Topics in market microstructure

Zovko, I.

Citation for published version (APA):
Chapter 3

The predictive power of zero intelligence in financial markets


3.1 Introduction

The traditional paradigm in economics is one of rational utility maximizing agents. Recognizing limitations in human cognition, economists have increasingly explored models in which agents have bounded rationality. We take this direction even further here by testing a model of trading in financial markets that drops agent rationality almost altogether. These results are particularly striking because the model predicts simple quantitative laws relating different properties of markets that are borne out well when tested against data.

While no one would dispute the fact that agents in financial markets behave strategically, and that for some purposes taking this into account is essential, we show in this paper that there are some problems where other factors may be more important. Previous work along these lines includes that of Becker (1962), who showed that
random agent behavior and a budget constraint is sufficient to guarantee the proper slope of supply and demand curves, and Gode and Sunder (1993), who demonstrated that if one replaces the students in a standard classroom economics experiment by zero-intelligence agents with a budget constraint, they perform surprisingly well. More specifically, the model we test here builds on earlier work on the double auction in financial economics (Mendelson, 1982; Cohen et al., 1985; Domowitz and Wang, 1994; Bollerslev et al., 1997) and physics (Bak et al., 1997; Eliezer and Kogan, 1998; Maslov, 2000; Slanina, 2001; Challet and Stinchcombe, 2001). (See also interesting subsequent work (Bouchaud et al., 2002, 2004)). The model makes the simple assumption that agents place orders to buy or sell at random (Daniels et al., 2003; Smith et al., 2003), subject to constraints imposed by current prices. While one might argue that tracking prices requires at least some intelligence, this is the minimal intelligence consistent with the assumptions of the model, which we will loosely refer to as “zero intelligence”. We show here that for certain problems such an approach can make surprisingly good quantitative predictions.

Another unusual aspect of the work presented here is the nature of the predictions that we test, which take the form of simple quantitative laws. These laws relate one set of market properties to another, placing restrictions on the allowed values of variables that are comparable to the ideal gas law of physics. They make quantitative predictions about magnitude and functional form, which are testable with only minimal auxiliary assumptions. This is in contrast to papers testing standard models based on rationality, which are typically forced to add strong auxiliary assumptions not contained in the original theoretical model, making the final results essentially qualitative. We present a brief review in the Supplementary Material (SM), Section 3.5.1.

3.1.1 Continuous double auction

The continuous double auction is the most widely used method of price formation in modern financial markets. The auction is called “double” because traders can submit orders both to buy and to sell and it is called “continuous” because they can do so at any time. Under the terminology we use here, an order that does not cross the opposite best price, and so does not result in an immediate transaction, is called a limit order. An example is a sell order with a higher price than any existing buy order. An order that does cross the opposite best price, and thus causes an immediate transaction, is called
CHAPTER 3. THE PREDICTIVE POWER OF ZERO INTELLIGENCE IN FINANCIAL MARKETS

Figure 3.1: A random process model of the continuous double auction. Stored limit orders are shown stacked along the price axis, with sell orders (supply) stacked above the axis at higher prices and buy orders (demand) stacked below the axis at lower prices. New sell limit orders are visualized as randomly falling down, and new buy orders as randomly “falling up”. New sell orders can be placed anywhere above the best buying price, and new buy orders anywhere below the best selling price. Limit orders can be removed spontaneously (e.g. because the agent changes her mind or the order expires) or they can be removed by market orders of the opposite type. This can result in changes in the best prices, which in turn alters the boundaries of the order placement process.

A market order\(^1\). Buy and sell limit orders accumulate in their respective queues, while buy and sell market orders cause transactions that remove limit orders. A limit order can also be removed from its queue by being cancelled, which can occur at any time. The lowest selling price offered at any point in time is called the best ask, \(a(t)\), and the highest buying price the best bid, \(b(t)\). The bid-ask spread \(s(t) \equiv a(t) - b(t)\) measures the gap between them. The best prices may change as new orders arrive or old orders are cancelled.

\(^1\)Real markets employ a host of different order types, which vary from market to market. However, by making appropriate decompositions (sometimes involving splitting an order into two pieces) it is always possible to break down the order flow into components that are effectively either market orders or limit orders.
CHAPTER 3. THE PREDICTIVE POWER OF ZERO INTELLIGENCE IN FINANCIAL MARKETS

3.1.2 Review of the model

The model that we test here (Daniels et al., 2003; Smith et al., 2003) was constructed to be the simplest possible sensible model of agent behavior in a continuous double auction. It assumes that two types of agents place orders randomly according to independent Poisson processes, as shown in Fig. 3.1. Impatient agents place market orders randomly with a Poisson rate of \( \mu \) shares per unit time. Patient agents, in contrast, place limit orders randomly in both price and time. Buy limit orders are placed uniformly anywhere in the semi-infinite interval \(-\infty < p < a(t)\), where \( p \) is the logarithm of the price, and similarly sell limit orders are placed uniformly anywhere in \( b(t) < p < \infty \). Both buying and selling limit orders arrive with the same Poisson rate density \( \alpha \), which is measured in shares per unit price per unit time. The log-price \( p \) is continuous and is independent of arrival time. Both limit and market orders are of constant size \( \sigma \) (measured in shares). Queued limit orders are cancelled according to a Poisson process, analogous to radioactive decay, with a fixed rate \( \delta \) per unit time. To keep the model as simple as possible, there are equal rates for buying and selling, and all of these processes are independent except for indirect coupling through the boundary conditions, as explained below.

As new orders arrive they may alter the best prices \( a(t) \) and \( b(t) \), which in turn changes the boundary conditions for subsequent limit order placement. For example, the arrival of a buy limit order inside the spread will alter the best bid \( b(t) \), which immediately alters the boundary condition for placing the next sell limit order. It is this feedback between order placement and price diffusion that makes this model interesting, and despite its apparent simplicity, very difficult to understand analytically. This model has been studied using simulation and with approximate analytic treatments based on mean field theory (Daniels et al., 2003; Smith et al., 2003).

Some readers may be puzzled by the use of a constant density over an infinite interval, which gives an infinite total arrival rate. The key is that the normalization is chosen to make the arrival rate in any given price interval finite. This is analogous to a model of snow falling and evaporating on an infinite plane: Though the total amount of snow arriving is infinite, the amount of snow falling in any given square during any given time is perfectly well-behaved. The situation here is much more complicated, due to the fact that market orders define a point removal process, and there are two kinds of “snow”, falling on overlapping and interacting intervals. Nonetheless, the basic trick of normalizing the density rather than the total is the same.
3.1.3 Predictions of the model

The rather radical assumption of a uniform limit order price density\textsuperscript{2} is made because it simplifies analysis, allowing the derivation of simple scaling laws relating the parameters to fundamental properties such as the average bid-ask spread. The mean value of the spread predicted based on a mean field theory analysis of the model (Daniels et al., 2003; Smith et al., 2003) is

$$\hat{s} = (\mu/\alpha)f(\sigma\delta/\mu).$$  \hspace{1cm} (3.1)

The nondimensional ratio $\epsilon \equiv \sigma\delta/\mu$ can be thought of as the ratio of removal by cancellation to removal by market orders, and plays an important role. $f(\epsilon)$ is a slowly varying, monotonically increasing function that can be approximated (Smith et al., 2003) as $f(\epsilon) = 0.28 + 1.86\epsilon^{3/4}$. The scaling law above is reasonable in that it predicts that the spread increases when there are more market orders or cancellations (which remove stored limit orders), and decreases with more limit orders (which fill the spread in more quickly). The dependence on $\mu/\alpha$ can be derived from dimensional analysis, under certain assumptions detailed in SM Section 3.5.2. However, the functional form of $f(\epsilon)$ is not obvious. One of the predictions of the model, that to our knowledge has not been hypothesized elsewhere in the literature, is that the order size $\sigma$ is an important determinant of the spread.

Another prediction of the model concerns the price diffusion rate, which drives the volatility of prices and is the primary determinant of financial risk. If we assume that prices make a random walk, then the diffusion rate measures the size and frequency of its increments. The variance $V$ of a random walk grows as $V(t) = Dt$, where $D$ is the diffusion rate and $t$ is time. This is the main free parameter in the Bachelier model of prices (Bachelier, 1964). While its value is essential for risk estimation and derivative pricing there is very little fundamental understanding of what actually determines it. In standard models it is often assumed to depend on “information arrival” (Clark, 1973), which has the disadvantage that it is impossible to measure directly. For our idealized model, numerical experiments indicate that the short term price diffusion rate is to a very good approximation given by the simple formula (Daniels et al., 2003; Smith et al., 2003)

$$\hat{D} = k\mu^{5/2}\delta^{1/2}\sigma^{-1/2}\alpha^{-2},$$  \hspace{1cm} (3.2)

where $k$ is a constant. This formula is reasonable in that it predicts that volatility increases with limit order removal (either by market

\textsuperscript{2}For an empirical investigation of the density of limit order placement see (Zovko and Farmer, 2002; Bouchaud et al., 2002)
CHAPTER 3. THE PREDICTIVE POWER OF ZERO INTELLIGENCE IN FINANCIAL MARKETS

orders or by cancellations) and decreases with limit order placement. The dependence on order size and the values of the scaling exponents are not so obvious. It has so far not been possible to derive this formula from theoretical considerations (though dimensional analysis was essential for guessing this functional form).

We would like to emphasize that the construction of the model and all the predictions derived from it were made prior to looking at the data. The model was constructed to be simple enough to be analytically tractable, and makes many strong assumptions. The assumption of random order placement leads to consequences that might be economically unreasonable in a rational setting, such as the existence of profit making opportunities. However, this is self-consistent with the assumption that the only intelligence the agents possess is the ability to mechanically adjust the prices of limit orders based on current best prices. Furthermore, simulations suggest that the arbitrage opportunities in this model are not risk-free, yielding only finite, risky profits\(^3\).

A useful concept is that of liquidity, which in this context can be defined as the availability of standing limit orders that allow trading to take place. The impatient market order traders are liquidity demanders, and the limit order traders are liquidity providers. The use of a zero-intelligence agent model makes it possible to study the flow of liquidity in and out of the market, and to study its interaction with price formation. This has not been properly addressed by models that attempt to fully treat agent rationality. Abandoning the assumption of rationality gives the ability to focus modeling effort on other problems, such as those addressed here.

### 3.2 Testing the scaling laws

#### 3.2.1 Data

We test this model with data from the electronic open limit order book of the London Stock Exchange (SETS), which includes about half of the total trading volume. We used data from eleven stocks in the period from August 1st 1998 to April 30th 2000, which includes 434 trading days and a total of roughly six million events. For all these stocks the number of total events exceeds 300,000 and was never less than 80 on any given day (where an event corresponds to an order placement or cancellation). Orders placed during the opening auction are removed to accommodate the fact that the model only applies for the continuous auction. See SM Section 3.5.3 for more details.

\(^3\)J. Rutt, J.D. Farmer, and J. Girard, work in progress.
3.2.2 Testing procedure

We test the model cross-sectionally over eleven stocks. For each stock we measure its average order flow rates and calculate the predicted average spread $\hat{\sigma}$ and diffusion rate $\hat{D}$ for that stock using equations 3.1 and 3.2. We then compare these predicted values to the actual values of the spread $\bar{\sigma}$ and diffusion rate $\bar{D}$ which we again measure from the data. The comparison is done via linear regressions of the predicted values against the actual, measured values. For a discussion see SM Sections 3.5.5 and 3.5.6.

Measurement of the parameters $\mu$ and $\sigma$ is straightforward: To measure $\mu$, for example, we simply compute the total number of shares of market orders and divide by time, or alternatively, we compute $\mu_t$ for each day and average; we get similar results in either case. However, a problem occurs in measuring the parameters $\alpha$ and $\delta$ due to the simplifying assumption of a uniform distribution of prices for limit order placement and a uniform cancellation rate. In the real data limit order placement and cancellation are concentrated near the best prices (Bouchaud et al., 2002; Zovko and Farmer, 2002). In order to cope with this we make an auxiliary assumption that order placement is uniform inside a price window around the best prices, and zero outside this window. We choose this price window $W$ to correspond to roughly 60% of limit orders away from the mid-price, and compute $\alpha$ by dividing the number of shares of limit orders placed inside this price window per unit time by $W$. We do this for each day and compute the average value of $\alpha$ for each stock. We similarly compute $\delta$ as the inverse of the average lifetime of orders cancelled inside the same price window $W$. See SM Section 3.5.5 for details.

The laws that we describe here do not make temporal predictions, but rather are restrictions of state variables. The ideal gas law, $PV = RT$, provides a good analogy. It predicts that pressure $P$, volume $V$, and temperature $T$ are constrained – any two of them determines the third. The gas constant $R$ is the only free parameter. In very much the same way, we are testing two relations between properties of order flows and properties of prices. We are not attempting to predict the temporal behavior of the order flows, only trying to see whether the restrictions between order flows and prices predicted by the model are valid. It is important to emphasize that while $\mu$, $\alpha$, $\delta$, and $\sigma$ can be viewed as free parameters of the model, they are not free parameters in the test of the model. Rather they are now variables, like $P$, $V$, and $T$ in the ideal gas law. The only free parameter is the price window $W$. We chose $W = 0.6$ as a prior; it turns out that it is also roughly the value that maximizes the goodness of fit, however, varying $W$ does not change the goodness of fit substantially.
3.2.3 Spread

To test Equation 3.1, we measure the average spread $\bar{s}$ across the full time period for each stock, and compare to the predicted average spread $\hat{s}$ based on order flows. Spread is measured as the average of $\log b(t) - \log a(t)$ (recall that in the model $p$ represents log price). The spread is measured after each event, with each event given equal weight. The opening auction is excluded.

To test our hypothesis that the predicted and actual values coincide, we perform a regression of the form $\log \bar{s} = A \log \hat{s} + B$. We took logarithms to do the regression because the spread is positive and the log of the spread is approximately normally distributed\(^4\). We use $A$ and $B$ for hypothesis testing. Based on the model we predict that the comparison should yield a straight line with $A = 1$ and $B = 0$. However, because of the degree of freedom in choosing the price interval $W$ as described above, the value of $B$ is somewhat arbitrary; varying $W$ through reasonable values changes $B$ significantly, with much less effect on $A$.

The least squares regression, shown together with the data comparing the predictions to the actual values in Fig. 3.2, gives $A = 0.99 \pm 0.10$ and $B = 0.06 \pm 0.29$. We thus strongly reject the null hypothesis that $A = 0$, indicating that the predictions are far better than random. More importantly, we are unable to reject the null hypothesis that $A = 1$. The regression has $R^2 = 0.96$, so the model explains most of the variance. Note that because of long-memory effects and cross-correlations between stocks the errors in the regression are larger than they would be for IID data (see SM Section 3.5.6).

3.2.4 Price diffusion rate

As for the spread, we compare the predicted price diffusion rate based on order flows to the actual price diffusion rate $\bar{D}$ for each stock averaged over the 21 month period, and regress the logarithm of the predicted vs. actual values, as shown in Fig. 3.3. See SM Section 3.5.5 for details on the diffusion rate and its estimation.

The regression gives $A = 1.33 \pm 0.25$ and $B = 2.43 \pm 1.75$. Thus, we again strongly reject the null hypothesis that $A = 0$. We are still unable to reject the null hypothesis that $A = 1$ with 95% confidence, though there is some suggestion that the real values increase faster than the predicted values. In any case, the predictions are at least a good approximation. Although the results are not as good as for the

\(^4\)An alternative would have been to take logarithms of each event and average the logarithms. We instead regard this as a test of the cross-sectional averages, and take logarithms of the cross-sectional values.
Figure 3.2: Regressions of predicted values based on order flow using equation 3.1 vs. actual values for the log spread. The dots show the average predicted and actual value for each stock averaged over the full 21 month time period. The solid line is a regression; the dashed line is the diagonal, representing the model’s prediction with $A = 1$ and $B = 0$.

spread, $R^2 = 0.76$, so the model still explains most of the variance.

3.3 Average market impact

Market impact is practically important because it is the dominant source of transaction costs for large trades, and conceptually important because it provides a convenient probe of the revealed supply and demand functions in the limit order book (see SM Section 3.5.7). When a market order of size $\omega$ arrives, if it removes all limit orders at the best bid or ask it will immediately change the midpoint price $m \equiv (a + b)/2$. We define the average market impact function $\phi$ in terms of the instantaneous logarithmic midpoint price shift $\Delta p$ conditioned on order size, $\phi(\omega) = E[\Delta p|\omega]$. $\Delta p$ is the difference between the price just before a market order arrives and the price just after it arrives (before any other events).

A long-standing mystery about market impact is that it is a
highly concave function of $\omega$ (Hausman et al., 1992; Farmer, 1996; Torre, 1997; Kempf and Korn, 1999; Plerou et al., 2002a; Bouchaud et al., 2002; Lillo et al., 2003; Gabaix et al., 2003a). This is unexpected since simple arguments would suggest that because of the multiplicative nature of returns, market impact should grow at least linearly (Smith et al., 2003). We know of no model that explains this. The model we are testing here predicts a concave average market impact function, with the concavity becoming more pronounced for small values of $\epsilon = \sigma \delta / \mu$. Intuitively, the concavity is due to the fact that limit orders near the best price are removed by transactions more rapidly than those far from the best price. As a result the average density of stored limit orders in the book increases moving away from the midpoint. An increase in density of limit orders implies a decreased price response to a market order of given size, resulting in a concave market impact function.

While the predictions of the model are qualitatively correct, from
a quantitative point of view the model predicts a larger variation with $\epsilon$ than what we actually observe. Nonetheless, the model is still quantitatively useful for understanding market impact, as described below.

A surprising regularity of the average market impact function is uncovered by simply plotting the data in the non-dimensional coordinates dictated by the model, as shown in Fig. 3.4. If we view market impact in standard dimensional units, such as British Pounds or shares, there is large variability from stock to stock; the story becomes much simpler in non-dimensional units.

When we plot the average market impact in standard dimensional coordinates, the behavior is highly variable from stock to stock. For example, in Fig. 3.4(b) we plot the average market impact $\phi(\omega) = E[\Delta p | \omega]$ as a function of the order size $\omega$ in units of British Pounds. We do this by binning together events with similar $\omega$ and plotting this vs. the corresponding mean price impact $\Delta p$ for each bin. The result varies widely from stock to stock. We have explored a variety of other ways for renormalizing the order size, as described in SM Section 3.5.7, but they all give similar results.

Plotting the data in non-dimensional units tells a simpler story. To do this we normalize the price shift and order size by appropriate dimensional scale factors based on the daily order flow rates. For the derivation of the non-dimensional coordinates used to do this see SM Section 3.5.2. This transforms the standard coordinates to non-dimensional coordinates as $^5 \Delta p \rightarrow \Delta p \cdot \alpha_t / \mu_t$ and $\omega \rightarrow \omega \cdot \delta_t / \mu_t$, where $\alpha_t$, $\mu_t$, and $\delta_t$ are the average parameters for day $t$. The data collapses onto roughly a single curve, as shown in Fig. 3.4(a). The variations from stock to stock are quite small; on average the corresponding bins for each stock deviate from each other by about 8%, roughly the size of the statistical sampling error. We do not find that this variation is statistically significant, through we should also say that such tests are complicated by the long-memory property of these time series and cross correlations between stocks, so that we do not consider the results fully reliable (see SM Section 3.5.6). In contrast, using standard dimensional coordinates the differences are easily shown to be highly statistically significant. This collapse illustrates that the non-dimensional coordinates dictated by the model provide substantial explanatory power: We can understand how the average market impact varies from stock to stock by a simple transformation of coordinates. Plotting in double logarithmic scale shows

---

$^5$Notice that $\Delta p$ has units of price but

$$[\Delta p \cdot \alpha/\mu] = \text{price} \cdot \frac{\text{shares} / (\text{price} \cdot \text{time})}{\text{shares} / \text{time}}$$

is without dimensions, i.e. the units cancel out.
that the curve of the collapse is roughly a power law of the form $\omega^{0.25}$ (see SM Section 3.5.7). This provides a more fundamental explanation for the empirically constructed collapse of average market impact for the New York Stock Exchange found earlier (Lillo et al., 2003).
Figure 3.4: The average market impact as a function of the mean order size. In (a) the price differences and order sizes for each transaction are normalized by the non-dimensional coordinates dictated by the model, computed on a daily basis. Most of the stocks collapse onto a single curve; there are a few that deviate, but the deviations are sufficiently small that given the long-memory nature of the data and the cross-correlations between stocks, it is difficult to determine whether these deviations are statistically significant. This means that we understand the behavior of the market impact as it varies from stock to stock by a simple transformation of coordinates. In (b), for comparison we plot the order size in units of British pounds against the average logarithmic price shift.
3.4 Conclusions

We have shown that the model we have presented here does a good job of predicting the average spread, and a decent job of predicting the price diffusion rate. Also, by simply plotting the data in non-dimensional coordinates we get a better understanding of the regularities of market impact. These results are remarkable because the underlying model largely drops agent rationality, instead focusing all its attention on the problem of understanding the constraints imposed by the continuous double auction.

It is worth comparing our results to those of previous empirical work. For example, Hasbrouck and Saar (Hasbrouck and Saar, 2002) find a positive correlation between volatility and the ratio of market orders to limit orders. They perform regressions of this ratio against volatility and several other dependent variables, and obtain goodness of fits similar to ours (except that in their case volatility was one of several independent variables, whereas in our case it was the dependent variable). They then discuss the results in terms of their consistency with three effects that one would expect from agent rationality. For example, one such effect is called “market order certainty”: When prices are more volatile, market orders become more attractive to a risk-averse rational agent, and so the fraction of market orders should increase. The observed positive correlations are consistent with this.

The model we test here offers an alternative explanation that does not depend on strategic choice. We also predict a positive correlation between volatility and the fraction of market orders (see equation 3.2) but for a different reason: An increase in the rate of market order submission reduces liquidity and thus increases price volatility. We certainly believe that agents respond in important ways to changing market conditions such as volatility, and indeed we have demonstrated this in previous work (Zovko and Farmer, 2002). Nonetheless, we argue that it is also necessary to understand the impact of agents’ actions on market conditions. By carefully treating the feedback in both directions between price formation and limit order pricing under minimal assumptions of rationality, this model provides a null hypothesis against which claims of rational behavior can be measured.

An important feature of the model we test here is its parsimony and falsifiability. Our model makes simultaneous quantitative predictions about volatility, spread, and market impact. We postulate specific functional forms for the relation between order flows and spread and volatility; while there are multiple variables involved, there is only one free parameter. Rationality-based theories, in contrast, rarely make predictions about magnitude or functional form,
and as a result their predictions are harder to test. Such tests generally require stronger auxiliary assumptions, such as imposed functional forms with multiple free parameters. Empirical studies that test such models often test only the sign of such effects, which often have a variety of alternative explanations. Our model makes sharper predictions, and is consequently more testable (Ziliak and McCloskey, 2004).

The approach taken here succeeds in part because it is less ambitious than a standard rationality-based model. This can be viewed as a divide and conquer strategy. Rather than attempting to explain the properties of the market from fundamental assumptions about utility maximization by individual agents, we divide the problem into two parts. The first (easier) problem, addressed here, is that of understanding the characteristics of the market given the order flows. The second (harder) problem, which remains to be investigated, is that of explaining why order flow varies as it does. Explaining variations in order flow involves behavioral and/or strategic issues that are likely to be much more difficult to understand. It is always desirable to solve easier problems first.

The model succeeds in part by reducing the problem to the measurement of the right variables. By measuring the rate of market order placement vs. limit order placement, and the rate of order cancellation, we are able to measure how patient or impatient traders are. The model makes quantitative predictions about how this affects other market properties. The agreement with the model indicates that patience is an important determinant of market behavior. Variations in patience might be explained by a rationality-based explanation in terms of information arrival, or a behavioral-based explanation driven by emotional response, but in either case it suggests that patience is a key factor.

These results have several practical implications. For market practitioners, understanding the spread and the market impact function is very useful for estimating transaction costs and for developing algorithms that minimize their effect. For regulators they suggest that it may be possible to make prices less volatile and lower transaction costs, if this is desired, by creating incentives for limit orders and disincentives for market orders. These scaling laws might also be used to detect anomalies, e.g. a higher than expected spread might be due to improper market maker behavior.

This is part of a broader research program that might be somewhat humorously characterized as the “low-intelligence” approach: We begin with minimally intelligent agents to get a good benchmark of the effect of market institutions, and once this benchmark is well-understood, add more intelligence, moving toward market efficiency. We thus start from almost zero rationality and work our way up, in
contrast to the canonical approach of starting from perfect rationality and working down.

The model we test here was constructed before looking at the data (Daniels et al., 2003; Smith et al., 2003), and was designed to be as simple as possible for analytic analysis. A more realistic (but necessarily more complicated) model would more closely mimic the properties of real order flows, which are price dependent and strongly correlated both in time and across price levels, or might incorporate elements of the strategic interactions of agents. An improved model would hopefully be able to capture more features of the data than those we have studied here. We know there are ways in which the current model is inappropriate, e.g., it allows arbitrage opportunities that do not exist in the real market. Nonetheless, as we have shown above, this extremely simple model does a good job of explaining some important properties of markets. For further discussion see SM Section 3.5.8.

How is it conceivable to successfully model a situation in which we know that agents engage in clever strategic behavior in terms of a model that completely neglects this? Perhaps a telephone exchange provides a good analogy: Even though each customer has a perfectly good reason for picking up the phone, communications engineers design exchanges by assuming they do so at random. Similarly, there are situations in markets where rational behavior can be treated in aggregate as though it were noise. The question is whether rational effects are more important or less important than stochastic effects. Rational effects are clearly important in determining overall price levels, but they may be dominated by random fluctuations in determining volatility. We do not mean to claim that market participants are unintelligent: Indeed, one of the virtues of this model is that it provides a benchmark to separate properties that are driven by the statistical mechanics of the market institution from those that are driven by the strategic behavior of agents. It suggests that institutions strongly shape our behavior, so that some of the properties of markets may depend more on the structure of institutions than on the rationality of individuals.
CHAPTER 3. THE PREDICTIVE POWER OF ZERO INTELLIGENCE IN FINANCIAL MARKETS

3.5 Supplementary Material

3.5.1 Literature review

We will begin by reviewing what we will call the "standard literature", which includes both purely empirical studies, as well as theoretical studies predicated on rationality models. We will then review the literature on random process models of the continuous double auction, which is more closely related to the model we test here.

Standard literature

The market microstructure literature focusing on the understanding of spread, volatility and market impact in financial markets is theoretically and empirically extensive. The theoretical analyses traditionally use the underlying paradigm of rational agents. Models of spread, starting with Demsetz (1968); Tinic (1972); Stoll (1978); Amihud and Mendelson (1980); Ho and Stoll (1981), have examined the possible determinants of spreads as a result of rational, utility-maximizing problem faced by the market makers. Models providing insight into the utility-maximizing response of agents to other various measures of market conditions such as volatility are for example Lo (2002) who investigate a simple model in which the log stock price is modeled as a Brownian motion diffusion process. Provided agents prefer a lower expected execution time, their model predicts a positive relationship between volatility and limit order placement. Copeland and Galai (1983); Glosten and Milgrom (1985); Easley and O’Hara (1987); Glosten (1995); Foucault (1999); Easley et al. (2001) examine asymmetric information effects on order placement. Andersen (1996) modifies the Glosten and Milgrom (1985) model with the stochastic volatility and information flow perspective. Other models of trading in limit order markets include Cohen et al. (1981); Angel (1994); Harris (1998); Chakravarty and Holden (1995); Seppi (1997); Rock (1990); Parlour and Seppi (2003); Parlour (1998); Foucault et al. (2001); Domowitz and Wang (1994).


Empirical research in volatility was initiated by statistical descriptions of the volatility process (Engle (1982); Bollerslev (1986), Bollerslev et al. (1992) for a survey), but has grown increasingly ambitious with multivariate structural models of the interaction between volatility and other economic variables. Positive correlation
of daily volume and volatility has been documented in Clark (1973); Epps and Epps (1976); Tauchen and Pitts (1983). Volume has been entered into ARCH specifications by Lamoureux and Lastrapes (1994). Other empirical investigations of volatility determinants and consequences include Gallant et al. (1992); Andersen (1996); Blume et al. (1991); Reiss and Werner (1994); Fleming et al. (2001); Hasbrouck (2003). A comprehensive study of the joint distribution or returns and volume is done by Gallant et al. (1992, 1993).

Random process models of continuous double auction

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 (1982), who modeled random order placement with periodic clearing. Cohen et al. (1985) developed a model of a continuous auction, modeling 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 vs. cancellation. Domowitz and Wang (1994) 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. (1997) on three weeks of data for the Deutschemark/U.S. 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 understand 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. (1997) which was developed by Eliezer and Kogan (1998) and by Tang and Tian (1999). They assume that limit orders are placed at a fixed distance from the midpoint, 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 (2000) introduced an alternative model that was solved analytically in the mean-field limit by Slanina (2001). 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 that 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 (2001), and independently by Daniels et al. (2003). 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 (1994), but allows for the possibility of order placement responding to price movement.

The model we test here was introduced by Daniels et al. (2003). 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 reference (Eliezer and Kogan, 1998), 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, the model studied here focuses 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. It demonstrates how dimensional analysis becomes a useful tool in an economic setting, and the analysis done in Daniels et al. and Smith et al. develops mean field theories to understand many relevant market properties. 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 reference Daniels et al. (2003), Bouchaud et al. (2002) 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. Chiarella and Iori (2002) have numerically studied fundamentalists and technical traders placing limit orders; a talk on this work by Giulia Iori in part inspired this model.

3.5.2 Dimensional analysis

Dimensional analysis can be used to simplify the study of this model and to make some approximate predictions about several of its properties. For a good reference on dimensional analysis see Barenblatt (1987).

There are three fundamental dimensional quantities in this model: shares, price, and time. There are five parameters. Because they have independent dimensions, when the dimensional constraints between the parameters are taken into account, this leaves only two independent degrees of freedom. It turns out that the order flow rates \( \mu, \alpha, \) and \( \delta \) are more important than the discreteness parameters \( \sigma \) and \( dp \), in the sense that the properties of the model are much more sensitive to variations in the order flow rates than they are to variations in \( \sigma \) or \( dp \). It therefore natural to construct non-dimensional units based on the order flow parameters alone. There are unique combinations of the three order flow rates with units of shares, price, and time. This gives characteristic scales for price, shares, and time, that are unique up to a constant: the characteristic number of shares \( N_c = \mu/\delta \), the characteristic price interval \( p_c = \mu/\alpha \), and the characteristic timescale \( t_c = 1/\delta \). These are the unique combinations of these three parameters with the correct dimensions.

These characteristic scales can be used to define non-dimensional coordinates based on the order flow rates. These are \( \hat{p} = p/p_c \) for price, \( \hat{N} = N/N_c \) for shares, and \( \hat{t} = t/t_c \) for time. The use of non-dimensional coordinates has the great advantage that it reduces the number of degrees of freedom from five to two. That is, instead of five independent parameters, we only have two independent parameters, which makes the model easier to understand.

The two irreducible degrees of freedom are naturally discussed in
CHAPTER 3. THE PREDICTIVE POWER OF ZERO INTELLIGENCE IN FINANCIAL MARKETS

terms of non-dimensional versions of the discreteness parameters. A non-dimensional scale parameter based on order size is constructed by dividing the typical order size $\sigma$ (with dimensions of shares) by the characteristic number of shares $N_c$. This gives the non-dimensional parameter $\epsilon \equiv \sigma/N_c = \delta\sigma/\mu$, which characterizes the granularity of the order flow. A non-dimensional scale parameter based on tick size is constructed by dividing the tick size $dp$ by the characteristic price, i.e. $dp/p_c = \alpha dp/\mu$. The usefulness of this is that the properties of the model only depend on the two non-dimensional parameters, $\epsilon$ and $dp/p_c$: Any variations of the parameters $\mu$, $\alpha$, and $\delta$ that keep these two non-dimensional parameters constant gives exactly the same market properties. One of the interesting results that emerges from analysis of the model is that the effect of the granularity parameter $\epsilon$ is generally much more important than the tick size $dp/p_c$. For a more detailed discussion, see reference Smith et al. (2003).

While we have investigated numerically the effect of varying the tick size in ref. Smith et al. (2003), for the purposes of comparing to data, here we simply take the limit $dp \to 0$, which provides a reasonable approximation.

3.5.3 The London Stock Exchange (LSE) data set

The London Stock Exchange is composed of two parts, the electronic open limit order book, and the upstairs quotation market, which is used to facilitate large block trades. During the time period of our dataset 40% to 50% of total volume was routed through the electronic order book and the rest through the upstairs market. It is believed that the limit order book is the dominant price formation mechanism of the London Stock Exchange: about 75% of upstairs trades happen between the current best prices in the order book (LSE bulletin, 2001). Our analysis involves only the data from the electronic order book. We chose to study this data set because we have a complete record of every action taken by every participating institution, allowing us to measure the order flows and cancellations and estimate all of the necessary parameters of our model.

We used data from the time period August 1st 1998 - April 30th 2000, which includes a total of 434 trading days and roughly six million events. We chose 11 stocks each having the property that the number of total number of events exceeds 300,000 and was never less than 80 on any given day. Some statistics about the order flow for each stock are given in table 3.1.

The trading day of the LSE starts at 7:50 with a roughly 10 minute long opening auction period (during the later part of the dataset the auction end time varies randomly by 30 seconds). Dur-
### Table 3.1: Summary statistics for stocks in the dataset.

<table>
<thead>
<tr>
<th>Stock Ticker</th>
<th>Num. Events (1000s)</th>
<th>Average Events (per day)</th>
<th>Effective Limit Orders (1000s)</th>
<th>Effective Market Orders (1000s)</th>
<th>Order Deletions (1000s)</th>
<th>Average Limit Order Size (shares)</th>
<th>Average Market Order Size (shares)</th>
<th># Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZN</td>
<td>608</td>
<td>1405</td>
<td>292</td>
<td>128</td>
<td>188</td>
<td>4,967</td>
<td>4,921</td>
<td>429</td>
</tr>
<tr>
<td>BARC</td>
<td>571</td>
<td>1318</td>
<td>271</td>
<td>128</td>
<td>172</td>
<td>7,370</td>
<td>6,406</td>
<td>433</td>
</tr>
<tr>
<td>CW</td>
<td>511</td>
<td>1184</td>
<td>244</td>
<td>134</td>
<td>134</td>
<td>12,671</td>
<td>11,151</td>
<td>434</td>
</tr>
<tr>
<td>GLXO</td>
<td>814</td>
<td>1885</td>
<td>390</td>
<td>225</td>
<td>138</td>
<td>8,927</td>
<td>8,573</td>
<td>434</td>
</tr>
<tr>
<td>LLOY</td>
<td>644</td>
<td>1485</td>
<td>302</td>
<td>159</td>
<td>120</td>
<td>11,376</td>
<td>11,690</td>
<td>434</td>
</tr>
<tr>
<td>ORA</td>
<td>314</td>
<td>884</td>
<td>153</td>
<td>57</td>
<td>129</td>
<td>9,502</td>
<td>8,597</td>
<td>434</td>
</tr>
<tr>
<td>RTR</td>
<td>408</td>
<td>978</td>
<td>201</td>
<td>127</td>
<td>164</td>
<td>10,065</td>
<td>9,243</td>
<td>431</td>
</tr>
<tr>
<td>SB</td>
<td>665</td>
<td>2056</td>
<td>195</td>
<td>195</td>
<td>100</td>
<td>12,358</td>
<td>12,157</td>
<td>426</td>
</tr>
<tr>
<td>SHEL</td>
<td>592</td>
<td>1387</td>
<td>277</td>
<td>156</td>
<td>156</td>
<td>44,165</td>
<td>30,133</td>
<td>434</td>
</tr>
<tr>
<td>VOD</td>
<td>940</td>
<td>2161</td>
<td>437</td>
<td>207</td>
<td>384</td>
<td>89,550</td>
<td>71,121</td>
<td>434</td>
</tr>
</tbody>
</table>

Fields from left to right: stock ticker symbol, total number of events (effective market orders + effective limit orders + order cancellations) in thousands, average number of events in a trading day, number of effective limit orders in thousands, number of effective market orders in thousands, number of order deletions in thousands, average limit order size in shares, average market order size in shares, number of trading days in the sample.
ing this time orders accumulate without transactions; then a clearing price for the opening auction is calculated, and all opening transactions take place at this price. Following the opening at 8:00 the market runs continuously, with orders matched according to price and time priority, until the market closes at 16:30. In the earlier part of the dataset, until September 22nd 1999, the market opening hour was 9:00. During the period we study there have been some minor modifications of the opening auction mechanism, but since we discard the opening auction data anyway this is not relevant.

Some stocks in our sample (VOD for example) have had stock price splits and tick price changes during the period of our sample. We take splits into account by transforming stock sizes and prices to pre-split values. In any case, since all measured quantities are in logarithmic units, of the form \( \log(p_1) - \log(p_2) \), the absolute price scale drops out. Our theory predicts that the tick size should change some of the quantities of interest, such as the bid-ask spread, but the predicted changes are small enough in comparison with the effect of other parameters that we simply ignore them (and base our predictions on the limit where the tick size is zero). Since granularity is much more important than tick size, this seems to be a good approximation.

3.5.4 Opening auction, real order types, time

Since the model does not take the opening auction into account, we simply neglect orders leading up to the opening auction, and base all our measurements on the remaining part of the trading day, when the auction is continuous.

In order to treat simply and in a unified manner the diverse types of orders traders can submit in a real market (for example, crossing limit orders, market orders with limiting price, ‘fill-or-kill, execute & eliminate) we use redefinitions based on whether an order results in an immediate transaction, in which case we call it an effective market order, or whether it leaves a limit order sitting in the book, in which case we call it an effective limit order. Marketable limit orders (also called crossing limit orders) are limit orders that cross the opposing best price, and so result in at least a partial transaction. The portion of the order that results in an immediate transaction is counted as an effective market order, while the non-transacted part (if any) is counted as an effective limit order. Orders that do not result in a transaction and do not leave a limit order in the book, such as for example, failed fill-or-kill orders, are ignored altogether. These have no effect on prices, and in any case, make up only a very small fraction of the order flow, typically less than 1%. Note that we drop the term “effective”, so that e.g. “market order” means “effective market order”.

43
A limit order can be removed from the book for many reasons, e.g. because the agent changes her mind, because a time specified when the order was placed has been reached, or because of the institutionally-mandated 30 day limit on order duration. We will lump all of these together, and simply refer to them as “cancellations”.

Our measure of time is based on the number of events, i.e., the time elapsed during a given period is just the total number of events, including effective market order placements, effective limit order placements, and cancellations. We call this event time. Price intervals are computed as the difference in the logarithm of prices, which is consistent with the model, in which all prices are assumed to be logarithmic in order to assure their positiveness.

### 3.5.5 Measurement of model parameters

We test the predictions of the model against real data cross-sectionally on eleven stocks. The parameters of the model are stated in terms of order arrival rates, cancellation rate, order size, and our dataset allows us to compute for each stock the average values of these parameters. As we explain here, these average rates are calculated as means of the daily values weighted by the daily number of events. An alternative would have been to calculate the mean values of the parameters over the entire 2 year period for each stock. While this works well for the parameters $\mu$ and $\sigma$, it does not work as well for $\alpha$ and $\delta$, as explained below.

**Measuring $\mu$ and $\sigma$**

The parameter $\mu_t$, which characterizes the average market order arrival rate on day $t$ is just the ratio of the number of shares of effective market orders (for both buy and sell orders) to the number of events during the trading day. Thus for $\mu$ it makes no difference whether we measure it across the whole period, or take a weighted average of daily values. This is also true for the average order size $\sigma_t$. One complication in measuring $\sigma$ is that the model assumes that the average size for limit orders and market orders is the same, whereas for the real data this is not strictly true. Nonetheless, as seen in Table 3.1, although the limit order size tends to be a bit larger than the market order size, it is still a fairly good approximation to take them to be the same. For the purposes of this analysis we use the limit order size to measure order size, based on theoretical arguments that this is more important than the market order size. In any case, this does not make a significant difference in the results.
CHAPTER 3. THE PREDICTIVE POWER OF ZERO INTELLIGENCE IN FINANCIAL MARKETS

Measuring $\alpha$ and $\delta$

The measuring of the cancellation rate $\delta_t$ and the limit order rate density $\alpha_t$ is more complicated, due to the highly simplified assumptions we made for the model. In contrast to our assumption of a constant density for placement of limit orders across the entire logarithmic price axis, real limit order placement is highly concentrated near the best prices. Roughly $\frac{2}{3}$ of all orders are placed either at the best price or inside the spread. Outside the spread the density of limit order placement falls off as a power law as a function of the distance from the best prices (Bouchaud et al., 2002; Zovko and Farmer, 2002). In addition, we have assumed a constant cancellation rate, whereas in reality orders placed near the best prices tend to be cancelled much faster than orders placed far from the best prices.

We cope with these problems by introducing an auxiliary assumption. Basically, we assume that order placement is constant inside an interval, and is zero outside that interval. This is described in more detail below.

In order to estimate the limit order rate density for day $t$, $\alpha_t$, we make an empirical estimate of the distribution of the relative price for effective limit order placement on each day. For buy orders we define the relative price as $\Delta = m - p$, where $p$ is the logarithm of the limit price and $m$ is the logarithm of the midquote price. Similarly for sell orders, $\Delta = p - m$. We then somewhat arbitrarily choose $Q_{t,\text{lower}}$ as the 2 percentile of the density of $\Delta$ corresponding to the limit orders arriving on day $t$, and $Q_{t,\text{upper}}$ as the 60 percentile of $\Delta$. Assuming constant density within this range, we calculate $\alpha_t$ as $\alpha_t = L/(Q_{t,\text{upper}} - Q_{t,\text{lower}})$ where $L$ is the total number of shares of effective limit orders within the price interval $(Q_{t,\text{lower}}, Q_{t,\text{upper}})$ on day $t$. The choice of $Q_{t,\text{upper}}$ is made in a compromise to include as much data as possible for statistical stability, but not so much as to include orders that are unlikely to ever be executed, and therefore unlikely to have any effect on prices.

Similarly, to cope with the fact that in reality the average cancellation rate $\delta$ decreases (Bouchaud et al., 2002) with the relative price $\Delta$, whereas in the model $\delta$ is assumed to be constant, we base our estimate for $\delta$ only on canceled limit orders within the range of the same relative price boundaries $(Q_{t,\text{lower}}, Q_{t,\text{upper}})$ defined above. We do this to be consistent in our choice of which orders are assumed to contribute significantly to price formation (orders closer to the best prices contribute more than orders that are further away). We then measure $\delta_t$, the cancellation rate on day $t$, as the inverse of the average lifetime of a canceled limit order in the above price range. Lifetime is measured in terms of number of events happening between the introduction of the order and its subsequent cancellation.
CHAPTER 3. THE PREDICTIVE POWER OF ZERO INTELLIGENCE IN FINANCIAL MARKETS

The parameter $Q_t^{\text{upper}}$ is referred to as $W$ in the main text. In other subsequent studies (to be reported elsewhere) we are able to set the parameter ($Q_t^{\text{lower}} = 0$, and to compute $\Delta$ relative to the opposite best price rather than the midprice, with negligiable differences in the results. The difference is that in the later studies we have a cleaner dataset. In this dataset there are some points that are clearly outliers, and it was convenient to introduce a lower cutoff for outlier removal. Thus, we do not feel that ($Q_t^{\text{lower}}$ is an important parameter for this analysis, and we have not discussed it in the text (where we have limited space).

The use of this procedure dictates that it is better to choose an av- erage of daily parameters rather than computing average parameters based on ratios of values for the whole period. Because the width of the interval over which orders are placed varies significantly in time. Moment-by-moment orderbook reconstruction makes it clear that the properties of the market tend to be relatively stationary during each day, changing more dramatically overnight. The order-flows on different days can be rather dissimilar. This non-stationarity of the order flow means that $\delta$ and $\alpha$ parameters calculation would perform poorly if we attempted to use an average price interval over the whole period. This would have the result that on some days we might count only a small fraction of the orderflow, excluding many orders that were important for price formation, while on other days we would include almost all the orders, including many that were not very relevant for price formation. This problem makes it natural to use daily averages of parameters.

This introduces the worry that daily variations in $W$ might be an important predictive variable, above and beyond its effect on changing $\alpha$ (which is consistent with the model). There is a tendency for the value of $W$ on a given day to track the spread, due to regulari-ties in order placement, and therefore to automatically have some correlation with the spread. We have done several studies, which will be reported in a future work, testing the importance of this effect. These show that while daily variations in $W$ do give additional predictability for the spread, other aspects of the model are substantially responsible for these results.

**Measuring the price diffusion rate**

The measurement of the price diffusion rate requires some discussion. We measure the intraday price diffusion by computing the mid-point price variance $V(\tau) = \text{Var}\{m(t+\tau) - m(t)\}$, for different time scales $\tau$. The averaging over $t$ includes all events that change the mid-point price. The plot of $V(\tau)$ against $\tau$ is called a diffusion curve and for an IID random walk is a straight line with slope $D$, the diffusion coefficient.
In our case, the computation of $D$ is as follows: For each day we compute the diffusion curve. In this way we avoid the overnight price changes which would bias our estimate. To the daily diffusion curve we then fit a straight line $V(\tau) = D\tau$ using least squares weighted by the square root of the number of observations for each value of $V(\tau)$. In fitting a straight line we are assuming IID mid-point price movement which is relatively well born out in the data. For an example of see Fig. 3.5. Averaging the daily diffusion rates we obtain the full sample estimate of the stock diffusion. We weigh the averaging by the number of events in each day. One must bear in mind that the price diffusion rate from day to day has substantial correlations, as illustrated in Fig. 3.6.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure3.5}
\caption{Illustration of the procedure for measuring the price diffusion rate for Vodafone (VOD) on August 4th, 1998. On the $x$ axis we plot the time $\tau$ in units of ticks, and on the $y$ axis the variance of mid-price diffusion $V(\tau)$. According to the hypothesis that mid-price diffusion is an uncorrelated Gaussian random walk, the plot should obey $V(\tau) = D\tau$. To cope with the fact that points with larger values of $\tau$ have fewer independent intervals and are less statistically significant, we use a weighted regression to compute the slope $D$.}
\end{figure}
Figure 3.6: Time series (top) and autocorrelation function (bottom) for daily price diffusion rate $D_t$ for Vodafone. Because of long-memory effects and the short length of the series, the long-lag coefficients are poorly determined; the figure is just to demonstrate that the correlations are quite large.

3.5.6 Estimating the errors for the regressions

The error bars presented in the text are based on a bootstrapping method. It may at first seem that the proper method would be to simply use White’s heteroskedasticity consistent estimators, however, we are driven to use this method for two reasons.

First, within each stock the daily values of the dependent variables display slowly decaying positive autocorrelation functions. Averaging the daily values to get an estimate of the stock-specific average may seem to remedy the autocorrelation problem. However, the autocorrelation is very persistent, almost to the scale of the length of our dataset, and the variables may indeed have long memory.\(^6\) This

\(^6\)It has recently been shown that order sign, order volume and liquidity as reflected by volume at the best price, are long-memory processes (Bouchaud
CHAPTER 3. THE PREDICTIVE POWER OF ZERO INTELLIGENCE IN FINANCIAL MARKETS

makes us suspicious about using standard statistics.

A second reason in using a bootstrapping method for inference is the fact that, in addition to possibly being long memory, the daily values of the variables are cross-correlated across stocks. (A high volatility in one stock on a particular day is likely to be associated with high volatility in other stocks.) These two reasons lead us to believe that using standard or White’s estimators would underestimate the regression errors.

The method we use is inspired by the variance plot method described in Beran (1994), Section 4.4. We divide the sample into blocks, apply the regression to each block, and then study the scaling of the deviation in the results as the blocks are made longer to coincide with the full sample. We divide the \( N \) daily data points for each stock into \( m \) disjoint blocks, each containing \( n \) adjacent days, so that \( n \approx N/m \). We use the same partition for each stock, so that corresponding blocks for each stock are contemporaneous. We perform an independent regression on each of the \( m \) blocks, and calculate the mean \( M_m \) and standard deviation \( \sigma_m \) of the \( m \) slope parameters \( A_i \) and intercept parameters \( B_i, i = 1, \ldots, m \). We then vary \( m \) and study the scaling as shown in Figs 3.7 and 3.8.

Figs 3.7(a) and (b) illustrate this procedure for the spread, and Figs 3.8(a) and (b) illustrate this for the price diffusion rate. Similarly, panels (c) and (d) in each figure show the mean and standard deviation for the intercept and slope as a function of the number of bins. As expected, the standard deviations of the estimates decreases as \( n \) increases. The logarithm of the standard deviation for the intercept and slope as a function of log \( n \) is shown in panels (e) and (f). For IID normally distributed data we expect a line with slope \( \gamma = -1/2 \); instead we observe \( \gamma > -1/2 \). For example for the spread \( \gamma \approx -0.19 \). \(|\gamma| < 1/2\) is an indication that this is a long memory process.

This method can be used to extrapolate the error for \( m = 1 \), i.e. the full sample. This is illustrated in panels (e) and (f) in each figure. The inaccuracy in these error bars is evident in the unevenness of the scaling. This is particularly true for the price diffusion rate. To get a feeling for the accuracy of the error bars, we estimate the standard deviation for the scaling regression assuming standard error, and repeat the extrapolation for the one standard deviation positive and negative deviations of the regression lines, as shown in panels (e) and (f) of Figs 3.7 and 3.8 The results are summarized in Table 3.2.

One of the effects that is evident in Figs 3.7(c-d) and 3.8(c-d) is that the slope coefficients tend to decrease as \( m \) increases. We believe this is due to the autocorrelation bias discussed in Section (3.5.6).

et al., 2004; Lillo and Farmer, 2003).
Figure 3.7: Subsample analysis of regression of predicted vs. actual spread. To get a better feeling for the true errors in this estimation (as opposed to standard errors which are certainly too small), we divide the data into subsamples (using the same temporal period for each stock) and apply the regression to each subsample. (a) (top left) shows the results for the intercept, and (b) (top right) shows the results for the slope. In both cases we see that progressing from right to left, as the subsamples increase in size, the estimates become tighter. (c) and (d) (next row) shows the mean and standard deviation for the intercept and slope. We observe a systematic tendency for the mean to increase as the number of bins decreases. (e) and (f) show the logarithm of the standard deviations of the estimates against log \( n \), the number of each points in the subsample. The line is a regression based on binnings ranging from \( m = N \) to \( m = 10 \) (lower values of \( m \) tend to produce unreliable standard deviations). The estimated error bar is obtained by extrapolating to \( n = N \). To test the accuracy of the error bar, the dashed lines are one standard deviation variations on the regression, whose intercepts with the \( n = N \) vertical line produce high and low estimates.
Figure 3.8: Subsample analysis of regression of predicted vs. actual price diffusion (see Fig. 6), similar to the previous figure for the spread. The scaling of the errors is much less regular than it is for the spread, so the error bars are less accurate.

<table>
<thead>
<tr>
<th>regression</th>
<th>estimated</th>
<th>standard</th>
<th>bootstrap</th>
<th>low</th>
<th>high</th>
</tr>
</thead>
<tbody>
<tr>
<td>spread intercept</td>
<td>0.06</td>
<td>0.21</td>
<td>0.29</td>
<td>0.25</td>
<td>0.33</td>
</tr>
<tr>
<td>spread slope</td>
<td>0.99</td>
<td>0.08</td>
<td>0.10</td>
<td>0.09</td>
<td>0.11</td>
</tr>
<tr>
<td>diffusion intercept</td>
<td>2.43</td>
<td>1.22</td>
<td>1.76</td>
<td>1.57</td>
<td>1.97</td>
</tr>
<tr>
<td>diffusion slope</td>
<td>1.33</td>
<td>0.19</td>
<td>0.25</td>
<td>0.23</td>
<td>0.29</td>
</tr>
</tbody>
</table>
CHAPTER 3. THE PREDICTIVE POWER OF ZERO INTELLIGENCE IN FINANCIAL MARKETS

3.5.7 Market impact

Relation of market impact to supply and demand schedules

The market impact function is closely related to the more familiar notions of supply and demand. We have chosen to measure average market impact in this paper rather than average relative supply and demand for reasons of convenience. Measuring the average relative supply and demand requires reconstructing the limit order book at each instant, which is both time consuming and error prone. The average market impact function, in contrast, can be measured based on a time series of orders and best bid and ask prices.

At any instant in time the stored queue of sell limit orders reveals the quantity available for sale at each price, thus showing the supply, and the stored buy orders similarly show the revealed demand. The price shift caused by a market order of a given size depends on the stored supply or demand through a moment expansion (Smith et al., 2003). Thus, the collapse of the market impact function reflects a corresponding property of supply and demand. Normally one would assume that supply and demand are functions of human production and desire; the results we have presented here suggest that on a short timescale in financial markets their form is dictated by the dynamical interaction of order accumulation, removal by market orders and cancellation, and price diffusion.

Alternative market impact collapse plots

We have demonstrated a good collapse of the market impact using nondimensional units. However, in deciding what “good” means, one should compare this to the best alternatives available. We compare to three such alternatives. In figure 3.9, the top left panel shows the collapse when using non-dimensional units derived from the model (repeated from the main text). The top right panel shows the average market impact when we instead normalise the order size by its sample mean. Order size is measured in units of shares and market impact is in log price difference. The bottom left panel attempts to take into account daily variations of trading volume, normalizing the order size by the average order size for that stock on that day. In the bottom right panel we use trade price to normalize the order sizes which are now in monetary units (British Pounds). We visually see that none of the alternative rescalings comes close to the collapse we obtain when using non-dimensional units; because of the much greater dispersion, the error bars in each case are much larger.
Figure 3.9: Market impact collapse under 4 kinds of axis rescaling. In each case we plot a normalised version of the order size on the horizontal axis vs. a (possibly normalised) average market impact \( \log(p_{t+1}) - \log(p_t) \) on the vertical axis. (a) (top left) collapse using non-dimensional units based on the model; (b) (top right) order size is normalised by its mean value for the sample. (c) (bottom right) order size is normalised the average daily volume. (d) (bottom right) Order size is multiplied by the current best midpoint price, making the horizontal axis the monetary value of the trade.
CHAPTER 3. THE PREDICTIVE POWER OF ZERO INTELLIGENCE IN FINANCIAL MARKETS

Error analysis for market impact

Assigning error bars to the average market impact is difficult because the absolute price changes $\Delta p$ have a slowly decaying positive autocorrelation function. This may be a long-memory process, although this is not as obvious as it is for other properties of the market, such as the volume and sign of orders (Bouchaud et al., 2004; Lillo and Farmer, 2003). The signed price changes $\Delta p$ have an autocorrelation function that rapidly decays to zero, but to compute market impact we sort the values into bins, and all the values in the bin have the same sign. One might have supposed that because the points entering a given bin are not sequential in time, the correlation would be sufficiently low that this might not be a problem. However, the autocorrelation is sufficiently strong that its effect is still significant, particularly for smaller market impacts, and must be taken into account.

To cope with this we assign error bars to each bin using the variance plot method described in, for example (Beran, 1994), Section 4.4. This is a more straightforward version of the method discussed in Section (3.5.6). The sample of size $N = 434$ is divided into $m$ subsamples of $n$ points adjacent in time. We compute the mean for each subsample, vary $n$, and compute the standard deviation of the means across the $m = N/n$ subsamples. We then make use of theorem 2.2 from (Beran, 1994) that states that the error in the $n$ sample mean of a long-memory process is $\hat{e} = \sigma n^{-\gamma}$, where $\gamma$ is a positive coefficient related to the Hurst exponent and $\sigma$ is the standard deviation. By plotting the standard deviation of the $m$ estimated intercepts as a function of $n$ we estimate $\gamma$ and extrapolate to $n = \text{sample length}$ to get an estimate of the error in the full sample mean. An example of an error scaling plot for one of the bins of the market impact is given in Fig. 3.10.

A central question about Fig. 3.9 is whether the data for different stocks collapse onto a single curve, or whether there are statistically significant idiosyncratic variations from stock to stock. From the results presented in Fig. 3.9 this is not completely clear. Most of the stocks collapse onto the curve for the pooled data (or the pooled data set with themselves removed). There are a few that appear to make statistically significant variations, at least if we assume that the mean value of the bins for different order size levels are independent. However, they are most definitely not independent, and this non-independence is difficult to model. In any case, the variations are always fairly small, not much larger than the error bars. Thus the collapse gives at least a good approximate understanding of the market impact, even if there are some small idiosyncratic variations it does not capture.
Figure 3.10: The variance plot procedure used to determine error bars for mean market impact conditional on order size. The horizontal axis $n$ denotes the number of points in the $m$ different samples, and the vertical axis is the standard deviation of the $m$ sample means. We estimate the error of the full sample mean by extrapolating $n$ to the full sample length.

**Market impact in log-log coordinates**

If we fit a function of the form $\phi(\omega) = K\omega^\beta$ to the market impact curve, we get $\beta = 0.26 \pm 0.02$ for buy orders and $\beta = 0.23 \pm 0.02$ for sell orders, as shown in Fig. 3.11. The functional form of the market impact we observe here is not in agreement with a recent theory by Gabaix et al. (2003a), which predicts $\beta = 0.5$. While the error bars given are standard errors, and are certainly too optimistic, it is nonetheless quite clear that the data are inconsistent with $\beta = 1/2$, as discussed in Ref. (Farmer and Lillo, 2004). This relates to an interesting debate: The theory for average market impact put forth by Gabaix et al. follows traditional thinking in economics, and postulates that agents optimize their behavior to maximize profits,
while the theory we test here assumes that they behave randomly, and that the form of the average market impact function is dictated by the statistical mechanics of price formation.

Figure 3.11: The average market impact vs. order size plotted on log-log scale. The upper left and right panels show buy and sell orders in non-dimensional coordinates; the fitted line has slope $\beta = 0.26 \pm 0.02$ for buy orders and $\beta = 0.23 \pm 0.02$ for sell orders. In contrast, the lower panels show the same thing in dimensional units, using British pounds to measure order size. Though the exponents are similar, the scatter between different stocks is much greater.

### 3.5.8 Extending the model

In the interest of full disclosure, and as a stimulus for future work, in this section we detail the ways in which the current model does not accurately match the data, and sketch possible improvements. This model was intended to describe a few average statistical properties of the market, some of which it describes surprisingly well. However, there are several aspects that it does not describe well. Fixing the problems with this models require a more sophisticated model of order flow, including a more realistic model of price dependence in order placement and cancellations (Bouchaud et al., 2002; Zovko
and Farmer, 2002), long-memory properties (Bouchaud et al., 2004; Lillo and Farmer, 2003) and the relationship of the different components of the order flow to each other. This is a much harder problem, and is likely to require a more complicated model. Members of our group are actively working on this problem. While this will certainly have many advantages over the current model, it will also have the disadvantage of introducing more free parameters and thereby complicating the scaling laws (and making the possibility for analytic results more remote).

One of the major ways in which this model is not realistic concerns price diffusion. Real price increments are roughly white, i.e., they are roughly uncorrelated. One might naively think that under IID Poisson order flow, price increments should also be IID. However, due to the coupling of boundary conditions for the buy market order/sell limit order process to those of the sell market order/buy limit order process, this is not the case. Because of the fact that supply and demand tend to build (i.e. the depth of standing limit orders increases) as one moves away from the center of the book, price reversals are more common than price changes in the same direction. As a result, the price increments generated by this model are more anti-correlated than those of real price series. This has an interesting consequence: If we add the assumption of market efficiency, and assume that real price increments must be white, it implies that real order flow should be positively autocorrelated in order to compensate for the anticorrelations induced by the continuous double auction. This has indeed subsequently been observed to be the case (Bouchaud et al., 2004; Lillo and Farmer, 2003).

One of the side effects of this anticorrelation of prices is that it implies that there exist arbitrage opportunities that can be taken advantage of by an intelligent agent. A separate study of these arbitrage opportunities makes it clear that they are not risk-free in the sense usually used in economics. That is, taking advantage of them necessarily involves taking risk, and they do not permit arbitrarily large profits – returns decrease with the size of investment and eventually go to zero. Exploring the nature of these arbitrage opportunities, and the effect that exploiting them has on prices is one of the directions in which this model can be improved (one that is being actively explored). However, we do not feel that the existence of such arbitrage opportunities (which in our opinion mimic those of real markets) presents a serious problem for the purposes for which we are using this model.

In the following bullets we summarize the main directions in which members of our group are working to improve this model.

- *Price diffusion.* The variance of real prices obeys the relation-
ship $\sigma^2(\tau) = D\tau^{2H}$ to a good approximation for all values of $\tau$, where $\sigma^2(\tau)$ is the variance of price changes or returns computed on timescale $\tau$. The Hurst exponent $H$ is close to and typically a little greater than 0.5. In contrast, under Poisson order flow, as already discussed above, due to the dynamics of the double continuous auction price formation process, prices make a strongly anti-correlated random walk. This means that the function $\sigma^2(\tau)$ is nonlinear. Asymptotically $H = 0.5$, but for shorter times $H < 0.5$. Alternatively, one can characterize this in terms of a timescale-dependent diffusion rate $D(\tau)$, so that the variance of prices increases as $\sigma^2(\tau) = D(\tau)\tau$. Refs. (Daniels et al., 2003; Smith et al., 2003) showed that the limits $\tau \to 0$ and $\tau \to \infty$ obey well-defined scaling relationships in terms of the parameters of the model. In particular, $D(0) \sim \mu^2 \delta/\alpha^2 \epsilon^{-1/2}$, and $D(\infty) \sim \mu^2 \delta/\alpha^2 \epsilon^{1/2}$. Interestingly, and for reasons we do not fully understand, the prediction of the short term diffusion rate $D(0)$ does a good job of matching the real data, as we have shown here, while $D(\infty)$ does a much poorer job.

- **Market efficiency.** The question of market efficiency is closely related to price diffusion. The anti-correlations mentioned above imply a market inefficiency. We are investigating the addition of “low-intelligence” agents to correct this problem.

- **Correlations in spread and price diffusion.** We have already discussed in Section (3.5.6) the problems that the autocorrelations in spread and price diffusion create for comparing the theory to the model on a daily scale. This is related to the fact that this model does not correctly capture either the fat tails of price fluctuations or the long-memory of volatility.

- **Lack of dependence on granularity parameter.** In Section (3.5.7) we discuss the fact that the model predicts more variation with the granularity parameter than we observe. Apparently the Poisson-based non-dimensional coordinates work even better than one would expect. This suggests that there is some underlying simplicity in the real data that we have not fully captured in the model.

Although in this paper we are stressing the fact that we can make a useful theory out of zero-intelligence agents, we are certainly not trying to claim that intelligence doesn’t play an important role in what financial agents do. Indeed, one of the virtues of this model is that it provides a benchmark to separate properties that are driven by the statistical mechanics of the market institution from those that are driven by conditional intelligent behavior.