A stochastic differential equation (SDE) is a differential equation in which one or more of the terms is a stochastic process, resulting in a solution which is itself a stochastic process. SDEs are used to model diverse phenomena such as fluctuating stock prices or physical systems subject to thermal fluctuations. Typically, SDEs incorporate random white noise which can be thought of as the derivative of Brownian motion (or the Wiener process); however, it should be mentioned that other types of random fluctuations are possible, such as jump processes.

Quantitative Model of Price Diffusion and Market Friction Based on Trading as a Mechanistic Random Process

Diffusion in a gas is the random motion of particles involved in the net movement of a substance from an area of high concentration to an area of low concentration. Each particle in a given gas continues to collide with other particles. In regions of the gas where the particle density is the highest, the particles bounce off each other and the boundary of their container at a greater rate than particles in less-dense regions.

Marcus G. Daniels,[1] J. Doyne Farmer,[1] László Gillemot,[1] Giulia Iori,[2] and Eric Smith[1]

[1]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501*

[2]*Mathematics Department, King's College London, Strand, London WC2R 2LS, United Kingdom*

(Received 8 January 2002; revised manuscript received 9 December 2002; published 13 March 2003)

We model trading and price formation in a market under the assumption that order arrival and cancellations are Poisson random processes. This model makes testable predictions for the most basic properties of markets, such as the diffusion rate of prices (which is the standard measure of financial risk) and the spread and price impact functions (which are the main determinants of transaction cost). Guided by dimensional analysis, simulation, and mean-field theory, we find scaling relations in terms of order flow rates. We show that even under completely random order flow the need to store supply and demand to facilitate trading induces anomalous diffusion and temporal structure in prices.

Price impact refers to the correlation between an incoming order (to buy o r to sell) and the subsequent price change. That a buy trade sho uld push the price up seems at first sight obvious and is easily demonstrat ed empirically,

DOI: 10.1103/PhysRevLett.90.108102 PACS numbers: 89.65.Gh, 05.40.Fb

The equation above characterizes the behavior of the continuous time stochastic process Xt as the sum of an ordinary Lebesgue integral and an Itō integral. A heuristic (but very helpful) interpretation of the stochastic differential equation is that in a small time interval of length δ the stochastic process Xt changes its value by an amount that is normally distributed with expectation μ(Xt, t) δ and variance σ(Xt, t)² δ and is independent of the past behavior of the process. This is so because the increments of a Wiener process are independent and normally distributed. The function μ is referred to as the drift coefficient, while σ is called the diffusion coefficient. The stochastic process Xt is called a diffusion process, and is usually a Markov process.

There have recently been efforts to apply physics methods to problems in economics [1]. This effort has yielded interesting empirical analyses and conceptual models. However, with the exception of refinements to option pricing theory, thus far it has had little success in producing theories that make falsifiable predictions about the most important properties of markets. In this Letter, we develop a mechanistic random process model of the continuous double auction, which is the standard method for trade matching in modern financial markets. This model differs from standard models in economics in that it makes no assumptions about agent rationality. The model makes falsifiable predictions based on parameters that can all be measured in real data, and preliminary results indicate that it has substantial explanatory power [2].

The random walk model was originally introduced by Bachelier to describe prices, five years before Einstein used it to model Brownian motion. Although this is one of the most widely used models of prices in financial economics, there is still no understanding of its most basic property, namely, its diffusion rate. We present a theory for how the diffusion rate of prices depends on the flow of orders into the market, deriving scaling relations based on dimensional analysis, mean-field theory, and simulation. We also make quantitative predictions of other basic market properties, such as the gap between the best prices for buying and selling, the density of stored demand vs price, and the impact of trading on prices.

Most modern financial markets operate continuously. The mismatch between buyers and sellers that typically exists at any given instant is solved via an order-based market with two basic kinds of orders. Impatient traders submit *market orders*, which are requests to buy or sell a given number of shares immediately at the best available price. More patient traders submit *limit orders*, which also state a limit price, corresponding to the worst allowable price for the transaction. Limit orders often fail to result in an immediate transaction, and are stored in a queue called the *limit order book*. Buy limit orders are called *bids*, and sell limit orders are called *offers* or *asks*. We will label the logarithm of the best (lowest) offering price $a(t)$ and the best (highest) bid price $b(t)$. There is typically a nonzero price gap between them, called the *spread* $s(t) = a(t) - b(t)$.

As market orders arrive, they are matched against limit orders of the opposite sign in order of price and arrival time. Because orders are placed for varying numbers of shares, matching is not necessarily one-to-one. For example, suppose the best offer is for 200 shares at \$60 and the next best is for 300 shares at \$60.25; a buy market order for 250 shares buys 200 shares at \$60 and 50 shares at \$60.25, moving the best offer $a(t)$ from \$60 to \$60.25. A high density per price of limit orders results in high *liquidity* for market orders; i.e., it implies a small price movement when a market order of a given size is placed.

The Poisson distribution is often useful for estimating the number of rare events in a large population over a unit of time. The rate for a Poisson distribution is the average number of occurrences in a mostly-xed population per unit of time. The parameter in the Poisson distribution is the rate { or how many rare events we expect to observe Using the rate, we can describe the probabili of observing exactly k rare events in a single unit of time. The mean and standard deviation of this distribution are "rate" and "sqrt of rate"

We analyze the queueing properties of such order-matching algorithms with the simple random order placement model shown in Fig. 1. All the order flows are modeled as Poisson processes. We assume that market orders in chunks of σ shares arrive at a rate of μ shares per unit time, with an equal probability for buy orders and sell orders. Similarly, limit orders in chunks of σ shares arrive at a rate of α shares per unit price and per unit time. Offers are placed with uniform probability at integer multiples of a tick size dp in the range $b(t) < p < \infty$, and similarly for bids on $-\infty < p < a(t)$. p represents the logarithm of the price, and dp is a logarithmic price interval [3]. (To avoid repetition the word *price* will henceforth refer to the logarithm of price.) When a market order arrives it causes a transaction. Under the assumption of constant order size, a buy market order removes an offer at price $a(t)$, and a sell market order removes a bid at price $b(t)$. Alternatively, limit orders can be removed spontaneously by being canceled or by expiring. We model this by letting any order be removed randomly with constant probability δ per unit time.

continuous uniform distribution or rectangular distribution is a family of symmetric probability distributions such that for each member of the family, all intervals of the same length on the distribution's support are equally probable. The support is defined by the two parameters, a and b, which are its minimum and maximum values

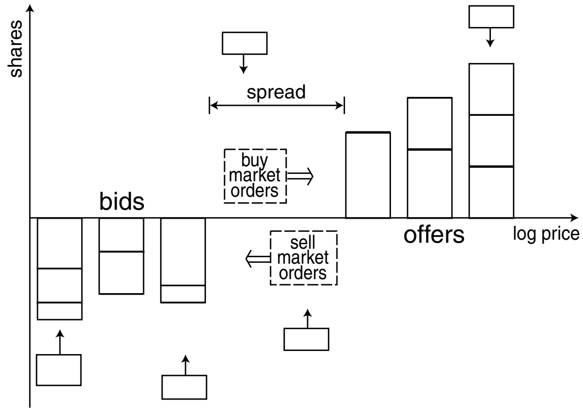

FIG. 1. Schematic of the order-placement process. Stored limit orders are shown "stacked" along the price axis, with bids (buy limit orders) negative and offers (sell limit orders) positive. New limit orders are visualized as "falling" randomly onto the price axis. New offers can be placed at any price greater than the best bid, and new bids can be placed at any price less than the best offer. Limit orders can be removed by random spontaneous deletion or by market orders of the opposite sign.

This order placement process is designed to permit an analytic solution. The model builds on previous work modeling the continuous double auction [4]. While the assumption of limit-order placement over an infinite interval is clearly unrealistic [5], it provides a tractable boundary condition for modeling the behavior of the limit-order book in the region of interest, near the midpoint price $m(t) = [a(t) + b(t)]/2$. It is also justified because limit orders placed far from the midpoint usually expire or are canceled before they are executed. For our analytic model, we use a constant order size σ. In simulations we also use variable order size, e.g., half-normal distributions with standard deviation $\sqrt{2/\pi}\sigma$, which gives similar results.

For simplicity in our model, we do not directly allow limit orders that cross the best price. For example, a buy limit order of size $x + y$ may have a limit price that is higher than the best ask, so that x shares immediately result in a trade, and y shares remain on the book. Such an order is indistinguishable from a market order for x shares immediately followed by a noncrossing limit order for y shares. By definition, our model implicitly allows such events, though it neglects the resulting correlation in order placement.

Dimensional analysis simplifies the analysis of this model and provides rough estimates of its scaling properties. The three fundamental dimensions are *shares*, *price*, and *time*. There are five parameters: three order flow rates and two discreteness parameters. The *order flow rates* are μ, the market order arrival rate, with dimensions of *shares per time*; α, the limit-order arrival rate per unit price, with dimensions of *shares per price per time*; and δ, the rate of limit-order decays, with dimensions of 1/*time*. The two *discreteness parameters* are the price tick size dp, with dimensions of *price*, and the order size σ, with dimensions of *shares*. Because there are five parameters and three dimensions, and because the dimensionality of the parameters is sufficiently rich, all the properties of the limit-order book can be described by functions of two independent parameters.

We perform the dimensional reduction by taking advantage of the fact that the effect of the order flow rates is primary to that of the discreteness parameters. This leads us to construct nondimensional units based on the order flow parameters alone, and take nondimensionalized versions of the discreteness parameters as the independent parameters whose effects remain to be understood. There are three order flow rates and three fundamental dimensions. Temporarily ignoring the discreteness parameters, there are unique combinations of the order flow rates with units of shares, price, and time. These define a characteristic number of shares $N_c = \mu/2\delta$, a characteristic price interval $p_c = \mu/2\alpha$, and a characteristic time scale $t_c = 1/\delta$. (The factors of 2 are a matter of convenience; they occur because we have defined the market order rate for either a buy or a sell order to be $\mu/2$.) These characteristic values can be used to define nondimensional coordinates $\hat{p} = p/p_c$ for price, $\hat{N} = N/N_c$ for shares, and $\hat{t} = t/t_c$ for time.

$\frac{\mu}{2\delta}\left[\frac{\text{shares}}{\text{time}}\frac{\text{time}}{1}\right]$

N_c [shares]

$\frac{\mu}{2\alpha}\left[\frac{\text{shares}}{\text{time}}\frac{\text{time}\cdot\text{price}}{\text{shares}}\right]$

p_c – price

$\frac{1}{\delta}\left[\frac{\text{time}}{1}\right]$

t_c = time

A nondimensional scale parameter based on order size is constructed by dividing the typical order size σ (which is measured in shares) by the characteristic number of shares N_c, i.e., $\epsilon \equiv \sigma/N_c = 2\delta\sigma/\mu$. ϵ characterizes the granularity of the orders stored in the limit-order book. A nondimensional scale parameter based on tick size is constructed by dividing by the characteristic price, i.e., $dp/p_c = 2\alpha dp/\mu$. The theoretical analysis and the simulations show that there is a sensible continuum limit as the tick size $dp \to 0$, in the sense that there is nonzero price diffusion and a finite spread. Furthermore, the dependence on tick size is usually weak, and for many purposes the limit $dp \to 0$ approximates the case of finite tick size fairly well.

Space constraints do not permit us to review the theoretical development of the model in this Letter; it is presented in detail in Ref. [6]. We write an approximate master equation for the number of shares at each price level p at time t, and then find a self-consistent mean-field theory steady-state solution. We also develop an independent interval approximation, borrowing methods from the study of reaction-diffusion phenomena [7]. We find that theory fits the simulation results accurately for large values of ϵ. For small values of ϵ, theory continues to capture the mean spread very well. The predictions of other properties remain qualitatively correct, but are no longer quantitatively accurate. The results we quote here are all from simulations; for the development of the theory and comparisons to simulation see Ref. [6].

Granularity $\epsilon = \frac{\sigma}{N_c}$

In physics and probability theory, mean field theory (MFT also known as self-consistent field theory) studies the behavior of large and complex stochastic models by studying a simpler model.

In the following, we explore the predictions of the model for the basic properties of markets. As already noted, neglecting the effects of the discreteness parameters gives three dimensional quantities and three parameters, which we call the *continuum approximation*. For the continuum approximation dimensional analysis alone yields simple estimates for the most relevant economic properties of the models. We have refined these estimates by simulation and mean-field theory, which take the effects of ϵ and dp/p_c into account. The results are summarized in Table I, and described in more detail below.

The bid-ask spread is the difference between the best price for buying and selling. It is an important determinant of transaction costs. The spread has dimensions of price and therefore scales under the continuum approximation as μ/α. Simulations and theory show that the spread varies as $(\mu/\alpha)f(\epsilon, dp/p_c)$, where f is a fairly flat function with $f(\epsilon, dp/p_c) \approx 1/2$ across much of the range of interest (see Ref. [6]).

Another interesting quantity is the average *depth profile* $n(p) = \langle n(p, t)\rangle$, which is the density of shares per price interval. The average depth profile is relatively small near the midpoint and increases to an asymptotic value far from the midpoint, as shown in Fig. 2. The approach to an asymptotic value is a consequence of our assumption of uniform order placement over an infinite range. It should be viewed as a convenient boundary condition for understanding the depth near $\hat{p} = 0$, where transactions occur. From dimensional analysis the asymptotic depth, which has units of shares/price is α/δ. This result is exact.

An important property of the depth profile is its slope near the origin, which determines the price response to the placement of a small market order. From continuum dimensional analysis, the slope of the average depth profile scales as $\lambda \sim \alpha^2/\mu\delta$. This is altered by effects due to the granularity of orders. For large ϵ, the depth profile is a concave function with nonzero values at $\hat{p} = 0$, whereas for small ϵ, $n(0) \approx 0$ and $n(\hat{p})$ is convex near $\hat{p} = 0$.

In addition to the spread, the price response for executing a market order is also a key factor determining transaction costs. It can be characterized by a price impact function $\Delta p = \phi(\omega, \tau, t)$, where Δp is the price shift at time $t + \tau$ caused by a market order ω at time t. Here we study the average instantaneous price impact $\phi(\omega) = \langle\phi(\omega, 0, t)\rangle$, where the average is taken over time t, and the limit $\tau \rightarrow 0$ corresponds to the change in quoted midpoint price immediately after a market order is placed. $\phi(\omega)$ can be understood in terms of the depth profile, as explained in Ref. [6]. Price impact causes market friction, since selling tends to drive the price up and buying tends to drive it down, so executing a circuit causes a loss. The price impact is closely related to the demand function, providing a natural starting point for theories of statistical or dynamical properties of markets [8]. A naive argument predicts that the price impact $\phi(\omega)$ should increase at least linearly [6]. In contrast, from empirical studies $\phi(\omega)$ for buy orders grows more slowly than linearly [9], and the most accurate measurements make it clear that it is strongly concave [2,10].

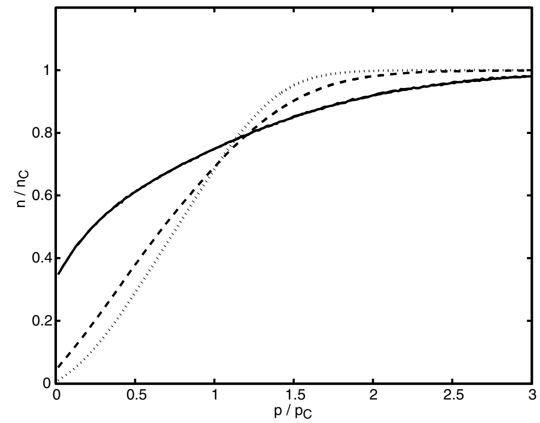

FIG. 2. The mean depth profile versus price in nondimensional coordinates $\hat{n} = np_c/N_c = n\delta/\alpha$ vs $\hat{p} = p/p_c = 2\alpha p/\mu$. The origin $p/p_c = 0$ corresponds to the midpoint. We show three different values of the nondimensional granularity parameter: $\epsilon = 0.2$ (solid line), $\epsilon = 0.02$ (dashed line), $\epsilon = 0.002$ (dotted line), all with tick size $dp = 0$.

The $\phi(\omega)$ predicted by our model is shown in Fig. 3. It approaches a linear function for large ϵ, but for smaller values of ϵ it is strongly concave, particularly near the midpoint. Plotting this on log-log scale, this function does not follow a pure power law. For example, for $\epsilon = 0.002$, the exponent is $\beta \approx 0.5$ for small orders, and $\beta \approx 0.2$ for larger orders. This is in agreement with the

TABLE I. Predictions of scaling of market properties vs order flow. The third column contains predictions from the continuum analysis, in which the discreteness parameters are ignored, and the fourth column gives more accurate predictions from theory and simulation. The functions f and g are the order of magnitude of one throughout the relevant ranges of variation of ϵ and dp/p_c.

Quantity	Dimensions	Continuum scaling relation	Scaling from simulation and theory
Asymptotic depth	shares/price	$d \sim \alpha/\delta$	$d = \alpha/\delta$
Spread	price	$s \sim \mu/\alpha$	$s = (\mu/\alpha)f(\epsilon, dp/p_c)$
Slope of depth profile	shares/price^2	$\lambda \sim \alpha^2/\mu\delta = d/s$	$\lambda = (\alpha^2/\mu\delta)g(\epsilon, dp/p_c)$
Price diffusion rate	price^2/time	$D \sim \mu^2\delta/\alpha^2$	$(\tau \rightarrow 0,\ dp \rightarrow 0)\ D_0 \sim \mu^2\delta/\alpha^2\epsilon^{-0.5}$
			$(\tau \rightarrow \infty,\ dp \rightarrow 0)\ D_\infty \sim \mu^2\delta/\alpha^2\epsilon^{0.5}$

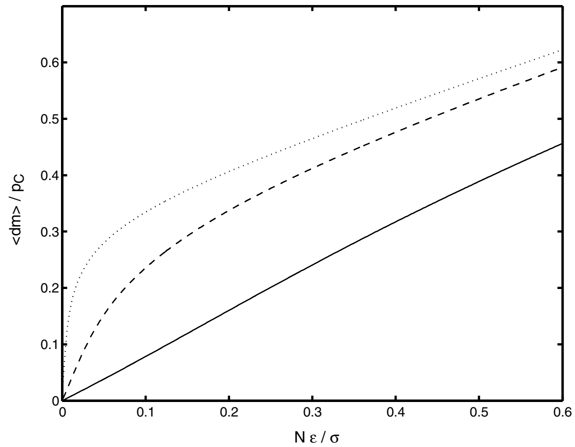

FIG. 3. The average price impact corresponding to the results in Fig. 2. The average instantaneous movement of the nondimensional midprice, $\langle dm \rangle / p_c$ caused by an order of size $N/N_c = N\epsilon/\sigma$. $\epsilon = 0.2$ (solid line), $\epsilon = 0.02$ (dashed line), $\epsilon = 0.002$ (dotted line).

best empirical measurements for the New York Stock Exchange and the London Stock Exchange [2,10].

The price diffusion rate is a property of central interest. In finance, it is typically characterized in terms of the standard deviation of prices at a particular time scale, which is referred to as *volatility*. Volatility is a measure of the uncertainty of price movements and is the standard way to characterize risk. We have made simulations of the variance of the change in the midpoint price at time scale τ, i.e., the variance of $m(t + \tau) - m(t)$. The slope is the diffusion rate, which at any fixed time scale is proportional to the square of the volatility. It appears that there are at least two time scales involved, with a faster diffusion rate for short time scales and a slower diffusion rate for long time scales. Such correlated diffusion is not predicted by mean-field analysis. Simulation results show that the diffusion rate is correctly described by the product of the estimate from continuum dimensional analysis $\mu^2\delta/\alpha^2$, and a τ-dependent power of the nondimensional granularity parameter $\epsilon = 2\delta\sigma/\mu$, as summarized in Table I. We cannot currently explain why this power is $-1/2$ for short term diffusion and $1/2$ for long-term diffusion.

This model contains numerous simplifying assumptions. Nevertheless, it is its very simplicity that allows us to make unambiguous predictions about the most basic properties of real markets. Our prediction for the price impact function agrees with the best empirical measurements on the New York and London Stock Exchanges [2,10], and suggests that concavity is a robust feature deriving from institutional structure rather than rationality. The fact that in both markets appropriate rescaling allows a collapse onto a function of the type we predict suggests the existence of universal supply and demand functions. Futhermore, the results indicate remarkable explanatory power for the average daily spread [2]. Even though we do not expect the predictions of this model to be exact in every detail, they provide a simple benchmark that can guide future improvements. Our model illustrates how the need to store supply and demand gives rise to interesting temporal properties of prices and liquidity even under assumptions of perfectly random order flow, and demonstrates the importance of making realistic models of market mechanisms.

We thank the McKinsey Corporation, Credit Suisse First Boston, Bob Maxfield, and Bill Miller for supporting this research, and Supriya Krishnamurthy for useful discussions.

[1] R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, England, 2000); J.-P. Bouchaud and M. Potters, *Theory of Financial Risk* (Cambridge University Press, Cambridge, England, 2000).

[2] M. Daniels, J. D. Farmer, L. Gillemot, P. Patelli, E. Smith, and I. Zovko (to be published).

[3] In real markets the price interval dp is linear rather than logarithmic; as long as $p \gg dp$ this is a good assumption.

[4] I. Domowitz and Jianxin Wang, J. Econ. Dyn. Control **18**, 29 (1994); T. Bollerslev, I. Domowitz, and J. Wang, J. Econ. Dyn. Control **21**, 1471 (1997); P. Bak, M. Paczuski, and M. Shubik, Physica (Amsterdam) **246A**, 430 (1997); D. Eliezer and I. I. Kogan, http://xxx.lanl.gov/cond-mat/9808240; S. Maslov, Physica (Amsterdam) **278A**, 571 (2000); G. Iori and C. Chiarella, Quant. Finance **2**, 346 (2002); F. Slanina, Phys. Rev. E **64**, 056136 (2001); D. Challet and R. Stinchcombe, Physica (Amsterdam) **300A**, 285 (2001).

[5] J.-P. Bouchaud, M. Mezard, and M. Potters, Quant. Finance **2**, 251 (2002); I. Zovko and J. D. Farmer, Quant. Finance **2**, 387 (2002).

[6] Eric Smith, J. Doyne Farmer, László Gillemot, and Supriya Krishnamurthy, xxx.lanl.gov/cond-mat/0210475.

[7] S. N. Majumdar, S. Krishnamurthy, and M. Barma, J. Stat. Phys. **99**, 1 (2000).

[8] J. D. Farmer, SFI working paper 98-12-117; J.-P. Bouchaud and R. Cont, Eur. Phys. J. B **6**, 543 (1998).

[9] J. A. Hausman and A.W. Lo, J. Finan. Economics **31**, 319 (1992); J. D. Farmer, http://www.predict.com/jdf/slippage.pdf (1996); N. Torre, *BARRA Market Impact Model Handbook* (BARRA Inc., Berkeley, CA, 1997); www.barra.com; A. Kempf and O. Korn, J. Fin. Markets **2**, 29 (1999); V. Plerou, P. Gopikrishnan, X. Gabaix, and H. E. Stanley, Phys. Rev. E **66**, 027104 (2002).

[10] F. Lillo, J. D. Farmer, and R. Mantegna, Nature (London) **421**, 129 (2003).