QUANTITATIVE FINANCE VOLUME 3 (2003) 481–514 RESEARCH PAPER
INSTITUTE OF PHYSICS PUBLISHING quant.iop.org

Statistical theory of the continuous double auction

Eric Smith[1], **J Doyne Farmer, László Gillemot and Supriya Krishnamurthy**

Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

E-mail: desmith@santafe.edu

Received 30 October 2002, in final form 16 July 2003
Published 18 September 2003
Online at stacks.iop.org/Quant/3/481

Abstract

Most modern financial markets use a continuous double auction mechanism to store and match orders and facilitate trading. In this paper we develop a microscopic dynamical statistical model for the continuous double auction under the assumption of IID random order flow, and analyse it using simulation, dimensional analysis, and theoretical tools based on mean field approximations. The model makes testable predictions for basic properties of markets, such as price volatility, the depth of stored supply and demand versus price, the bid–ask spread, the price impact function, and the time and probability of filling orders. These predictions are based on properties of order flow and the limit order book, such as share volume of market and limit orders, cancellations, typical order size, and tick size. Because these quantities can all be measured directly there are no free parameters. We show that the order size, which can be cast as a non-dimensional granularity parameter, is in most cases a more significant determinant of market behaviour than tick size. We also provide an explanation for the observed highly concave nature of the price impact function. On a broader level, this work suggests how stochastic models based on zero intelligence agents may be useful to probe the structure of market institutions. Like the model of perfect rationality, a stochastic zero intelligence model can be used to make strong predictions based on a compact set of assumptions, even if these assumptions are not fully believable.

1. Introduction

This section provides background and motivation, a description of the model, and some historical context for work in this area. Section 2 gives an overview of the phenomenology of the model, explaining how dimensional analysis applies in this context, and presenting a summary of numerical results. Section 3 develops an analytic treatment of the model, explaining some of the numerical findings of section 2. We conclude in section 4 with a discussion of how the model may be enhanced to bring it closer to real life markets, and some comments comparing the approach taken here to standard models based on information arrival and valuation.

1.1. Motivation

In this paper we analyse the continuous double auction trading mechanism under the assumption of random order flow, developing a model introduced in [1], which is in turn based on a line of earlier work [2–11]. This analysis produces quantitative predictions about the most basic properties of markets, such as volatility, depth of stored supply and demand,

[1] Author to whom any correspondence should be addressed.

1469-7688/03/060481+34$30.00 PII: S1469-7688(03)55402-1

Maureen O'Hara defines market microstructure as "the study of the process and outcomes of exchanging assets under a specific set of rules. While much of economics abstracts from the mechanics of trading, microstructure theory focuses on how specific trading mechanisms affect the price formation process.
Microstructure deals with issues of market structure and design, price formation and price discovery, transaction and timing cost, information and disclosure, and market maker and investor behavior.

Transaction costs are of two types: processing costs and market impact costs. Processing costs are commissions paid to brokers, taxes, and fees. These costs are known and easily measurable. Market impact costs consist of the halfspread and additional market impact, if the order exceeds the quote depth for the security to be purchased or sold. The market impact can be seen as the price concession we must pay to liquidity providers in order to accommodate our trade. Market impact is intrinsically difficult to measure and represents the bulk of the total transaction costs. As a rule, that which is not easily measurable or quantifiable tends to receive less of our attention
http://www.barra.com/newsletter/nl166/sywtcnl166.asp
In financial markets, market impact is the effect that a market participant has when it buys or sells an asset. It is the extent to which the buying or selling moves the price against the buyer or seller, i.e. upward when buying and downward when selling. It is closely related to market liquidity; in many cases "liquidity" and "market impact" are synonymous.

the bid–ask spread, the price impact, and probability and time to fill[2]. These predictions are based on the rate at which orders flow into the market, and other parameters of the market, such as order size and tick size. The predictions are falsifiable, based on parameters that can all be measured independently of the quantities of interest. This extends the original random walk model of Bachelier [12] by providing a basis for the diffusion rate of prices. The model also provides a possible explanation for the highly concave nature of the price impact function. Even though some of the assumptions of the model are too simple to be literally true, the model provides a foundation onto which more realistic assumptions may easily be added.

The model demonstrates the importance of financial institutions in setting prices, and how solving a necessary economic function such as providing liquidity can have unanticipated side-effects. In a world of imperfect rationality and imperfect information, the task of demand storage necessarily causes persistence. Under perfect rationality all traders would instantly update their orders with the arrival of each piece of new information, but this is clearly not true for real markets. The limit order book, which is the queue used for storing unexecuted orders, has a long memory when there are persistent orders. It can be regarded as a device for storing supply and demand, somewhat like how a capacitor is a device for storing charge. We show that even under completely random IID order flow, the price process displays anomalous diffusion and interesting temporal structure. The converse is also interesting: for prices to be effectively random, incoming order flow must be non-random, in just the right way to compensate for the persistence (see the remarks in section 4.4.)

This work is also of interest from a fundamental point of view because it suggests an alternative approach to doing economics. The assumption of perfect rationality has been popular in economics because it provides a parsimonious model that makes strong predictions. In the spirit of Becker [13] and Gode and Sunder [14], we show that the opposite extreme of zero intelligence random behaviour provides another reference model that also makes very strong predictions. Like perfect rationality, zero intelligence is an extreme simplification that is obviously not literally true. But as we show here, it provides a useful tool for probing the behaviour of financial institutions. The resulting model may easily be extended by introducing simple boundedly rational behaviours. We also differ from standard treatments in that we do not attempt to understand the properties of prices from fundamental assumptions about utility. Rather, we split the problem in two. We attempt to understand how prices depend on order flow rates, leaving the problem of what determines these order flow rates for the future.

[2] The theoretical development presented here was done in advance of looking at the data, but as this paper is going to press, tests based on London Stock Exchange data have shown that some of the predictions of this model are extremely good. On a roughly annual timescale, regressions based on this model explain about 96% of the variance across different stocks of the bid–ask spread, 85% of the variance of the price diffusion rate, and the non-dimensional coordinates defined by the model produce a good collapse for the price impact function [25].

One of our main results concerns the average price impact function. The liquidity for executing a market order can be characterized by a price impact function $\Delta p = \phi(\omega, \tau, t)$. Δp is the shift in the logarithm of the price at time $t + \tau$ caused by a market order of size ω placed at time t. Understanding price impact is important for practical reasons such as minimizing transaction costs, and also because it is closely related to an excess demand function[3], providing a natural starting point for theories of statistical or dynamical properties of markets [15, 16]. A naive argument predicts that the price impact $\phi(\omega)$ should increase at least linearly. This argument goes as follows: fractional price changes should not depend on the scale of price. Suppose buying a single share raises the price by a factor $k > 1$. If k is constant, buying ω shares in succession should raise it by k^{ω}. Thus, if buying ω shares all at once affects the price at least as much as buying them one at a time, the ratio of prices before and after impact should increase at least exponentially. Taking logarithms implies that the price impact as we have defined it above should increase at least linearly[4].

In contrast, from empirical studies $\phi(\omega)$ for buy orders appears to be concave [17–22]. Lillo *et al* [22] have shown that for stocks in the NYSE the concave behaviour of the price impact is quite consistent across different stocks. Our model produces concave price impact functions that are in qualitative agreement with these results.

Our work also demonstrates the value of physics techniques for economic problems. Our analysis makes extensive use of dimensional analysis, the solution of a master equation through a generating functional, and a mean field approach that is commonly used to analyse non-equilibrium reaction–diffusion systems and evaporation–deposition problems.

1.2. Background: the continuous double auction

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 order. 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*, or *quotes* which also state a limit price, corresponding to the worst allowable price for the transaction. (Note that the word 'quote' can be used either to refer to the limit price or to the limit order itself.) 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 use the logarithmic price $a(t)$ to denote the position of the best (lowest) offer and $b(t)$ for the position of the best (highest) bid. These are also called the *inside quotes*. There is

[3] In financial models it is common to define an excess demand function as demand minus supply; when the context is clear the modifier 'excess' is dropped, so demand refers to both supply and demand.

[4] This has practical implications. It is common practice to break up orders in order to reduce losses due to market impact. With a sufficiently concave market impact function, in contrast, it is cheaper to execute an order all at once.

typically a non-zero price gap between them, called the *spread* $s(t) = a(t) - b(t)$. Prices are not continuous, but rather have discrete quanta called *ticks*. Throughout this paper, all prices will be expressed as logarithms, and to avoid endless repetition, the word *price* will mean the logarithm of the price. The minimum interval that prices change on is the *tick size* dp (also defined on a logarithmic scale; note that this is not true for real markets). Note that dp is *not* necessarily infinitesimal.

As market orders arrive they are matched against limit orders of the opposite sign in order of first price and then arrival time, as shown in figure 1. 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 of limit orders per price results in high *liquidity* for market orders, i.e., it decreases the price movement when a market order is placed. Let $n(p, t)$ be the stored density of limit order volume at price p, which we will call the *depth profile* of the limit order book at any given time t. The total stored limit order volume at price level p is $n(p, t)\,dp$. For unit order size the shift in the best ask $a(t)$ produced by a buy market order is given by solving the equation

discreet quanta of price

dp, value of price in this price, min interval

next price shift in price / depth profile / tick size

$$\omega = \sum_{p=a(t)}^{p'} n(p, t)\,dp \qquad (1)$$

density lim vol at p / min interval that price change on

for p'. The shift in the best ask, $p' - a(t)$, is the instantaneous price impact for buy market orders. A similar statement applies for sell market orders, where the price impact can be defined in terms of the shift in the best bid. (Alternatively, it is also possible to define the price impact in terms of the change in the mid-point price.)

We will refer to a buy limit order whose limit price is greater than the best ask, or a sell limit order whose limit price is less than the best bid, as a *crossing limit order* or *marketable limit order*. Such limit orders result in immediate transactions, with at least part of the order immediately executed.

1.3. The model

This model introduced in [1] is designed to be as analytically tractable as possible while capturing key features of the continuous double auction. All the order flows are modelled as Poisson processes. We assume that market orders arrive in chunks of σ shares, at a rate of μ shares per unit time. The market order may be a 'buy' order or a 'sell' order with equal probability. (Thus the rate at which buy orders or sell orders arrive individually is $\mu/2$.) Limit orders arrive in chunks of σ shares as well, at a rate α shares per unit price and per unit time for buy orders and also for sell orders. Offers are placed with uniform probability at integer multiples of a tick size dp in the range of price $b(t) < p < \infty$, and similarly for bids on $-\infty < p < a(t)$. 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 if it was the last offer at that price, moves the best ask up to the next occupied

a(t) / b(t)

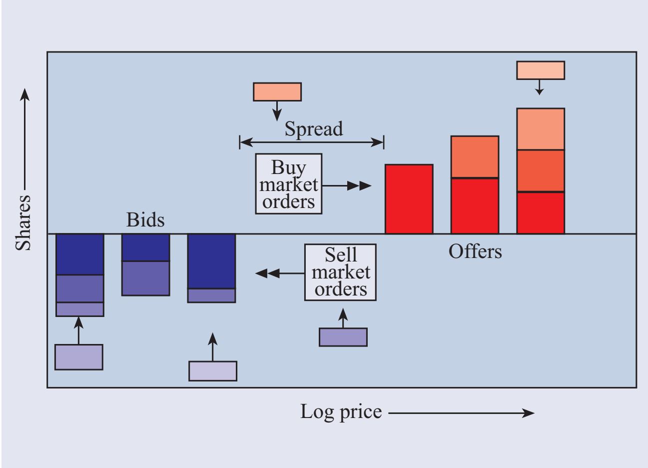

Figure 1. A schematic illustration of the continuous double auction mechanism and our model of it. Limit orders are stored in the limit order book. We adopt the arbitrary convention that buy orders are negative and sell orders are positive. As a market order arrives, it has transactions with limit orders of the opposite sign, in order of price (first) and time of arrival (second). The best quotes at prices $a(t)$ or $b(t)$ move whenever an incoming market order has sufficient size to fully deplete the stored volume at $a(t)$ or $b(t)$. Our model assumes that market order arrival, limit order arrival, and limit order cancellation follow a Poisson process. New offers (sell limit orders) can be placed at any price greater than the best bid, and are shown here as 'raining down' on the price axis. Similarly, new bids (buy limit orders) can be placed at any price less than the best offer. Bids and offers that fall inside the spread become the new best bids and offers. All prices in this model are logarithmic.

price tick. Similarly, a sell market order removes a bid at price $b(t)$, and if it is the last bid at that price, moves the best bid down to the next occupied price tick. In addition, limit orders may also be removed spontaneously by being cancelled or by expiring, even without a transaction having taken place. We model this by letting them be removed randomly with constant probability δ per unit time.

While the assumption of limit order placement over an infinite interval is clearly unrealistic, it provides a tractable boundary condition for modelling the behaviour of the limit order book near the mid-point price $m(t) = (a(t) + b(t))/2$, which is the region of interest since it is where transactions occur. Limit orders far from the mid-point are usually cancelled before they are executed (we demonstrate this later in figure 5), and so far from the mid-point, limit order arrival and cancellation have a steady state behaviour characterized by a simple Poisson distribution. Although under the limit order placement process the total number of orders placed per unit time is infinite, the order placement per unit price interval is bounded and thus the assumption of an infinite interval creates no problems. Indeed, it guarantees that there are always an infinite number of limit orders of both signs stored in the book, so the bid and ask are always well defined and the book never empties. (Under other assumptions about limit order placement this is not necessarily true, as we later demonstrate in figure 30.) We are also considering versions of the model involving more realistic order placement functions; see the discussion in section 4.3.

Limit order placement process ⇓ infinit, bounded

total no. of orders unit TIME

total no. of orders unit PRICE

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"

$-\infty < p < a(t)$ $b(t) < p < \infty$

Market Order: bid(t) | ask(t)

Market Maker: ask a(t) | bid b(t) | Limit order crossing - A Crossing is a trade where a Trading Participant i s both the buyer and the seller.

In this model, to keep things simple, we are using the conceptual simplification of *effective market orders* and *effective limit orders*. When a crossing limit order is placed part of it may be executed immediately. The effect of this part on the price is indistinguishable from that of a market order of the same size. Similarly, given that this market order has been placed, the remaining part is equivalent to a non-crossing limit order of the same size. Thus a crossing limit order can be modelled as an effective market order followed by an effective (non-crossing) limit order[5]. Working in terms of effective market and limit orders affects data analysis: the effective market order arrival rate μ combines both pure market orders and the immediately executed components of crossing limit orders, and similarly the limit order arrival rate α corresponds only to the components of limit orders that are not executed immediately. This is consistent with the boundary conditions for the order placement process, since an offer with $p \leqslant b(t)$ or a bid with $p \geqslant a(t)$ would result in an immediate transaction, and thus would be effectively the same as a market order. Defining the order placement process with these boundary conditions realistically allows limit orders to be placed anywhere inside the spread.

Another simplification of this model is the use of logarithmic prices, both for the order placement process and for the tick size $\mathrm{d}p$. This has the important advantage that it ensures that prices are always positive. In real markets price ticks are linear, and the use of logarithmic price ticks is an approximation that makes both the calculations and the simulation more convenient. We find that the limit $\mathrm{d}p \to 0$, where tick size is irrelevant, is a good approximation for many purposes. We find that tick size is less important than other parameters of the problem, which provides some justification for the approximation of logarithmic price ticks.

Assuming a constant probability for cancellation is clearly *ad hoc*, but in simulations we find that other assumptions with well-defined timescales, such as constant duration time, give similar results. 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 $\sqrt{\pi/2}\sigma$, which ensures that the mean value remains σ. As long as these distributions have thin tails, the differences do not qualitatively affect most of the results reported here, except in a trivial way. As discussed in section 4.3, decay processes without well-defined characteristic times and size distributions with power law tails give qualitatively different results and will be treated elsewhere.

Even though this model is simply defined, the time evolution is not trivial. One can think of the dynamics as being composed of three parts:

(1) the buy market order/sell limit order interaction, which determines the best ask;
(2) the sell market order/buy limit order interaction, which determines the best bid; and
(3) the random cancellation process.

[5] In assigning independently random distributions for the two events, our model neglects the correlation between market and limit order arrival induced by crossing limit orders.

Processes (1) and (2) determine each other's boundary conditions. That is, process (1) determines the best ask, which sets the boundary condition for limit order placement in process (2), and process (2) determines the best bid, which determines the boundary conditions for limit order placement in process (1). Thus processes (1) and (2) are strongly coupled. It is this coupling that causes the bid and ask to remain close to each other, and guarantees that the spread $s(t) = a(t) - b(t)$ is a stationary random variable, even though the bid and ask are not. It is the coupling of these processes through their boundary conditions that provides the non-linear feedback that makes the price process complex.

CHECK HOW BID AND ASK ARE GENERETED

1.4. Summary of prior work

There are two independent lines of prior work, one in the financial economics literature, and the other in the physics literature. The models in the economics literature are directed toward econometrics and treat the order process as static. In contrast, the models in the physics literature are mostly conceptual toy models, but they allow the order process to react to changes in prices, and are thus fully dynamic. Our model bridges this gap. This is explained in more detail below.

The first model of this type that we are aware of in the economics literature was due to Mendelson [2], who modelled random order placement with periodic clearing. Cohen *et al* [3] developed a model of a continuous auction, modelling limit orders, market orders, and order cancellation as Poisson processes. However, they only allowed limit orders at two fixed prices, buy orders at the best bid, and sell orders at the best ask. This assumption allowed them to use standard results from queuing theory to compute properties such as the expected number of stored limit orders, the expected time to execution, and the relative probability of execution versus cancellation. Domowitz and Wang [4] extended this to multiple price levels by assuming arbitrary order placement and cancellation processes (which can take on any value at each price level). They assume that these processes are fixed in time, and do not respond to changes in the best bid or ask. This allows them to derive the distribution of the spread, transaction prices, and waiting times for execution. This model was tested by Bollerslev *et al* [5] on three weeks' worth of data for the Deutschmark/US Dollar exchange rate. They showed that it does a good job of predicting the distribution of the spread. However, since the prices are pinned, the model does not make a prediction about price diffusion, and this also creates errors in the predictions of the spread and stored supply and demand.

The models in the physics literature, which appear to have been developed independently, differ in that they address price dynamics. That is, they incorporate the feedback between order placement and price formation, allowing the order placement process to change in response to changes in prices. These models have mainly been conceptual toy models designed to explain the anomalous diffusion properties of prices (a property that all of these models fail to reproduce, as explained later). This line of work begins with a paper by Bak *et al* [6] which was developed by Eliezer and Kogan [7] and by Tang [8]. They assume that limit orders are placed at a fixed distance from the mid-point, and that the limit prices of these orders are then randomly shuffled until they result in transactions. It is the random shuffling that causes price diffusion. This assumption, which we feel is unrealistic, was made to take advantage of the analogy to a standard reaction–diffusion model in the physics literature. Maslov [9] introduced an alterative model that was solved analytically in the mean field limit by Slanina [10]. Each order is randomly chosen to be either a buy or a sell with equal probability, and either a limit order or a market order with equal probability. If a limit order, it is randomly placed within a fixed distance of the current price. Both the Bak *et al* model and that of Maslov result in anomalous price diffusion, in the sense of the Hurst exponent $H = 1/4$ (in contrast to standard diffusion, which has $H = 1/2$, or real prices which tend to have $H > 1/2$). In addition, the Maslov model unrealistically requires equal probabilities for limit and market order placement; otherwise the inventory of stored limit orders either goes to zero or grows without bound. A model adding a Poisson order cancellation process was proposed by Challet and Stinchcombe [11], and independently by Daniels *et al* [1]. Challet and Stinchcombe showed that this results in $H = 1/4$ for short times, but asymptotically gives $H = 1/2$. The Challet and Stinchcombe model, which posits an arbitrary, unspecified function for the relative position of limit order placement, is quite similar to that of Domowitz and Wang [4], but allows for the possibility of order placement responding to price movement.

The model studied in this paper was introduced by Daniels *et al* [1]. Like other physics models, it treats the feedback between order placement and price movement. It has the advantage that it is defined in terms of five scalar parameters, and so is parsimonious and can easily be tested against real data. Its simplicity enables a dimensional analysis, which gives approximate predictions about many of the properties of the model. Perhaps most important is the use to which the model is put: with the exception of [7], work in the physics literature has focused almost entirely on the anomalous diffusion of prices. While interesting and important for refining risk calculations, from a practical point of view this is a second order effect. In contrast, we focus on first order effects of primary interest to market participants, such as the bid–ask spread, volatility, depth profile, price impact, and the probability and time to fill an order. We demonstrate how dimensional analysis becomes a useful tool in an economic setting, and develop mean field theories to explain the properties of the model. Many of the important properties of the model can be stated in terms of simple scaling relations in terms of the five parameters.

Subsequent to [1], Bouchaud *et al* [23] demonstrated that they can derive a simple equation for the depth profile, by making the assumption that prices execute a random walk and introducing an additional free parameter. In this paper we show how to do this from first principles without introducing a free parameter. Iori and Chiarella [24] have numerically studied fundamentalists and technical traders placing limit orders; a talk on this work by Iori in part inspired this model.

2. Overview of predictions of the model

In this section we give an overview of the phenomenology of the model. Because this model has five parameters, understanding all their effects would generally be a complicated problem in and of itself. This task is greatly simplified by the use of dimensional analysis, which reduces the number of independent parameters from five to two. Thus, before we can even review the results, we need to first explain how dimensional analysis applies in this setting. One of the surprising aspects of this model is that one can derive several powerful results using the simple technique of dimensional analysis alone.

Unless otherwise mentioned the results presented in this section are based on simulations. These results are compared to theoretical predictions in section 3.

2.1. Dimensional analysis

Because dimensional analysis is not commonly used in economics we first present a brief review. For more details see Bridgman [26].

Dimensional analysis is a technique that is commonly used in physics and engineering to reduce the number of independent degrees of freedom by taking advantage of the constraints imposed by dimensionality. For sufficiently constrained problems it can be used to guess the answer to a problem without doing a full analysis. The idea is to write down all the factors that a given phenomenon can depend on, and then find the combination that has the correct dimensions. For example, consider the problem of the period of a pendulum: the period T has dimensions of *time*. Obvious candidates that it might depend on are the mass of the bob m (which has units of *mass*), the length l (which has units of *distance*), and the acceleration of gravity g (which has units of $distance/time^2$). There is only one way to combine these to produce something with dimensions of *time*, i.e. $T \sim \sqrt{l/g}$. This determines the correct formula for the period of a pendulum up to a constant. Note that it makes it clear that the period does not depend on the mass, a result that is not obvious *a priori*. We were lucky in this problem because there were three parameters and three dimensions, with a unique combination of the parameters having the right dimensions; in general dimensional analysis can only be used to reduce the number of free parameters through the constraints imposed by their dimensions.

For this problem the three fundamental dimensions in the model are *shares*, *price*, and *time*. Note that by *price*, we mean the logarithm of price; as long as we are consistent, this does not create problems with the dimensional analysis. There are five parameters: three rate constants 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 $1/time$. These play a role similar to rate constants in physical problems. The two *discreteness parameters* are the price tick size $\mathrm{d}p$, with dimensions of *price*, and the order size σ, with dimensions of *shares*. This is summarized in table 1.

3 DiM
· Shares
· price
· time
5 PARAM
→3 rate const
ORDER FLOW RATE
· $\mu\left[\frac{\text{shares}}{\text{time}}\right]$ market order arrival
· $\alpha\left[\frac{\text{shares}}{\text{price}\cdot\text{time}}\right]$ Limit order arrival
· $\delta\left[\frac{1}{\text{time}}\right]$ rate of limit order decays

→2 discreteness parameters
· dp[price] tick size
· σ [shares] order size

Table 1. The five parameters that characterize this model. α, μ and δ are order flow rates, and $\mathrm{d}p$ and σ are discreteness parameters.

Parameter	Description	Dimensions
α	Limit order rate	Shares/(price time)
μ	Market order rate	Shares/time
δ	Order cancellation rate	1/time
$\mathrm{d}p$	Tick size	Price
σ	Characteristic order size	Shares

3 order flow rates

discreteness parameters

Table 2. Important characteristic scales and non-dimensional quantities. We summarize the characteristic share size, price and times defined by the order flow rates, as well as the two non-dimensional scale parameters $\mathrm{d}p/p_c$ and ϵ that characterize the effect of finite tick size and order size. Dimensional analysis makes it clear that all the properties of the limit order book can be characterized in terms of functions of these two parameters.

Parameter	Description	Expression
N_c	Characteristic number of shares	$\mu/2\delta$ (Shares)
p_c	Characteristic price interval	$\mu/2\alpha$ [price]
t_c	Characteristic time	$1/\delta$ [time]
$\mathrm{d}p/p_c$	Non-dimensional tick size	$2\alpha\,\mathrm{d}p/\mu$
ϵ	Non-dimensional order size	$2\delta\sigma/\mu$

Dimensional analysis can be used to reduce the number of relevant parameters. Because there are five parameters and three dimensions (*price, shares, time*), and because in this case the dimensionality of the parameters is sufficiently rich, the dimensional relationships reduce the degrees of freedom, so all the properties of the limit order book can be described by functions of two parameters. It is useful to construct these two parameters so that they are non-dimensional. $(\frac{dp}{p_c} ; \epsilon)$

We perform the dimensional reduction of the model by guessing that the effect of the order flow rates is primary to that of the discreteness parameters. This leads us to construct non-dimensional units based on the order flow parameters alone, and take non-dimensionalized versions of the discreteness parameters as the independent parameters whose effects remain to be understood. As we will see, this is justified by the fact that many of the properties of the model depend only weakly on the discreteness parameters. We can thus understand much of the richness of the phenomenology of the model through dimensional analysis alone.

There are three order flow rates and three fundamental dimensions. If we temporarily ignore 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 timescale $t_c = 1/\delta$. This is summarized in table 2. The factors of two occur because we have defined the market order rate for either a buy or a sell order to be $\mu/2$. We can thus express everything in the model in non-dimensional terms by dividing by N_c, p_c, or t_c as appropriate, e.g. to measure shares in non-dimensional units $\hat{N} = N/N_c$, or to measure price in non-dimensional units $\hat{p} = p/p_c$.

depth profili => stored density n(p, t) of limit order volume at price p,

The value of using non-dimensional units is illustrated in figure 2. Figure 2(*a*) shows the average depth profile for three different values of μ and δ with the other parameters held fixed. When we plot these results in dimensional units the results look quite different. However, when we plot them in terms of non-dimensional units, as shown in figure 2(*b*), the results are indistinguishable. As explained below, because we have kept the non-dimensional order size fixed, the collapse is perfect. Thus, the problem of understanding the behaviour of this model is reduced to studying the effect of tick size and order size.

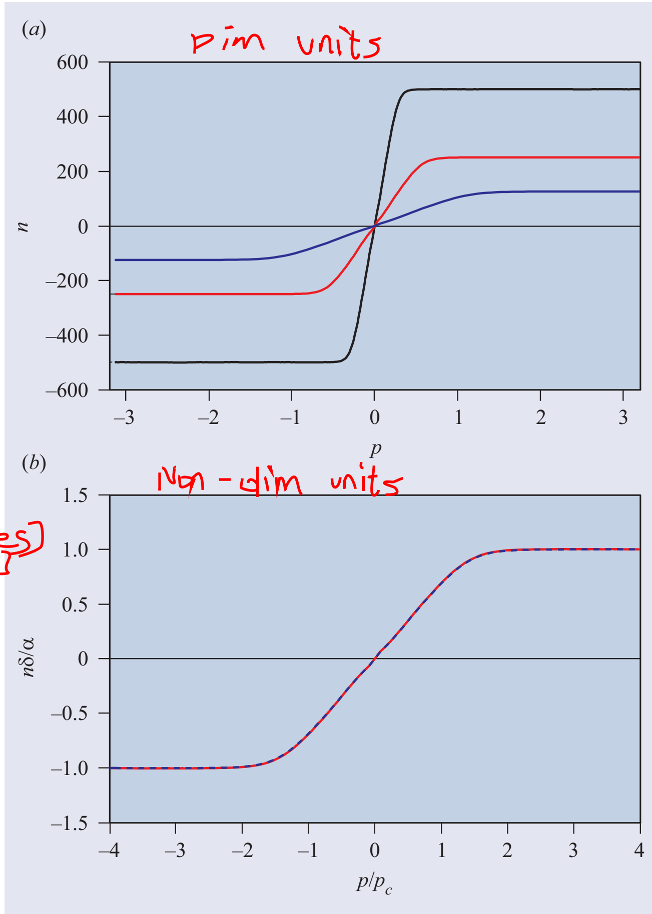

Figure 2. The usefulness of non-dimensional units. (*a*) We show the average depth profile for three different parameter sets. The parameters $\alpha = 0.5$, $\sigma = 1$, and $\mathrm{d}p = 0$ are held constant, while δ and μ are varied. The curve types are: black: $\delta = 0.001$, $\mu = 0.2$; red: $\delta = 0.002$, $\mu = 0.4$; and blue: $\delta = 0.004$, $\mu = 0.8$. (*b*) The same, but plotted in non-dimensional units. All three curves collapse on top of each other. The horizontal axis has units of *price*, and so has non-dimensional units $\hat{p} = p/p_c = 2\alpha p/\mu$. The vertical axis has units of n shares/price, and so has non-dimensional units $\hat{n} = np_c/N_c = n\delta/\alpha$. Because we have chosen the parameters to keep the non-dimensional order size ϵ constant, the collapse is perfect. Varying the tick size has little effect on the results other than making them discrete.

To understand the effect of tick size and order size it is useful to do so in non-dimensional terms. The non-dimensional scale parameter based on tick size is constructed by dividing by the characteristic price, i.e. $\mathrm{d}p/p_c = 2\alpha\,\mathrm{d}p/\mu$. The theoretical analysis and the simulations show that there is a sensible continuum limit as the tick size $\mathrm{d}p \to 0$, in the sense that there is non-zero price diffusion and a finite spread. Furthermore, the dependence on tick size is weak, and for

$N_c = \dfrac{\text{market order rate } \mu/}{2 \cdot \text{decay limit order rate } \delta}$ $\quad t_c = \dfrac{1}{\text{decay limit order rate}}$

$p_c = \dfrac{\text{market order rate } \mu/}{2 \cdot \text{limit order rate } \alpha}$

asymptotic analysis is a method of describing limiting behavior || The simplest example is, when considering a function f(n), there is a need to describe its properties when n becomes very large. Thus, if f(n) = n2+3n, the term 3n becomes insignificant compared to n2 when n is very large. The function "f(n) is said to be asymptotically equivalent to n2 as n → ∞", and this is written symbolically as f(n) ~ n2 || ->In applied mathematics, asymptotic analysis is used to build numerical methods to approximate equation solutions. | ->in computer science in the analysis of algorithms, considering the performance of algorithms when applied to very large input datasets. | ->the behavior of physical systems when they are very large, an example being Statistical mechanics. | ->in accident analysis when identifying the causation of crash through count modeling with large number of crash counts in a given time and space.

many purposes the limit $\mathrm{d}p \to 0$ approximates the case of finite tick size fairly well. As we will see, working in this limit is essential for getting tractable analytic results.

A non-dimensional scale parameter based on order size is constructed by dividing the typical order size (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 'chunkiness' of the orders stored in the limit order book. As we will see, ϵ is an important determinant of liquidity, and it is a particularly important determinant of volatility. In the continuum limit $\epsilon \to 0$ there is no price diffusion. This is because price diffusion can occur only if there is a finite probability for price levels outside the spread to be empty, thus allowing the best bid or ask to make a persistent shift. If we let $\epsilon \to 0$ while the average depth is held fixed the number of individual orders becomes infinite, and the probability that spontaneous decays or market orders can create gaps outside the spread becomes zero. This is verified in simulations. Thus the limit $\epsilon \to 0$ is always a poor approximation to a real market. ϵ is a more important parameter than the tick size $\mathrm{d}p/p_c$. In the mean field analysis in section 3, we let $\mathrm{d}p/p_c \to 0$, reducing the number of independent parameters from two to one, and in many cases find that this is a good approximation.

depth profili => stored density n(p, t) of limit order volume at price p,

The order size σ can be thought of as the order *granularity*. Just as the properties of a beach with fine sand are quite different from that of one populated by fist sized boulders, a market with many small orders behaves quite differently from one with a few large orders. N_c provides the scale against which the order size is measured, and ϵ characterizes the granularity in relative terms. Alternatively, $1/\epsilon$ can be thought of as the annihilation rate from market orders expressed in units of the size of spontaneous decays. Note that in non-dimensional units the number of shares can also be written as $\hat{N} = N/N_c = N\epsilon/\sigma$.

The construction of the non-dimensional granularity parameter illustrates the importance of including a spontaneous decay process in this model. If $\delta = 0$ (which implies $\epsilon = 0$) there is no spontaneous decay of orders, and depending on the relative values of μ and α, generically either the depth of orders will accumulate without bound or the spread will become infinite. As long as $\delta > 0$, in contrast, this is not a problem.

For some purposes the effects of varying tick size and order size are fairly small, and we can derive approximate formulae using dimensional analysis based only on the order flow rates. For example, in table 3 we give dimensional scaling formulae for the average spread, the market order liquidity (as measured by the average slope of the depth profile near the mid-point), the volatility, and the asymptotic depth (defined below). Because these estimates neglect the effects of discreteness, they are only approximations of the true behaviour of the model, which do a better job of explaining some properties than others. Our numerical and analytical results show that some quantities also depend on the granularity parameter ϵ and to a weaker extent on the tick size $\mathrm{d}p/p_c$. Nonetheless, the dimensional estimates based on order flow alone provide a good starting point for understanding market behaviour. A comparison to more precise formulae derived from theory and simulations is given in table 4.

Table 3. Estimates from dimensional analysis for the scaling of a few market properties based on order flow rates alone. α is the limit order density rate, μ is the market order rate, and δ is the spontaneous limit order removal rate. These estimates are constructed by taking the combinations of these three rates that have the proper units. They neglect the dependence on the order granularity ϵ and the non-dimensional tick size $\mathrm{d}p/p_c$. More accurate relations from simulation and theory are given in table 4.

Quantity	Dimensions	Scaling relation
Asymptotic depth	Shares/price	$d \sim \alpha/\delta$
Spread	Price	$s \sim \mu/\alpha$
Slope of depth profile	Shares/price^2	$\lambda \sim \alpha^2/\mu\delta = d/s$
Price diffusion rate	Price^2/time	$D_0 \sim \mu^2\delta/\alpha^2$

Table 4. The dependence of market properties on model parameters based on simulation and theory, with the relevant figure numbers. These formulae include corrections for order granularity ϵ and finite tick size $\mathrm{d}p/p_c$. The formula for asymptotic depth from dimensional analysis in table 3 is exact with zero tick size. The expression for the mean spread is modified by a function of ϵ and $\mathrm{d}p/p_c$, though the dependence on them is fairly weak. For the liquidity λ, corresponding to the slope of the depth profile near the origin, the dimensional estimate must be modified because the depth profile is no longer linear (mainly depending on ϵ) and so the slope depends on price. The formulae for the volatility are empirical estimates from simulations. The dimensional estimate for the volatility from table 3 is modified by a factor of $\epsilon^{-0.5}$ for the early time price diffusion rate and a factor of $\epsilon^{0.5}$ for the late time price diffusion rate.

Quantity	Scaling relation	Figure
Asymptotic depth	$d = \alpha/\delta$	3
Spread	$s = (\mu/\alpha) f(\epsilon, \mathrm{d}p/p_c)$	10, 24
Slope of depth profile	$\lambda = (\alpha^2/\mu\delta) g(\epsilon, \mathrm{d}p/p_c)$	3, 20, 21
Price diffusion ($\tau \to 0$)	$D_0 = (\mu^2\delta/\alpha^2)\epsilon^{-0.5}$	11, 14(*c*)
Price diffusion ($\tau \to \infty$)	$D_\infty = (\mu^2\delta/\alpha^2)\epsilon^{0.5}$	11, 14(*c*)

An approximate formula for the mean spread can be derived by noting that it has dimensions of *price*, and the unique combination of order flow rates with these dimensions is μ/α. While the dimensions indicate the scaling of the spread, they cannot determine multiplicative factors of order unity. A more intuitive argument can be made by noting that inside the spread removal due to cancellation is dominated by removal due to market orders. Thus the total limit order placement rate inside the spread, for either buy or sell limit orders αs, must equal the order removal rate $\mu/2$, which implies that spread is $s = \mu/2\alpha$. As we will see later, this argument can be generalized and made more precise within our mean field analysis which then also predicts the observed dependence on the granularity parameter ϵ. However, this dependence is rather weak and only causes a variation of roughly a factor of two for $\epsilon < 1$ (see figures 10 and 24), and the factor of $1/2$ derived above is a good first approximation. Note that this prediction of the mean spread is just the characteristic price p_c.

It is also easy to derive the mean *asymptotic depth*, which is the density of shares far away from the mid-point. The asymptotic depth is an artificial construct of our assumption of order placement over an infinite interval; it should be regarded as providing a simple boundary condition so that we can study the behaviour near the mid-point price. The mean asymptotic depth has dimensions of *shares/price*, and is therefore given by α/δ. Furthermore, because removal by market orders is insignificant in this regime, it is determined by the balance between order placement and decay, and far from the mid-point the depth at any given price is Poisson distributed. This result is exact.

because → we are talking about boundaries not about middle far away from the spread

The average slope of the depth profile near the mid-point is an important determinant of liquidity, since it affects the expected price response when a market order arrives. The slope has dimensions of *shares/price*^{2}, which implies that in terms of the order flow rates it scales roughly as $\alpha^2/\mu\delta$. This is also the ratio of the asymptotic depth to the spread. As we will see later, this is a good approximation when $\epsilon \sim 0.01$, but for smaller values of ϵ the depth profile is not linear near the mid-point, and this approximation fails.

The last two entries in table 4 are empirical estimates for the price diffusion rate D, which is proportional to the square of the volatility. That is, for normal diffusion, starting from a point at $t = 0$, the variance v after time t is $v = Dt$. The volatility at any given timescale t is the square root of the variance at timescale t. The estimate for the diffusion rate based on dimensional analysis in terms of the order flow rates alone is $\mu^2\delta/\alpha^2$. However, simulations show that short time diffusion is much faster than long time diffusion, due to negative autocorrelations in the price process, as shown in figure 11. The initial and the asymptotic diffusion rates appear to obey the scaling relationships given in table 4. Though our mean field theory is not able to predict this functional form, the fact that early and late time diffusion rates are different can be understood within the framework of our analysis, as described in section 3.5. Anomalous diffusion of this type implies negative autocorrelations in mid-point prices. Note that we use the term 'anomalous diffusion' to imply that the diffusion rate is different on short and long timescales. We do not use this term in the sense that it is normally used in the physics literature, i.e. that the long time diffusion is proportional to t^γ with $\gamma \neq 1$ (for long times $\gamma = 1$ in our case).

2.2. Varying the granularity parameter ϵ

We first investigate the effect of varying the order granularity ϵ in the limit $\mathrm{d}p \to 0$. As we will see, the granularity has an important effect on most of the properties of the model, and particularly on depth, price impact and price diffusion. The behaviour can be divided into three regimes, roughly as follows:

- *Large* ϵ, *i.e.* $\epsilon \gtrsim 0.1$. This corresponds to a large accumulation of orders at the best bid and ask, nearly linear market impact, and roughly equal short and long time price diffusion rates. This is the regime where the mean field approximation used in the theoretical analysis works best.
- *Medium* ϵ, *i.e.* $\epsilon \sim 0.01$. In this range the accumulation of orders at the best bid and ask is small and near the mid-point price the depth profile increases nearly linearly with price. As a result, as a crude approximation the price impact increases as roughly the square root of the order size.
- *Small* ϵ, *i.e.* $\epsilon \lesssim 0.001$. The accumulation of orders at the best bid and ask is very small, and near the mid-point the depth profile is a convex function of price. The price impact is very concave. The short time price diffusion rate is much greater than the long time price diffusion rate.

Since the results for bids are symmetric with those for offers about $p = 0$, for convenience we only show the results for offers, i.e. buy market orders and sell limit orders. In this subsection prices are measured relative to the mid-point, and simulations are in the continuum limit where the tick size $\mathrm{d}p \to 0$. The results in this section are from numerical simulations. Also, bear in mind that far from the mid-point the predictions of this model are not valid due to the unrealistic assumption of an order placement process with an infinite domain. Thus the results are potentially relevant to real markets only when the price p is at most a few times as large as the characteristic price p_c.

2.2.1. Depth profile. The *mean depth profile*, i.e. the average number of shares per price interval, and the mean cumulative depth profile are shown in figure 3, and the standard deviation of the cumulative profile is shown in figure 4. Since the depth profile has units of *shares/price*, non-dimensional units of depth profile are $\hat{n} = np_c/N_c = n\delta/\alpha$. The cumulative depth profile at any given time t is defined as

$$N(p,t) = \sum_{\tilde{p}=0}^{p} n(\tilde{p},t)\,\mathrm{d}p. \tag{2}$$

This has units of shares and so in non-dimensional terms is $\hat{N}(p) = N(p)/N_c = 2\delta N(p)/\mu = N(p)\epsilon/\sigma$.

$1/\epsilon$

In the high ϵ regime the annihilation rate due to market orders is low (relative to $\delta\sigma$), and there is a significant accumulation of orders at the best ask, so the average depth is much greater than zero at the mid-point. The mean depth profile is a concave function of price. In the medium ϵ regime the market order removal rate increases, depleting the average depth near the best ask, and the profile is nearly linear over the range $p/p_c \leqslant 1$. In the small ϵ regime the market order removal rate increases even further, making the average depth near the ask very close to zero, and the profile is a convex function over the range $p/p_c \leqslant 1$.

The standard deviation of the depth profile is shown in figure 4. We see that the standard deviation of the cumulative depth is comparable to the mean depth, and that as ϵ increases, near the mid-point there is a similar transition from convex to concave behaviour.

The uniform order placement process seems at first glance one of the most unrealistic assumptions of our model, leading to depth profiles with a finite asymptotic depth (which also implies that there are an infinite number of orders in the book). However, orders far away from the spread in the asymptotic region almost never get executed and thus do not affect the market dynamics. To demonstrate this in figure 5 we show

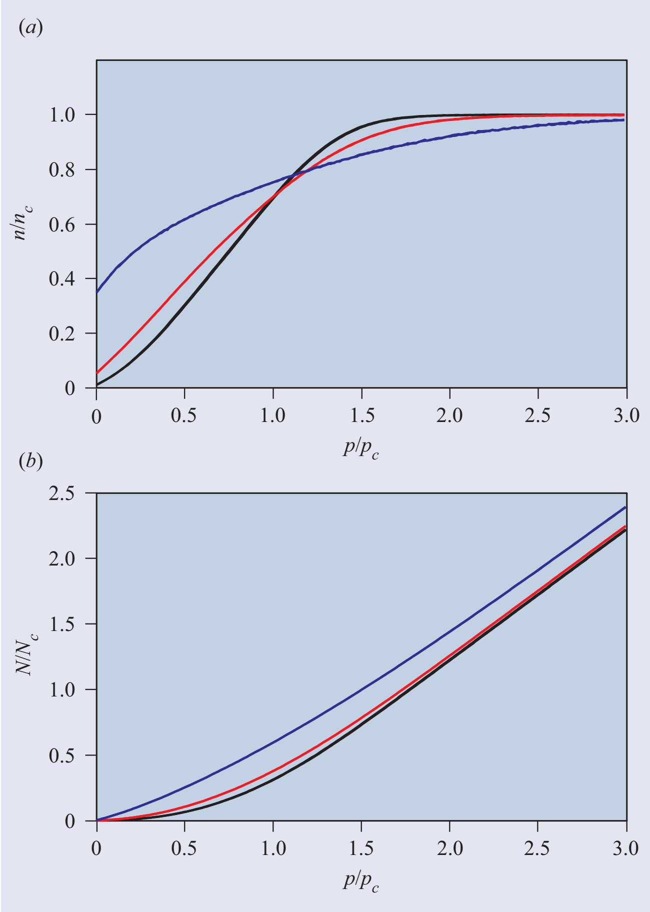

Figure 3. The mean depth profile and cumulative depth versus $\hat{p} = p/p_c = 2\alpha p/\mu$. The origin $p/p_c = 0$ corresponds to the mid-point. *(a)* The average depth profile n in non-dimensional coordinates $\hat{n} = np_c/N_c = n\delta/\alpha$. *(b)* The non-dimensional cumulative depth $N(p)/N_c$. We show three different values of the non-dimensional granularity parameter: $\epsilon = 0.2$ (blue), $\epsilon = 0.02$ (red), $\epsilon = 0.002$ (black), all with tick size $\mathrm{d}p = 0$.

the comparison between the limit order depth profile and the depth n_e of only those orders which eventually get executed[6]. The density n_e of executed orders decreases rapidly as a function of the distance from the mid-price. Therefore we expect that near the mid-point our results should be similar to alternative order placement processes, as long as they also lead to an exponentially decaying profile of executed orders (which is what we observe above). However, to understand the behaviour further away from the mid-point we are also working on enhancements that include more realistic order placement processes grounded on empirical measurements of market data, as summarized in section 4.3.

2.2.2. Liquidity for market orders: the price impact function. In this subsection we study the *instantaneous price impact* function $\phi(t, \omega, \tau \to 0)$. This is defined as the (logarithm of the) mid-point price shift immediately after the arrival of a market order in the absence of any other events.

[6] Note that the ratio n_e/n is not the same as the probability of filling orders (figure 12) because in that case the price p/p_c refers to the distance of the order from the mid-point at the time when it was placed.

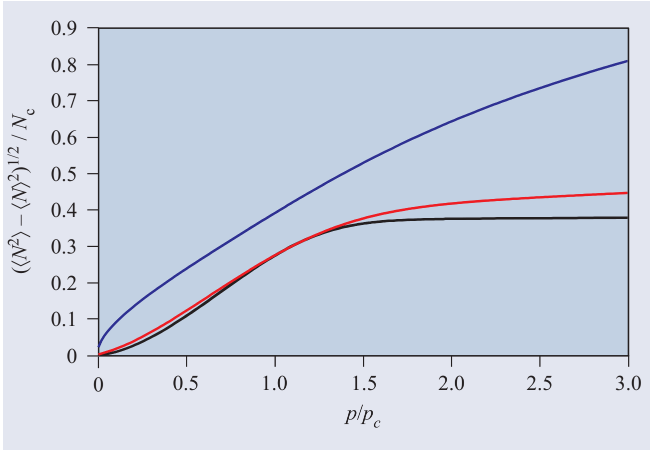

Figure 4. Standard deviation of the non-dimensionalized cumulative depth versus non-dimensional price, corresponding to figure 3.

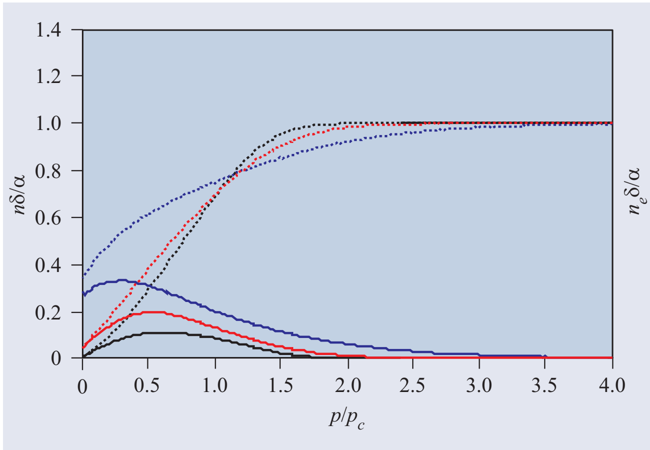

Figure 5. A comparison between the depth profiles and the effective depth profiles as defined in the text, for different values of ϵ. Solid curves refer to the effective depth profiles n_e and dotted curves correspond to the depth profiles.

This should be distinguished from the asymptotic price impact $\phi(t, \omega, \tau \to \infty)$, which describes the permanent price shift. While the permanent price shift is clearly very important, we do not study it here. The reader should bear in mind that all prices p, $a(t)$, etc are logarithmic.

The price impact function provides a measure of the liquidity for executing market orders. (The liquidity for limit orders, in contrast, is given by the probability of execution, studied in section 2.2.5.) At any given time t, the instantaneous $(\tau = 0)$ price impact function is the inverse of the cumulative depth profile. This follows immediately from equations (1) and (2), which in the limit $\mathrm{d}p \to 0$ can be replaced by the continuum transaction equation:

$$\omega = N(p, t) = \int_0^p n(\tilde{p}, t)\,\mathrm{d}\tilde{p}. \tag{3}$$

This equation makes it clear that at any fixed t the price impact can be regarded as the inverse of the cumulative depth profile $N(p, t)$. When the fluctuations are sufficiently small

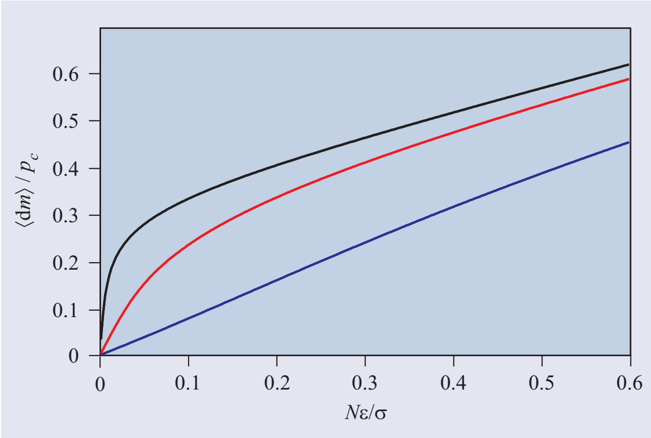

Figure 6. The average price impact corresponding to the results in figure 3. The average instantaneous movement of the non-dimensional mid-price, $\langle \mathrm{d}m \rangle / p_c$, caused by an order of size $N/N_c = N\epsilon/\sigma$. $\epsilon = 0.2$ (blue), $\epsilon = 0.02$ (red), $\epsilon = 0.002$ (black).

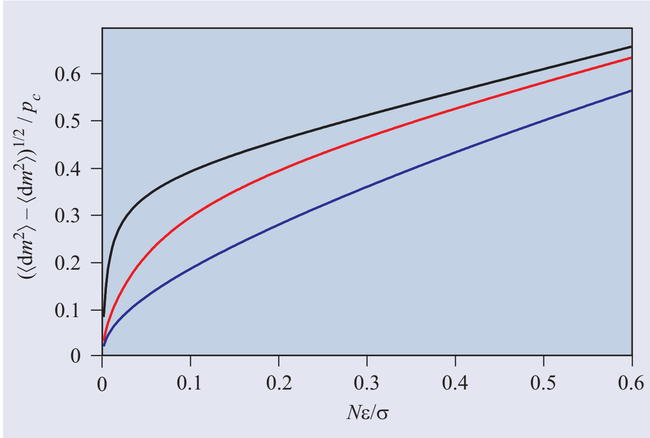

Figure 7. The standard deviation of the instantaneous price impact $\mathrm{d}m/p_c$ corresponding to the means in figure 6, as a function of normalized order size $\epsilon N/\sigma$. $\epsilon = 0.2$ (blue), $\epsilon = 0.02$ (red), $\epsilon = 0.002$ (black).

we can replace $n(p,t)$ by its mean value $n(p) = \langle n(p,t) \rangle$. In general, however, the fluctuations can be large, and the average of the inverse is not equal to the inverse of the average. There are corrections based on higher order moments of the depth profile, as given in the moment expansion derived in appendix A.1. Nonetheless, the inverse of the mean cumulative depth provides a qualitative approximation that gives insight into the behaviour of the price impact function. (Note that everything becomes much simpler using medians, since the median of the cumulative price impact function is exactly the inverse of the median price impact, as derived in appendix A.1.)

Mean price impact functions are shown in figure 6 and the standard deviation of the price impact is shown in figure 7. The price impact exhibits very large fluctuations for all values of ϵ: the standard deviation has the same order of magnitude as the mean or is even greater for small $N\epsilon/\sigma$ values. Note that these are actually *virtual price impact* functions. That is, to explore the behaviour of the instantaneous price impact for a wide range of order sizes, we periodically compute the price impact that an order of a given size would have caused at that instant, if it had been submitted. We have checked that real price impact curves are the same, but they require a much longer time to accumulate reasonable statistics.

One of the interesting results in figure 6 is the scale of the price impact. The price impact is measured relative to the characteristic price scale p_c, which as we have mentioned earlier is roughly equal to the mean spread. As we will argue in relation to figure 8, the range of non-dimensional shares shown on the horizontal axis spans the range of reasonable order sizes. This figure demonstrates that throughout this range the price is the order of magnitude (and typically less than) the mean spread size.

Due to the accumulation of orders at the ask in the large ϵ regime, for small p the mean price impact is roughly linear. This follows from equation (3) under the assumption that $n(p)$ is constant. In the medium ϵ regime, under the assumption that the variance in depth can be neglected, the mean price impact should increase as roughly $\omega^{1/2}$. This follows from equation (3) under the assumption that $n(p)$ is linearly increasing and $n(0) \approx 0$. (Note that we see this as a crude approximation, but there can be substantial corrections caused by the variance of the depth profile.) Finally, in the small ϵ regime the price impact is highly concave, increasing much more slowly than $\omega^{1/2}$. This follows because $n(0) \approx 0$ and the depth profile $n(p)$ is convex.

To get a better feel for the functional form of the price impact function, in figure 8 we numerically differentiate it versus log order size, and plot the result as a function of the appropriately scaled order size. (Note that because our prices are logarithmic, the vertical axis already incorporates the logarithm.) If we were to fit a local power law approximation to the function at each price, this would correspond to the exponent of that power law near that price. Notice that the exponent is almost always less than one, so the price impact is almost always concave. Making the assumption that the effect of the variance of the depth is not too large, so that equation (3) is a good assumption, the behaviour of this figure can be understood as follows: for $N/N_c \approx 0$ the price impact is dominated by $n(0)$ (the constant term in the average depth profile) and so the logarithmic slope of the price impact is always close to one. As N/N_c increases, the logarithmic slope is driven by the shape of the average depth profile, which is linear or convex for smaller ϵ, resulting in concave price impact. For large values of N/N_c, we reach the asymptotic region where the depth profile is flat (and where our model is invalid by design). Of course, there can be deviations from this behaviour caused by the fact that the mean of the inverse depth profile is not in general the inverse of the mean, i.e. $\langle N^{-1}(p) \rangle \neq \langle N(p) \rangle^{-1}$ (see appendix A.1).

To compare to real data, note that $N/N_c = N\epsilon/\sigma$. N/σ is just the order size in shares in relation to the average order size, so by definition it has a typical value of one. For the London Stock Exchange, we have found that typical values of ϵ are in the range 0.001–0.1. For a typical range of order sizes from 100–100 000 shares, with an average size of 10 000 shares, the meaningful range for N/N_c is therefore roughly 10^{-5} to

The general form of a power law relating x and y is log y = b+a log x. If we raise the base of the logarithm (which doesn't actually matter), say e, to the values on both sides of this equation, we get y = ebea log x = ebxa. Since eb is just "some constant," let us replace it by constant c. Thus, a power law can be written as y = cxa for some constants a and c.

In statistics, a power law is a functional relationship between two quantities, where one quantity varies as a power of another. For instance, the number of cities having a certain population size is found to vary as a power of the size of the population. Empirical power-law distributions hold only approximately or over a limited range.

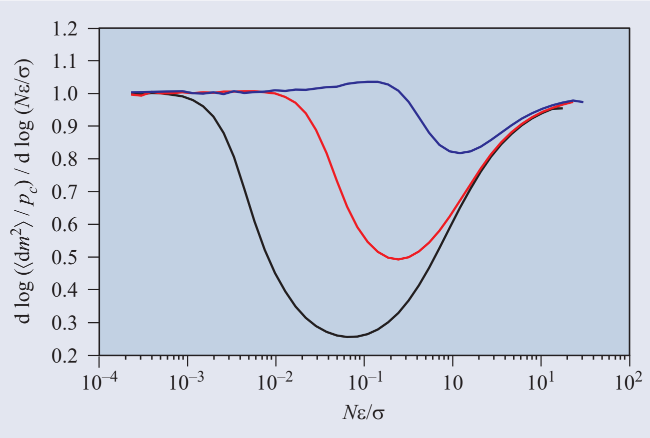

Figure 8. Derivative of the non-dimensional mean mid-price movement, with respect to the logarithm of the non-dimensional order size $N/N_c = N\epsilon/\sigma$, obtained from the price impact curves in figure 6.

1. In this range, for small values of ϵ the exponent can reach values as low as 0.2. This offers a possible explanation for the previously mysterious concave nature of the price impact function, and contradicts the linear increase in price impact based on the naive argument presented in the introduction.

2.2.3. Spread. The probability density of the spread is shown in figure 9. This shows that the probability density is substantial at $s/p_c = 0$. (Remember that this is in the limit $dp \to 0$.) The probability density reaches a maximum at a value of the spread approximately $0.2p_c$, and then decays. It might seem surprising at first that it decays more slowly for large ϵ, where there is a large accumulation of orders at the ask. However, it should be borne in mind that the characteristic price $p_c = \mu/\alpha$ depends on ϵ. Since $\epsilon = 2\delta\sigma/\mu$, by eliminating μ this can be written as $p_c = 2\sigma\delta/(\alpha\epsilon)$. Thus, holding the other parameters fixed, large ϵ corresponds to small p_c, and vice versa. So in fact, the spread is very small for large ϵ, and large for small ϵ, as expected. The figure just shows the small corrections to the large effects predicted by the dimensional scaling relations.

For large ϵ the probability density of the spread decays roughly exponentially moving away from the mid-point. This is because for large ϵ the fluctuations around the mean depth are roughly independent. Thus the probability for a market order to penetrate to a given price level is roughly the probability that all the ticks smaller than this price level contain no orders, which gives rise to an exponential decay. This is no longer true for small ϵ. Note that for small ϵ the probability distribution of the spread becomes insensitive to ϵ, i.e. the non-dimensionalized distribution for $\epsilon = 0.02$ is nearly the same as that for $\epsilon = 0.002$.

It is apparent from figure 9 that in non-dimensional units the mean spread increases with ϵ. This is confirmed in figure 10, which displays the mean value of the spread as a function of ϵ. The mean spread increases monotonically with ϵ. It depends on ϵ as roughly a constant (equal to approximately 0.45 in non-dimensional coordinates) plus a

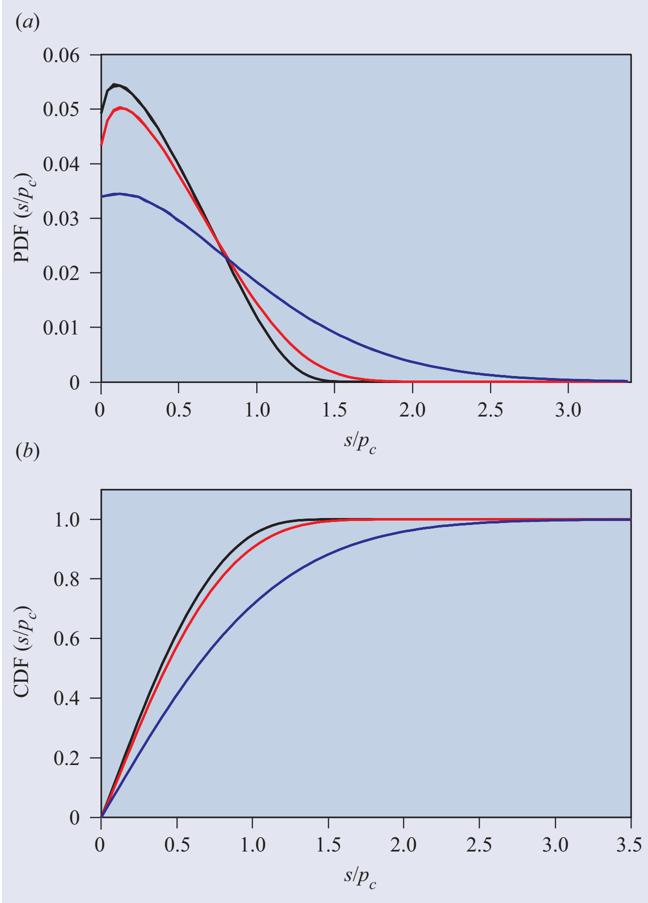

Figure 9. The probability density function (*a*), and cumulative distribution function (*b*) of the non-dimensionalized bid–ask spread s/p_c, corresponding to the results in figure 3. $\epsilon = 0.2$ (blue), $\epsilon = 0.02$ (red), $\epsilon = 0.002$ (black).

linear term whose slope is rather small. We believe that for most financial instruments $\epsilon < 0.3$. Thus the variation in the spread caused by varying ϵ in the range $0 < \epsilon < 0.3$ is not large, and the dimensional analysis based only on rate parameters given in table 4 is a good approximation. We get an accurate prediction of the ϵ dependence across the full range of ϵ from the independent interval approximation (IIA) technique derived in section 3.7, as shown in figure 24.

2.2.4. Volatility and price diffusion. The price diffusion rate, which is proportional to the square of the volatility, is important for determining risk and is a property of central interest. From dimensional analysis in terms of the order flow rates the price diffusion rate has units of *price*^{2}/*time*, and so must scale as $\mu^2\delta/\alpha^2$. We can also make a crude argument for this as follows: the dimensional estimate of the spread (see table 4) is $\mu/2\alpha$. Let this be the characteristic step size of a random walk, and let the step frequency be the characteristic time $1/\delta$ (which is the average lifetime for a share to be cancelled). This argument also gives the above estimate for the diffusion rate. However, this is not correct in the presence of negative autocorrelations in the step sizes. The numerical results make it clear that there are important ϵ

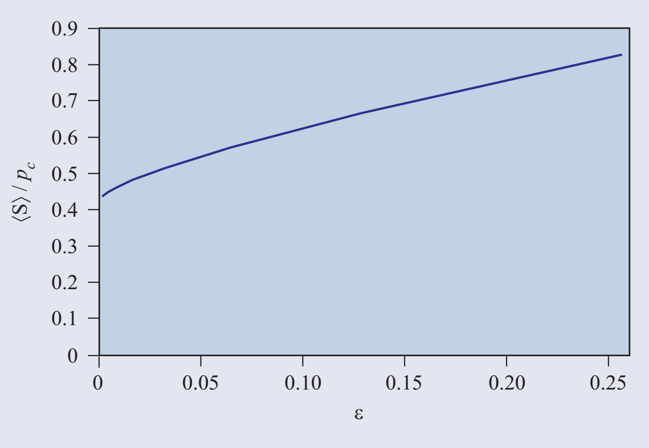

Figure 10. The mean value of the spread in non-dimensional units $\hat{s} = s/p_c$ as a function of ϵ. This demonstrates that the spread only depends weakly on ϵ, indicating that the prediction from dimensional analysis given in table 3 is a reasonable approximation.

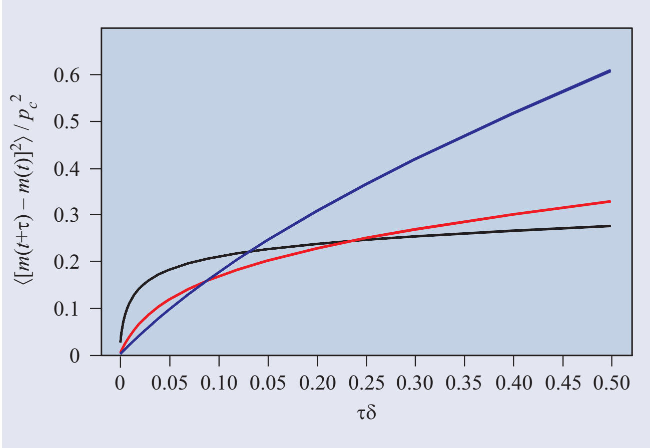

Figure 11. The variance of the change in the non-dimensionalized mid-point price versus the non-dimensional time delay interval $\tau\delta$. For a pure random walk this would be a straight line whose slope is the diffusion rate, which is proportional to the square of the volatility. The fact that the slope is steeper for short times comes from the non-trivial temporal persistence of the order book. The three cases correspond to figure 3: $\epsilon = 0.2$ (blue), $\epsilon = 0.02$ (red), $\epsilon = 0.002$ (black).

dependent corrections to this result, as demonstrated below.

In figure 11 we plot simulation results for the variance of the change in the mid-point price at timescale τ, $\text{var}(m(t+\tau) - m(t))$. The slope is the diffusion rate, which at any fixed timescale is proportional to the square of the volatility. It appears that there are at least two timescales involved, with a faster diffusion rate for short timescales and a slower diffusion rate for long timescales. Such anomalous 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 dimensional analysis based on order flow parameters alone, $\mu^2\delta/\alpha^2$, and a τ dependent power of the non-dimensional granularity parameter $\epsilon = 2\delta\sigma/\mu$, as summarized in table 4. We cannot currently explain why this power is $-1/2$ for short term diffusion and $1/2$ for long term diffusion. However, a qualitative understanding can be gained based on the conservation law that we derive in section 3.3. A discussion of how this relates to price diffusion is given in section 3.5.

Note that the temporal structure in the diffusion process also implies non-zero autocorrelations of the mid-point price $m(t)$. This corresponds to weak negative autocorrelations in price differences $m(t) - m(t-1)$ that persist for timescales until the variance versus τ becomes a straight line. The timescale depends on parameters, but is typically of the order of 50 market order arrival times. This temporal structure implies that there exists an arbitrage opportunity which, when exploited, would make prices more random and the structure of the order flow non-random.

2.2.5. Liquidity for limit orders: probability and time to fill. The liquidity for limit orders depends on the probability that they will be filled, and the time to be filled. This obviously depends on price: limit orders close to the current transaction prices are more likely to be filled quickly, while those far away have a lower likelihood of being filled. Figure 12 plots the probability Γ of a limit order being filled versus the non-dimensionalized price at which it was placed (as with all the figures in this section, this is shown in the mid-point price centred frame). Figure 12 shows that in non-dimensional coordinates the probability of filling close to the bid for sell limit orders (or the ask for buy limit orders) decreases as ϵ increases. For large ϵ, this is less than 1 even for negative prices. This says that even for sell orders that are placed close to the best bid there is a significant chance that the offer is deleted before being executed. This is not true for smaller values of ϵ, where $\Gamma(0) \approx 1$. Far away from the spread the fill probabilities as a function of ϵ are reversed, i.e. the probability for filling limit orders increases as ϵ increases. The crossover point where the fill probabilities are roughly the same occurs at $p \approx p_c$. This is consistent with the depth profile in figure 3 which also shows that depth profiles for different values of ϵ cross at about $p \sim p_c$.

Similarly figure 13 shows the average time τ taken to fill an order placed at a distance p from the instantaneous mid-price. Again we see that though the average time is larger at larger values of ϵ for small p/p_c, this behaviour reverses at $p \sim p_c$.

2.3. Varying tick size dp/p_c

The dependences on discrete tick size dp/p_c of the cumulative distribution function for the spread, instantaneous price impact and mid-price diffusion are shown in figure 14. We chose an unrealistically large value of the tick size, with $dp/p_c = 1$, to show that, even with very coarse ticks, the qualitative changes in behaviour are typically relatively minor.

Figure 14(*a*) shows the cumulative density function of the spread, comparing $dp/p_c = 0$ and 1. It is apparent from this figure that the spread distribution for coarse ticks 'effectively integrates' the distribution in the limit $dp \to 0$. That is, at integer tick values the mean cumulative depth profiles roughly match, and in between integer tick values, for coarse ticks the probability is smaller. This happens for the obvious reason

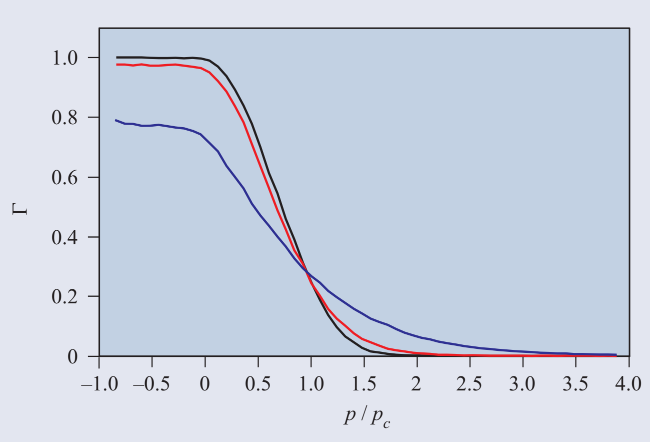

Figure 12. The probability Γ for filling a limit order placed at a price p/p_c where p is calculated from the instantaneous mid-price at the time of placement. The three cases correspond to figure 3: $\epsilon = 0.2$ (blue), $\epsilon = 0.02$ (red), $\epsilon = 0.002$ (black).

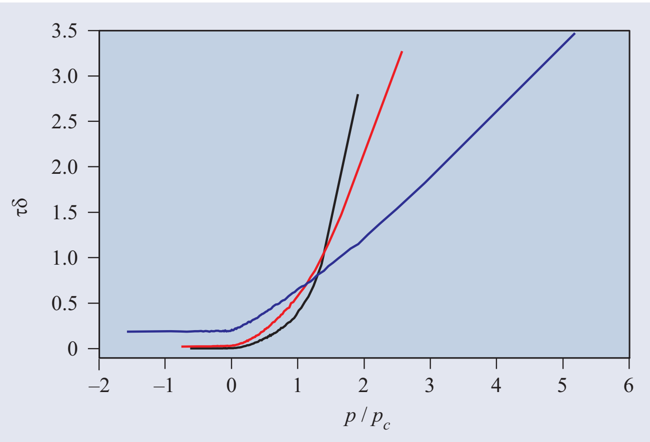

Figure 13. The average time τ non-dimensionalized by the rate δ, to fill a limit order placed at a distance p/p_c from the instantaneous mid-price.

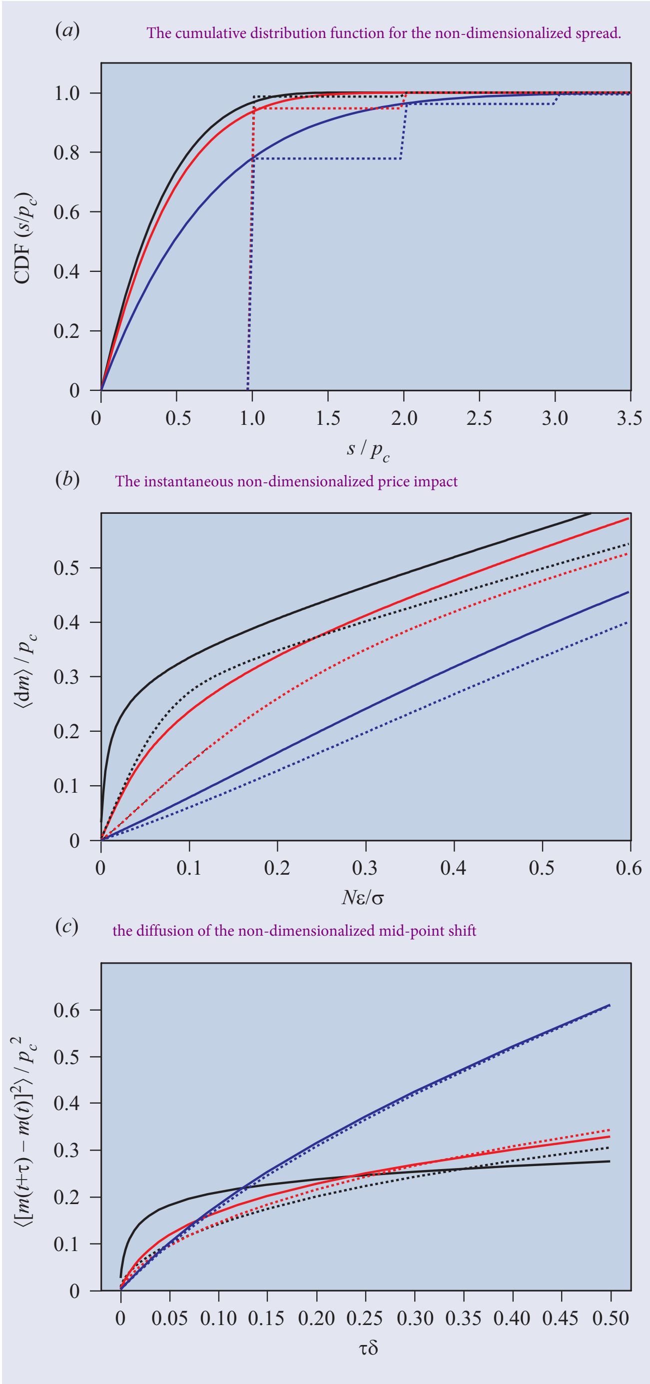

Figure 14. Dependence of market properties on tick size. Solid curves are $dp/p_c \to 0$; dotted curves are $dp/p_c = 1$. Cases correspond to figure 3, with $\epsilon = 0.2$ (black), $\epsilon = 0.02$ (red), $\epsilon = 0.002$ (blue). (*a*) The cumulative distribution function for the non-dimensionalized spread. (*b*) The instantaneous non-dimensionalized price impact, while (*c*) is the diffusion of the non-dimensionalized mid-point shift, corresponding to figure 11.

that coarse ticks quantize the possible values of the spread, and place a lower limit of one tick on the value the spread can take. The shift in the mean spread from this effect is not shown, but it is consistent with this result; there is a constant offset of roughly 1/2 tick.

The alteration in the price impact is shown in figure 14(*b*). Unlike the spread distribution, the average price impact varies continuously. Even though the tick size is quantized, we are averaging over many events and the probability of the price impact of each tick size is a continuous function of the order size. Large tick size consistently lowers the price impact. The price impact rises more slowly for small p, but is then similar except for a downward translation.

The effect of coarse ticks is less trivial for mid-price diffusion, as shown in figure 14(*c*). At $\epsilon = 0.002$, coarse ticks remove most of the rapid short term volatility of the mid-point, which in the continuous price case arises from price fluctuations smaller than $dp/p_c = 1$. This lessens the negative autocorrelation of mid-point price returns, and reduces the anomalous diffusion. At $\epsilon = 0.2$, where both early volatility and late negative autocorrelation are smaller, coarse ticks have less effect. The net result is that the mid-price diffusion becomes less sensitive to the value of ϵ as tick size increases, and there is less anomalous price diffusion.

3. Theoretical analysis

3.1. Summary of analytic methods

We have investigated this model analytically using two approaches. The first one is based on a master equation, given in section 3.6. This approach works best in the mid-point centred frame. Here we attempt to solve directly for the average number of shares at each price tick as a function of price. The mid-point price makes a random walk with a non-stationary distribution. Thus the key to finding a stationary analytic solution for the average depth is to use comoving price coordinates, which are centred on a reference point near the centre of the book, such as the mid-point or the best bid. In the first approximation, fluctuations about the mean depth at adjacent prices are treated as independent. This allows us to replace the distribution over depth profiles with a simpler probability density over occupation numbers n at each p and t. We can take a continuum limit by letting the tick size $\mathrm{d}p$ become infinitesimal. With finite order flow rates, this gives vanishing probability for the existence of more than one order at any tick as $\mathrm{d}p \to 0$. This is described in detail in section 3.6.3. With this approach we are able to test the relevance of correlations as a function of the parameter ϵ as well as predict the functional dependence of the cumulative distribution of the spread on the depth profile. It is seen that correlations are negligible for large values of ϵ ($\epsilon \sim 0.2$) while they are very important for small values ($\epsilon \sim 0.002$).

1.

Our second analytic approach which we term the IIA is most easily carried out in the bid centred frame and is described in section 3.7. This approach uses a different representation, in which the solution is expressed in terms of the empty intervals between non-empty price ticks. The system is characterized at any instant of time by a set of intervals $\{\ldots x_{-1}, x_0, x_1, x_2 \ldots\}$ where for example x_0 is the distance between the bid and the ask (the spread), x_{-1} is the distance between the second buy limit order and the bid, and so on (see figure 15). Equations are written for how a given interval varies in time. Changes to adjacent intervals are related, giving us an infinite set of coupled non-linear equations. However, using a mean field approximation we are able to solve the equations, albeit only numerically. Besides predicting how the various intervals (for example the spread) vary with the parameters, this approach also predicts the depth profiles as a function of the parameters. The predictions from the IIA are compared to data from numerical simulations, in section 3.7.2. They match very well for large ϵ and less well for smaller values of ϵ. The IIA can also be modified to incorporate various extensions to the model, as mentioned in section 3.7.2.

2
JIA
independent
interval
approximation

In both approaches, we use a mean field approximation to get a solution. The approximation basically lies in assuming that fluctuations in adjacent intervals (which might be adjacent price ranges in the master equation approach or adjacent empty intervals in the IIA) are independent. Also, both approaches are most easily tractable only in the continuum limit $\mathrm{d}p \to 0$, when every tick has at most only one order. They may however be extended to general tick size as well. This is explained in the appendix for the master equation approach.

Because correlations are important for small ϵ, both methods work well mostly in the large ϵ limit, though qualitative aspects of small ϵ behaviour may also be gleaned from them. Unfortunately, at least based on our preliminary investigation of London Stock Exchange data, it seems that it is this small ϵ limit that real markets may tend more towards. So our approximate solutions may not be as useful as we would like. Nonetheless, they do provide some conceptual insights into what determines depth and price impact.

In particular, we find that the shape of the mean depth profile depends on a single parameter ϵ, and that the relative sizes of its first few derivatives account for both the order size dependence of the market impact, and the renormalization of the mid-point diffusivity. A higher relative rate of market versus limit orders depletes the centre of the book, though less than the classical estimate predicts. This leads to more concave impact (explaining figure 8) and faster short term diffusivity. However, the orders pile up more quickly (versus classically non-dimensionalized price) with distance from the mid-point, causing the rapid early diffusion to suffer larger mean reversion. These are the effects shown in figure 11. We will elaborate on the above remarks in the following sections; however, the qualitative relation of impact to mid-point autocorrelation supplies a potential interpretation of data, which may be more robust than details of the model assumptions or its quantitative results.

Both of the treatments described above are approximations. We can derive an exact global conservation law of order placement and removal whose consequences we elaborate in section 3.3. This conservation law must be respected in any sensible analysis of the model, giving us a check on the approximations. It also provides some insight into the anomalous diffusion properties of this model.

3.2. Characterizing limit order books: dual coordinates

We begin with the assumption of a *price space*. Price is a dimensional quantity, and the space is divided into bins of length $\mathrm{d}p$ representing the ticks, which may be finite or infinitesimal. Prices are then discrete or continuous valued, respectively.

Statistical properties of interest are computed from temporal sequences or ensembles of *limit order book configurations*. If n is the variable used to denote the number of shares from limit orders in some bin $(p, p + \mathrm{d}p)$ at the beginning t of an elementary time interval, a configuration is specified by a function $n(p, t)$. It is convenient to take n positive for sell limit orders, and negative for buy limit orders. Because the model dynamics precludes crossing limit orders, there is in general a highest instantaneous buy limit order price, called the bid $b(t)$, and a lowest sell limit order price, the ask $a(t)$, with $b(t) < a(t)$ always. The *mid-point price*, defined as $m(t) \equiv [a(t) + b(t)]/2$, may or may not be the price of any actual bin, if prices are discrete ($m(t)$ may be a half-integer multiple of $\mathrm{d}p$). These quantities are illustrated in figure 15.

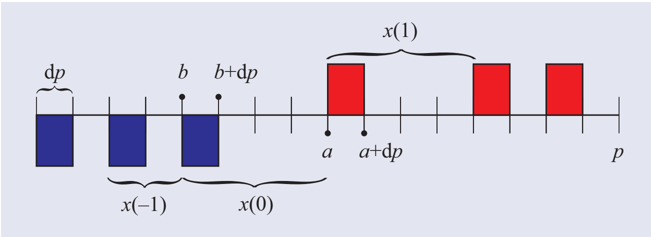

Figure 15. The price space and order profile. $n(p, t)$ has been chosen to be 0 or ± 1, a restriction that will be convenient later. Price bins are labelled by their lower boundary price, and intervals $x(N)$ will be defined below.

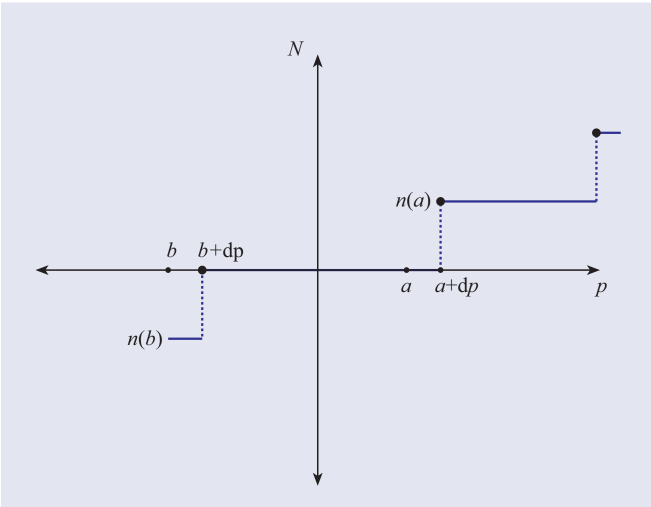

Figure 16. The accumulated order number $N(p, t)$. $N(a, t) \equiv 0$, because contributions from all bins cancel in the two sums. N remains zero down to $b(t) + \mathrm{d}p$, because there are no uncancelled, non-zero terms. $N(b, t)$ becomes negative, because the second sum in equation (4) now contains $n(b, t)$, not cancelled by the first.

An equivalent specification of a limit order book configuration is given by the cumulative order count

$$N(p, t) \equiv \sum_{-\infty}^{p-\mathrm{d}p} |n(p, t)| - \sum_{-\infty}^{a-\mathrm{d}p} |n(p, t)|, \tag{4}$$

where $-\infty$ denotes the lower boundary of the price space, whose exact value must not affect the results. (Because by definition there are no orders between the bid and ask, the bid could equivalently have been used as the origin of summation. Because price bins will be indexed here by their lower boundaries, though, it is convenient here to use the ask.) The absolute values have been placed so that N, like n, is negative in the range of buy orders and positive in the range of sells. The construction of $N(p, t)$ is illustrated in figure 16.

In many cases of either sparse orders or infinitesimal $\mathrm{d}p$, with fixed order size (which we may as well define to be one share) there will be either zero or one share in any single bin, and equation (4) will be invertible to an equivalent specification of the limit order book configuration

$$p(N, t) \equiv \max\{p | N(p, t) = N\}, \tag{5}$$

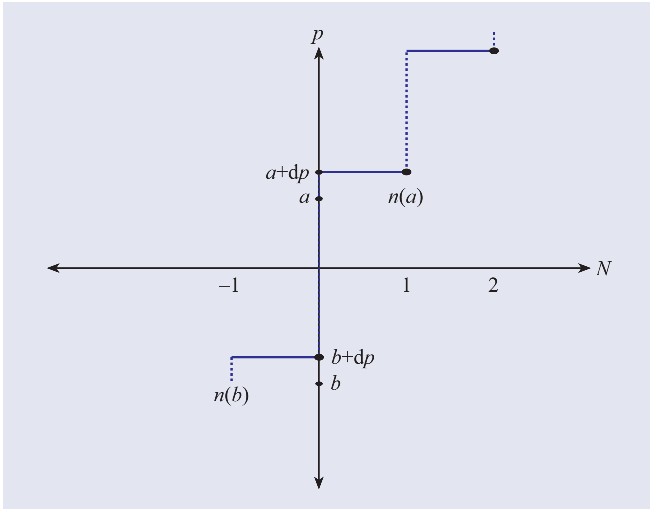

Figure 17. The inverse function $p(N, t)$. The function is in general defined only on discrete values of N, so this domain is only invariant when order size is fixed, a convenience that will be assumed below. Between the discrete domain, and the definition of p as a maximum, the inverse function effectively interpolates between vertices of the reflected image of $N(p, t)$, as shown by the dotted line.

shown in figure 17. (Strictly, the inversion may be performed for any distribution of order sizes, but the resulting function is intrinsically discrete, so its domain is only invariant when order size is fixed. To give $p(N, t)$ the convenient properties of a well-defined function on an invariant domain, this will be assumed below.)

With definition (5), $p(0, t) \equiv a(t)$, $p(-1, t) \equiv b(t)$, and one can define the intervals between orders as

$$x(N, t) \equiv p(N, t) - p(N - 1, t). \tag{6}$$

Thus $x(0, t) = a(t) - b(t)$, the instantaneous bid–ask spread. The lowest values of $x(N, t)$ bracketing the spread are shown in figure 15. For symmetric order placement rules, probability distributions over configurations will be symmetric under either $n(p, t) \to -n(-p, t)$, or $x(N, t) \to x(-N, t)$. Coordinates N and p furnish a dual description of configurations, and n and x are their associated differences. The master equation approach of section 3.6 assumes independent fluctuation in n while the IIA of section 3.7 assumes independent fluctuation in x. (In this section, it will be convenient to use the abbreviation $x(N, t) \equiv x_N(t)$.)

3.3. Frames and marginals

The $x(N, t)$ specification of limit order book configurations has the property that its distribution is stationary under the dynamics considered here. The same is not true for $p(N, t)$ or $n(p, t)$ directly, because bid, mid-point, and ask prices undergo a random walk, with a renormalized diffusion coefficient. Stationary distributions for n variables can be obtained in *comoving frames*, of which there are several natural choices.

The *bid centred configuration* is defined as

$$n_{\mathrm{b}}(p, t) \equiv n(p - b(t), t). \tag{7}$$

Ergodic theory (ergon work, hodos way) is a branch of mathematics that studies dynamical systems with an invariant measure and related problems. Its initial development was motivated by problems of statistical physics.
A central concern of ergodic theory is the behavior of a dynamical system when it is allowed to run for a long time. The first result in this direction is the Poincaré recurrence theorem, which claims that almost all points in any subset of the phase space eventually revisit the set. More precise information is provided by various ergodic theorems which assert that, under certain conditions, the time average of a function along the trajectories exists almost everywhere and is related to the space average. Two of the most important theorems are those of Birkhoff (1931) and von Neumann which assert the existence of a time average along each trajectory. For the special class of ergodic systems, this time average is the same for almost all initial points: statistically speaking, the system that evolves for a long time "forgets" its initial state. Stronger properties, such as mixing and equidistribution, have also been extensively studied.

If an appropriate rounding convention is adopted in the case of discrete prices, a *mid-point centred configuration* can also be defined, as

$$n_{\mathrm{m}}(p,t) \equiv n(p-m(t),t). \tag{8}$$

The mid-point centred configuration has qualitative differences from the bid centred configuration, which will be explored below. Both give useful insights into the order distribution and diffusion processes. The ask centred configuration, $n_{\mathrm{a}}(p,t)$, need not be considered if order placement and removal are symmetric, because it is the mirror image of $n_{\mathrm{b}}(p,t)$.

The *spread* is defined as the difference $s(t) \equiv a(t) - b(t)$, and is the value of the ask in bid centred coordinates. In mid-point centred coordinates, the ask appears at $s(t)/2$.

The configurations n_{b} and n_{m} are dynamically correlated over short time intervals, but evolve ergodically in periods longer than finite characteristic correlation times. Marginal probability distributions for these can therefore be computed as time averages, either as functions on the whole price space, or at discrete sets of prices. Their marginal mean values at a single price p will be denoted $\langle n_{\mathrm{b}}(p)\rangle$, $\langle n_{\mathrm{m}}(p)\rangle$, respectively.

These means are subject to global balance constraints, between total order placement and removal in the price space. Because all limit orders are placed above the bid, the bid centred configuration obeys a simple balance relation:

$$\frac{\mu}{2} = \sum_{p=b+\mathrm{d}p}^{\infty} (\alpha - \delta\langle n_{\mathrm{b}}(p)\rangle). \tag{9}$$

Equation (9) says that buy market orders must account, on average, for the difference between all limit orders placed, and all decays. After passing to non-dimensional coordinates below, this will imply an inverse relation between corrections to the classical estimate for diffusivity at early and late times, discussed in section 3.5. In addition, this conservation law plays an important role in the analysis and determination of the $x(N,t)$, as we will see later in the text.

The mid-point centred averages satisfy a different constraint:

$$\frac{\mu}{2} = \alpha\frac{\langle s\rangle}{2} + \sum_{p=b+\mathrm{d}p}^{\infty} (\alpha - \delta\langle n_{\mathrm{m}}(p)\rangle). \tag{10}$$

Market orders in equation (10) account not only for the excess of limit order placement over evaporation at prices above the mid-point, but also the 'excess' orders placed between $b(t)$ and $m(t)$. Since these always lead to mid-point shifts, they ultimately appear at positive comoving coordinates, altering the shape of $\langle n_{\mathrm{m}}(p)\rangle$ relative to $\langle n_{\mathrm{b}}(p)\rangle$. Their rate of arrival is $\alpha\langle m-b\rangle = \alpha\langle s\rangle/2$. These results are also confirmed in simulations.

3.4. Factorization tests

Whether in the bid centred frame or the mid-point centred frame, the probability distribution function for the entire configuration $n(p)$ is too difficult a problem to solve in its entirety. However, an approximate master equation can be formed for n independently at each p if all joint probabilities factor into independent marginals, as

$$\Pr(\{n(p_i)\}_i) = \prod_i \Pr(n(p_i)), \tag{11}$$

where Pr denotes, for instance, a probability density for n orders in some interval around p.

Whenever orders are sufficiently sparse that the expected number in any price bin is simply the probability that the bin is occupied (up to a constant of proportionality), the independence assumption implies a relation between the cumulative distribution for the spread of the ask and the mean density profile. In units where the order size is one, the relation is

$$\Pr(s/2 < p) = 1 - \exp\left(-\sum_{p'=b+\mathrm{d}p}^{p-\mathrm{d}p} \langle n_{\mathrm{m}}(p')\rangle\right). \tag{12}$$

This relation is tested against simulation results in figure 18. One can observe that there are three regimes.

A high ϵ regime is defined when the mean density profile at the mid-point $\langle n_{\mathrm{m}}(0)\rangle \lesssim 1$, and strongly concave downward. In this regime, the approximation of independent fluctuations is excellent, and a master equation treatment is expected to be useful. Intermediate ϵ is defined by $\langle n_{\mathrm{m}}(0)\rangle \ll 1$ and nearly linear, and the approximation of independence is marginal. Large ϵ is defined by $\langle n_{\mathrm{m}}(0)\rangle \ll 1$ and concave upward, and the approximation of independent fluctuations is completely invalid. These regimes of validity correspond also to the qualitative ranges noted already in section 2.2. validity

In the bid centred frame however, equation (12) never seems to be valid for any range of parameters. We will discuss later why this might be so. For the present therefore, the master equation approach is carried out in the mid-point centred frame. Alternatively, the mean field theory of the separations is most convenient in the bid centred frame, so that frame will be studied in the dual basis. The relation of results in the two frames, and via the two methods of treatment, will provide a good qualitative, and for some properties quantitative, understanding of the depth profile and its effect on impacts.

It is possible in a modified treatment to match certain features of simulations at any ϵ, by limited incorporation of correlated fluctuations. However, the general master equation will be developed independently of these, and tested against simulation results at large ϵ, where its defining assumptions are well met.

3.5. Comments on renormalized diffusion

A qualitative understanding of why the diffusivity is different over short and long timescales, as well as why it may depend on ϵ, may be gleaned from the following observations.

First, global order conservation places a strong constraint on the classically non-dimensionalized density profile in the bid centred frame. We have seen that at $\epsilon \ll 1$, the density profile becomes concave upward near the bid, accounting for an increasing fraction of the allowed 'remainder area' as

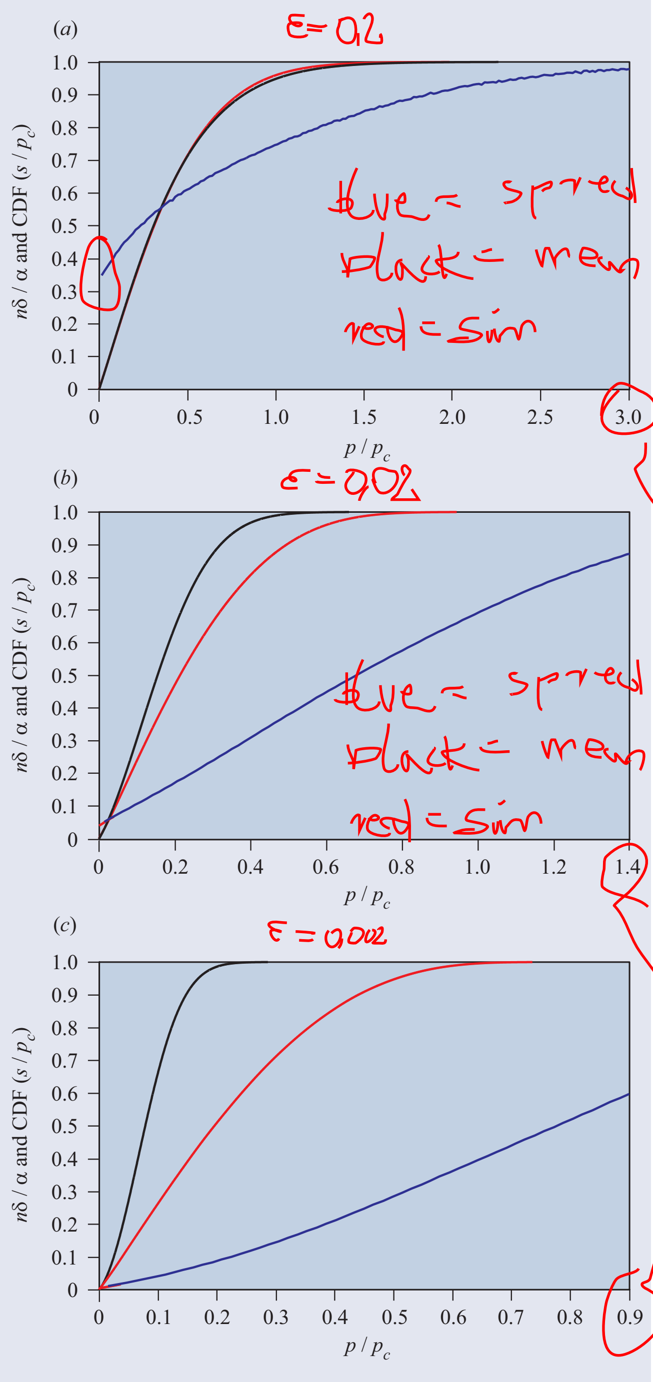

Figure 18. CDFs $\Pr(s/2 < p)$ from simulations (red), mean density profile $\langle n_{\mathrm{m}}(p)\rangle$ from simulations (blue), and computed CDF of spread (black) from $\langle n_{\mathrm{m}}(p)\rangle$, under the assumption of uncorrelated fluctuations, at three values of ϵ. (*a*) $\epsilon = 0.2$ (low market order rate); the approximation is very good. (*b*) $\epsilon = 0.02$ (intermediate market order rate); the approximation is marginal. (*c*) $\epsilon = 0.002$ (high market order rate); the approximation is very poor.

$\epsilon \to 0$ (see figures 3 and 28). Since this remainder area is fixed at unity, it can be conserved only if the density profile approaches one more quickly with increasing price. Low density at low price appears to lead to more frequent persistent steps in the effective short term random walk, and hence large short term diffusivity. However, increased density far from the bid indicates less impact from market orders relative to the relaxation time of the Poisson distribution, and thus a lower long time diffusivity.

The qualitative behaviour of the bid centred density profile is the same as that of the mid-point centred profile, and this is expected because the spread distribution is stationary, rather than diffusive. In other words, the only way the diffusion of the bid or ask can differ from that of the mid-point is for the spread to either increase or decrease for several succeeding steps. Such autocorrelation of the spread cannot accumulate with time if the spread itself is to have a stationary distribution. Thus, the shift in the mid-point over some time interval can only differ from that of the bid or ask by at most a constant, as a result of a few correlated changes in the spread. This difference cannot grow with time, however, and so does not affect the diffusivity at long times.

Indeed, both of the predicted corrections to the classical estimate for diffusivity are seen in simulation results for mid-point diffusion. The simulation results, however, show that the implied autocorrelations change the diffusivity by factors of $\sqrt{\epsilon}$, suggesting that these corrections require a more subtle derivation than the one attempted here. This will be evidenced by the difficulty of obtaining a source term $\mathcal{S}$ in the density coordinates section 3.6, which satisfied both the global order conservation law, and the proper zero price boundary condition, in the mid-point centred frame.

An interesting speculation is that the subtlety of these correlations also causes the density $n(p,t)$ in bid centred coordinates not to approximate the mean field condition at any of the parameters studied here, as noted in section 3.4. Since short term and long term diffusivity corrections are related by a hard constraint, the difficulty of producing the late time density profile should match that of producing the early time profile. The mid-point centred profile is potentially easier, in that the late time complexity must be matched by a combination of the early time density profile and the scaling of the expected spread. It appears that the complex scaling is absorbed in the spread, as per figures 10 and 24, leaving a density that can be approximately calculated with the methods used here.

3.6. Master equations and mean field approximations

There are two natural limits in which functional configurations may become simple enough to be tractable probabilistically, with analytic methods. They correspond to mean field theories in which fluctuations of the dual differentials of either $N(p,t)$ or $p(N,t)$ are independent. In the first case, probabilities may be defined for any density $n(p,t)$ independently at each p, and in the second for the separation intervals $x(N,t)$ at each N. The mean field theory from the first approximation will be solved in section 3.6.1, and that from the second in section 3.7. As mentioned above, because the fluctuation independence approximation is only usable in a mid-point centred frame, $n(p,t)$ will refer always to this frame. $x(N,t)$ is well defined without reference to any frame.

3.6.1. A number density master equation. If share number fluctuations are independent at different p, a density $\pi(n, p, t)$ may be defined, which gives the probability of finding n orders in bin $(p, p + \mathrm{d}p)$, at time t. The normalization condition defining π as a probability density is

$$\sum_n \pi(n, p, t) = 1, \tag{13}$$

for each bin index p and at every t. The index t will be suppressed henceforth in the notation since we are looking for time independent solutions.

Supposing an arbitrary density of order book configurations $\pi(n, p)$ at time t, the stochastic dynamics of the configurations causes probability to be redistributed according to the master equation

$$\begin{aligned}
\frac{\partial}{\partial t}\pi(n, p) = {} & \frac{\alpha(p)\,\mathrm{d}p}{\sigma}[\pi(n - \sigma, p) - \pi(n, p)] \\
& + \frac{\delta}{\sigma}[(n + \sigma)\pi(n + \sigma, p) - n\pi(n, p)] \\
& + \frac{\mu(p)}{2\sigma}[\pi(n + \sigma, p) - \pi(n, p)] \\
& + \sum_{\Delta p} P_+(\Delta p)[\pi(n, p - \Delta p) - \pi(n, p)] \\
& + \sum_{\Delta p} P_-(\Delta p)[\pi(n, p + \Delta p) - \pi(n, p)].
\end{aligned} \tag{14}$$

Here $\partial\pi(n, p)/\partial t$ is a continuum notation for $[\pi(n, p, t + \delta t) - \pi(n, p, t)]/\delta t$, where δt is an elementary time step, chosen short enough that at most one event alters any typical configuration. Equation (14) represents a general balance between additions and removals, without regard to the meaning of n. Thus, $\alpha(p)$ is a function that must be determined self-consistently with the choice of frame. As an example of how this works, in a bid centred frame, $\alpha(p)$ takes a fixed value $\alpha(\infty)$ at all p, because the deposition rate is independent of position and frame shifts. The mid-point centred frame is more complicated, because depositions below the mid-point cause shifts that leave the deposited order above the mid-point. The specific consequence for $\alpha(p)$ in this case will be considered below. $\mu(p)/2$ is, similarly, the rate of market orders surviving to cancel limit orders at price p. $\mu(p)/2$ decreases from $\mu(0)/2$ at the ask (for buy market orders, because μ total orders are divided evenly between buys and sells) to zero as $p \to \infty$, as market orders are screened probabilistically by intervening limit orders. $\alpha(\infty)$ and $\mu(0)$ are thus the parameters α and μ of the simulation.

The lines of equation (14) correspond to the following events. The term proportional to $\alpha(p)\,\mathrm{d}p/\sigma$ describes depositions of discrete orders at that rate (because α is expressed in *shares per price per time*), which raise configurations from $n - \sigma$ to n shares at price p. The term proportional to δ comes from deletions and has the opposite effect, and is proportional to n/σ, the number of *orders* that can independently decay. The term proportional to $\mu(p)/2\sigma$ describes market order annihilations. For general configurations, the preceding three effects may lead to shifts of the origin by arbitrary intervals Δp, and $P_\pm$ are for the moment unknown distributions over the frequency of those shifts. They must be determined self-consistently with the configuration of the book which emerges from any solution to equation (14).

A limitation of the simple product representation of frame shifts is that it assumes that whole order book configurations are transported under $p \pm \Delta p \to p$, independently of the value of $n(p)$. As long as fluctuations are independent, this is a good approximation for orders at all p which are not either the bid or the ask, either before or after the event that causes the shift. The correlations are never ignorable for the bins which are the bid and ask, though, and there is some distribution of instances in which any p of interest plays those parts. Approximate methods for incorporating those correlations will require replacing the product form with a sum of products conditioned on states of the order book, as will be derived below.

The important point is that the order flow dependence of equation (14) is independent of these self-consistency requirements, and may be solved by the use of generating functionals at general $\alpha(p)$, $\mu(p)$, and $P_\pm$. The solution, exact but not analytically tractable at general $\mathrm{d}p$, will be derived in closed form in the next subsection. It has a well-behaved continuum limit at $\mathrm{d}p \to 0$, however, which is analytically tractable, so that a special case will be considered in the following subsection.

λ - Slope of depth profile Shares/price2

3.6.2. Solution by generating functional. The moment generating functional for π is defined for a parameter $\lambda \in [0, 1]$ as

$$\Pi(\lambda, p) \equiv \sum_{n/\sigma=0}^{\infty} \lambda^{n/\sigma}\pi(n, p). \tag{15}$$

Introducing a shorthand form for its value at $\lambda = 0$,

$$\Pi(0, p) = \pi(0, p) \equiv \pi_0(p), \tag{16}$$

while the normalization condition (13) for probabilities gives

$$\Pi(1, p) = 1, \qquad \forall p. \tag{17}$$

By definition of the average of $n(p)$ in the distribution π, denoted as $\langle n(p)\rangle$,

$$\left.\frac{\partial}{\partial\lambda}\Pi(\lambda, p)\right|_{\lambda\to 1} = \frac{\langle n(p)\rangle}{\sigma}, \tag{18}$$

and because Π will be regular in some sufficiently small neighbourhood of $\lambda = 1$, one can make the expansion

$$\Pi(\lambda, p) = 1 + (\lambda - 1)\frac{\langle n(p)\rangle}{\sigma} + \mathcal{O}(\lambda - 1)^2. \tag{19}$$

Multiplying equation (14) by $\lambda^{n/\sigma}$ and summing over n (and suppressing the argument p in the notation everywhere; $\alpha(p)$ or $\alpha(0)$ will be used where the distinction of the function from its boundary value is needed), the stationary solution for Π must satisfy

$$\begin{aligned}
0 = {} & \frac{\lambda - 1}{\sigma}\left\{\alpha\,\mathrm{d}p\,\Pi - \delta\sigma\frac{\partial\Pi}{\partial\lambda} - \frac{\mu}{2\lambda}(\Pi - \pi_0)\right\}\Bigg|_{(\lambda,p)} \\
& + \sum_{\Delta p} P_+(\Delta p)[\Pi(\lambda, p - \Delta p) - \Pi(\lambda, p)] \\
& + \sum_{\Delta p} P_-(\Delta p)[\Pi(\lambda, p + \Delta p) - \Pi(\lambda, p)].
\end{aligned} \tag{20}$$

Only the symmetric case with no net drift will be considered here for simplicity, which requires $P_+(\Delta p) = P_-(\Delta p) \equiv P(\Delta p)$. In a Fokker–Planck expansion, the (unrenormalized) diffusivity of whatever reference price is used as coordinate origin is related to the distribution P by

$$D \equiv \sum_{\Delta p} P(\Delta p)\,\Delta p^2. \tag{21}$$

The rate at which shift events happen is

$$R \equiv \sum_{\Delta p} P(\Delta p), \tag{22}$$

and the mean shift amount appearing at linear order in derivatives (relevant at $p \to 0$) is

$$\langle \Delta p \rangle \equiv \frac{\sum_{\Delta p} P(\Delta p)\,\Delta p}{\sum_{\Delta p} P(\Delta p)}. \tag{23}$$

Anywhere in the interior of the price range (where p is not at any stage the bid, ask, or a point in the spread), equation (20) may be written as

$$\left\{ \frac{\partial}{\partial \lambda} - \frac{D}{\delta(\lambda - 1)} \frac{\partial^2}{\partial p^2} - \frac{\alpha\,\mathrm{d}p - \mu/2\lambda}{\delta\sigma} \right\} \Pi = \frac{\mu}{2\delta\sigma\lambda}\pi_0. \tag{24}$$

Evaluated at $\lambda \to 1$, with the use of the expansion (19), this becomes

$$\left(1 - \frac{D}{\delta}\frac{\partial^2}{\partial p^2}\right)\langle n \rangle = \frac{\alpha\,\mathrm{d}p}{\delta} - \frac{\mu}{2\delta}(1 - \pi_0). \tag{25}$$

At this point it is convenient to specialize to the case $\mathrm{d}p \to 0$, wherein the eligible values of any $\langle n(p)\rangle$ become just σ and zero. The expectation is then related to the probability of zero occupancy (at each p) as

$$\langle n \rangle = \sigma[1 - \pi_0], \tag{26}$$

yielding immediately

$$\frac{\alpha\,\mathrm{d}p}{\delta} = \left[\frac{\mu}{2\delta\sigma} + \left(1 - \frac{D}{\delta}\frac{\mathrm{d}^2}{\mathrm{d}p^2}\right)\right]\langle n \rangle. \tag{27}$$

Equation (27) defines the general solution $\langle n(p)\rangle$ for the master equation (14), in the continuum limit $2\alpha\,\mathrm{d}p/\mu \to 0$. The shift distribution $P(\Delta p)$ appears only through the diffusivity D, which must be solved self-consistently, along with the otherwise arbitrary functions α and μ. The more general solution at large $\mathrm{d}p$ is carried out in appendix B.1.

A first step toward non-dimensionalization may be taken by writing equation (27) in the form (re-introducing the indexing of the functions)

$$\frac{\alpha(p)}{\alpha(\infty)} = \left[\frac{\mu(p)}{\mu(0)} + \epsilon\left(1 - \frac{D}{\delta}\frac{\mathrm{d}^2}{\mathrm{d}p^2}\right)\right]\frac{1}{\epsilon}\frac{\delta\langle n\rangle}{\alpha\,\mathrm{d}p}. \tag{28}$$

Far from the mid-point, where only depositions and cancellations take place, orders in bins of width $\mathrm{d}p$ are Poisson distributed with mean $\alpha(\infty)\,\mathrm{d}p/\delta$. Thus, the asymptotic value of $\delta\langle n\rangle/\alpha(\infty)\,\mathrm{d}p$ at large p is unity. This is consistent with a limit for $\alpha(p)/\alpha(\infty)$ of unity, and a limit for the screened $\mu(p)/\mu(0)$ of zero. The reason for grouping the non-dimensionalized number density with $1/\epsilon$, together with the proper normalization of the characteristic price scale, will come from examining the decay of the dimensionless function $\mu(p)/\mu(0)$.

3.6.3. Screening of the market order rate.

In the context of independent fluctuations, equation (26) implies a relation between the mean density and the rate at which market orders are screened as price increases. The effect of a limit order, resident in the price bin p when a market order survives to reach that bin, is to prevent its arriving at the bin at $p + \mathrm{d}p$. Though the nature of the shift induced, when such annihilation occurs, depends on the comoving frame being modelled, the change in the number of orders surviving is independent of frame, and is given by

$$\mathrm{d}\mu = -\mu(1 - \pi_0) = -\mu\langle n\rangle/\sigma. \tag{29}$$

Equation (29) may be rewritten as

$$\frac{\mathrm{d}\log(\mu(p)/\mu(0))}{\mathrm{d}p} = -\frac{1}{\epsilon}\left(\frac{2\alpha(\infty)}{\mu(0)}\right)\left(\frac{\delta\langle n(p)\rangle}{\alpha(\infty)\,\mathrm{d}p}\right), \tag{30}$$

identifying the characteristic scale for prices as $p_c = \mu(0)/2\alpha(\infty) \equiv \mu/2\alpha$. Writing $\hat{p} \equiv p/p_c$, the function that screens market orders is the same as the argument of equation (28), and will be denoted as

$$\frac{1}{\epsilon}\frac{\delta\langle n(p)\rangle}{\alpha(\infty)\,\mathrm{d}p} \equiv \psi(\hat{p}). \tag{31}$$

Defining a non-dimensionalized diffusivity $\beta \equiv D/\delta p_c^2$, equation (27) can then be put in the form

$$\frac{\alpha(p)}{\alpha(\infty)} = \left[\frac{\mu(p)}{\mu(0)} + \epsilon\left(1 - \beta\frac{\mathrm{d}^2}{\mathrm{d}\hat{p}^2}\right)\right]\psi, \tag{32}$$

with

$$\frac{\mu(p)}{\mu(0)} \equiv \varphi(\hat{p}) = \exp\left(-\int_0^{\hat{p}} \mathrm{d}\hat{p}'\,\psi(\hat{p}')\right). \tag{33}$$

3.6.4. Verifying the conservation laws.

Since nothing about the derivation so far has made explicit use of the frame in which n is averaged, the combination of equation (32) with (33) respects the conservation laws (9) and (10), if appropriate forms are chosen for the deposition rate $\alpha(p)$.

For example, in the bid centred frame, $\alpha(p)/\alpha(\infty) = 1$ everywhere. Multiplying equation (32) by $\mathrm{d}\hat{p}$ and integrating over the whole range from the bid to $+\infty$, we recover the non-dimensionalized form of equation (9):

$$\int_0^{\infty} \mathrm{d}\hat{p}\,(1 - \epsilon\psi) = 1, \tag{34}$$

iff we are careful with one convention. The integral of the diffusion term formally produces the first derivative $\mathrm{d}\psi/\mathrm{d}\hat{p}|_0^\infty$. We must regard this as a true first derivative, and consider its evaluation at zero continued far enough below the bid to capture the identically zero first derivative of the sell order depth profile.

In the mid-point centred frame, the correct form for the source term should be $\alpha(\hat{p})/\alpha(\hat{\infty}) = 1 + \Pr(\hat{s}/2 \geqslant \hat{p})$, whatever the expression for the cumulative distribution function. Recognizing that the integral of the CDF is, by parts, the mean value of $\hat{s}/2$, the same integration of equation (32) gives

$$\int_0^\infty \mathrm{d}\hat{p}\,(1-\epsilon\psi) = 1 - \frac{\langle\hat{s}\rangle}{2}, \tag{35}$$

the non-dimensionalized form of equation (10). Again, this works only if the surface contribution from integrating the diffusion term vanishes.

Neither of these results required the assumption of independent fluctuations, though that will be used below to give a simple approximate form for $\Pr(\hat{s}/2 \geqslant \hat{p}) \approx \varphi(\hat{p})$. They therefore provide a check that the extinction form (33) propagates market orders correctly into the interior of the order book distribution, to respect global conservation. They also check the consistency of the intuitively plausible form for α in the mid-point centred frame. The detailed form is then justified whenever the assumption of independent fluctuations is checked to be valid.

3.6.5. Self-consistent parametrization. The assumption of independent fluctuations of $n(p)$ used above to derive the screening of market orders is equivalent to a specification of the CDF of the ask. Market orders are only removed between prices p and $p + \mathrm{d}p$ in those instances when the ask is at p. Therefore

$$\Pr(\hat{s}/2 \geqslant \hat{p}) = \varphi(\hat{p}), \tag{36}$$

the continuum limit of equation (12). Together with the form $\alpha(\hat{p})/\alpha(\hat{\infty}) = 1 + \Pr(\hat{s}/2 \geqslant \hat{p})$, equation (32) becomes

$$1+\varphi = -\left[\frac{\mathrm{d}\varphi}{\mathrm{d}\hat{p}} + \epsilon\left(1 - \beta\frac{\mathrm{d}^2}{\mathrm{d}\hat{p}^2}\right)\frac{\mathrm{d}\log\varphi}{\mathrm{d}\hat{p}}\right]. \tag{37}$$

(If the assumption of independent fluctuations were valid in the bid centred frame, it would take the same form, but with φ removed on the left-hand side.)

To consistently use the diffusion approximation, with the realization that for $p = 0$, $n\pi(n, p - \Delta p) = 0$ for essentially all Δp in equation (14), it is necessary to set the Fokker–Planck approximation to $\psi(0 - \langle\Delta p\rangle) = 0$ as a boundary condition. Non-dimensionalized, this gives

$$\left.\frac{\beta}{2}\frac{\mathrm{d}^2\psi}{\mathrm{d}\hat{p}^2}\right|_0 = \frac{R}{\delta}\left(\langle\Delta\hat{p}\rangle\frac{\mathrm{d}}{\mathrm{d}\hat{p}} - 1\right)\psi\bigg|_0, \tag{38}$$

where R is the rate at which shifts occur (equation (22)). In the solutions below, the curvature will typically be much smaller than $\psi(0) \sim 1$, so it will be convenient to enforce the simpler condition

$$\langle\Delta\hat{p}\rangle\left.\frac{\mathrm{d}\psi}{\mathrm{d}\hat{p}}\right|_0 - \psi(0) \approx 0, \tag{39}$$

and verify that it is consistent once solutions have been evaluated.

Self-consistent expressions for β and $\langle\Delta p\rangle$ are then constructed as follows. Given an ask at some position a (in the mid-point centred frame), there is a range from $-a$ to a in which sell limit orders may be placed, which will induce positive mid-point shifts. The shift amount is half as great as the distance from the bid, so the measure for shifts $\mathrm{d}P_+(\Delta p)$ from sell limit order addition inherits a term $2\alpha(0)(\mathrm{d}\Delta p)\Pr(a \geqslant \Delta p)$, where the last factor counts the instances with asks large enough to admit shifts by Δp. There is an equal contribution to $\mathrm{d}P_-$ from addition of buy limit orders. Symmetry requires that for every positive shift due to an addition, there is a negative shift due to evaporation with equal measure, so the contribution from buy limit order removal should equal that for sell limit order addition. When these contributions are summed, the measures for positive and negative shifts both equal

$$\mathrm{d}P_\pm(\Delta p) = 4\alpha(\infty)(\mathrm{d}\Delta p)\Pr(a \geqslant \Delta p). \tag{40}$$

Equation (40) may be inserted into the continuum limit of the definition (21) for D, and then non-dimensionalized to give

$$\beta = \frac{4}{\epsilon}\int_0^\infty \mathrm{d}\Delta\hat{p}\,(\Delta\hat{p})^2\varphi(\Delta\hat{p}), \tag{41}$$

where the mean field substitution of $\varphi(\Delta\hat{p})$ for $\Pr(a \geqslant \Delta p)$ has been used. Similarly, the mean shift amount used in equation (39) is

$$\langle\Delta\hat{p}\rangle = \frac{\int_0^\infty \mathrm{d}\Delta\hat{p}\,(\Delta\hat{p})\varphi(\Delta\hat{p})}{\int_0^\infty \mathrm{d}\Delta\hat{p}\,\varphi(\Delta\hat{p})}. \tag{42}$$

A fit of equation (37) to simulations, using these self-consistent measures for shifts, is shown in figure 19. This solution is actually a compromise between approximations with opposing ranges of validity. The diffusion equation using the mean order depth describes non-zero transport of limit orders through the mid-point, an approximation inconsistent with the correlations of shifts with states of the order book. This approximation is a small error only at $\epsilon \to 0$. On the other hand, both the form of α and the self-consistent solutions for $\langle\Delta\hat{p}\rangle$ and β made use of the mean field approximation, which we saw was only valid for $\epsilon \lesssim 1$. The two approximations appear to create roughly compensating errors in the intermediate range $\epsilon \sim 0.02$.

3.6.6. Accounting for correlations. The numerical integral implementing the diffusion solution actually does not satisfy the global conservation condition that the diffusion term integrate to zero over the whole price range. Thus, it describes diffusive transport of orders through the mid-point, and as such also does not have the right $\hat{p} = 0$ boundary condition. The effective absorbing boundary represented by the pure diffusion solution corresponds roughly to the approximation made by Bouchaud *et al* [23]. It differs from theirs, though, in that their method of images effectively approximates the region of the spread as a point, whereas equation (32) actually resolves the screening of market orders as the spread fluctuates.

Treating the spread region—roughly defined as the range over which market orders are screened—as a point is consistent with treating the resulting coarse grained 'mid-point' as an absorbing boundary. If the spread is resolved, however, it is not consistent for diffusion to transport any finite number density through the mid-point, because the mid-point is strictly always in the centre of an open set with no orders, in a continuous price space. The correct behaviour in a neighbourhood of the 'fine

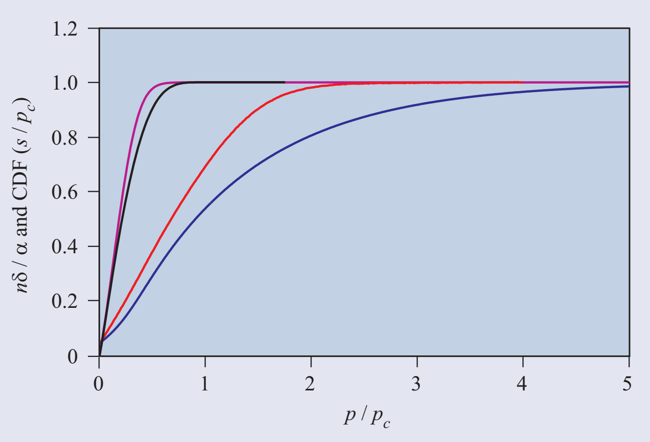

Figure 19. A fit of the self-consistent solution with the diffusivity term to simulation results for the mid-point centred frame. The blue curve is the analytic solution for the mean number density, and the red curve is the simulation result, at $\epsilon = 0.02$. The purple curve is the analytic prediction for the cumulative distribution function $\Pr(\hat{s}/2 \leqslant \hat{p})$, and the black curve is the simulation result.

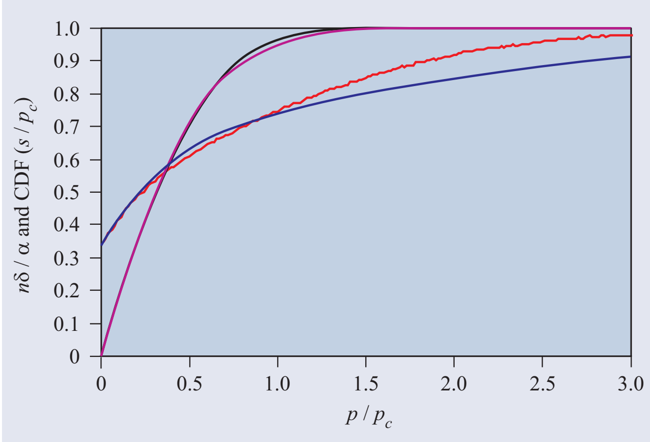

Figure 20. Reconstruction with source terms $\mathcal{S}$ that approximately account for correlated fluctuations near the mid-point. $\epsilon = 0.2$. The red curve is the averaged order book depth from simulations, and the blue curve is the mean field result. The purple curve is the simulated CDF for $\hat{s}/2$, and the green curve is the mean field result. (These two are so close that they cannot be resolved in this plot.) The black curve is the CDF that would be produced from the simulated depth, if the mean field approximation were exact.

grained mid-point' can be obtained by explicitly accounting for the correlation of the state of orders, with the shifts that are produced when market or limit order additions occur.

We expect the problem of recovering both the global conservation law and the correct $\hat{p} = 0$ boundary condition to be difficult, as it should be responsible for the non-trivial corrections to short term and long term diffusion mentioned earlier. We have found, however, that by explicitly sacrificing the global conservation law, we can incorporate the dependence of shifts on the position of the ask, in an interesting range around the mid-point. At general ϵ, the corrections to diffusion reproduce the mean density over the main support of the CDF of the spread. While the resulting density does not predict that CDF (due to correlated fluctuations), it resembles the real density closely enough that the independent CDFs of the two are similar.

3.6.7. Generalizing the shift induced source terms.

Non-dimensionalizing the generating functional master equation (20) and keeping leading terms in $\mathrm{d}p$ at $\lambda \to 1$, we get

$$
\begin{aligned}
\frac{\alpha(\hat{p})}{\alpha(\hat{\infty})} = {} & \left(\frac{\mu(\hat{p})}{\mu(\hat{0})} + \epsilon\right)\psi(\hat{p}) \\
& - \int \mathrm{d}P_{+}(\Delta\hat{p})\,[\psi(\hat{p} - \Delta\hat{p}) - \psi(\hat{p})] \\
& - \int \mathrm{d}P_{-}(\Delta\hat{p})\,[\psi(\hat{p} + \Delta\hat{p}) - \psi(\hat{p})]
\end{aligned} \tag{43}
$$

where $\mathrm{d}P_{\pm}(\Delta\hat{p})$ is the non-dimensionalized measure that results from taking the continuum limit of $P_{\pm}$ in the variable $\Delta\hat{p}$.

Equation (43) is inaccurate because the number of orders shifted into or out of a price bin p, at a given spread, may be identically zero, rather than the unconditional mean value ψ. We take that into account by replacing the last two lines of equation (43) with lists of source terms, whose forms depend on the position of the ask, weighted by the probability density for that ask. Independent fluctuations are assumed by using equation (36).

It is convenient at this point to denote the replacement of the last two lines of equation (43) with the notation $\mathcal{S}$, yielding

$$
\frac{\alpha(\hat{p})}{\alpha(\hat{\infty})} = \left(\frac{\mu(\hat{p})}{\mu(\hat{0})} + \epsilon\right)\psi - \mathcal{S}. \tag{44}
$$

The global conservation laws for orders would be satisfied if $\int \mathrm{d}\hat{p}\,\mathcal{S} = 0$.

The source term $\mathcal{S}$ is derived approximately in appendix B.2. The solution to equation (44) at $\epsilon = 0.2$, with the simple diffusion source term replaced by the evaluations (B.29)–(B.37), is compared to the simulated order book depth and spread distribution in figure 20. The simulated $\langle n(p)\rangle$ satisfies equation (35), showing what is the correct 'remainder area' below the line $\langle n\rangle \equiv 1$. The numerical integral deviates from that value by the incorrect integral $\int \mathrm{d}\hat{p}\,\mathcal{S} \neq 0$. However, most of the probability for the spread lies within the range where the source terms $\mathcal{S}$ are approximately correct, and as a result the distribution for $\hat{s}/2$ is predicted fairly well.

Even where the mean field approximation is known to be inadequate, the source terms defined here capture most of the behaviour of the order book distribution in the region that affects the spread distribution. Figure 21 shows the comparison to simulations for $\epsilon = 0.02$, and figure 22 that for $\epsilon = 0.002$. Both cases fail to reproduce the distribution for the spread, and also fail to capture the large $\hat{p}$ behaviour of ψ. However, they approximate ψ at small $\hat{p}$ well enough that the resulting distribution for the spread is close to what would be produced by the simulated ψ if fluctuations were independent.

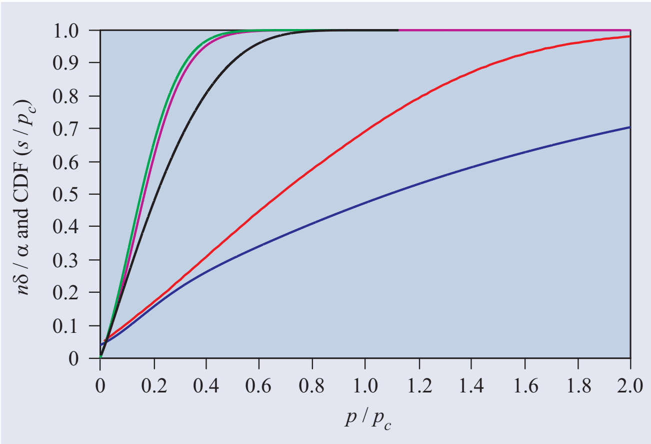

Figure 21. Reconstruction with correlated source terms for $\epsilon = 0.02$. The colour scheme is the same as in figure 20.

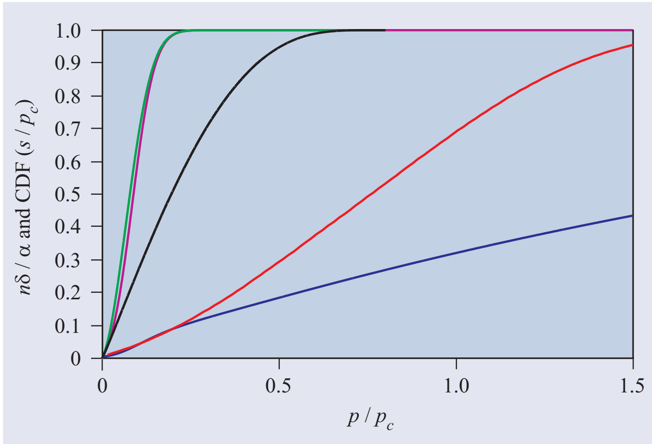

Figure 22. Reconstruction with correlated source terms for $\epsilon = 0.002$. The colour scheme is the same as in figure 20.

3.7. A mean field theory of order separation intervals: the independent interval approximation

A simplifying assumption that is in some sense dual to independent fluctuations of $n(p)$ is independent fluctuations in the intervals $x(N)$ at different N. Here we develop a mean field theory for the order separation intervals in this model. From this, we will also be able to make an estimate of the depth profiles for any value of the parameters. For convenience of notation we will use x_N to denote $x(N)$.

Limit order placements are considered to take place strictly on sites which are not occupied. This is the same level of approximation as made in the previous section. The time step is normalized to unity, as above, so that rates are equal to probabilities after one update of the whole configuration. The rates α and μ used in this section correspond to $\alpha(\infty)$ and $\mu(0)$ as defined earlier.

As shown in figure 15 the configuration is entirely specified instant by instant if the instantaneous values of the order separation intervals are known.

Consider now how these intervals might change due to various processes. For the spread x_0, these processes and the corresponding change in x_0 are listed below.

(1) $x_0 \to x_0 + x_1$ with rate $(\delta + \mu/2\sigma)$ (when the ask either evaporates or is deleted by a market order).

(2) $x_0 \to x_0 + x_{-1}$ with rate $(\delta + \mu/2\sigma)$ (when the bid either evaporates or is deleted by the corresponding market order).

(3) $x_0 \to x'$ for any value $1 \leqslant x' \leqslant x_0 - 1$, when a sell limit order is deposited anywhere in the spread. The rate for any single deposition is $\alpha\,\mathrm{d}p/\sigma$, so the cumulative rate for *some* deposition is $\alpha\,\mathrm{d}p(x_0 - 1)/\sigma$. (The -1 comes from the prohibition against depositing on occupied sites.)

(4) Similarly $x_0 \to x_0 - x'$ for any $1 \leqslant x' \leqslant x_0 - 1$, when a buy limit order is deposited in the spread, also with cumulative rate $\alpha\,\mathrm{d}p(x_0 - 1)/\sigma$.

(5) Since the above processes describe all possible single event changes to the configuration, the probability that it remains unchanged in a single time step is $1 - 2\delta - \mu/\sigma - 2\alpha\,\mathrm{d}p(x_0 - 1)/\sigma$.

In all that follows, we will put $\sigma = 1$ without loss of generality. If we know x_0, x_1, and x_{-1} at time t, the expected value at time $t + \mathrm{d}t$ is then

$$x_0(t + \mathrm{d}t) = x_0(t)[1 - 2\delta - \mu - 2\alpha(x_0 - 1)] + (x_0 + x_1)\left(\delta + \frac{\mu}{2}\right) + (x_0 + x_{-1})\left(\delta + \frac{\mu}{2}\right) + (\alpha_0\,\mathrm{d}p)x_0(x_0 - 1). \tag{45}$$

Here, $x_i(t)$ represents the value of the interval averaged over many realizations of the process evolved up to time t.

Again representing the finite difference as a time derivative, the change in the expected value, given x_0, x_1 and x_{-1}, is

$$\frac{\mathrm{d}x_0}{\mathrm{d}t} = (x_1 + x_{-1})\left(\delta + \frac{\mu}{2}\right) - (\alpha\,\mathrm{d}p)x_0(x_0 - 1). \tag{46}$$

Were it not for the quadratic term arising from deposition, equation (46) would be a linear function of x_0, x_1 and x_{-1}. However, we now need an approximation for $\langle x_0^2\rangle$, where the angle brackets represent an average over realizations as before or equivalently a time average in the steady state. Let us for the moment assume that we can approximate $\langle x_0^2\rangle$ by $a\langle x_0\rangle^2$, where a is some as yet undetermined constant to be determined self-consistently. We will make this approximation for all the x_k. This is clearly not entirely accurate because the PDF of x_k could depend on k (as indeed it does; we will comment on this a little later). However, as we will see this is still a very good approximation.

We will therefore make this approximation in equation (46) and everywhere below, and look for steady state solutions when the x_k have reached a time independent average value.

It then follows that

$$(\delta + \mu/2)(x_1 + x_{-1}) = a\alpha\,\mathrm{d}p\,x_0(x_0 - 1). \tag{47}$$

The interval x_k may be thought of as the inverse of the density at a distance $\sum_{j=0}^{k-1} x_j$ from the bid. That is, $x_i \approx 1/\langle n(\sum_{j=0}^{i-1} x_j\,\mathrm{d}p)\rangle$, the dual to the mean depth, at least

Table 5. Events that can change the value of x_1, with their rates of occurrence.

Case	Rate	Range
$x_1 \to x_2$	$(\delta + \mu/2)$	
$x_1 \to (x_1 + x_2)$	δ	
$x_1 \to x'$	$\alpha\,\mathrm{d}p$	$x' \in (1, x_0 - 1)$
$x_1 \to x_1 - x'$	$\alpha\,\mathrm{d}p$	$x' \in (1, x_1 - 1)$

at large i. It therefore makes sense to introduce a normalized interval

$$\hat{x}_i \equiv \epsilon \frac{\alpha}{\delta} x_i \,\mathrm{d}p = \frac{x_i \,\mathrm{d}p}{p_c} \approx \frac{1}{\psi(\sum_{j=0}^{i-1} \hat{x}_j)}, \tag{48}$$

the mean field inverse of the normalized depth ψ. In this non-dimensionalized form, equation (47) becomes

$$(1+\epsilon)(\hat{x}_1 + \hat{x}_{-1}) = a\hat{x}_0(\hat{x}_0 - \mathrm{d}\hat{p}), \tag{49}$$

where $\mathrm{d}\hat{p} = \mathrm{d}p/p_c$.

Since the depth profile is symmetric about the origin, $\hat{x}_1 = \hat{x}_{-1}$. From the equations, it can be seen that this ansatz is self-consistent and extends to all higher $\hat{x}_i$. Substituting this in equation (49) we get

$$(1+\epsilon)\hat{x}_1 = \frac{a}{2}\hat{x}_0(\hat{x}_0 - \mathrm{d}\hat{p}) = (1+\epsilon)\hat{x}_{-1}. \tag{50}$$

Proceeding to the change of x_1, the events that can occur, with their probabilities, are shown in table 5, with the remaining probability that x_1 remains unchanged.

The differential equation for the mean change of x_1 can be derived along previous lines and becomes

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = \left(2\delta + \frac{\mu}{2}\right)x_2 - \left(\delta + \frac{\mu}{2}\right)x_1 + \alpha\,\mathrm{d}p\left[\frac{x_0(x_0-1)}{2} - \frac{x_1(x_1-1)}{2} - x_1(x_0-1)\right]. \tag{51}$$

Note that in the above equations, the mean field approximation consists of assuming that terms like $\langle x_0 x_1\rangle$ are approximated by the product $\langle x_0\rangle\langle x_1\rangle$. This is thus an 'independent interval' approximation.

Non-dimensionalizing equation (51) and combining the result with equation (50) gives the stationary value for x_2 from x_0 and x_1,

$$(1+2\epsilon)\hat{x}_2 = \frac{a}{2}\hat{x}_1(\hat{x}_1 - \mathrm{d}\hat{p}) + \hat{x}_1(\hat{x}_0 - \mathrm{d}\hat{p}). \tag{52}$$

Following the same procedure for general k, the non-dimensionalized recursion relation is

$$(1+k\epsilon)\hat{x}_k = \frac{a}{2}\hat{x}_{k-1}(\hat{x}_{k-1} - \mathrm{d}\hat{p}) + \hat{x}_{k-1}\sum_{i=0}^{k-2}(\hat{x}_i - \mathrm{d}\hat{p}). \tag{53}$$

3.7.1. Asymptotes and conservation rules. Far from the bid or ask, $\hat{x}_k$ must go to a constant value, which we denote as $\hat{x}_\infty$. In other words, for large k, $\hat{x}_{k+1} \to \hat{x}_k$. Taking the difference between equation (53) for $k+1$ and k in this limit gives the identification

$$\epsilon\hat{x}_\infty = \hat{x}_\infty(\hat{x}_\infty - \mathrm{d}\hat{p}), \tag{54}$$

or $\hat{x}_\infty = \epsilon + \mathrm{d}\hat{p}$. Apart from the factor of $\mathrm{d}\hat{p}$, arising from the exclusion of deposition on already occupied sites, this agrees with the limit $\psi(\infty) \to 1/\epsilon$ found earlier. In the continuum limit $\mathrm{d}\hat{p} \to 0$ at fixed ϵ, these are the same.

From the large k limit of equation (53), one can also solve easily for the quantity $S_\infty \equiv \sum_{i=0}^{\infty}(\hat{x}_i - \hat{x}_\infty)$, which is related to the bid centred order conservation law mentioned in section 3.3. Dividing by a factor of $\hat{x}_\infty$ at large k,

$$(1+k\epsilon) = \frac{a}{2}(\hat{x}_\infty - \mathrm{d}\hat{p}) + \sum_{i=0}^{k-2}(\hat{x}_i - \mathrm{d}\hat{p}), \tag{55}$$

or, using equation (54) and rewriting the sum on the right-hand side as $\sum_{i=0}^{k-2}(\hat{x}_i - \hat{x}_\infty) + \sum_{i=0}^{k-2}\hat{x}_\infty - \mathrm{d}\hat{p}$,

$$1 + \left(1 - \frac{a}{2}\right)\epsilon = S_\infty. \tag{56}$$

The interpretation of S_∞ is straightforward. There are $k+1$ orders in the price range $\sum_{i=0}^{k} x_i$. Their decay rate is $\delta(k+1)$, and the rate of annihilation from market orders is $\mu/2$. The rate of additions, up to an uncertainty about what should be considered the centre of the interval, is $(\alpha\,\mathrm{d}p)\sum_{i=0}^{k}(x_i - 1)$ in the bid centred frame (where effective α is constant and additions on top of previously occupied sites are forbidden). Equality of addition and removal is the bid centred order conservation law (again), in the form

$$\frac{\mu}{2} + \delta(k+1) = \alpha\,\mathrm{d}p\sum_{i=0}^{k}(x_i - 1). \tag{57}$$

Taking k large, non-dimensionalizing, and using equation (54), equation (57) becomes

$$1 = S_\infty. \tag{58}$$

Table 6. Theoretical results versus results from simulations for S_∞.

ϵ	S_∞ from theory	S_∞ from MCS
0.66	1	1.000
0.2	1	1.000
0.04	1	0.998
0.02	1	1.000

This conservation law is indeed respected to a remarkable accuracy in Monte Carlo simulations of the model as indicated in table 6.

In order that the equation for the x_k obey this exact conservation law, we require equation (56) to be equal to equation (58). We can hence now self-consistently set the value of $a = 2$.

The value of a implies that we have now set $\langle x_k^2\rangle \sim 2\langle x_k\rangle^2$. This would be strictly true if the probability distribution function of the interval x_k were exponentially distributed for all k. This is generally a good approximation for large k for any ϵ. Figure 23 shows the numerical results from Monte Carlo simulations of the model, for the probability distribution function for three intervals x_0, x_1, and x_5 at $\epsilon = 0.1$. The functional forms for $P(x_0)$ and $P(x_1)$ are better approximated

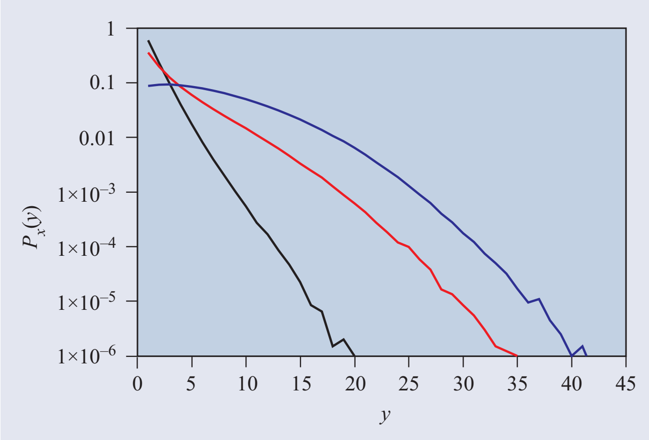

Figure 23. The probability distribution functions $P_x(y)$ versus y for the intervals $x = x_0$, x_1 and x_5 at $\epsilon = 0.1$, on a semi-log scale. The blue curve is for x_0, the red one for x_1, and the black one for x_5. The functional form of the distribution changes from a Gaussian to an exponential.

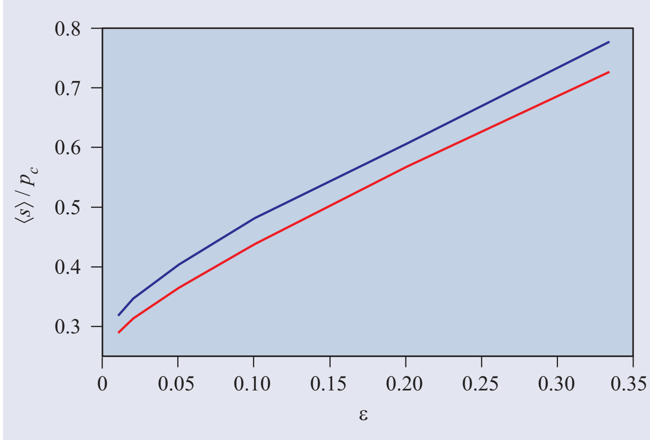

Figure 24. The mean value of the spread in non-dimensional units $\hat{s} = s/p_c$ as a function of ϵ. The numerical value above (blue) is compared with the theoretical estimate below (red).

by a Gaussian than an exponential. However, $P(x_5)$ is clearly an exponential.

Equation (58) has an important consequence for the short term and long term diffusivities, which can also be seen in simulations, as mentioned in earlier sections. The non-dimensionalization of the diffusivity D with the rate parameters suggests a classical scaling of the diffusivity

$$D \sim p_c^2 \delta = \frac{\mu^2}{4\alpha^2}\delta. \tag{59}$$

As mentioned earlier, it is observed from simulations that the locally best short time fit to the actual diffusivity of the mid-point is $\sim\sqrt{1/\epsilon}$ times the estimate (59), and the long time diffusivity is $\sim\sqrt{\epsilon}$ times the classical estimate. While we do not yet know how to derive this relation analytically, the fact that early and late time renormalizations must have this qualitative relation can be argued from the conservation law (58).

S_∞ is the area enclosed between the actual density and the asymptotic value. Increases in $1/\epsilon$ (descaled market order rate) deplete orders near the spread, diminishing the mean depth at small $\hat{p}$, and induce the upward curvature seen in figure 3, and even more strongly in figure 28 below. As noted above, they cause more frequent shifts (more than compensating for the slight decrease in average step size), and increase the classically descaled diffusivity β. However, as a result, this increases the fraction of the area in S_∞ accumulated near the spread, requiring that the mean depth at larger $\hat{p}$ increase to compensate (see figure 28). The resulting steeper approach to the asymptotic depth at prices greater than the mean spread, and the larger negative curvature of the distribution, are fitted by an effective diffusivity that *decreases* with increasing $1/\epsilon$. Since the distribution further from the mid-point represents the imprint of market order activity further in the past, this effective diffusivity describes the long term evolution of the distribution. The resulting anticorrelation of the small $\hat{p}$ and large $\hat{p}$ effective diffusion constants implied by conservation of the area S_∞ is exactly consistent with their respective $\sim\sqrt{1/\epsilon}$ and $\sim\sqrt{\epsilon}$ scalings. The general idea here is to connect diffusivities at short and long timescales to the depth profile near the spread and far away from the spread respectively. The conservation law for the depth profile then implies a connection between these two diffusivities.

3.7.2. Direct simulation in interval coordinates.

The set of equations determined by the general form (53) is ultimately parametrized by the single input $\hat{x}_0$. The correct value for $\hat{x}_0$ is determined when the $\hat{x}_k$ are solved for recursively, by requiring convergence to $\hat{x}_\infty$. We do this recursion numerically, in the same way as we did to solve the differential equation for the normalized mean density $\psi(\hat{p})$.

In figure 24 we compare the numerical result for $\hat{x}_0$ with the analytical estimate generated as explained above. The results are surprisingly good throughout the entire range. Although the theoretical value consistently underestimates the numerical value, the functional form is captured accurately.

In figure 25, the values of x_k for all k are compared to the values determined directly from simulations.

Figure 26 shows the same data on a semilog scale for $\hat{x}_k/\hat{x}_\infty - 1$, showing the exponential decay at large argument characteristic of a simple diffusion solution. The IIA is clearly a good approximation for large ϵ. However, for small ϵ it starts deviating significantly from the simulations, especially for large k.

The values of x_k computed from the IIA can be very directly used to get an estimate of the price impact. The price impact, as defined in earlier sections, can be thought of as the change in the position of the mid-point (or the bid), consecutive to a certain number of orders being filled. Within the framework of the simplified model that we study here, this is simply the quantity $\langle \Delta m\rangle = (1/2)\sum_{k'=1}^{k} x_{k'}$, for k orders. The factor of 1/2 comes from considering the change in the position of the mid-point and not the bid. Figure 27 shows $\langle \Delta m\rangle$ non-dimensionalized by p_c plotted as a function of the number of orders (multiplied by ϵ), for three different values

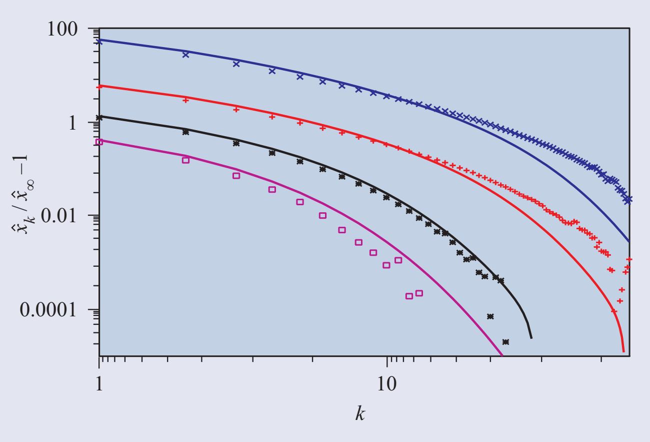

Figure 25. Four pairs of curves for the quantity $\hat{x}_k/\hat{x}_\infty - 1$ versus k. The value of ϵ increases from top to bottom ($\epsilon = 0.02, 0.04, 0.2, 0.66$). In each pair of curves, the points are obtained from simulations while the solid curve is the prediction of equation (53) evaluated numerically. The difference between the numerics and mean field results increases as ϵ decreases, especially for large k.

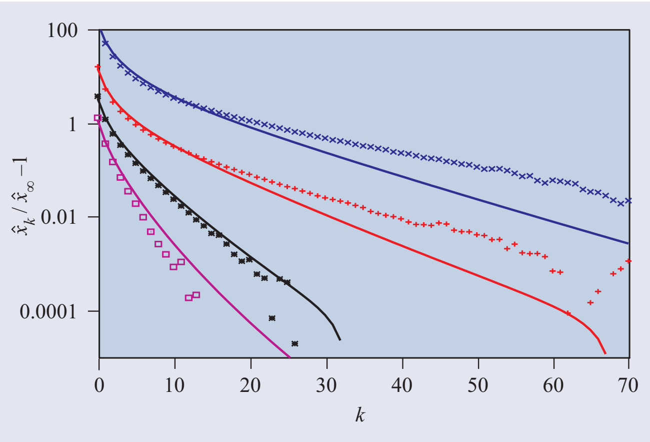

Figure 26. The same plot as in figure 25 but on a semi-log scale to show the exponential decay at large k.

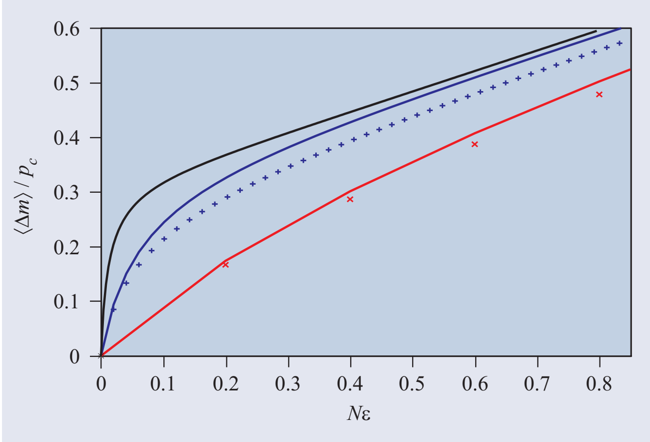

Figure 27. Three pairs of curves for the quantity $\langle \Delta m \rangle / p_c$ versus $N\epsilon$ where $\langle \Delta m \rangle = (1/2) \sum_{k=1}^{N} x_k$. The value of ϵ increases from top to bottom ($\epsilon = 0.002, 0.02, 0.2$). In each pair of curves, the points are obtained from simulations while the solid curve is the prediction of the IIA. For $\epsilon = 0.002$, we show only the theoretical prediction. The theory captures the functional form of the price impact curves for different ϵ. Quantitatively, it is better for larger epsilon, as remarked earlier.

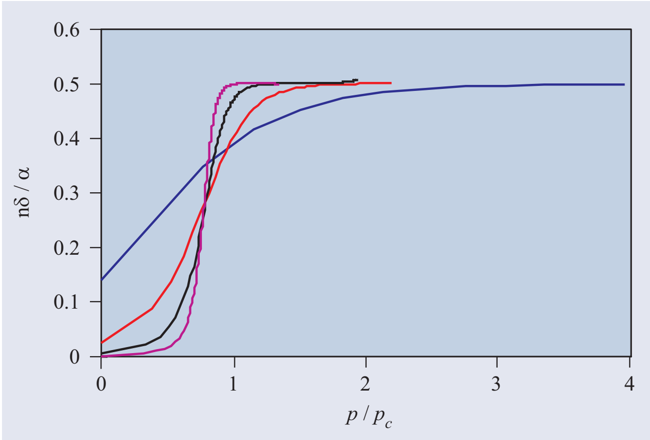

Figure 28. Density profiles for different values of ϵ ranging over the values 0.2, 0.02, 0.004, 0.001, obtained from the IIA.

of ϵ. Again, the theory matches quite well with the numerics, qualitatively. For large ϵ the agreement is quantitative as well.

The simplest approximation to the density profiles in the mid-point centred frame is to continue to approximate the mean density as $1/x_k$, but to regard that density as evaluated at position $x_0/2 + \sum_{k=1}^{i} x_k$. This clearly is not an adequate treatment in the range of the spread, both because the intervals are discrete, whereas mean ψ is continuous, and because the density profiles satisfy different global conservation laws associated with non-constancy of α. For large k however, this approximation might hold. The mean field values (only) corresponding to a plot of $\epsilon\psi(\hat{p})$ versus $\hat{p}$ are shown in figure 28. Here the theoretically estimated x_k at different parameter values are used to generate the depth profile using the procedure detailed above.

A comparison of the theoretically estimated profiles with the results from Monte Carlo simulations of the model is shown in figure 29. As is evident, the theoretical estimate for the density profile is better for large ϵ than small ϵ.

We can also generalize the above analysis to when the order placement process is no longer uniform. In particular it has been found that a power law order placement process is relevant [23, 27]. We carry out the above analysis for when $\alpha = \Delta_0^{\beta}/(\Delta + \Delta_0)^{\beta}$ where Δ is the distance from the current bid and Δ_0 determines the 'shoulder' of the power law. We find an interesting dependence of the existence of solutions on β. In particular we find that for $\beta > 1$, Δ_0 needs to be larger than some value (which depends on β as well as other parameters of the model such as μ and δ) for solutions of the IIA to exist. This might be interpreted as a market order wiping out the entire book, if the exponent is too large. When solutions exist, we find that the depth profile has a peak, consistent with the findings of [23]. In figure 30 the depth profiles for three different values of Δ_0 are plotted.

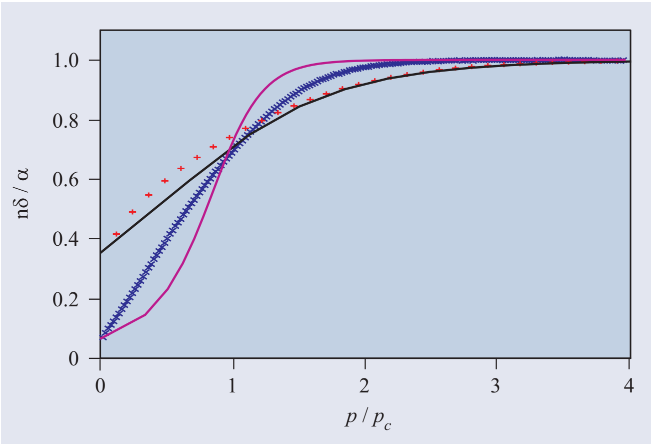

Figure 29. Density profiles from Monte Carlo simulation (points) and the IIA (curves). Plus signs and the black curve are for $\epsilon = 0.2$, while crosses and the purple curve are for $\epsilon = 0.02$.

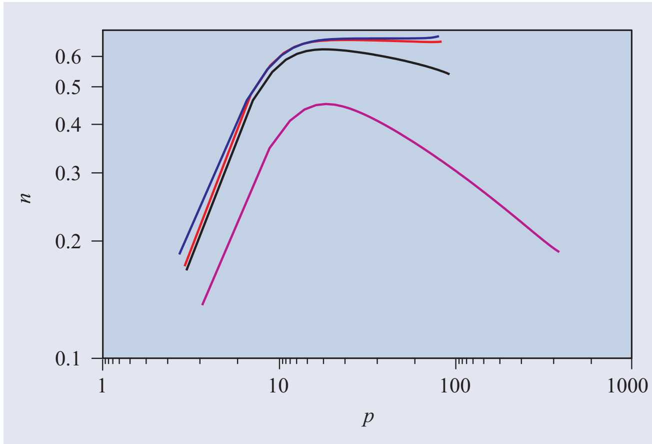

Figure 30. Density profiles for a power law order placement process for different values of Δ_0.

4. Concluding remarks

4.1. Conceptual shifts

One of the key properties of the model that we have discussed here is its robustness. In contrast, an earlier model due to Maslov [9] has the undesirable property that small changes in its structure causes qualitative changes in its properties. An important difference in our model, and that of Challet and Stinchcombe [11], is the addition of order cancellation[7]. While superficially only the introduction of an additional characteristic timescale, this is essential to make the properties of the model reasonable. Without evaporation, on average the arrival of market orders must exactly balance limit orders for any non-trivial or non-singular steady state to be possible. A systematic excess of market orders will eventually clear the book, while an excess of limit orders will eventually freeze the mid-price because of infinite order accumulation. The power laws that emerge in Maslov's model [9, 10] suggest that the unstable barrier between these two degenerate limits is a fine tuned critical point, and is not a robust phenomenon—a small change in parameters, such as a small alteration in the relative market and limit order rates, will destroy them. Thus, for this model to be reasonable one must also explain why a real market would tune itself to remain critical. In contrast, the properties of the model that we analyse here are robust under small variations in parameters, but do not explain the asymptotic power law properties of real data.

Real financial data often display a mixture of scale free and scale dependent behaviour. The probability of price fluctuations of a given size, for example, has a distribution with a finite standard deviation, defining a characteristic scale, at the same time that large fluctuations display power laws [29], and are therefore scale free. Because the model discussed here assumes that order flows are Poisson, it addresses only the finite scale behaviour, and does not address the power law behaviour. Preliminary indications suggest that this model is quite successful in capturing finite scale behaviour [25]. One of the challenges for models of this type is to explain both the scale dependent and the scale free properties at the same time. This will require better models of real order flow[8].

Another important and difficult future direction in modelling limit order books is to reconcile the empirically well-established concavity of the price impact function with uncorrelated price returns. A key lesson learned from the analytic treatment of this model is that the price impact function and the correlation of price returns are intimately related. That is, under Poisson order flow price impact functions are automatically concave (which is good because it qualitatively agrees with properties of real data), but this concavity also drives anticorrelated price returns (whereas real price returns are roughly uncorrelated). This is discussed in section 3.5. A key question is whether these properties can be reconciled using an unconditional random process model of order flow. We believe that this reconciliation may be possible using the right model of correlated order flow, but it is also possible that it will require a better understanding of strategic agent behaviour, such as order placement processes that are conditioned on time varying properties such as price volatility or the current state of the limit order book.

as a price statetgy -> reinfocment learning? will it change anytnig?

4.2. Successes and failures of the theoretical analysis

The theoretical analysis that we have presented here is quite successful in some respects, but there are still some open issues. We have presented two dual theoretical treatments, one a mean field theory cast in terms of the depth of the order book (using a master equation), and the other a mean field theory for the price interval separation between orders (in the limit of zero tick size). Depth and order separation are dual coordinates, and the two approaches have strengths and weaknesses that we review below. The theoretical analysis has clearly been useful

[7] Order cancellation was already present in the 1985 model of Cohen *et al* [3], which made use of results from queuing theory. The physics literature has added insight through the analogy to evaporation–deposition processes.

[8] Challet and Stinchcombe [11] claim that their model displays power price tails and clustered volatility. We have studied this carefully in our model, and while we have seen scaling behaviour across finite ranges that could be mistaken for power law behaviour, more careful investigation shows that this does not persist in the limit. We believe that robust asymptotic power law behaviour in models of this type must be driven by non-Poisson order flows.

The general form of a power law relating x and y is log y = b+a log x. If we raise the base of the logarithm (which doesn't actually matter), say e, to the values on both sides of this equation, we get y = ebea log x = ebxa. Since eb is just "some constant," let us replace it by constant c. Thus, a power law can be written as y = cxa for some constants a and c.

in showing which properties of the model can be explained simply through dimensional analysis, which can be explained by mean field theory, and which are intrinsically driven by fluctuations that are not captured by mean field theory.

Dimensional analysis already provides several predictions that roughly match the simulations. For example, it gives a rough estimate of the price diffusion rates. The prediction that the spread should scale as μ/α provides a surprisingly good estimate, both for the simulations reported here, and for real data [25]. This estimate is refined by the IIA, which predicts a correction in terms of a slowly varying dimensionless function $f(\epsilon)$. It does an excellent job of predicting the simulation results for $f(\epsilon)$ (see figure 24), and thus is very successful in explaining the average properties of the bid–ask spread. The unconditional master equation approach, in contrast, does a rather poor job, but the corrections that are needed to improve it are insightful in many other respects, as explained below.

One of the important contributions of the theory is in providing an understanding of the granularity parameter ϵ, which in general is a more important determinant of market properties than tick size. Granularity effects, due to finite order size, cause significant non-linearities in properties such as mean depth, market impact and price diffusion. The master equation treatment makes it clear why these non-linearities are particularly strong for small ϵ. In particular, it makes it clear that the correlated fluctuations between price changes and depth near the mid-point must be properly taken into account. Once this is done (see section 3.6.7), the master equation does a much better job of matching the simulation results, particularly for larger values of ϵ. This is important because it also gives qualitative insight into why ϵ is such an important parameter, and why granularity drives concave price impact, which drives anticorrelated price diffusion.

While we have provided a qualitative understanding for why price fluctuations are anticorrelated for Poisson order flow, we have failed to provide a quantitative understanding of this phenomenon, which appears to depend on fluctuations that are out of reach of mean field theory. In addition, we do not explain fluctuations in the mean depth of the order book as well as we would like to. Since depth fluctuations affect the market impact, as explained in appendix A, these problems produce a systematic offset in the predicted mean market impact. These are both interesting problems for future study.

4.3. Future enhancements

As we have mentioned above, the zero intelligence, IID order flow model should be regarded as just a starting point from which to add more complex behaviour. We are considering several enhancements to the order flow process whose effects we intend to discuss in future papers. The enhancements include:

- *Trending of order flow.* We have demonstrated that IID order flow necessarily leads to non-IID prices. The converse is also true: non-IID order flow is necessary for IID prices. In particular, the order flow must contain trends; i.e. if order flow has recently been skewed toward buying, it is more likely to continue to be skewed toward buying. If we assume perfect market efficiency, in the sense that prices are a random walk, this implies that there must be trends in order flow.
- *Power law placement of limit prices.* For both the London Stock Exchange and the Paris Bourse, the distribution of the limit price relative to the best bid or ask appears to decay as a power law [23, 27]. Our investigations of this show that this can have an important effect. Exponents larger than one result in order books with a finite number of orders. In this case, depending on other parameters, there is a finite probability that a single market order can clear the entire book (see section 3.7.2).
- *Power law or log-normal order size distribution.* Real order placement processes have order size distributions that appear to be roughly like a log-normal distribution with a power law tail [28]. This has important effects on the fluctuations in liquidity.
- *Non-Poisson order cancellation process.* When considered in real time, order placement cancellation does not appear to be Poisson [11], so better models are needed.
- *Conditional order placement.* Agents may conditionally place larger market orders when the book is deeper, causing the market impact function to grow more slowly. We intend to measure this effect and incorporate it into our model.
- *Feedback between order flow and prices.* In reality there are feedbacks between order flow and price movements beyond the feedback in the reference point for limit order placement built into this model. This can induce bursts of trading, causing order flow rates to speed up or slow down, and give rise to clustered volatility.

better order needed

The last item is just one of many examples of how one can surely improve the model by making order flow conditional on available information. However, we believe it is important to first gain an understanding of the properties of simple unconditional models, and then build on this foundation to get a fuller understanding of the problem.

4.4. Comparison to standard models based on valuation and information arrival

In the spirit of Becker [13] and Gode and Sunder [14], we assume a simple, zero intelligence model of agent behaviour and show that the market institution exerts considerable power in shaping the properties of prices. While not disputing that agent behaviour might be important, our model suggests that, at least on the short timescale, many of the properties of the market are dictated by the market institution, and in particular the need to store supply and demand. Our model is stochastic and fully dynamic, and makes predictions that go beyond the realm of experimental economics, giving quantitative predictions about fundamental properties of a real market. We have developed what were previously conceptual toy models in the physics literature into a model with testable explanatory power.

This raises questions about the comparison to standard models based on the response of valuations to news. The idea that news might drive changes in order flow rates is compatible with our model. That is, news can drive changes in order flow, which in turn cause the best bid or ask price to change. But notice that in our model there are no assumptions about valuations. Instead, everything depends on order flow rates. For example, the diffusion rate of prices increases as the $5/2$ power of market order flow rate, and thus volatility, which depends on the square root of the diffusion rate, increases as the $5/4$ power. Of course, order flow rates can respond to information; an increase in market order rate indicates added impatience, which might be driven by changes in valuation. But changes in long term valuation could equally well cause an increase in limit order flow rate, which *decreases* volatility. Valuation *per se* does not determine whether volatility will increase or decrease. Our model says that volatility does not depend directly on valuations, but rather on the urgency with which they are felt, and the need for immediacy in responding to them.

Understanding the shape of the price impact function was one of the motivations that originally set this project into motion. The price impact function is closely related to supply and demand functions, which have been central aspects of economic theory since the 19th century. Our model suggests that the shape of price impact functions in modern markets is significantly influenced not so much by strategic thinking as by an economic fundamental: the need to store supply and demand in order to provide liquidity. *A priori* it is surprising that this requirement alone may be sufficient to dictate at least the broad outlines of the price impact curve.

Our model offers a 'divide and conquer' strategy for understanding fundamental problems in economics. Rather than trying to ground our approach directly on assumptions of utility, we break the problem into two parts. We provide an understanding of how the statistical properties of prices respond to order flow rates, and leave the problem open of how order flow rates depend on more fundamental assumptions about information and utility. Order flow rates have the significant advantage that, unlike information, utility, or the cognitive powers of an agent, they are directly measurable. We hope that by breaking the problem into two pieces, and partially solving the second piece, we can ultimately help provide a deeper understanding of how markets work.

Acknowledgments

We would like to thank the McKinsey Corporation, Credit Suisse First Boston, the McDonnel Foundation, Bob Maxfield and Bill Miller for supporting this research. We would also like to thank Paolo Patelli, R Rajesh, Spyros Skouras and Ilija Zovko for helpful discussions, and Marcus Daniels for valuable technical support.

Appendix A. Relationship of price impact to cumulative depth

An important aspect of markets is the immediate liquidity, by which we mean the immediate response of prices to incoming market orders. When a market order enters, its execution range depends both on the spread and on the depth of the orders in the book. These determine the sequence of transaction prices produced by that order, as well as the instantaneous market impact. Long term liquidity depends on the longer term response of the limit order book, and is characterized by the price impact function $\phi(\omega, \tau)$ for values of $\tau > 0$. Immediate liquidity affects short term volatility, and long term liquidity affects volatility measured over longer timescales. In this section we address only short term liquidity. We address volatility on longer timescales in section 2.2.4.

We characterize liquidity in terms of either the depth profile or the price impact. The *depth profile* $n(p, t)$ is the number of shares n at price p at time t. For many purposes it is convenient to think in terms of the *cumulative depth profile* N, which is the sum of n values up (or down) to some price. For convenience we establish a reference point at the centre of the book where we define $p \equiv 0$ and $N(0) \equiv 0$. The reference point can be either the mid-point quote, or the best bid or ask. We also study the *price impact function* $\Delta p = \phi(\omega, \tau, t)$, where Δp is the shift in price at time $t + \tau$ caused by an order of size ω placed at time t. Typically we define Δp as the shift in the mid-point price, though it is also possible to use the best bid or ask (equation (1)).

The price impact function and the depth profile are closely related, but the relationship is not trivial. $N(\Delta p)$ gives us the average total number of orders up to a distance Δp away from the origin, whereas, in order to calculate the price impact, what we need is the average shift Δp caused by a *fixed* number of orders. Making the identifications $p = \Delta p$, $N = \omega$, and choosing a common reference point, the instantaneous price impact is the inverse of the instantaneous cumulative depth, i.e. $\phi(\omega, 0, t) = N^{-1}(\omega, t)$. This relationship is clearly true instant by instant. However, it is not true for averages, since the mean of the inverse is not in general equal to the inverse of the means; i.e. $\langle \phi \rangle \neq \langle N \rangle^{-1}$. This is highly relevant here, since because the fluctuations in these functions are huge, our interest is primarily in their statistical properties, and in particular the first few moments.

A relationship between the moments can be derived as follows.

A.1. Moment expansion

There is some subtlety in how we relate the market impact to the cumulative order count $N(p, t)$. One eligible definition of market impact Δp is the movement of the mid-point, following the placement of an order of size ω. If we define the reference point so that $N(a, t) \equiv 0$, and the market order is a buy, this definition puts $\omega(\Delta p, t) = N(a + 2\Delta p, t) - N(a, t)$. In words, the mid-point shift is half the shift in the best offer. An alternative choice would be to let $\omega(\Delta p, t) = N(\Delta p, t)$, which would include part of the instantaneous spread in the definition of impact in mid-point centred coordinates, or none of it in ask centred coordinates. The issue of how impact is related to $N(p, t)$ is separate from whether the best ask is set equal to the reference point for prices, and may be chosen differently to answer different questions.

Under any such definition, however, the impact Δp is a monotonic function of ω in every instance, so either may be taken as the independent variable, along with the index t that labels the instance. We wish to account for the differences in instance averages $\langle\,\rangle$ of ω and Δp, regarded respectively as the dependent variables, in terms of the fluctuations of either.

In spite of the fact that the density $n(p,t)$ is a highly discontinuous variable in general, monotonicity of the cumulative $N(p,t)$ enables us to picture a power series expansion for $\omega(p,t)$ in p, with coefficients that fluctuate in time. The simplest such expansion that captures much of the behaviour of the simulated output is

$$\omega(p,t) = a(t) + b(t)p + \frac{c(t)}{2}p^2, \tag{A.1}$$

if p is regarded as the independent variable, or

$$p(\omega,t) = \frac{-b(t) + \sqrt{b^2(t) + 2c(t)(\omega - a(t))}}{c(t)}, \tag{A.2}$$

if ω is. While the variable $a(t)$ would seem unnecessary since ω is zero at $p = 0$, empirically we find that simultaneous fits to both ω and ω^2 at low order can be made better by incorporating the additional freedom of fluctuations in a.

We imagine splitting each t dependent coefficient into its mean, and a zero mean fluctuation component, as

$$a(t) \equiv \bar{a} + \delta a(t), \tag{A.3}$$

$$b(t) \equiv \bar{b} + \delta b(t), \tag{A.4}$$

and

$$c(t) \equiv \bar{c} + \delta c(t). \tag{A.5}$$

The fluctuation components will in general depend on ϵ. The values of the mean and second moment of the fluctuations can be extracted from the mean distributions $\langle\omega\rangle$ and $\langle\omega^2\rangle$. The mean values come from the linear expectation:

$$\langle\omega(0)\rangle = \bar{a}, \tag{A.6}$$

$$\left.\frac{\partial\langle\omega(p)\rangle}{\partial p}\right|_{p=0} = \bar{b}, \tag{A.7}$$

and

$$\left.\frac{\partial^2\langle\omega(p)\rangle}{\partial p^2}\right|_{p=0} = \bar{c}. \tag{A.8}$$

Given these, the fluctuations then come from the quadratic expectation as

$$\langle\omega^2(0)\rangle = \bar{a}^2 + \langle\delta a^2\rangle, \tag{A.9}$$

$$\left.\frac{\partial\langle\omega^2(p)\rangle}{\partial p}\right|_{p=0} = 2\bar{a}\bar{b} + 2\langle\delta a\,\delta b\rangle, \tag{A.10}$$

$$\left.\frac{\partial^2\langle\omega^2(p)\rangle}{\partial p^2}\right|_{p=0} = 2(\bar{b}^2 + \bar{a}\bar{c}) + 2\langle\delta b^2 + \delta a\,\delta c\rangle, \tag{A.11}$$

$$\left.\frac{\partial^3\langle\omega^2(p)\rangle}{\partial p^3}\right|_{p=0} = 6(\bar{b}\bar{c} + \langle\delta b\,\delta c\rangle), \tag{A.12}$$

and

$$\left.\frac{\partial^4\langle\omega^2(p)\rangle}{\partial p^4}\right|_{p=0} = 6(\bar{c}^2 + \langle\delta c^2\rangle). \tag{A.13}$$

When ω is given a specific definition in terms of the cumulative distribution, its averages become averages over the density in the order book.

The values of the moments as obtained above may then be used in a derivative expansion of the inverse function (A.2), making the prediction for the averaged impact

$$\begin{aligned}\langle p(\omega)\rangle = \bar{p} &+ \frac{1}{2}\overline{\frac{\partial^2 p}{\partial a^2}}\langle\delta a^2\rangle + \frac{1}{2}\overline{\frac{\partial^2 p}{\partial b^2}}\langle\delta b^2\rangle + \frac{1}{2}\overline{\frac{\partial^2 p}{\partial c^2}}\langle\delta c^2\rangle \\ &+ \overline{\frac{\partial^2 p}{\partial a\,\partial b}}\langle\delta a\,\delta b\rangle + \overline{\frac{\partial^2 p}{\partial a\,\partial c}}\langle\delta a\,\delta c\rangle + \overline{\frac{\partial^2 p}{\partial b\,\partial c}}\langle\delta b\,\delta c\rangle,\end{aligned} \tag{A.14}$$

where an overbar denotes the evaluation of the function (A.2) or its indicated derivative at $b(t) = \bar{b}$, $c(t) = \bar{c}$ and ω. The fluctuations $\langle\delta b^2\rangle$ and $\langle\delta a\,\delta c\rangle$ cannot be determined independently from equation (A.11). However, in keeping with this fact, their coefficient functions in equation (A.14) are identical, so the inversion remains fully specified.

If we denote by $\bar{Z}$ the radical

$$\bar{Z} \equiv \sqrt{\bar{b}^2 + 2\bar{c}(\omega - \bar{a})}, \tag{A.15}$$

the various partial derivative functions in equation (A.14) evaluate to

$$\frac{1}{2}\overline{\frac{\partial^2 p}{\partial a^2}} = \frac{-\bar{c}}{2\bar{Z}^3}, \tag{A.16}$$

$$\overline{\frac{\partial^2 p}{\partial a\,\partial b}} = \frac{\bar{b}}{\bar{Z}^3}, \tag{A.17}$$

$$\frac{1}{2}\overline{\frac{\partial^2 p}{\partial b^2}} = \overline{\frac{\partial^2 p}{\partial a\,\partial c}} = \frac{\omega - \bar{a}}{\bar{Z}^3}, \tag{A.18}$$

$$\overline{\frac{\partial^2 p}{\partial b\,\partial c}} = \frac{1}{\bar{c}^2} - \frac{\bar{b}}{\bar{Z}\bar{c}^2} - \frac{(\omega - \bar{a})\bar{b}}{\bar{c}\bar{Z}^3}, \tag{A.19}$$

and

$$\frac{1}{2}\overline{\frac{\partial^2 p}{\partial c^2}} = \frac{\bar{Z} - \bar{b}}{\bar{c}^3} - \frac{\omega - \bar{a}}{2\bar{c}^2\bar{Z}} - \frac{(\omega - \bar{a})^2}{2\bar{c}\bar{Z}^3}. \tag{A.20}$$

Plugging these into equation (A.14) gives the predicted mean price impact, compared to the actual mean in figure A.1. Here the measure used for price impact is the movement of the ask from buy market orders. The cumulative order distribution is computed in ask centred coordinates, eliminating the contribution from the half-spread in the p coordinate. The inverse of the mean cumulative distribution (black), which corresponds to $\bar{p}$ in equation (A.14), clearly underestimates the actual mean impact (blue). However, the corrections from only second order fluctuations in a, b and c account for much of the difference at all values of ϵ.

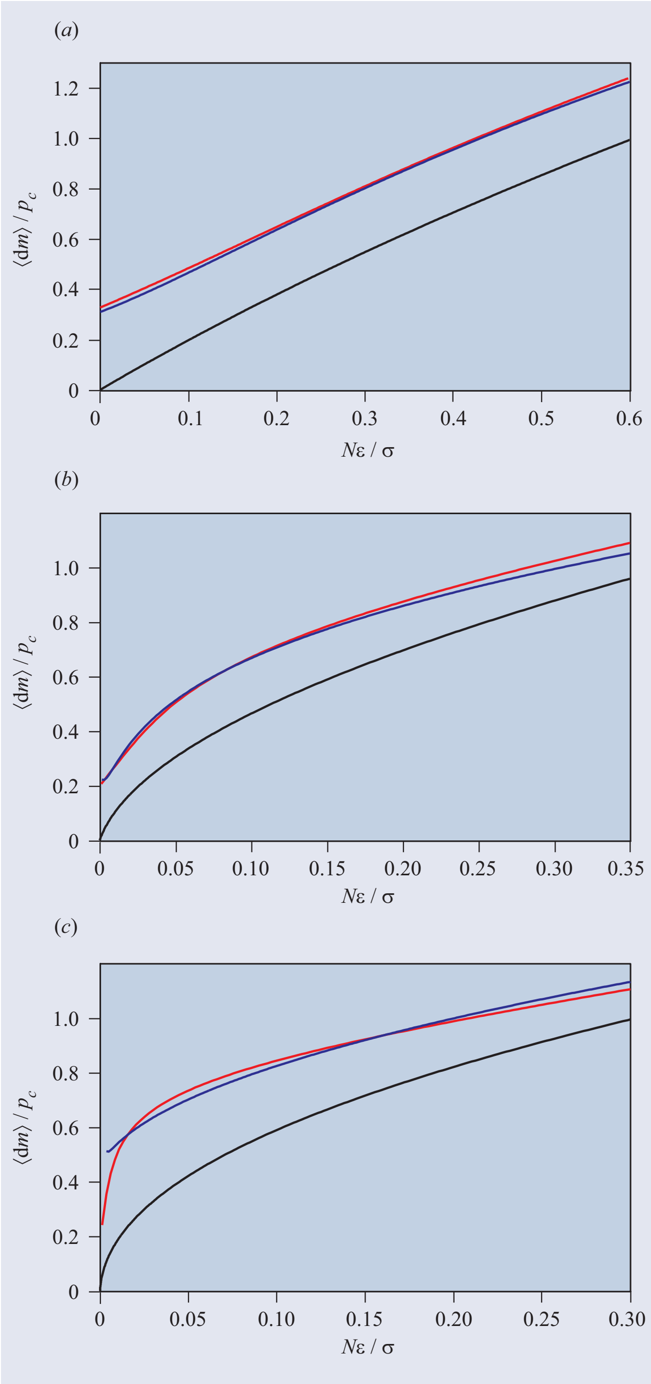

Figure A.1. Comparison of the inverse mean cumulative order distribution $\bar{p}$ (black), to the actual mean impact (blue), and the second order fluctuation expansion ((A.14), red). (*a*) $\epsilon = 0.2$. (*b*) $\epsilon = 0.02$. (*c*) $\epsilon = 0.002$.

A.2. Quantiles

Another way to characterize the relationship between depth profile and market impact is in terms of their quantiles (the fraction greater than a given value; for example the median is the 0.5 quantile). Interestingly, the relationship between quantiles is trivial. Letting $Q_r(x)$ be the rth quantile of x, because the cumulative depth $N(p)$ is a non-decreasing function with inverse $p = \phi(N)$, we have the relation

$$Q_r(\phi) = (Q_{1-r}(N))^{-1}. \tag{A.21}$$

This provides an easy and accurate way to compare depth and price impact when the tick size is sufficiently small. However, when the tick size is very coarse, the quantiles are in general not very useful, because unlike the mean, the quantiles do not vary continuously, and only take on a few discrete values.

As we have argued in the previous section, in non-dimensional coordinates all of the properties of the limit order book are described by the two dimensionless scale factors ϵ and $\mathrm{d}p/p_c$ (see table 2). When expressed in dimensionless coordinates, any property, such as depth, spread or price impact, can only depend on these two parameters. This reduces the search space from five dimensions to two, which greatly simplifies the analysis. Any results can easily be re-expressed in dimensional coordinates from the definitions of the dimensionless parameters.

Appendix B. Supporting calculations in density coordinates

The following two subsections provide details for the master equation solution in density coordinates. The first provides the generating functional solution for the density at general $\mathrm{d}p$, and the second the approximate source term for correlated fluctuations.

B.1. Generating functional at general bin width

As in the main text, α and μ represent the functions of p everywhere in this subsection, because the boundary values do not propagate globally. Equation (24) can be solved by assuming there is a convergent expansion in (formal) small D/δ,

$$\Pi \equiv \sum_j \left(\frac{D}{\delta}\right)^j \Pi_j, \tag{B.1}$$

and it is convenient to embellish the shorthand notation as well, with

$$\Pi_j(0, p) \equiv \pi_{0j}(p). \tag{B.2}$$

It follows that the expected number also expands as

$$\langle n \rangle \equiv \sum_j \left(\frac{D}{\delta}\right)^j \langle n \rangle_j. \tag{B.3}$$

Order by order in D/δ, equation (24) requires

$$\left(\frac{\partial}{\partial \lambda} - \frac{\alpha\,\mathrm{d}p - \mu/2\lambda}{\delta\sigma}\right)\Pi_j = \frac{\mu}{2\delta\sigma\lambda}\pi_{0j} + \frac{\partial^2}{\partial p^2}\frac{\Pi_{j-1}}{(\lambda - 1)}. \tag{B.4}$$

Because the Π_j have been introduced in order to be chosen homogeneous of degree zero in D/δ, the normalization condition requires that

$$\Pi_0(1, p) = 1, \qquad \Pi_{j\neq 0}(1, p) = 0, \qquad \forall p. \tag{B.5}$$

The implied recursion relations for expected occupation numbers are

$$\langle n\rangle_0 = \frac{\alpha\,\mathrm{d}p}{\delta} - \frac{\mu}{2\delta}(1-\pi_{00}), \tag{B.6}$$

at $j = 0$, and

$$\langle n\rangle_j = \frac{\mu}{2\delta}\pi_{0j} + \frac{\partial^2}{\partial p^2}\langle n\rangle_{j-1} \tag{B.7}$$

otherwise.

Equation (B.4) is solved immediately by use of an integrating factor, to give the recursive integral relation

$$\Pi_j(\lambda) = \pi_{0j}\left[1 + \frac{\alpha\,\mathrm{d}p}{\delta\sigma}\mathcal{I}(\lambda)\right] + \mathcal{I}(\lambda)\left\langle\!\!\left\langle \frac{\partial^2}{\partial p^2}\frac{\Pi_{j-1}}{\lambda-1}\right\rangle\!\!\right\rangle_\lambda, \tag{B.8}$$

where

$$\mathcal{I}(\lambda) \equiv \lambda\int_0^1 \mathrm{d}z\,\mathrm{e}^{(\alpha\,\mathrm{d}p/\delta\sigma)(1-z)} z^{(\mu/2\delta\sigma)}, \tag{B.9}$$

and

$$\left\langle\!\!\left\langle \frac{\partial^2}{\partial p^2}\frac{\Pi_{j-1}}{\lambda-1}\right\rangle\!\!\right\rangle_\lambda \equiv \frac{\lambda}{\mathcal{I}(\lambda)}\int_0^1 \mathrm{d}z\,\mathrm{e}^{(\alpha\,\mathrm{d}p/\delta\sigma)(1-z)} z^{(\mu/2\delta\sigma)}\,\frac{\partial^2}{\partial p^2}\frac{\Pi_{j-1}(\lambda z)}{\lambda z-1}. \tag{B.10}$$

The surface condition (B.5) provides the starting point for this recursion, by giving at $j = 0$

$$\pi_{00} = \frac{1}{1 + (\alpha\,\mathrm{d}p/\delta\sigma)\mathcal{I}(1)}. \tag{B.11}$$

Given forms for α and μ, equation (B.6) may be solved directly from equation (B.11), and extended by equation (B.7) to solve for $\langle n(p)\rangle$. More generally, equations (B.9), (B.10) and (B.11) may be solved to any desired order numerically, to obtain the fluctuation characteristics of $n(p)$. Finding the solution becomes difficult, however, when α and μ must be related self-consistently to the solutions for Π. The special case $\mathrm{d}p \to 0$ admits a drastic simplification, in which the whole expansion for $\langle n(p)\rangle$ may be directly summed, to recover the result in the main text. In this limit, one gets a single differential equation in p which is solvable by numerical integration. The existence and regularity of this solution demonstrates the existence of a continuum limit on the price space, and can be simulated directly by allowing orders to be placed at arbitrary real valued prices.

B.1.1. Recovering the continuum limit for prices.

In the limit that the dimensionless quantity $\alpha\,\mathrm{d}p/\delta\sigma \to 0$, equation (B.9) simplifies to

$$\mathcal{I}(\lambda) \to \frac{\lambda}{1+\mu/2\delta\sigma} + \mathcal{O}(\mathrm{d}p), \tag{B.12}$$

from which it follows that

$$\langle n\rangle_0 \to \frac{\alpha\,\mathrm{d}p/\delta}{1+\mu/2\delta\sigma} + \mathcal{O}(\mathrm{d}p^2). \tag{B.13}$$

The important simplification given by vanishing $\mathrm{d}p$, as will be seen below, is that the expansion (19) collapses, at leading order in $\mathrm{d}p$, to

$$\Pi_0(\lambda) \to 1 + (\lambda-1)\frac{\langle n\rangle_0}{\sigma} + \mathcal{O}(\mathrm{d}p^2). \tag{B.14}$$

Equation (B.14) is used as the input to an inductive hypothesis

$$\Pi_{j-1}(\lambda) \to \Pi_{j-1}(1) + (\lambda-1)\frac{\langle n\rangle_{j-1}}{\sigma} + \mathcal{O}(\mathrm{d}p^2), \tag{B.15}$$

(note that $\langle n\rangle_{j-1} \sim \mathcal{O}(\mathrm{d}p)$, $\Pi_{j-1}(1) =$ either 1 or 0), which, with equation (B.10), then recovers the condition at j:

$$\Pi_j(\lambda) \to (\lambda-1)\mathcal{I}(1)\frac{\mathrm{d}^2}{\mathrm{d}p^2}\frac{\langle n\rangle_{j-1}}{\sigma} + \mathcal{O}(\mathrm{d}p^2). \tag{B.16}$$

Using equation (19) at $\lambda \to 1$, and equation (B.12) for $\mathcal{I}$, gives the recursion for the number density

$$\langle n\rangle_{j\neq 0} \to \frac{1}{1+\mu/2\delta\sigma}\frac{\mathrm{d}^2}{\mathrm{d}p^2}\langle n\rangle_{j-1} + \mathcal{O}(\mathrm{d}p^2). \tag{B.17}$$

The sum (B.3) for $\langle n\rangle$ is then

$$\langle n\rangle = \sum_j\left(\frac{D}{\delta}\frac{1}{1+\mu/2\delta\sigma}\frac{\mathrm{d}^2}{\mathrm{d}p^2}\right)^j\langle n\rangle_0. \tag{B.18}$$

Using equation (B.13) for $\langle n\rangle_0$ and rearranging terms, equation (B.18) is equivalent to

$$\langle n\rangle = \frac{1}{1+\mu/2\delta\sigma}\sum_j\left(\frac{D}{\delta}\frac{\mathrm{d}^2}{\mathrm{d}p^2}\frac{1}{1+\mu/2\delta\sigma}\right)^j\frac{\alpha\,\mathrm{d}p}{\delta}. \tag{B.19}$$

The series expansion in the price Laplacian is formally the geometric sum

$$(1+\mu/2\delta\sigma)\langle n\rangle = \left[1 - \frac{D}{\delta}\frac{\mathrm{d}^2}{\mathrm{d}p^2}\frac{1}{1+\mu/2\delta\sigma}\right]^{-1}\frac{\alpha\,\mathrm{d}p}{\delta}, \tag{B.20}$$

which can be inverted to give equation (27), a relation that is local in derivatives.

B.2. Cataloguing correlations

A correct source term $\mathcal{S}$ must correlate the incidences of zero occupation with the events producing shifts. It is convenient to separate these into the four independent types of deposition and removal.

First we consider removal of buy limit orders, which generates a negative shift of the mid-point. Let $\hat{a}'$ denote the position of the ask after the shift. Then all possible shifts $\Delta\hat{p}$ are related to a given price bin $\hat{p}$ and $\hat{a}'$ in one of three ordering cases, shown in table B.1. For each case, the source term corresponding to $[\psi(\hat{p}-\Delta\hat{p}) - \psi(\hat{p})]$ in equation (43) is given, together with the measure of order book configurations for which that case occurs. The mean field assumption (36) is used to estimate these measures.

As argued when defining β in the simpler diffusion approximation for the source terms, the measure of shifts from

Table B.1. Contributions to 'effective P_-' from removal of a buy limit order, conditioned on the position of the ask relative to p.

Case	Source	Probability
$\Delta\hat{p} \leqslant \hat{a}' < \hat{p}$	$\psi(\hat{p}-\Delta\hat{p})-\psi(\hat{p})$	$\varphi(\Delta\hat{p})-\varphi(\hat{p})$
$\Delta\hat{p} \leqslant \hat{p} < \hat{a}' \leqslant \hat{p}+\Delta\hat{p}$	$0-\psi(\hat{p})$	$\varphi(\hat{p})-\varphi(\hat{p}+\Delta\hat{p})$
$\hat{p} < \Delta\hat{p} < \hat{a}' \leqslant \hat{p}+\Delta\hat{p}$	$0-\psi(\hat{p})$	$\varphi(\Delta\hat{p})-\varphi(\hat{p}+\Delta\hat{p})$

Table B.2. Contributions to 'effective P_+' from removal of a sell limit order, conditioned on the position of the ask relative to p.

Case	Source	Probability
$\Delta\hat{p} \leqslant \hat{a}' < \hat{p}$	$\psi(\hat{p}+\Delta\hat{p})-\psi(\hat{p})$	$\varphi(\Delta\hat{p})-\varphi(\hat{p})$
$\Delta\hat{p} \leqslant \hat{p} < \hat{a}' \leqslant \hat{p}+\Delta\hat{p}$	$\psi(\hat{p}+\Delta\hat{p})-0$	$\varphi(\hat{p})-\varphi(\hat{p}+\Delta\hat{p})$
$\hat{p} < \Delta\hat{p} < \hat{a}' \leqslant \hat{p}+\Delta\hat{p}$	$\psi(\hat{p}+\Delta\hat{p})-0$	$\varphi(\Delta\hat{p})-\varphi(\hat{p}+\Delta\hat{p})$

Table B.3. Contributions to 'effective P_-' from addition of a sell limit order, conditioned on the position of the ask relative to p.

Case	Source	Probability
$\Delta\hat{p} \leqslant \hat{a} < \hat{p}-\Delta\hat{p}$	$\psi(\hat{p}-\Delta\hat{p})-\psi(\hat{p})$	$\varphi(\Delta\hat{p})-\varphi(\hat{p}-\Delta\hat{p})$
$\Delta\hat{p} \leqslant \hat{p}-\Delta\hat{p} < \hat{a} \leqslant \hat{p}$	$0-\psi(\hat{p})$	$\varphi(\hat{p}-\Delta\hat{p})-\varphi(\hat{p})$
$\hat{p}-\Delta\hat{p} < \Delta\hat{p} < \hat{a} \leqslant \hat{p}$	$0-\psi(\hat{p})$	$\varphi(\Delta\hat{p})-\varphi(\hat{p})$

Table B.4. Contributions to 'effective P_+' from addition of a buy limit order, conditioned on the position of the ask relative to p.

Case	Source	Probability
$\Delta\hat{p} \leqslant \hat{a}' < \hat{p}$	$\psi(\hat{p}+\Delta\hat{p})-\psi(\hat{p})$	$\varphi(\Delta\hat{p})-\varphi(\hat{p})$
$\Delta\hat{p} \leqslant \hat{p} < \hat{a}' \leqslant \hat{p}+\Delta\hat{p}$	$\psi(\hat{p}+\Delta\hat{p})-0$	$\varphi(\hat{p})-\varphi(\hat{p}+\Delta\hat{p})$
$\hat{p} < \Delta\hat{p} < \hat{a}' \leqslant \hat{p}+\Delta\hat{p}$	$\psi(\hat{p}+\Delta\hat{p})-0$	$\varphi(\Delta\hat{p})-\varphi(\hat{p}+\Delta\hat{p})$

removal of either buy or sell limit orders should be symmetric with that of their addition within the spread, which is $2\,\mathrm{d}\Delta\hat{p}$ for either type, in cases when the shift $\pm\Delta\hat{p}$ is consistent with the value of the spread. The only change in these more detailed source terms is replacement of the simple $\Pr(a \geqslant \Delta p)$ with the entries in the third column of table B.1. When the $\Delta\hat{p}$ cases are integrated over their range as specified in the first column and summed, the result is a contribution to $\mathcal{S}$ of

$$\int_0^{\hat{p}} 2\,\mathrm{d}\Delta\hat{p}\,\psi(\hat{p}-\Delta\hat{p})[\varphi(\Delta\hat{p})-\varphi(\hat{p})] - \int_0^{\infty} 2\,\mathrm{d}\Delta\hat{p}\,\psi(\hat{p})[\varphi(\Delta\hat{p})-\varphi(\hat{p}+\Delta\hat{p})]. \quad \text{(B.21)}$$

Sell limit order removals generate another sequence of cases, symmetric with the buys, but inducing positive shifts. The cases, source terms and frequencies are given in table B.2. Their contribution to $\mathcal{S}$, after integration over $\Delta\hat{p}$, is then

$$\int_0^{\infty} 2\,\mathrm{d}\Delta\hat{p}\,\psi(\hat{p}+\Delta\hat{p})[\varphi(\Delta\hat{p})-\varphi(\hat{p}+\Delta\hat{p})] - \int_0^{\hat{p}} 2\,\mathrm{d}\Delta\hat{p}\,\psi(\hat{p})[\varphi(\Delta\hat{p})-\varphi(\hat{p})]. \quad \text{(B.22)}$$

Order addition is treated similarly, except that $\hat{a}$ denotes the position of the ask before the event. Sell limit order additions generate negative shifts, with the cases shown in table B.3. Integration over $\Delta\hat{p}$ consistent with these cases gives the negative shift contribution to $\mathcal{S}$:

$$\int_0^{\hat{p}/2} 2\,\mathrm{d}\Delta\hat{p}\,\psi(\hat{p}+\Delta\hat{p})[\varphi(\Delta\hat{p})-\varphi(\hat{p}-\Delta\hat{p})] - \int_0^{\hat{p}} 2\,\mathrm{d}\Delta\hat{p}\,\psi(\hat{p})[\varphi(\Delta\hat{p})-\varphi(\hat{p}+\Delta\hat{p})]. \quad \text{(B.23)}$$

The corresponding buy limit order addition cases are given in table B.4, and their positive shift contribution to $\mathcal{S}$ turns out to be the same as that from removal of sell limit orders (B.22).

Writing the source as a sum of two terms $\mathcal{S} \equiv \mathcal{S}_{\text{buy}} + \mathcal{S}_{\text{sell}}$, the combined contribution from buy limit order additions and removals is

$$\mathcal{S}_{\text{buy}}(\hat{p}) = \int_0^{\hat{p}} 2\,\mathrm{d}\Delta\hat{p}\,[\psi(\hat{p}-\Delta\hat{p})-\psi(\hat{p})][\varphi(\Delta\hat{p})-\varphi(\hat{p})] - \int_0^{\infty} 2\,\mathrm{d}\Delta\hat{p}\,[\psi(\hat{p}+\Delta\hat{p})-\psi(\hat{p})] \times [\varphi(\Delta\hat{p})-\varphi(\hat{p}+\Delta\hat{p})]. \quad \text{(B.24)}$$

The corresponding source term from sell order addition and removal is

$$\mathcal{S}_{\text{sell}}(\hat{p}) = \int_0^{\hat{p}/2} 2\,\mathrm{d}\Delta\hat{p}\,\psi(\hat{p}-\Delta\hat{p})[\varphi(\Delta\hat{p})-\varphi(\hat{p}-\Delta\hat{p})] - 2\int_0^{\hat{p}} 2\,\mathrm{d}\Delta\hat{p}\,\psi(\hat{p})[\varphi(\Delta\hat{p})-\varphi(\hat{p})] + \int_0^{\infty} 2\,\mathrm{d}\Delta\hat{p}\,\psi(\hat{p}+\Delta\hat{p})[\varphi(\Delta\hat{p})-\varphi(\hat{p}+\Delta\hat{p})]. \quad \text{(B.25)}$$

The forms (B.24) and (B.25) do not lead to $\int \mathrm{d}\hat{p}\,\mathcal{S} = 0$, and correcting this presumably requires distributing the orders erroneously transported through the mid-point by the diffusion term, to interior locations where they then influence long time diffusion autocorrelation. These source terms manifestly satisfy $\mathcal{S}(0) = 0$, though, and that determines the intercept of the average order depth.

B.2.1. Getting the intercept right. Evaluating equation (44) with $\alpha(\hat{p})/\alpha(\hat{\infty}) = 1 + \Pr(\hat{s}/2 \geqslant \hat{p})$, at $\hat{p} = 0$, gives the boundary value of the non-dimensionalized, mid-point centred, mean order density

$$\psi(0) = \frac{2}{1+\epsilon}, \tag{B.26}$$

which dimensionalizes to

$$\frac{\langle n(0)\rangle}{\sigma\,\mathrm{d}p} = \frac{2\alpha(\infty)/\sigma}{\mu(0)/2\sigma + \delta}. \tag{B.27}$$

Equation (B.27) for the *total* density is the same as the form (B.6) produced by the diffusion solution for the *zeroth order* density, as should be the case if diffusion no longer transports orders through the mid-point. This form is verified in simulations, with mid-point centred averaging.

Interestingly, the same argument for the bid centred frame would simply omit the φ from $\alpha(0)/\alpha(\infty)$, predicting that

$$\psi(0) = \frac{1}{1+\epsilon}, \tag{B.28}$$

a result which is *not* confirmed in simulations. Thus, in addition to not satisfying the mean field approximation, the bid centred density average appears to receive some diffusive transport of orders all the way down to the bid.

B.2.2. Fokker–Planck expanding correlations. Equations (B.24) and (B.25) are not directly easy to use in a numerical integral. However, they can be Fokker–Planck expanded to terms with behaviour comparable to the diffusion equation, and the correct behaviour near the mid-point. Doing so gives the non-dimensional expansion of the source term $\mathcal{S}$ corresponding to the diffusion contribution in equation (32):

$$\mathcal{S} = \mathcal{R}(\hat{p})\psi(\hat{p}) + \mathcal{P}(\hat{p})\frac{\mathrm{d}\psi(\hat{p})}{\mathrm{d}\hat{p}} + \epsilon\beta(\hat{p})\frac{\mathrm{d}^2\psi(\hat{p})}{\mathrm{d}\hat{p}^2}. \tag{B.29}$$

The rate terms in equation (B.29) are integrals defined as

$$\begin{aligned}\mathcal{R}(\hat{p}) = &\int_0^\infty 2\,\mathrm{d}\Delta\hat{p}\,[\varphi(\Delta\hat{p}) - \varphi(\hat{p}+\Delta\hat{p})] \\ &- 2\int_0^{\hat{p}} 2\,\mathrm{d}\Delta\hat{p}\,[\varphi(\Delta\hat{p}) - \varphi(\hat{p})] \\ &+ \int_0^{\hat{p}/2} 2\,\mathrm{d}\Delta\hat{p}\,[\varphi(\Delta\hat{p}) - \varphi(\hat{p}-\Delta\hat{p})],\end{aligned} \tag{B.30}$$

$$\begin{aligned}\mathcal{P}(\hat{p}) = 2&\int_0^\infty 2\,\mathrm{d}\Delta\hat{p}\,\Delta\hat{p}\,[\varphi(\Delta\hat{p}) - \varphi(\hat{p}+\Delta\hat{p})] \\ &- \int_0^{\hat{p}} 2\,\mathrm{d}\Delta\hat{p}\,\hat{p}[\varphi(\Delta\hat{p}) - \varphi(\hat{p})] \\ &- \int_0^{\hat{p}/2} 2\,\mathrm{d}\Delta\hat{p}\,\Delta\hat{p}\,[\varphi(\Delta\hat{p}) - \varphi(\hat{p}-\Delta\hat{p})],\end{aligned} \tag{B.31}$$

and

$$\begin{aligned}\epsilon\beta(\hat{p}) = &\int_0^\infty 2\,\mathrm{d}\Delta\hat{p}\,(\Delta\hat{p})^2[\varphi(\Delta\hat{p}) - \varphi(\hat{p}+\Delta\hat{p})] \\ &+ \tfrac{1}{2}\int_0^{\hat{p}} (\Delta\hat{p})^2\,2\,\mathrm{d}\Delta\hat{p}\,[\varphi(\Delta\hat{p}) - \varphi(\hat{p})] \\ &+ \tfrac{1}{2}\int_0^{\hat{p}/2} 2\,\mathrm{d}\Delta\hat{p}\,(\Delta\hat{p})^2[\varphi(\Delta\hat{p}) - \varphi(\hat{p}-\Delta\hat{p})].\end{aligned} \tag{B.32}$$

All of the coefficients (B.30), (B.31) vanish manifestly as $\hat{p} \to 0$, and at large $\hat{p}$, $\mathcal{R}, \mathcal{P} \to 0$, while $\epsilon\beta(\hat{p}) \to 4\int_0^\infty \mathrm{d}\Delta\hat{p}\,(\Delta\hat{p})^2\varphi(\Delta\hat{p})$, recovering the diffusion constant (41) of the simplified source term. However, they are still not convenient for numerical integration, being non-local in φ.

The exponential form (33) is therefore exploited to approximate φ, in the region where its value is largest, with the expansion

$$\varphi(\hat{p} \pm \Delta\hat{p}) \approx \varphi(\hat{p})\varphi(\Delta\hat{p})\mathrm{e}^{\pm\hat{p}\,\Delta\hat{p}\,\partial\psi/\partial\hat{p}|_0}. \tag{B.33}$$

In the range where the mean field approximation is valid, φ is dominated by the constant term $\psi(0)$, and even the factors $\mathrm{e}_0^{\pm\hat{p}\,\Delta\hat{p}\,\partial\psi/\partial\hat{p}|}$ can be approximated as unity. This leaves the much simplified expansions

$$\begin{aligned}\mathcal{R}(\hat{p}) = &[1-\varphi(\hat{p})]\mathcal{I}_0(\infty) - 2[\mathcal{I}_0(\hat{p}) - 2\hat{p}\varphi(\hat{p})] \\ &+ [1-\varphi(\hat{p})]\mathcal{I}_0(\hat{p}/2),\end{aligned} \tag{B.34}$$

$$\begin{aligned}\mathcal{P}(\hat{p}) = &2[1-\varphi(\hat{p})]\mathcal{I}_1(\infty) - [\mathcal{I}_1(\hat{p}) - \hat{p}^2\varphi(\hat{p})] \\ &- [1-\varphi(\hat{p})]\mathcal{I}_1(\hat{p}/2),\end{aligned} \tag{B.35}$$

and

$$\begin{aligned}\epsilon\beta(\hat{p}) = &[1-\varphi(\hat{p})]\mathcal{I}_2(\infty) + \tfrac{1}{2}[\mathcal{I}_2(\hat{p}) - \tfrac{2}{3}\hat{p}^3\varphi(\hat{p})] \\ &+ [1-\varphi(\hat{p})]\mathcal{I}_2(\hat{p}/2).\end{aligned} \tag{B.36}$$

In equations (B.34)–(B.36),

$$\mathcal{I}_j(\hat{p}) \equiv \int_0^{\hat{p}} 2\,\mathrm{d}\Delta\hat{p}\,(\Delta\hat{p})^j\varphi(\Delta\hat{p}), \tag{B.37}$$

for $j = 0, 1, 2$. These forms (B.34)–(B.36) are inserted in equation (B.29) for $\mathcal{S}$ to produce the mean field results compared to simulations in figures 20–22.

References

[1] Daniels M G, Farmer J D, Iori G and Smith E 2003 Quantitative model of price diffusion and market friction based on trading as a mechanistic random process *Phys. Rev. Lett.* **90** 108102

[2] Mendelson H 1982 Market behavior in a clearing house *Econometrica* **50** 1505

[3] Cohen K J, Conroy R M and Maier S F 1985 Order flow and the quality of the market *Market Making and the Changing Structure of the Securities Industry* ed Y Amihud, T Ho and R Schwartz (Lexington, MA: Lexington Books)

[4] Domowitz I and Wang J 1994 Auctions as algorithms *J. Econ. Dynamics Control* **18** 29

[5] Bollerslev T, Domowitz I and Wang J 1997 Order flow and the bid–ask spread: an empirical probability model of screen-based trading *J. Econ. Dynamics Control* **21** 1471

[6] Bak P, Paczuski M and Shubik M 1997 Price variations in a stock market with many agents *Physica* A **246** 430

[7] Eliezer D and Kogan I I 1998 Scaling laws for the market microstructure of the interdealer broker markets *Preprint* cond-mat/9808240

[8] Tang L-H and Tian G-S 1999 Reaction–diffusion-branching models of stock price fluctuations *Physica* A **264** 543

[9] Maslov S 2000 Simple model of a limit order-driven market *Physica* A **278** 571

[10] Slanina F 2001 Mean-field approximation for a limit order driven market model *Phys. Rev.* E **64** 056136

[11] Challet D and Stinchcombe R 2001 Analyzing and modelling 1+1d markets *Physica* A **300** 285

[12] Bachelier L 1900 Théorie de la spéculation
Cootner P H 1964 *The Random Character of Stock Prices* (Cambridge, MA: MIT Press) (reprint)

[13] Becker G 1962 *J. Political Economy* **70** 1

[14] Gode D and Sunder S 1993 Allocative efficiency of markets with zero intelligence traders: markets as a partial substitute for individual rationality *J. Political Economy* **101** 119

[15] Farmer J D 2002 Market force, ecology, and evolution *Indust. Corp. Chang.* **11** 895 (SFI working paper 98-12-117)

[16] Bouchaud J-P and Cont R 1998 A Langevin approach to stock market fluctuations and crashes *Eur. Phys. J.* B **6** 543

[17] Hausman J A and Lo A W 1992 An ordered probit analysis of transaction stock prices *J. Financial Economics* **31** 319

[18] Farmer J D 1996 Slippage 1996 *Prediction Company Internal Technical Report* http://www.predict.com/jdf/slippage.pdf

[19] Torre N 1997 *BARRA Market Impact Model Handbook* (Berkeley, CA: BARRA Inc.) www.barra.com

[20] Kempf A and Korn O 1998 Market depth and order size *University of Mannheim Technical Report*

[21] Plerou V, Gopikrishnan P, Gabaix X and Stanley H E 2002 Quantifying stock price response to demand fluctuations *Phys. Rev.* E **66** 027104

[22] Lillo F, Farmer J D and Mantegna R N 2003 Collapse of price impact function *Nature* **421** 129

[23] Potters M and Bouchaud J-P 2003 Statistical properties of stock order books: empirical results and models *Physica* A **324** 133

[24] Chiarella C and Iori G 2002 A simulation analysis of the microstructure of double auction markets *Quant. Finance* **2** 346

[25] Farmer J D, Patelli P and Zovko I I 2003 The predictive power of zero intelligence in financial markets *Preprint* cond-mat/0309233

[26] Bridgman P W 1922 *Dimensional Analysis* (New Haven, CT: Yale University Press)

[27] Zovko I I and Farmer J D 2002 The power of patience: a behavioral regularity in limit order placement *Quant. Finance* **2** 387

[28] Maslov S and Mills M 2001 Price fluctuations from the order book perspective—empirical facts and a simple model *Physica* A **299** 234

[29] Mantegna R N and Stanley H E 1999 *Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge, MA: Cambridge University Press)