed-interval returns, the most well-known being Hasbrouck's vector autoregressive [2] model. Yet, the connection between order flows across different assets is particularly relevant when considering indexes, futures, and options, which cannot be considered in isolation and in which supply and demand are driven by events occurring in another market—the market for the components of the index or the underlying asset of the derivative. Such multiasset limit order book models are very relevant for applications and their development represents a challenge.

More remains to be done to bridge the gap between microstructure models—which hinge on various behavioral assumptions about market participants' strategies—and stochastic models, which are more operational and data driven. The theoretical input from microstructure models may be used to design new stochastic models whose ingredients will have a better economic interpretation. The study of heavy-traffic limits and diffusion limits of discrete order book models seems to be a promising direction of research. This approach restores analytical tractability into the models and allows to connect microstructural features to statistical properties of prices and order flow, which are observable and may be used in the empirical validation of models.

We have focused in this survey on stochastic models and related statistical techniques for high-frequency data. Other techniques, though less present in the research literature (machine learning methods, wavelet-based methods, nonparametric pattern recognition methods, large-scale data mining techniques, and methods from cryptography) have been extensively applied by practitioners for prediction and pattern detection in high-frequency data. These methods have contributed in some cases to the development of successful trading algorithms, which may or may not be generalizable to other data sets, but it remains to be seen whether they can also provide interesting fundamental insights into the dynamics of supply, demand, and price in financial markets.

AUTHOR
Rama Cont (Rama.Cont@columbia.edu) received an engineering degree from Ecole Polytechnique, an M.Sc. degree in theoretical physics from Ecole Normale Superieure (Paris), and a doctorate degree in theoretical physics from Université de Paris XI-Orsay (1998). From 1998 to 2005, he was a research scientist in applied mathematics at Ecole Polytechnique (France). He is an associate professor in operations research and director of the Center for Financial Engineering at Columbia University, New York, and a

CNRS research scientist at Laboratoire des Probabilités et Modèles Aléatoires, Université Pierre and Marie Curie (Paris VI). His research focuses on stochastic analysis, quantitative methods in finance, statistical modeling of financial time series, high-frequency data, and market microstructure modeling. He is the author of *Financial Modeling with Jump Processes* (2003) and the editor-in-chief of the *Encyclopedia of Quantitative Finance* (2010) and has extensive consulting experience related to quantitative modeling in finance, derivatives, and risk management. He was awarded the 2010 Louis Bachelier Prize (Grand Prix Louis Bachelier de l'Academie des Sciences) from the French Academy of Sciences and the Natixis Foundation for Quantitative Research for his research on mathematical modeling in finance.

REFERENCES

[1] J.-P. Bouchaud, D. Farmer, and F. Lillo, "How markets slowly digest changes in supply and demand," in *Handbook of Financial Markets: Dynamics and Evolution*, T. Hens and K. Schenk-Hoppe, Eds. New York: Elsevier, 2008, pp. 57–130.

[2] J. Hasbrouck, *Empirical Market Microstructure*. New York: Oxford Univ. Press, 2007.

[3] J.-P. Bouchaud, M. Mézard, and M. Potters, "Statistical properties of stock order books: Empirical results and models," *Quant. Finance*, vol. 2, no. 4, pp. 251–256, 2002.

[4] E. Smith, J. D. Farmer, L. Gillemot, and S. Krishnamurthy, "Statistical theory of the continuous double auction," *Quant. Finance*, vol. 3, no. 6, pp. 481–514, 2003.

[5] J. D. Farmer, L. Gillemot, F. Lillo, S. Mike, and A. Sen, "What really causes large price changes?," *Quant. Finance*, vol. 4, pp. 383–397, Aug. 2004.

[6] B. Biais, L. Glosten, and C. Spatt, "Market microstructure: A survey of microfoundations, empirical results and policy implications," *J. Financial Markets*, vol. 8, no. 2, pp. 217–264, 2005.

[7] C. Parlour, "Price dynamics in limit order markets," *Rev. Financial Stud.*, vol. 11, no. 4, pp. 789–816, 1998.

[8] T. Foucault, O. Kadan, and E. Kandel, "Limit order book as a market for liquidity," *Rev. Financial Stud.*, vol. 18, no. 4, pp. 1171–1217, 2005.

[9] L. Glosten, "Is the limit order book inevitable?," *J. Finance*, vol. 49, no. 4, pp. 1127–1161, 1994.

[10] A. Kyle, "Continuous auctions and insider trading," *Econometrica*, vol. 53, no. 6, pp. 1315–1335, Nov. 1985.

[11] I. Rosu, "A dynamic model of the limit order book," *Rev. Financial Stud.*, vol. 22, no. 11, pp. 4601–4641, 2009.

[12] M. Avellaneda and S. Stoikov, "High-frequency trading in a limit order book," *Quant. Finance*, vol. 8, no. 3, pp. 217–224, 2008.

[13] R. Almgren and N. Chriss, "Optimal execution of portfolio transactions," *J. Risk*, vol. 3, no. 2, pp. 5–39, 2000.

[14] D. Bertsimas and A. Lo, "Optimal control of execution costs," *J. Financial Markets*, vol. 1, no. 1, pp. 1–50, 1998.

[15] A. Obizhaeva and J. Wang, "Optimal trading strategy and supply/demand dynamics," Working Paper, Cambridge, MA: MIT, 2006.

[16] A. Alfonsi, A. Schied, and A. Schulz, "Optimal execution strategies in limit order books with general shape function," *Quant. Finance*, vol. 10, no. 2, pp. 143–157, 2010.

[17] CFTC and SEC, "Findings regarding the market events of May 6, 2010," *CFTC, Rep. Joint Advisory Committee on Emerging Regulatory Issues*, Washington, DC, 2010.

[18] A. Kirilenko, A. Kyle, M. Samadi, and T. Tuzun, "The flash crash: The impact of high frequency trading on an electronic market," *SSRN Working Paper*, 2010.

[19] N. Hautsch, *Modelling Irregularly Spaced Financial Data (Lecture Notes in Economics and Mathematical Systems, no. 539)*. Berlin: Springer-Verlag, 2004.

[20] R. Cont, "Empirical properties of asset returns: Stylized facts and statistical issues," *Quant. Finance*, vol. 1, no. 2, pp. 223–236, 2001.

[21] R. F. Engle, "The econometrics of ultra-high frequency data," *Econometrica*, vol. 68, no. 1, pp. 1–22, 2000.

[22] R. Cont and P. Tankov, *Financial Modelling with Jump Processes*. London, U.K.: Chapman & Hall, 2003.

[23] R. Engle and J. Russell, "Autoregressive conditional duration: A new model for irregularly spaced transaction data," *Econometrica*, vol. 66, no. 5, pp. 1127–1162, 1998.

[24] E. Ghysels and J. Jasiak, "GARCH for irregularly spaced financial data: The ACD-GARCH model," *Stud. Nonlinear Dynamics Econometrics*, vol. 2, no. 4, pp. 133–149, 1998.

[25] L. Bauwens and N. Hautsch, "Modelling financial high frequency data using point processes," in *Handbook of Financial Time Series*, T. G. Andersen, R. A. Davis, J.-P. Kreiß, and T. Mikosch, Eds. Berlin: Springer-Verlag, 2009, pp. 953–979.

[26] C. Lee and M. Ready, "Inferring trade direction from intraday data," *J. Finance*, vol. 46, no. 2, pp. 733–746, 1991.

[27] R. Cont, A. Kukanov, and S. Stoikov, "The price impact of order book events," Working Paper, New York: Columbia Univ., 2010.

[28] R. Cont, S. Stoikov, and R. Talreja, "A stochastic model for order book dynamics," *Oper. Res.*, vol. 58, no. 3, pp. 549–563, 2010.

[29] L. E. Harris and V. Panchapagesan, "The information content of the limit order book: Evidence from NYSE specialist trading decisions," *J. Financial Markets*, vol. 8, no. 1, pp. 25–67, Feb. 2005.

[30] C. Gourieroux, J. Jasiak, and G. L. Fol, "Intra-day market activity," *J. Financial Markets*, vol. 2, no. 3, pp. 193–226, Aug. 1999.

[31] B. Hollifield, R. A. Miller, and P. Sandas, "Empirical analysis of limit order markets," *Rev. Econ. Stud.*, vol. 71, no. 4, pp. 1027–1063, Oct. 2004.

[32] J. Large and J. Field, "Pro-rata matching in one-tick markets," *Oxford Man Inst.*, Working Paper, 2008-40, Center for Financial Studies, Frankfurt, Germany: Goeth-Universität.

[33] H. Mendelson, "Market behavior in a clearing house," *Econometrica*, vol. 50, no. 6, pp. 1505–1524, Nov. 1982.

[34] R. Cont and A. de Larrard. (2010). Price dynamics in a Markovian limit order market. [Online]. Available: http://ssrn.com/abstract=1735338

[35] A. Andersen, R. Cont, and E. Vinkovskaya, "A point process model for the high-frequency dynamics of a limit order book," New York: Columbia Univ., *Financial Eng. Report 2010–2008*, 2010.

[36] P. Hewlett, "Clustering of order arrivals, price impact and trade path optimisation," Working Paper, Oxford, UK: OCIAM, 2006.

[37] B. Biais, P. Hillion, and C. Spatt, "An empirical analysis of the order flow and order book in the Paris Bourse," *J. Finance*, vol. 50, no. 5, pp. 1655–1689, 1995.

[38] W. Whitt, *Stochastic Process Limits*. Berlin: Springer-Verlag, 2002.

[39] R. Cont and A. de Larrard, "Order book dynamics in liquid markets: Heavy traffic approximations and diffusion limits," Working Paper, Universite Pierre and Marie Curie, France, 2010.

[40] J. M. Harrison and V. Nguyen, "Brownian models of multiclass queueing networks: Current status and open problems," *Queueing Syst.*, vol. 13, no. 1, pp. 5–40, 1993.

[41] M. Avellaneda, S. Stoikov, and J. Reed, "Forecasting prices from level-i quotes in the presence of hidden liquidity," SSRN Working Paper, 2010.

[42] J. Hasbrouck, "Measuring the information content of stock trades," *J. Finance*, vol. 46, no. 1, pp. 179–207, 1991.

[43] M. Evans and R. Lyons, "Order flow and exchange rate dynamics," *J. Political Economy*, vol. 110, no. 1, pp. 170–180, 2002.

[44] X. Gabaix, P. Gopikrishnan, V. Plerou, and H. Stanley, "A theory of power-law distributions in financial market fluctuations," *Nature*, vol. 423, no. 6937, pp. 267–270, 2003.

[45] D. Keim and A. Madhavan, "The upstairs market for large-block transactions: Analysis and measurement of price effects," *Rev. Financ. Stud.*, vol. 9, no. 1, pp. 1–36, 1996.

[46] A. Kempf and O. Korn, "Market depth and order size," *J. Financ. Markets*, vol. 2, no. 1, pp. 29–48, Feb. 1999.

[47] N. Torre and M. Ferrari, "The market impact model," *BARRA Res. Insights*, 1999. [Online]. Available: http://www.mscibarra.com/research/articles/barra/Market_Impact_Model.pdf

[48] M. Potters and J. P. Bouchaud, "More statistical properties of order books and price impact," *Physica A*, vol. 324, no. 1, pp. 133–140, 2003.

[49] R. F. Engle, R. Ferstenberg, and J. Russell, "Measuring and modeling execution cost and risk," Working Paper No. FIN-06-044, New York: New York Univ., 2006.

[50] R. Almgren, C. Thum, E. Hauptmann, and H. Li, "Equity market impact," *Risk*, vol. 18, pp. 57–62, 2005.

[51] A. Obizhaeva, "Portfolio transitions and stock price dynamics," Working Paper, Univ. of Maryland, 2009.

[52] R. F. Engle and A. Lunde, "Trades and quotes: A bivariate point process," *J. Financ. Econ.*, vol. 1, no. 2, pp. 159–188, 2003.

[53] Z. Eisler, J.-P. Bouchaud, and J. Kockelkoren, "The price impact of order book events: Market orders, limit orders and cancellations," Working Paper, Paris: Capital Fund Management, 2010.

[54] V Plerou, P Gopikrishnan, X Gabaix, and H Stanley, "Quantifying stock-price response to demand fluctuations," *Phys.Rev. E*, vol. 66, no. 2, pp. 027104-1–027104-4, 2002.

[55] T. G. Andersen and T. Bollerslev, "Towards a unified framework for high and low frequency return volatility modeling," *Stat. Neerlandica*, vol. 52, no. 3, pp. 273–302, 1998.

[56] T.G. Andersen and T. Bollerslev, "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts," *Int. Econ. Rev.*, vol. 39, no. 4, pp. 885–905, 1998.

[57] T. G. Andersen, T. Bollerslev, and N. Meddahi, "Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities," *Econometrica*, vol. 73, no. 1, pp. 279–296, 2005.

[58] R. Cont and A. de Larrard, "Linking volatility with order flow: Price dynamics in limit order markets," Working Paper, Universite Pierre and Murie Curie, 2011.

[SP]