A new approach to understanding the market impact of large trading orders

J. Doyne Farmer,1,2,† Austin Gerig,1,‡ Fabrizio Lillo,3,1,§ and Henri Waelbroeck4,¶

1Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501
2LUISS Guido Carli, Viale Pola 12, 00198, Roma, Italy
3Dipartimento di Fisica e Tecnologie Relative, viale delle Scienze I-90128, Palermo, Italy
4Pipeline Financial Group, Inc., 60 E. 42nd St., Suite 624, New York NY 10165

Market impact is the change in price due to initiating a trade. In this paper we develop a new theory for average market impact based on properties of order flow, efficiency of price returns and other empirically testable assumptions. Our approach differs from previous efforts in that our results do not depend on assumptions about the functional form of utility or other ad hoc assumptions. We argue that the concave form of market impact is driven by market efficiency as follows: Large trading orders are executed incrementally, which imparts positive autocorrelations into order flow. To be compatible with efficiency the resulting predictability implies that returns must be asymmetric – returns of the same sign as the large order being executed must be smaller than those of the opposite sign. As the large order develops it becomes more and more predictable, which makes same-signed returns smaller and and thus makes the market impact concave. Under our theory the market impact depends on several factors, including the volume distribution of large orders, the information set used by market participants, and the way in which hidden orders couple to price volatility. In this paper we assume no coupling to volatility and show that depending on which of these assumptions is used it is possible to get linear, power law, or logarithmic impact functions. We present empirical results using data from the London Stock Exchange that are suggestive of logarithmic impact, but more data is needed to determine this with certainty.

PRELIMINARY DRAFT.
PLEASE DO NOT DISTRIBUTE.

* This paper is the first of two papers in which we develop a theory of market impact for large trading orders. We have divided this into two parts because there are two basic lines of attack with different solutions, depending on the assumptions about scale and information revelation, as discussed in Section V.

†jdf@santafe.edu
‡gerig@santafe.edu
§lillo@santafe.edu
¶henri.waelbroeck@pipelinefinancial.com
I. INTRODUCTION

Market impact measures the expected price change induced by initiating a trade\(^1\). One of the puzzles of finance is that the market impact as a function of trading volume consistently has a highly concave functional form, i.e. its derivative is a decreasing function of volume. The main goal of this paper is to explain this, and to present a quantitative theory for market impact and its dependence on trading volume. Previous approaches to deriving market impact have been based on utility maximization, which has the problem that the impact function derived depends on the functional form assumed for the utility function. Empirical evidence does not support a unique correct functional form for utility, which calls into question the results of any theory that depends on this. In contrast, we derive a market impact function here based on a no-arbitrage argument and other assumptions that are empirically testable, in the style of Black and Scholes. This has the important advantage that we produce a functional form without effectively assuming it at the outset, making our theory falsifiable.

We wish to make it clear that this paper is not the final word on market impact. As we describe in detail in the rest of the paper, there are several points in the argument where there is a choice of several possible assumptions, and the correct answer depends on this choice. Determining which of these choices is correct will have to wait for more comprehensive empirical verification and more complete theoretical development. Our empirical results are preliminary and are intended to demonstrate the plausibility of what we have done and indicate an approach; obtaining a definitive answer will require analyzing much more data. Nonetheless, despite the caveats above, in this paper we make contribute to the literature on this subject by presenting a new approach. A future paper will investigate an alternative set of assumptions within this basic framework.

Understanding market impact is important conceptually for what it tells us about the aggregation of information about trades into prices, and also for what it tells us about the underlying behavior of supply and demand. It is closely related to the demand elasticity of price, originally introduced by Alfred Marshall\(^2\). Knowing market impact (or demand elasticity) does not allow one to compute absolute price levels, but it does make it possible to answer the hypothetical question, “How much would I move the price if I were to make a trade?” From an information point of view, trades should be informative, in the sense that in a world in which each agent has different information, a trade and its incorporation into the price make this information public. One of our main accomplishments here is in developing a theory for how information about order flow is incorporated into prices. In particular, we explore two different models for the information that market participants can use to predict order flow and show that these lead to very different functional forms for

\(^1\) “Initiation” refers to the party who is immediately responsible for a trade taking place. For example in a continuous double auction the initiating party is the one who places the order that causes an immediate transaction. This can also be defined in other market structures where actions are taken sequentially.

\(^2\) For asynchronous market clearing, in which parties change their supply or demand functions one at a time, it is possible to show that the market impact for a series of trades is linearly proportional to the demand elasticity of price averaged over a sequence of trades (Farmer, Gerig and Lillo, in progress).
Understanding market impact also has important practical implications. Practitioners care about understanding market impact because it reduces profits. Because market impact increases with trading size it places a limit on fund size. By adversely moving the prices at which transactions are made it can turn a profitable strategy into a losing strategy. This is particularly problematic for high-frequency trading, and is the reason why savvy hedge funds close once they reach a critical size. The functional form matters: If market impact increases rapidly with volume, a fund’s size is severely limited; if it increases slowly, the fund can grow much larger. Of course liquidity in markets varies enormously depending on context (see e.g. Gillemot, Farmer and Lillo, 2006), but what we are interested in here is the market impact as a function of volume averaging over different market conditions.

The functional form of market impact is one of the unexplained puzzles in finance. There have been many empirical studies of market impact. All of them have observed a concave function of volume, i.e. that the derivative of market impact is a decreasing function. The functional forms that have been reported to give a good fit to the data vary from study to study. We believe that much of this variation comes about because these studies actually measure different things. Some of them measure the market impact of a single trade made in an order book, some measure the aggregate impact of sequential trades in an order book, some of them measure block trades, and many of them measure a mixture of all three. We believe that these follow different laws and need to be analyzed separately. This paper is one of a series in which we develop theories for market impact for each of the above cases.

In this paper we develop a theory for the total market impact of large trades that are executed sequentially in small pieces. We call such trades hidden orders; they are also sometimes called trading packages. The strategic reasons for incremental execution were originally analyzed by Kyle (1985), who developed a model for a trader with inside information, and showed that if market makers are efficient this trader will split her trade into pieces and the information will be gradually incorporated into the price as each trade is made. Our theory extends Kyle’s original model in several directions. First, we base our model on empirically measurable assumptions. Second, we show that sequential execution of a large order introduces predictable structure into order flow that has to be removed in the stream of corresponding prices in order to guarantee efficiency. The importance of the predictability of order flow for price formation was presaged by Hasbrouck (1988, 1991), who stressed that only the unpredictable component of order flow can affect market impact, and was further developed by Bouchaud et al. (2004) and Lillo and Farmer (2004). Our model leads to quantitative predictions about the scale and functional form of the market impact of large trades. We develop predictions for market impact under two different assumptions about the predictability of order flow that can be viewed as extreme cases. Most importantly, our model produces falsifiable predictions.

We follow in the footsteps of many attempts to model price formation. We have already mentioned Kyle (1985), who assumed a normal distribution for reservation prices and derived a linear market impact function. Keim and Madhavan (1996) modeled block trades based

on the cost of searching for counter-parties, and under the assumption that this increased as a power law in the number of counter-parties, derived a power law impact function. Evans and Lyons (2002) developed a theory for interdealer and public trading, and under the assumption that the public’s demand function is linear derived a linear impact function. Gabaix et al. (2003, 2006) assumed that block traders have an additive risk aversion term in their utility function of the form $V^{\delta/2}$, where $V$ is the variance of their profits. Under the assumption of rational utility maximization and a random walk for prices, they showed that the impact of large trades should scale as $V^{\delta/2}$, where $V$ is the trading volume. If $\delta = 1$, i.e. first order risk aversion, their theory predicts impact increases as the square root of volume. Each of these theories adds insight, but suffer from the problem that the functional form of the answer depends on the functional form assumed in originally setting up the model.

Our model adds to this literature by developing a new theory based on assumptions that are empirically verifiable, which makes predictions that can be tested without introducing any parameters that cannot be independently measured. We first present empirical evidence showing that the predictability of order flow is mainly due to sequential execution of hidden orders. Under the assumption that hidden orders are random we show how the predictability of order flow depends on the size distribution of hidden orders. Under the assumption that prices are efficient we then derive the market impact function. The model depends on the probability distribution for the size of hidden orders, which we argue for large orders is well-approximated by a power law, and on the information participants have about hidden orders. We propose two possible models that in a sense bracket the space of possibilities for inferring the status of hidden orders: One of them assumes participants use a linear time series model based on public information about order flow; the other assumes a more detailed structural break model based on private information. These lead to different testable predictions for the functional form of market impact, which then allow us to infer the information structure; we find that the detailed model provides a better fit.

We also provide preliminary empirical tests of our model. Empirical studies of the market impact of hidden orders are difficult to perform because linking realized trades together requires information about the identity of trading parties. Previous studies by Chan and Lakonishok (1993, 1995), and Torre (1997) were based on data from a brokerage that made it possible to explicitly track the orders of each client. Our approach follows the lead of Vaglica et al. (2007), who have developed a method for reconstructing large orders using information about brokerage codes. We use data from the London Stock Exchange containing codes identifying the member of the exchange submitting each trading order, and develop a simple algorithm for linking together a series of realized trades into underlying hidden orders. This approach has the disadvantage that it is impossible to classify hidden orders unambiguously, as there is always the possibility that two parties are trading simultaneously through the same member of the exchange. In fact we are able to estimate the probability that this happens, and so estimate the misclassification rate. Our approach has the advantage that it allows us to classify all trading orders for the whole market, and so gives us more data to work with than we would have if we were restricted to the clients of a single brokerage.

II. OUTLINE

Since the arguments in this theory do not follow the standard template of most economics theory papers, and since the arguments are complicated, we first present an overview of the arguments which can act as a guide for understanding how the sections of the paper fit
together.

In Section III we present the key empirical facts that the theory depends on. We review what is known about the predictability of order flow and argue that the main driver of the predictability of order flow is the sequential execution of hidden orders whose size is distributed as a power law. We present several pieces of empirical evidence for this: (1) In several different stock markets signed order flow is observed to be a long-memory process whose autocorrelation exponent decays as a power law with exponent $\gamma \approx 0.5$. (2) Volume distributions in block markets are power laws with tail exponent $\alpha \approx 1.5$. (3) A previously developed theory for incremental execution of hidden orders predicts that under the assumption of a uniform execution rate $\alpha = 1 + \gamma$, in agreement with the data. (4) New results show that for data with brokerage codes order flow trades made under the same brokerage code have long-memory and those made under different brokerage codes do not.

In Section IV we present the basic set up of the model. Under the assumption that the predictability of order flow comes from hidden orders we derive a relationship connecting the persistence of hidden order flow to the predictability for realized order flow. We then derive the consequences of this on returns, and show that it leads to an asymmetry in expected returns. This in turn is used to derive a general formula for market impact. The basic reason for concave market impact can be understood as follows: As a hidden order develops its probability of continuing does not remain constant. This implies that the ratio of returns of the same sign (as the hidden order) to those of the opposite sign decreases. This makes the market impact concave.

The previous argument allows us to compute the ratio of buy and sell returns, but it does not allow us to determine their scale. In Section V we discuss several assumptions about how the overall scale of returns varies during the course of a hidden order. We discuss three possible assumptions: (1) The scale does not change during a hidden order, (2) the response to buying and selling are symmetric, and (3) the scale is set by arbitrage conditions between liquidity providers and liquidity domanders. We show that (1) and (2) give approximately the same results. In contrast, because (3) is more complicated than (1) and (2) and because it requires the introduction of other conditions, we do not develop here, deferring it to a future paper.

In Section VI we develop two different ideas about how participants model the predictability of order flow. The first assumes that participants use a linear time series model to predict order flow based on public information; the second assumes more detailed information about hidden orders, including the underlying distribution of sizes, the number of hidden orders present and how much each order has executed so far. We derive the market impact function for each. We first do this under what we believe to be the correct assumption for hidden order volumes, namely that their sizes are power law distributed, and then to develop an alternative null hypothesis, under the assumption that the distribution is a stretched exponential. The results are not obvious. Under the detailed model, for example, an exponential distribution of hidden order volume leads to a linear market impact function, a stretched exponential leads to a power law market impact, and a power law hidden order distribution leads to a logarithmic impact function.

In Section VII we present some preliminary empirical results testing these models. We first describe the method we use for approximately identifying hidden orders using only information about exchange membership and discuss the advantages and disadvantages of this method relative to other possibilities.

In Section VIII we discuss some of the strengths and weaknesses of the present theory. The
problems include (1) the arbitrariness of the symmetry or constant volatility assumption, (2) the possibility that we need to incorporate other sources of price change, and (3) concerns about the information revelation incentives suggested by these results. In particular, these results suggest that there are circumstances where it is to the advantage of a liquidity taker to reveal information about his or her trading. We discuss ways to fix these problems.

Finally in Section IX we summarize and provide a broader perspective on these results, and discuss future work.

III. PREDICTABILITY OF ORDER FLOW

Our central thesis in this paper is that the juxtaposition of the predictability of order flow and market efficiency determine the average market impact. In this section we review the empirical evidence for predictability in order flow, emphasizing that the predictability is strong and highly regular, and that the dominant cause of this predictability is order splitting.

A. Summary of empirical observations

Positive serial autocorrelation of signed order flow has been observed by many authors in many different markets, including for the Paris Bourse by Biais, Hillion and Spratt (1995), in foreign exchange markets by Danielsson and Payne (2001), and the NYSE by Ellul et al. (2005) and Yeo (2006). However the first autocorrelation, which was emphasized in these papers, only captures a small part of the story. In the London, Paris, and New York stock exchanges it is empirically observed that the signs of trades in stock markets obey a long-memory process [Bouchaud et al. (2004), Lillo and Farmer (2004)]. This means that the autocorrelation function $C(\tau)$ of the signs $\epsilon_t$ of transactions decays in time as $C(\tau) \sim \tau^{-\gamma}$, where $0 < \gamma < 1$. The autocorrelation function of a long-memory process is not integrable in $\tau$ between 0 and $+\infty$ and, as a consequence, such processes do not have a typical time scale. Long-memory processes can be characterized by the exponent $\gamma$ of the autocorrelation function or equivalently in terms of the Hurst exponent $H = 1 - \gamma/2$. The measurement of long-memory is very robust; in London, for example, every stock examined shows long-memory at highly statistically significant levels (Lillo and Farmer, 2004). Whether long-memory also exists in other types of markets (FX, interest rates, interest rates, commodities, etc.) is not known.

B. Order splitting is the dominant cause of order flow predictability

Why is order flow so predictable? Parlour (1998) developed a theory for predictability of order flow based on strategic considerations that predicted that the signs of successive limit

---

4 In economics it has unfortunately become common to use the term “long-memory” to refer to fractionally integrated processes. We use the term in its more general sense to mean any process whose autocorrelation function is non-integrable (Beran, 1994). This can include processes with structure breaks, such as that studied by Ding, Engle and Granger (1993).
orders should be anti-correlated, but in fact this is just the opposite of what is observed\(^5\). Another possible explanation is herding, which refers to the idea that investors influence each other and so tend to act in unison. While this may play a role in introducing weak autocorrelations, as we will demonstrate in a moment, this is a relatively minor effect.

Instead we find that the dominant cause of the predictability of order flow is order splitting. Large desired trades, which we call **hidden orders**, are broken up into small pieces and executed over a sustained period of time through a series of **realized orders**, causing positive autocorrelations in the realized order flow. The strategic motivations for order splitting were originally demonstrated by Kyle (1985). Based on a few empirically well-supported assumptions, Lillo, Mike and Farmer (2005) developed a quantitative theory for order splitting and showed that it is in good agreement with the data. They assume that the signs of hidden orders are chosen in an IID manner. Large hidden orders of size \( V \) are split into \( N \) pieces of constant size \( \bar{v} = V/N \) and executed at a uniform rate, independent of \( V \). Under the simplest version of the theory, at any given time there are a fixed number \( M \) of active hidden orders. Hidden orders are selected for execution at random, and as each execution happens its size is decremented by \( \bar{v} \). When the size goes to zero a new hidden order with a random size and sign takes its place. Under the assumption that the size distribution of hidden orders scales asymptotically for large sizes as \( P(V > x) \sim x^{-\alpha} \), where \( \alpha \) is a positive constant, they showed that in the limit \( \tau \to \infty \) the resulting autocorrelation \( C(\tau) \) of realized order signs is asymptotically a power law \( C(\tau) \sim \tau^{-\gamma} \), with \( \gamma = \alpha - 1 \).

The theory is in good agreement with the data. The assumption of a power law tail with tail exponent \( \alpha \approx 1.5 \) for the hidden order size distribution is originally motivated by the observations of trades in the NYSE\(^7\) by Gopikrishnan et al. (2000). This same behavior was observed in the LSE by Lillo, Mike and Farmer, where it is possible to separate block trades from order book trades. The distribution of block trades is consistent with the power law hypothesis with \( \alpha \approx 1.5 \), but the distribution of on-book market trades is not – they have a much thinner tail. The autocorrelation function for realized trades predicted by the theory matches well with that observed, and in particular the predicted value of the exponent of the autocorrelation function under the predicted relationship \( \gamma = \alpha - 1 \) is satisfied by the data. Empirical analysis of the Spanish Stock Exchange using data with brokerage codes has shown that \( V \sim N \), i.e. uniform execution rate is a good assumption (Vaglica, 2007), and we have also verified this in the LSE.

We now present additional evidence for the hypothesis that the dominant cause of long-memory is order splitting. This analysis takes advantage of the fact that we have the membership code associated with each trade (see the discussion of the data in Section VII A). We assume that individual hidden orders are executed primarily through a single member of the LSE and compare the autocorrelation function for trades with the same membership code to those with different membership codes. Figure 1 shows the result for AZN; the results for other stocks are essentially the same. The autocorrelation functions of the signs of trading orders from the same membership code and all membership codes both decay roughly as

\(^5\) In the LSE we observe that the signs of orders are positively correlated, regardless of whether they are limit or market orders. This was also observed by Yeo (2006).

\(^6\) For convenience \( N \) is an integer and \( V \) is in increments of \( \bar{v} \). This has only a minor effect on the results.

\(^7\) The NYSE data includes a mixture of order book trades and block trades; the much larger block trades dominate the tail behavior, and since they provide an alternative mechanism for executing large orders, can be considered as a proxy for hidden orders.
FIG. 1: Autocorrelation of trade signs as a function of transaction sequence for transactions with same brokerage code (green triangles), different brokerage code (red x), and all transactions irrespective of brokerage code (blue circles), plotted on double logarithmic scale. To suppress statistical fluctuations autocorrelation coefficients with nearby lags are averaged.

a power law, as indicated by their approximation to a straight line on log-log scale. More rigorous statistical tests based on the methods used in Lillo and Farmer (2004) confirm this. In contrast, the autocorrelation function for realized order signs from different brokerages decays rapidly, and is clearly not a power law. By lag 10 there are already negative values; because we are using a logarithmic scale we cannot even plot them. In contrast, for the data based on the same brokerage codes, all autocorrelations are positive out to lags of 1000.

To summarize, there are three empirical results supporting the hypothesis that order splitting is the primary cause of long-memory in trading signs. These are (1) observations of trading volume distributions consistent with power laws in block markets with exponent $\alpha \approx 1.5$; (2) agreement of observed long-memory of order-book transactions with the predicted relation $\alpha = \gamma + 1$; and (3) disappearance of long-memory for orders with different brokerage codes. While there are likely to be other factors that contribute to the predictability of order flow, order splitting of hidden orders whose volume is drawn from a power law-tailed distribution seems to be the dominant cause.

C. Reconciling efficiency with long-memory

The long-memory of order flow presents a challenge for market efficiency: Order flow is highly predictable using simple models, but prices cannot be. Using a simple linear time series model based on past order flow it is possible to make accurate predictions of future order flow. Since the order flow follows a long-memory process the prediction accuracy decays
slowly as a power law in time, in contrast to the rapid exponential decay of a standard ARMA process. If prices were similarly predictable there would be huge arbitrage opportunities. This is exacerbated by the non-integrability of the autocorrelation, which implies that if each trade \( v_i \) has a fixed permanent response \( f(v_i) \), the aggregate cumulative response due to an infinite sequence of past trades should be infinite.

Two different solutions to this puzzle have been proposed, one due to Bouchaud et al. (2004, 2006), and the other by Lillo and Farmer (2004). The suggestion of Bouchaud et al. is that the impact of each trade decays to zero. To be more specific, they assume that the market impact is additive, and each individual term can be written as the product of a volume dependent component \( f \) (which is a fixed function) and a time dependent component \( G \). The net price change felt at time \( t \) by trades from an arbitrarily chosen initial time \( 1 \) to a later time \( t \) is

\[
\log p_t - \log p_1 = \sum_{i=1}^t \epsilon_i f(v_i) G(t-i) + n'_t, \tag{1}
\]

where \( n'_t \) is uncorrelated noise. If \( G(\tau) \) has a permanent part, i.e. \( \lim_{\tau \to \infty} G(\tau) > 0 \), then because the sign sequence \( \epsilon_i \) is long-memory the asymptotic market impact of any given trade is infinite. In contrast, providing \( G \) goes to zero as \( G(\tau) \sim \tau^{-\phi} \) for large \( \tau \), where

\[
\phi = (1 - \gamma)/2 = H - 1/2, \tag{2}
\]

they show the asymptotic market impact will remain finite. With an appropriate choice of constants (which is satisfied by the data) this is less than the spread.

The alternate suggestion of Lillo and Farmer is that the price response to trades has a permanent component but varies in time (i.e. \( f \) is no longer a fixed function). To prevent the long-memory from being transmitted to the price there must be an asymmetry between the price response of buying and selling, that depends on the history of order flow. That is, if buy orders are more likely then the impact of buy orders diminishes in relation to the impact of sell orders, so that the market remains efficient. In fact, as originally postulated by Farmer et al. (2006) and shown by Gerig (2007), these two approaches are equivalent. The basic reason is that under the imposition of efficiency the structure of the long-memory asymptotically causes the time varying permanent impacts to cancel, so that the summer impacts behave as if they were decaying to zero.

In this paper we take these insights further in several directions. First, we show that the constraint of efficiency in the face of long-memory can be used to deduce the functional form of the market impact caused by a sequence of trades. Second, we show that a linear time series model is not necessarily the best model for order flow, and that changing the model of order flow can fundamentally change the predicted impact function.

### IV. THEORY OF MARKET IMPACT UNDER RETURN EFFICIENCY

As we have already emphasized, the fact that hidden orders are only executed incrementally makes order flow extremely predictable. Enforcing market efficiency in the face of this predictability imposes strong constraints on the response of prices that make it possible to derive the market impact function. In this section we first derive a relation between the structure of the hidden order process and the predictability of order flow. We then impose efficiency and derive a general relation between the predictability of hidden orders and returns.
For the sake of clarity we develop this theory at a microscopic level, i.e. we formulate the model at the level of individual transactions. This has the disadvantage of requiring close attention to detail, but it has the advantages of making everything very specific and most importantly, it allows us to match everything precisely to empirical data.

A. Basic set up of the model

Following Lillo, Mike, and Farmer (2005) we assume that all buying and selling decisions correspond to hidden orders of size $V$ that are executed through a series of realized orders in a continuous double auction. The signs of hidden orders, $\epsilon = +1$ for buy and $\epsilon = -1$ for sell, are chosen at random in an IID manner and their sizes are drawn from a distribution $P(V)$. If a hidden order is sufficiently small it is executed immediately, but if it is too large it is broken up into $N$ realized orders and executed incrementally. For convenience we assume that all realized orders have the same size $\bar{v}$, and that hidden orders come only in discrete sizes $N\bar{v}$, where $N$ is an integer. (This simplifies the analysis and does not substantially affect any of the conclusions). We will say that a hidden order is active during the period it is being executed; $M(t)$ hidden orders can be active at the same time. While a given hidden order $j$ is active, for many purposes it is useful to treat its execution as a Poisson process in which a realized order has probability $\pi_j$ of originating from $j$. We will call $\pi_j$ the participation rate. Under the Poisson approximation the average time between executions of a given order $j$ is $\theta_j = 1/\pi_j$. Since we are using transaction time $\theta_j - 1$ is just the average number of intervening executions coming from other hidden orders. Even though the signs of hidden orders are IID, as we will describe in detail in the next section, order flow can be persistent. Information about the status of a hidden order, such as the number of executions $n_j$ that have already occurred, can make it possible to compute the probability $P(n_j)$ that it will continue, i.e. that $N_j > n_j$. As we will make clearer in Section VI, this depends on the unconditional distribution $P(V)$ of hidden order sizes. We will often drop the subscript $j$ when it is obvious.

As a simplifying assumption we assume that all hidden orders are executed with market orders. Since hidden orders are typically executed by a mixture of market and limit orders this is a fairly restrictive assumption. All these results can be generalized to apply to hidden orders with more general execution strategies, but this introduces some complications that we would rather avoid for the moment. So for now we assume pure market order executions, which implies a one-to-one correspondence between realized orders and transactions.

Another important aspect of the set-up of the model is the way in which we measure returns. We work in transaction time $t$, which is incremented as $t' = t + 1$ every time a transaction occurs. Prices are measured in terms of the logarithmic midprice $p = \log(1/2(a+b))$, where $a$ is the best offered selling price (the best ask) and $b$ is the best offered buying price (the best bid). The return $r_t = p_t - p_{t-1}$ is measured as the logarithmic price difference from immediately after a transaction is received until immediately after the next transaction is received. This means that returns include the immediate impact of any price change caused by the transaction, as well as any subsequent changes due to quote driven price changes, i.e. those due to the receipt of limit orders inside the spread or cancellation of all orders at the best bid or offer.

It is also useful to measure returns in units of hidden order executions. Letting $t_0$ be
the time of the decision to begin executing a hidden order\(^8\), and \(t_n\) be the time of the \(n^{th}\) execution, the hidden order return \(\rho_n = p_n - p_{n-1}\), where \(p_n\) is the logarithmic midprice at time \(t_n\). (As before we snapshot the price immediately after the execution occurs).

Of course none of our conclusions depend on these details, but to develop the model it is necessary to be specific.

B. Predictability of order flow due to hidden orders

Information about the presence or absence of hidden orders makes order flow predictable. We now derive a general relationship between the participation rate \(\pi\), the hidden order continuation probability \(\mathcal{P}\), and the predictability of the order flow. Let \(p^+_t = P(\epsilon_t = \epsilon_j)\) be the probability that the sign \(\epsilon_t\) of a realized order at time \(t\) is the same as the sign \(\epsilon_j\) of hidden order \(j\), and let \(p^-_t\) be the probability that it has the opposite sign. By assumption the signs of any two hidden orders \(j\) and \(k\) are independent. For convenience also assume that the unconditional probabilities of buy and sell orders are the same (the results are easily generalized to avoid this, but this complicates the calculation and only trivially alters the results). Suppressing the ubiquitous time indices, this allows us to write \(p^+_t\) in the form

\[
p^+ = \left(\pi + \frac{1}{2}(1 - \pi)\right)\mathcal{P} + \frac{1}{2}(1 - \mathcal{P}) = \frac{1}{2} (1 + \pi \mathcal{P}). \tag{3}
\]

This can understood by examining the first expression, starting with the leftmost term: By definition the probability that a hidden order continues is \(\mathcal{P}\). If it continues then there is a probability \(\pi\) that it has a transaction (which automatically has the same sign), and probability \(1 - \pi\) that another hidden order has a transaction, in which case there is a probability of one half of having the same sign. The probability that the order does not continue is \(1 - \mathcal{P}\), in which case there is probability one half of having the same sign.

C. Consequences of market efficiency on returns

We can now impose information efficiency and derive a relationship for expected returns. Let \(r^+_t = E_{t-1}[\epsilon_j r_t | \epsilon_t = \epsilon_j]\) be the expected return given that the realized trade observed at time \(t\) has the same sign as the hidden order, and similarly let \(r^-_t = -E_{t}[\epsilon_j r_t | \epsilon_t \neq \epsilon_j]\) be the opposite case. Let \(\nu\) be a proxy of volatility, which we will take to be \(\nu = p^+r^+ + p^-r^-\). Providing \(r^+\) and \(r^-\) are both positive this is also the absolute value of returns (which is useful for data analysis).

Informational market efficiency in its strongest form requires that \(E_{t-1}[r_t] = 0\); we will allow for the possibility that the market is not perfectly efficient, so that \(E_{t-1}[r_t] = \Delta\), where \(\Delta\) is the deviation from perfect efficiency. When combined with the definition of volatility

---

\(^8\) For empirical analysis we have the problem that \(t_0\) is unobservable. This is not a serious problem, however, if one assumes that there is no information revelation prior to the execution of the first order, in which case we simply take \(t_0 = t_1 - 1\). Choosing earlier times, e.g. \(t_0 = t_1 - \theta\), does not make much difference.
and the conservation of probability this gives a simple system of three linear equations

\[
\begin{align*}
    p^+ r^+ - p^- r^- &= \Delta, \\
    p^+ r^+ + p^- r^- &= \nu, \\
    p^+ + p^- &= 0.
\end{align*}
\]  

(4)

Solving for the conditional expected returns \( r^+ \) and \( r^- \) gives

\[
\begin{align*}
    r^+ &= \frac{\nu + \Delta}{2p^+}, \\
    r^- &= \frac{\nu - \Delta}{2(1-p^+)}.
\end{align*}
\]  

(5)

Note that the condition that \( \nu \) corresponds to the absolute value of returns is met as long as \( \nu > \Delta \), which guarantees that \( r^- > 0 \).

When there is predictability in order flow, i.e. when \( p^+ \neq 1/2 \), in an efficient market \( \Delta = 0 \) and the conditional expected returns \( r^+ \) and \( r^- \) are different. Their ratio becomes:

\[
\frac{r^+}{r^-} = \frac{1 - p^+}{p^+} = \frac{1 - \hat{\epsilon}}{1 + \hat{\epsilon}}.
\]  

(6)

where we have introduced the transaction sign predictor \( \hat{\epsilon} = E[\epsilon] \) and used the relation \( p^+ = (1 + \hat{\epsilon})/2 \). This equation specifies the average asymmetry of price response for transactions that have the same sign as the hidden order vs. those that have different signs. We will use this equation when testing if markets are efficient with respect to knowledge of the predictor \( \hat{\epsilon} \). There are many possible mechanisms that can produce the asymmetry specified in Eq. 6. One such cause is an asymmetry in the liquidity provided for buying vs. selling, for example due to an asymmetry in the depth at the best bid vs. the best offer. Another possible cause is an asymmetry in liquidity taking, for example because buyers submit smaller market orders than sellers. Yet another possible cause is quote driven price changes that are influenced by predictability of order flow. It is clear that all three of these causes play a role. While these differences are interesting from a microstructure point of view they are not important for the theory developed here.

Using Eq. 3 we can now compute the unconditional \textit{a posteriori} expected return conditioned on taking the expectation at time \( t \) and finding that hidden order \( j \) is still active.

\[
E_t[r_t] = \epsilon_j \left[ \pi r^+ + \frac{1 - \pi}{2} (r^+ - r^-) \right] = \epsilon_j \left[ \frac{\nu \pi (1 - P) + \Delta (1 - \pi^2 P)}{1 - (\pi P)^2} \right].
\]  

(7)

The middle expression can be interpreted as follows: Assuming hidden order \( j \) is a buy order, there is a probability \( \pi \) that the hidden order \( j \) generates the transaction, in which case the order has the same sign and the average return is \( r^+ \), and there is a probability \( (1 - \pi) \) that another order generates the transaction, in which case there is an equal probability \( 1/2 \) for either generating a buy order with average return \( r^+ \) or a sell order with average return \( -r^- \). If the hidden order is a sell order the factor \( \epsilon_j \) flips all the signs.

### D. Hidden order vs. transaction returns

It is useful to derive a relation between the expected transaction by transaction returns \( E_t[r_t] \) and the expected hidden order returns \( E_n[\rho_n] \). Letting \( \tau_n = t_{n+1} - t_n \) be the time
interval between hidden order executions, to derive a simple relation we assume that \( \tau_n \) is independent of \( r_t \), or alternatively that \( E[r_t] \) is roughly constant from \( t_{n-1} \) to \( t_n \). Note that \( \tau_n \) is a stochastic variable. Under the Poisson approximation this gives

\[
E_n[\rho_n] = \sum_{t=t_n+1}^{t_{n+1}} E_t[r_t] = E[\tau_n]E_t[r_t],
\]

\[
= \sum_{\tau=1}^{\infty} \tau \pi (1 - \pi)^{\tau-1} E_t[r_t],
\]

\[
= \theta E_t[r_t].
\]

This is not very sensitive to the Poisson arrival hypothesis; for example, if we assume that hidden orders are executed at periodic intervals of constant length \( \theta \) we get the same answer.

E. Predicted market impact

We define the total impact of an hidden order as the return from time \( t_0 \) and time \( t_N \), which is the time immediately after the last transaction of the hidden order. We define \( \rho_0 \) the impact between time \( t_0 \) and time \( t_1 \) and combining Eqs. 7 and 8 the a posteriori expected impact of a hidden order \( E[R] \) is

\[
E[R|N] = E[\rho_0] + \sum_{n=1}^{N-1} E[\rho_n] = E[\rho_0] + \epsilon_j \sum_{n=1}^{N-1} \nu(n) \frac{1 - \mathcal{P}(n)}{1 - (\pi \mathcal{P}(n))^2} + \frac{\epsilon_j}{\pi} \sum_{n=1}^{N-1} \Delta(n) \frac{1 - \pi^2 \mathcal{P}(n)}{1 - (\pi \mathcal{P}(n))^2}.
\]

We have assumed that \( \pi \) is constant but all the other terms can vary with \( n \). As we will demonstrate later, for the stocks we study we find that most of the time \( \pi < 0.1 \), making it is a good approximation to neglect terms containing \( \pi^2 < 0.01 \). Thus the impact is approximately

\[
E[R|N] \approx E[\rho_0] + \epsilon_j \sum_{n=1}^{N-1} \nu(n)(1 - \mathcal{P}(n)) + \frac{\epsilon_j}{\pi} \sum_{n=1}^{N-1} \Delta(n).
\]

V. POSSIBLE ASSUMPTIONS ABOUT SCALE OF PRICE RESPONSE

In the previous section we developed a theory for market impact under the assumption that returns are efficient with respect to an information set. As we showed, this determines the ratio of returns \( (r^+/r^-) \), but it does not determine the scale. As the hidden order develops it is possible that returns overall become bigger or smaller, which can change the functional form of the resulting market impact. In this section we propose three possible assumptions that can determine the scale of impact as a function of \( n \), the position in the hidden order. These are (1) constant volatility, (2) symmetric response, and (3) arbitrage between liquidity takers and liquidity providers of the expected average transaction price. We show that (1) and (2) are asymptotically equivalent, but that (3) predicts rising volatility as a function of \( n \), and leads to a different functional form.
A. Constant volatility

One possible assumption is that \( \nu(n) \) is constant. A possible argument for this hypothesis is that the volatility should not be strongly coupled to hidden orders, and that volatility (which is closely related to \( \nu \)) should remain constant as the hidden order develops. If we assume that \( \nu \) is constant, i.e. \( \nu(n) = \nu_0 \), and that the market is efficient, Equation 10 gives the simple form

\[
E[R|N] \approx E[\rho_0] + \epsilon_j \nu_0 \sum_{n=1}^{N-1} (1 - P(n)).
\]  

(11)

B. Symmetry

Another natural assumption is that the price response is symmetric, i.e. satisfying the dual conditions

\[
\begin{align*}
  r^+(n) &= r_0 (1 - \hat{\epsilon}) \\
  r^-(n) &= r_0 (1 + \hat{\epsilon}).
\end{align*}
\]

(12)

(13)

This says that the increase or decrease in the response for trades of the same sign is equal to the decrease or increase in the response to trades of the opposite sign. Notice that the ratio \((r^+/r^-)\) satisfies Eq. 6, as it must for market efficiency to hold. This assumption uniquely requires a linear response of \( r^+ \) and \( r^- \) to the predictor \( \hat{\epsilon} \). Later in this paper we look at data to determine if these relations hold empirically.

Adding Eqs. 12 and 12 together gives \( r^+ + r^- = 2r_0 \). When this additional constraint is taken with Eq. 4, we can solve for the volatility, giving

\[
\nu = r_0 [1 - (\pi \mathcal{P})^2] + \Delta \pi \mathcal{P}.
\]  

(14)

Substituting this into Eq.9 and using \( E[\rho_0] = \epsilon_j r_0 \) we obtain the simple form

\[
E[R|N] = \epsilon_j \left[ r_0 + \frac{1}{\pi} \sum_{n=1}^{N-1} \Delta(n) + r_0 \sum_{n=1}^{N-1} (1 - \mathcal{P}(n)) \right],
\]  

(15)

without making any approximations. This has the consequence that predictability of order flow influences volatility, i.e. if the market is efficient \( \Delta = 0 \) and volatility drops as \((\pi \mathcal{P}(n))^2\) as the hidden order develops. The volatility can be considered approximately constant, however, because the term \((\pi \mathcal{P}(n))^2\) is empirically very small. Under the approximations that \( \Delta \) and \( \pi \mathcal{P}(n) \) are small Eq. 15 reduces to Eq. 11. In other words, the assumptions that volatility is not affected by a hidden order or that the price response to hidden orders is symmetric result in approximately the same market impact.

The symmetry assumption is clearly a much weaker condition than market efficiency, and may or may not hold depending on other factors. There is some justification for symmetry on a ceteris paribus basis using an argument in the style of physics. Symmetry is a fundamental principle in physics, where it is common to argue that unless there is something at hand that violates symmetry, it should be enforced. For example, conservation of momentum results from the assumption that the laws of physics should be invariant under translation in space. The question here is whether there are any effects that break the symmetry between the
response to buy orders vs. that of sell orders. Prices are inherently asymmetric, in the sense that the price can never drop below zero, whereas the price can get arbitrarily large. However, $E[R]$ the expected logarithmic return, which is unbounded in both directions and therefore inherently symmetric. There is, however, an asymmetry induced by finite tick size – measured in logarithmic terms the tick size is smaller for an upward price move than a downward price move. Thus ceteris paribus we expect the price response to be symmetric unless the price in ticks is sufficiently small that symmetry is strongly violated. Of course, there may be other factors that break symmetry, such as inventory effects for market makers, short-selling rules, or asymmetries in psychological responses to buying vs. selling. Thus symmetry is not a strong argument, and we only expect it to be a crude approximation if it holds at all.

C. Arbitrage based on average execution price

Another possible assumption is that there is no net arbitrage between liquidity takers and liquidity providers. We have already imposed the condition that the price be efficient; we now add the condition ...

VI. IMPACT UNDER DIFFERENT MODELS OF ORDER FLOW

Under the assumption of efficiency the market impact is determined by the predictability of order flow, which in turn depends on the information available to market participants and the models they use to interpret this information. The calculations of the previous section show that this depends on the participation rate $\pi$, as well as the power of a given model to correctly predict $P(n)$, the probability that a hidden order will continue. In this section we investigate two different models for predicting order flow, which are based on different information sets.

1. Linear time series model based on order signs. Participants observe publicly available information about order flow and use standard linear time series models to predict it.

2. Detailed model of hidden order execution. Participants know how many hidden orders are active and know how much each has executed so far (which is typically private information). They do not know the identities of who is trading a particular hidden orders or its size, but they do know the unconditional distribution from which the sizes are drawn.

We view these two strategies as extremes that should bracket the range of possibilities. At the end of this section we also sketch another possible model that is in between these two, in which participants use publicly available information about order flow as in (1), but use a structural break model to detect hidden orders. We conjecture that this gives asymptotic results closer to model (2).

We now examine these two order flow models in more detail. In each case the calculation of market impact is reduced to understanding the behavior of $P(n)$. 
A. Linear time series model

A natural method for predicting order flow is with a linear time series model based on public information about realized trading order signs. This indirectly provides information about the presence and persistence of hidden orders and results in a clear prediction for market impact. For convenience we will assume an autoregressive (AR) model of the form

$$\epsilon_t = \text{sign} \left[ \sum_{i=1}^{\infty} a_i \epsilon_{t-i} + \eta_t \right],$$  \hspace{1cm} (16)

as proposed by Lillo and Farmer (2004). $\epsilon_t = \pm 1$ is the sign of the order at time $t$, $\eta_t$ is uncorrelated noise, $\text{sign} = 1$ if its argument is positive and $-1$ if it is not, and $a_i$ are real numbers that can be estimated on historical data using standard methods (we will use ordinary least squares). The probability for a buy order at time $t$ is $p^+_t = (E_t - 1[\epsilon_t] + 1)/2$, and with an AR model can be estimated as

$$\hat{p}^+_t = \frac{1}{2} \left[ \sum_{k=1}^{\infty} a_k \epsilon_{t-k} + 1 \right].$$  \hspace{1cm} (17)

Because the signs are a long-memory process and because for the stocks we study here there is a great deal of data, the coefficients $a_k$ can be reasonably accurately estimated up to very large values of $k$, e.g. $k \sim 500$. For conceptual purposes we assume that the sum can be taken to infinity, though in reality one can only estimate a finite number of lags.

As discussed in Section III we assume that the predictability of order flow is entirely due to the presence of hidden orders. Assume hidden order $j$ has been active for $n\theta$ transactions and that all other hidden orders are equally likely to be buy or sell. For convenience consider a buy order; the calculation for a sell order can be done by simply flipping all the signs. If the participation rate of order $j$ is $\pi$ the expected sign $E_t[\epsilon_t] = 1\pi + 0(1-\pi) = \pi$. Using Eq. 3 gives

$$p^+ = \frac{1}{2} (1 + \pi \mathcal{P}) = \frac{1}{2} \left[ \sum_{k=1}^{n\theta} a_k \pi + 1 \right],$$  \hspace{1cm} (18)

which implies

$$\mathcal{P}(n) = \sum_{k=1}^{n\theta} a_k.$$  \hspace{1cm} (19)

Using Eq. 15 with the symmetry assumption we have

$$E[R|N] \approx \epsilon_j \left[ r_0 + r_0 \sum_{n=1}^{N-1} \left( 1 - \sum_{k=1}^{n\theta} a_k \right) + \frac{1}{\pi} \sum_{n=1}^{N-1} \Delta(n) \right].$$  \hspace{1cm} (20)

To get a better understanding of the behavior for large $n$ we can simplify this considerably by passing to the continuum limit and making use of asymptotic properties of linear time series models for long-memory processes. As argued in Section III there is good evidence that the distribution of hidden order lengths $N$ is typically a power law $P(N) \sim N^{-\alpha}$ with $\alpha < 2$. Under these assumptions, as shown by Lillo, Mike and Farmer (2005), the autocorrelation $C(\tau)$ decays as a power law $C(\tau) \sim \tau^{-\gamma}$ for large $\tau$, where $\gamma = \alpha - 1$. We are not able to derive the $a_k$ for this model and we choose to approximate the sign time series with the sign
of a FARIMA process. It is known (Beran, 1994) that for large \( k \) a FARIMA process has 
\[ a_k \approx \phi k^{-\phi-1}, \]
where \( \phi = (1 - \gamma)/2 \). For large \( k \) we can pass into the continuum limit and estimate the sum in Eq. 19 by an integral. This gives
\[ P(n) = 1 - (n\theta)^{-\phi}. \]

The total impact can be computed using Eq. 9 or one of its approximations – the differences are not large. Assuming perfect efficiency \( (\Delta = 0) \) and under the symmetric approximation from Eq. 15 the asymptotic impact is
\[ E[R|N] = \epsilon_j r_0 \left[ 1 + \sum_{i=1}^{N-1} \left( 1 - \left( n\theta \right)^{-\phi} \right) \right]. \]

In order to convert the sum in integral we use the approximation
\[ \sum_{n=1}^{N-1} f(n) \approx \int_0^{N-1} f(x + 1/2) \, dx \]
and we obtain
\[ E[R|N] = r_0 \left( 1 + \frac{2^{\phi-1}_{\theta^{-\phi}}}{1 - \phi} \left[ (2N - 1)^{1-\phi} - 1 \right] \right) \sim N^{1-\phi} \]

Thus the market impact asymptotically increases with the length of the hidden order as \( N^{1-\phi} \). A typical decay exponent for the autocorrelation of order signs is \( \gamma \approx 0.5 \) [Lillo and Farmer (2004), Bouchaud et al. (2004)], which means that \( \phi \approx 0.25 \). This means that according to the linear time series model the impact should increase as roughly the \( 3/4 \) power of the order size. An interesting property of this solution is that the size of the impact decays as \( \theta^{-\phi} \). This means that the slower an order is executed, the less impact it has, and in the limit as the order is executed infinitely slowly the impact goes to zero.

We have derived this based on the properties of a linear time series model to predict order flow, as originally suggested by Lillo and Farmer (2004), but it is also possible to derive the same result based on the hypothesis of a power law decay of purely temporary impact, as originally suggested by Bouchaud et al. (2004). Either approach gives the same result.

B. Detailed model of hidden order executions

We now investigate the possibility that market participants have more detailed information about hidden orders, including what would normally be considered private information. We assume that they know:

- The number of active hidden orders \( M(t) \) and the participation rate \( \pi \) of each order.
- The correct distribution of hidden orders sizes \( P(V) \), or equivalently of hidden order lengths \( P(N) \).
- The number of realized orders \( n_j \) that have been executed so far for each hidden order \( j \), where \( 1 \leq n_j \leq N_j \).

We also assume that they do not know the size of any particular hidden order \( j \), and thus they do not know when the execution of order \( j \) will end until just after it has actually ended (see Section IV D). They also do not know who originates each order, so there is no distinction between informed and uninformed trades.
1. Stretched exponential distribution of hidden order sizes

To illustrate how the distribution of hidden order sizes determines the shape of impact function we first consider a stretched exponential distribution, of the form

\[ P(V) = \frac{\beta}{\Gamma(1/\beta, 1)} e^{-V^{\beta}}, \] (25)

where \( \Gamma(a, z) \) is the incomplete Gamma function and we have assumed \( V \geq 1 \). We do not believe this is the correct functional form. We know this because this leads to an autocorrelation function for realized order signs that decays as \( C(\tau) \sim \exp(-\tau^{\beta})\tau^{-1} \), i.e. it asymptotically decays as a stretched exponential. In contrast, empirical work for London, New York, and Paris markets makes it very clear that the autocorrelation asymptotically decays as a power law \( \tau^{-\gamma} \), with \( \gamma < 1 \) [Lillo and Farmer (2004), Bouchaud et al. (2004)]. Nonetheless, analysis of the stretched exponential is useful because it gives insight into how the tail behavior of the hidden order distribution causes the concavity of the market impact function.

As before we assume that \( V \sim N \), i.e. that the length of a hidden order is proportional to its size. For a stretched exponential the cumulative distribution is

\[ P(N_j > N) = \int_N^\infty P(N) \, dN = \frac{\Gamma(\frac{1}{\beta}, N^{\beta})}{\Gamma(\frac{1}{\beta}, 1)}. \] (26)

Once an order has already had \( n \) executions, the probability that it will continue is

\[ P(n) = \frac{P(N > n + 1)}{P(N_j > n)} = \frac{\Gamma(\frac{1}{\beta}, (n + 1)^{\beta})}{\Gamma(\frac{1}{\beta}, n^{\beta})}. \] (27)

By using the asymptotic expansion of the Gamma function we get the asymptotic expansion

\[ P(n) \sim \exp\left[-(n + 1)^{\beta} + n^{\beta}\right] \left(1 + \frac{1}{n}\right)^{1-\beta} \] (28)

The function \( P(n) \) is increasing and converging to 1 for \( \beta < 1 \), decreasing and converging to 0 for \( \beta > 1 \), and constant with the value \( e^{-1} \) for \( \beta = 1 \). This shows how the predictability of order flow depends on the distribution \( P(V) \) of hidden order sizes: If \( P(V) \) has tails that are heavier than those of an exponential, then the longer a hidden order goes on, the more likely it is to continue. If its tails are thinner than those of an exponential, the longer it goes on, the less likely it is to continue. The boundary case is when \( P(V) \) is exponential, in which case the likelihood for a hidden order to continue is independent of its size.

The impact can be written

\[ E[R|N] = \epsilon_j r_0 \left(1 + \sum_{n=1}^{N-1} (1 - P(n)) \right) = \epsilon_j r_0 \left(1 + \sum_{n=1}^{N-1} \frac{P(n)}{P(x > n)} \right) \] (29)

By using the above expressions

\[ E[R|N] = \epsilon_j r_0 \left(1 + \beta \sum_{n=1}^{N-1} \frac{e^{-n^{\beta}}}{\Gamma(1/\beta, n^{\beta})} \right) \approx \epsilon_j r_0 \left(1 + \beta \int_0^{N-1} \frac{e^{-(n+1/2)^{\beta}}}{\Gamma(1/\beta, (n + 1/2)^{\beta})} \, dn \right) \approx N^\beta \] (31)
The impact is concave for $\beta < 1$, convex for $\beta > 1$, and linear for $\beta = 1$. Thus we see that once again the exponential distribution is the boundary case: If the distribution of hidden order sizes is exponential the impact is linear. If the tails are thinner than those of an exponential the impact is convex, and if the tails are fatter than those of an exponential, the impact is concave.

2. Power law distribution of hidden order sizes

A more realistic assumption is that hidden order sizes are drawn from a power law distribution, of the form $P(V_j > V) = KV^{-\beta}$. As discussed in Section III, while this is unlikely to be a good hypothesis for small $V$, the evidence supporting it for large $V$ is very strong. Given that a hidden order has already had $n$ realized transactions, the probability $P(n)$ that it will continue is

$$P(n) = \frac{\sum_{t=n+1}^{\infty} Kt^{-(1+\alpha)}}{\sum_{t=n}^{\infty} Kt^{-(1+\alpha)}} = \frac{\zeta(1 + \alpha, n + 1)}{\zeta(1 + \alpha, n)}$$

where $\zeta$ is the Riemann Zeta function. The total impact can be computed using Eq. 9 or one of its approximations. Using Eq. 15 with the symmetry assumption we obtain

$$E[R|N] = \epsilon_j \left( r_0 + r_0 \sum_{n=1}^{N-1} \left( 1 - \frac{\zeta(1 + \alpha, n + 1)}{\zeta(1 + \alpha, n)} \right) + \frac{1}{\pi} \sum_{n=1}^{N-1} \Delta(n) \right).$$

In the large $N$ limit this can be written much more simply. For large $n$ we can approximate $P(n)$ as

$$P(n) = \frac{\zeta(1 + \alpha, n + 1)}{\zeta(1 + \alpha, n)} \approx \left( \frac{n}{n+1} \right)^\alpha.$$ 

Assuming perfect efficiency $\Delta = 0$ and neglecting small terms, from Eq. 11 the asymptotic impact is

$$E[R|N] = \epsilon_j r_0 \left( 1 + \sum_{n=1}^{N-1} \left( 1 - \left( \frac{n}{n+1} \right)^\alpha \right) \right) \approx \epsilon_j r_0 \left( 1 + \alpha \sum_{n=1}^{N-1} \frac{1}{n} \right)$$

By replacing the sum with an integral according to Eq. 23 we have

$$E[R|N] = \epsilon_j r_0 (1 + \alpha \log(2N - 1)) \sim \log N.$$ 

The predicted impact thus asymptotically grows as a logarithm, i.e. it grows much slower than under the linear time series model. This solution also has the interesting property that the impact is independent of the participation rate, i.e. it is the same no matter how quickly the trade is executed. This obviously has limits, i.e. one would expect this to break down if a trade is executed so quickly that it drains all the liquidity in the order book.

C. Return distribution

Under the assumption that average impact $E[R|N]$ makes the principal size-dependent contribution to the impact of individual hidden orders, the combination of the distribution
$P(N)$ for the length of hidden orders and the impact $E[R|N]$ determines the distribution of returns $P(R)$. Since $f(N) = E[R|N]$ is monotonic we can write

$$p(R) = p(N)/f'(N)$$

where $p(R)$ and $p(N)$ are the probability densities for $R$ and $N$. For example, if we assume $P(N) \sim N^{-(1+\alpha)}$, and $R = r_0 \log N$ (an assumption we will justify in the next section), this implies

$$P(R) \sim \frac{N^{-(1+\alpha)}}{(1+N)^{-1}} \sim N^{-\alpha}.$$ 

Solving for $N$ in terms of $R$ gives $N = \exp(R/r_0) - 1$, and plugging this into the equation above gives

$$P(R) \sim (\exp(R/r_0) - 1)^{-\alpha}.$$ 

When $R$ is small, we have $\exp(R/r_0) \approx 1 + R/r_0$, and this becomes

$$P(R) \approx (R/r_0)^{-\alpha},$$

which would suggest extremely heavy tails. However, $R/r_0$ is not small; $r_0$ is the order of the spread, and for large $N$, $R$ is considerably larger than this. Thus, the more relevant limit is when $\exp(R/r_0) \gg 1$, in which case this becomes

$$P(R) \sim \exp(-\alpha R/r_0),$$

i.e. an exponential. This should not be surprising, since the logarithm strongly truncates the tails.

### D. Discussion of information sets in the two models

The linear time series model and the detailed model represent extremes, chosen to bracket the space of possibilities. They are simply convenient, computable cases. It is likely that participants can beat the linear model, at the same time that it seems doubtful that they could gather the private information needed for the detailed model. The question of which is closer to reality can only be answered empirically, as addressed in Section VII.

How might liquidity providers gather the information needed for the detailed information model? In addition to the signs of orders, which are publicly available in the LSE and many other automated continuous double auction markets, hidden orders generate price changes that provide useful confirming information. It is also possible that many traders do not do a good job of concealing their orders, and indeed we have heard anecdotal evidence of traders executing their hidden orders at regular intervals of time and using the same size for all their revealed orders, making it easy to detect a pattern of sustained buying or selling. Furthermore, in the LSE most people trade in both the off-book market and the on-book market, and while brokers in the off-book market do not usually make the names of their clients public, to do business they necessarily release information that can be informative about the presence of large buyers or sellers. So, while no one is likely to have the full information set required for the detailed model, liquidity providers may be able to glean enough information in many situations to detect the presence of large hidden orders.
E. Sketch of a structural break model using public information

The linear time series model that we have analyzed does not make optimal use of the information contained in order signs. This is apparent because it operates on a fixed window of lagged values, and thus is not sufficiently responsive to sudden stopping and starting of sign imbalances due to the beginning and ending of hidden orders. An alternative is a structure break model that detects sudden changes in the probability of buyer vs. seller initiated orders. We are working toward developing such a model, but are not ready to report any results now. For the moment we will just sketch the basic idea and make an estimate of its performance.

The sequence of order signs can be regarded as a Bernoulli process with a bias $p^+$ that shifts discontinuously by $\pi/2$ as a hidden order starts or stops. The problem of detecting the presence of hidden orders is thus reduced to the problem of detecting an intermittently biased coin, for which there is a large literature. To get a feeling of the possible performance of such an algorithm, consider a simple detection method that examines the recent sequence of signs from time $t = t_0 - L$ to $t = t_0$, where $t_0$ is the present. It breaks this into two subsequences, the first on $[t_0 - L, t_0 - L + n]$ and the second on $[t_0 - n + 1, t_0]$. If a hidden order begins execution at time $t_0 - n$, under the assumption that the rest of the order flow is unbiased there will be a bias in order signs. One can detect such a bias by performing a $t$-test on the two intervals. By examining all possible values of $n$ and searching for the one with the largest $t$-statistic it should be possible to detect a structure break and estimate the participation rate and starting time $n$ of the hidden order. Under the null hypothesis of an unbiased coin, under the binomial distribution a hidden order with a participation rate of 10% will typically generate a detection with a confidence of one standard deviation after the observation of 100 transactions and with two standard deviation at 400 transactions. With $\theta = 1/\pi = 10$ this corresponds to $n = 10$ for one standard deviation or $n = 40$ for two standard deviations. This should be sufficient to detect long hidden orders. While this particular algorithm is not optimal (e.g. because it does not deal well with the presence of multiple hidden orders), it illustrates the principle that a structure break model may be able to extract the essential information needed to detect long hidden orders. Thus, even if the detailed model is not literally true, it may be a reasonable approximation for understanding the behavior of large orders.

VII. EMPIRICAL RESULTS

In this section we present empirical results to test the theory that we have developed above. These results are reported for 6 stocks and depend on an algorithm we have developed to determine hidden orders (which we describe below). Because of the small sample size and because of the determination of hidden orders can be problematic, these results should not be considered definite.

A. A brief note about the data

We study six stocks traded on the London Stock Exchange AZN (AstraZeneca), BSY (British Sky Broadcasting Group), LLOY (Lloyds TSB Group), PRU (Prudential Plc), RTO (Rentokil Initial), and VOD (Vodafone Group). The choice of these six stocks is somewhat
arbitrary, and is largely determined by the fact that we have carefully cleaned these data and believe that we have a reliable record of almost every order placement (see discussion below). AZN, LLOY, and VOD are among the most heavily traded names; see Table I for the trading volume for each stock. The data is from the on-book exchange (SETS) only, and is for the period from May 2, 2000 to December 31, 2002. There are time-stamp issues with orders traded off-book, and therefore it is difficult to determine the impact of these orders. Also, many off-book trades are eventually offloaded in the on-book exchange – if we included off-book data, many trades would be analyzed in duplicate. For these reasons, we have chosen to use only on-book data. The dataset contains a complete record of all order placements, so we are able to determine the signs of orders unambiguously. We consistently use AZN to illustrate our results, and present results from other stocks or make comments in the text when the results are significantly different from those obtained with AZN.

These data also have codes attached to each trading order indicating the LSE member firm through which it was submitted\(^9\). The number of member firms of the LSE is of order 100, but they vary in terms of total activity; a substantial portion of the trading volume for a stock can be focused within only a fraction of the firms. Membership in the exchange does not identify the individual trading accounts, and in most cases members are acting as brokerages, handling the trades for an unknown number of clients. Since we do not know who these are we will refer to such clients as “agents”.

While the data we get from the exchange is generally quite reliable, the time stamps associated with orders are only accurate to the second, and the correct sequencing of orders posted in the same second is not guaranteed. This often leads to inconsistencies such as execution against orders that do not yet exist when reconstructing the order book. We have resequenced the data to avoid such inconsistencies. While this resequencing is not always unique, it is at least plausible. We do not think this matters for most of the results presented here.

### B. Detection of hidden orders

To test our theoretical predictions about market impact we need to identify individual hidden orders. Doing this accurately requires information about trading accounts. We do not have such information, but we do have membership codes (that have been anonymized) making it possible to identify which orders are made by the same member of the exchange. This provides enough information to separate most hidden orders, particularly large ones. Although our method of doing this is not perfectly accurate, it passes several consistency criteria that suggest that it is good enough in key respects to provide a preliminary test of our theory and allow us to measure the market impact of large orders. A previous algorithm for identifying hidden orders was introduced by Vaglica et al. (2007).

Our algorithm is very simple. We assume that all the transactions of a given hidden order are made by the same member of the exchange, that they are of the same sign, and that they are within \(\theta_{\text{max}} = 100\) transactions of each other\(^10\). The algorithm proceeds by

---

9 In the original data set these codes were randomly shuffled every month, but Tom O’Brien of the LSE has graciously provided us with a key that allowed us to unscramble the codes.

10 The value \(\theta_{\text{max}} = 100\) was chosen by varying \(\theta_{\text{max}}\) and choosing the value that minimized the autocorrelation of the hidden order sequence, as described in Section VII D.
applying the criteria to the first 100 orders, labeling all orders with no matches as hidden orders of length one, and lumping together others transactions as being part of the same hidden order according to the criteria above. We then examine each transaction successively, either adding it to any pre-existing hidden order that satisfies the criteria (including those of length one) or designating it as a new hidden order of length one. Under these criteria it is possible to have multiple hidden orders active at the same time. It is also possible to have two hidden orders of opposite signs active at the same time under the same membership code.

We need to make several caveats: (1) Our algorithm assumes that all agents submit their orders through a single member of the exchange. This is certainly not strictly true – agents are known to split their orders across several brokerages. Nonetheless, the fact that orders submitted through different brokerages do not have long-memory, as shown in Figure 1, indicates that agents submit most of their orders through the same member of the exchange. (2) Many different agents trade through the same member of the exchange, and more than one of them may be actively trading with that member at the same time. Thus the algorithm is in some cases lumping together two or more hidden orders. See our analysis below, which provides estimates of the error rate. (3) It is common to use a mixture of market and limit orders to execute a hidden order, but we are studying only hidden orders made up purely of market orders. This sometimes results in erroneous splitting of larger hidden orders. To understand how serious this problem is, in studies not reported here we have relaxed the requirement that all orders have the same sign. When we do this we find that the execution strategies of hidden orders are strongly clumped into three groups: Pure market orders, pure limit orders, and 50-50 mixtures of market and limit orders, with few large orders having other mixture ratios. This indicates that pure market order strategies are fairly common, so that splitting due to mixed orders is not as important a problem as it might be otherwise. (4) The algorithm imposes an upper bound $\theta_{max}$ which will artificially split hidden orders if any two executions are separated by more than $\theta_{max}$ transactions. We analyze this effect quantitatively in the next section. (5) We are only considering the on-book market; the same agent may also trade in the off-book market. During this period roughly fifty percent of order volume is executed in the off-book market. The off-book market is not anonymous, and brokers strongly discourage order splitting, which suggests that off-book market trades are less relevant than on-book trades for understanding the behavior of hidden orders.

All of these effects introduce problems into our analysis that the reader should bear in mind. In a future study we intend to do a careful comparison of several different methods of hidden order classification, including the method of Vaglica et al. (2007) and a method based on hidden Markov models.

C. Estimate of the error rate of the algorithm

As discussed above our reconstruction makes errors both in falsely merging hidden orders from different agents using the same member of the exchange to execute their order, and

\footnote{If the same agent uses a mixture of market and limit orders, the limit orders will have the opposite sign, even though their execution results in an inventory accumulation of the same sign. This research is done in collaboration with Javier Vicente, and will be reported separately.}
in falsely splitting large hidden orders. In this section we make estimates of both of these effects.

To get an estimate of false merging we compute the fraction of the time that each member has at least one active hidden order. For example, according to the results of our algorithm, for AZN the two most active members have active hidden orders in process about 20% of the time. Thus, under the hypothesis that each individual hidden order is executed according to a Poisson process, there is a 4% chance that two orders executed by one of these members will overlap. Relating the Poisson process assumption will lead to higher overlap rates, so this estimate is optimistic, but it suggests that the order of magnitude of the overlap effect is not prohibitively large.

Similarly, under the Poisson process assumption, since \((1 - \pi)\) is the probability that a hidden order with participation rate \(\pi\) does not have an execution on any given transaction, the probability that a very long hidden order will go for \(\theta_{max}\) transactions without being falsely split by our algorithm is \(p_s = (1 - \pi)^{\theta_{max}}\). The typical length at which splitting becomes likely is therefore \(L_{max} = 1/p_s = 1/(1 - \pi)^{\theta_{max}}\). For AZN, for example, \(\pi \approx 0.05\) implies that with \(\theta_{max} = 100\), \(L_{max} \approx 170\). We thus expect the tail of the distribution of hidden orders to be artificially truncated by our reconstruction algorithm at roughly this length.

D. Consistency checks

Several consistency checks give insight into the performance of our algorithm. One is the autocorrelation function of the signs of hidden orders. If the assumption that the signs of hidden orders are IID is correct, and our reconstruction method is sufficiently accurate, then we should recover uncorrelated hidden order signs. We compute the autocorrelation function of reconstructed hidden order signed sizes, putting them in sequence based on the time when each hidden order begins. For \(\tau > 0\) the coefficients of the autocorrelation function are all close to zero, but with a slight negative bias. This can be seen for the stock AZN in Figure 2(a). The cumulative of the autocorrelation function is also plotted and appears to level out at around a lag of 400 transactions. We report the average of the first 100 autocorrelation coefficients for each stock in Table I. As seen in the table, the coefficients are negative but very close to zero. This should be compared with the strong positive autocorrelations observed for realized orders in Figure 1. This indicates that the assumption of IID hidden order signs is reasonable and that our algorithm is not making large systematic errors in signing and sequencing hidden orders.

Even if the true sequence of hidden order signs is indeed IID, the problems of falsely merging or splitting hidden orders will affect the autocorrelation of hidden order signs observed in the reconstruction. Falsely merging orders induces a negative autocorrelation. Assuming an IID sequence, this is because nearby orders of the same sign are compressed into a single order, thus causing a tendency for the sequence of signs to alternate. Similarly, splitting orders induces a positive autocorrelation, due to the fact that a single order is replicated to become two or more nearby orders of the same sign. The net autocorrelation of the reconstructed hidden order sequence is thus a combination of these two effects, added to whatever autocorrelation might exist in the true sequence of hidden orders.

Our primary goal here is to measure market impact, and the possibility that our algorithm artificially splits or merges orders potentially distorts the impact. Since the impact is a concave function of order size, artificially splitting a large order will tend to assign improperly
small impacts to large orders and similarly merging orders tends to assign improperly large impacts to large orders. Thus it is important the error rates for these are reasonably small.

We have shown that the distribution of hidden order volumes is an important determinant of the asymptotic market impact, and argued that the cumulative distribution of hidden order lengths is asymptotically a power law for large $N$ with exponent $\alpha = \gamma + 1$. We can check whether our reconstruction is consistent with this hypothesis based on the empirical histogram of hidden order lengths, as shown in Figure 2(b). This figure is plotted in double logarithmic scale so that a power law appears as a straight line. The tail exponent $\alpha$ of the distribution is calculated using a Hill estimator\(^\text{12}\), and the resulting estimate, $\alpha = 1.6$, is drawn as a straight line in the plot. In Table I we present the estimated values of $\alpha$ for each stock and we show values of $\gamma$ measured by computing the Hurst exponent of signs for the realized order flow\(^\text{13}\). As seen in the table, the equation $\alpha = \gamma + 1$ is consistent for AZN and VOD, and $\alpha$ is slightly larger than $\gamma + 1$ for the other stocks. This suggests our hidden order reconstruction algorithm produces hidden order sizes that are consistent with the predicted distribution of sizes. In performing this reconstruction we have the problem that for small values of $N$ there is no reason to expect a power law, and for $N > 170$ we have predicted that the algorithm will truncate the distribution. This leaves very little dynamic range to test our original assumption. We certainly do not mean to suggest that the plot in Figure 2(b) confirms the power law hypothesis (which we think is well-supported by other evidence), but the fact that the exponent is consistent with our prediction indicates that this reconstruction is as consistent as we could hope given the artificial splitting of large orders by the reconstruction algorithm. Nonetheless, we still see a few hidden orders with lengths as long as $N = 2000$, corresponding to a timescale of roughly $N\theta = 40,000$ transactions, or roughly 40 trading days. Note that the algorithm of Vaglica et al. (2007), in contrast, shows results consistent with a power law even for large sizes, and produces longer hidden orders.

\(^{12}\) The Hill estimator is calculated as $\hat{\xi} = 1 + n/\sum_{i=1}^{n} \log(x_i/x_{min})$, where $x_i$ represents empirical data and $n$ is the number of observations $x_i \geq x_{min}$. $n$ is set such that 0.5\% to 1\% of the data is included in the estimate.

\(^{13}\) The exact relation is $\gamma = 2(1 - H)$. This procedure is more accurate than estimating the exponent directly from the autocorrelation function (see Lillo and Farmer, 2004).
FIG. 2: Four diagnostics for our hidden order reconstruction algorithm applied to the stock AZN. (a), the autocorrelation function of signed hidden order size ($\epsilon N$). By grouping transactions into hidden orders, the order flow no longer exhibits long memory. (b) $P(N > x)$, the probability that the length of a reconstructed hidden order is greater than a given value $x$, plotted on double logarithmic scale. For comparison a line of slope 1.6 is shown, corresponding to the asymptotic power law predicted from the relation $\alpha = \gamma + 1$ (based on measurements of $\gamma$ from the realized order flow; see Section III B). (c) is a histogram of the number $M(t)$ of hidden orders that are active at the same time, compared to a normal distribution. (d) is the cumulative probability distribution $P(\tau > x)$ for the interval $\tau$ between successive executions of hidden orders, plotted on semi-logarithmic scale and compared to an exponential distribution.

Figure 2(c) is a normalized histogram of the number of hidden orders that are active at a time. This is very well fit by a normal distribution with mean $E[M] = 14.3$ and standard deviation 2.7. Thus on average there are about 14 active hidden orders for AZN (the average number of orders for the other stocks are reported in Table I). For comparison, Lillo, Mike and Farmer estimated this by comparing the empirical autocorrelation function of realized orders to that predicted by their model, giving an estimate $E[M] \approx 21$. The reconstructed distribution and the mean value estimated from Figure 2(c) is sensitive to $\theta_{max}$, but it is reassuring that these estimates are of the same order of magnitude.

Finally, Figure 2(d) shows the cumulative probability distribution $P(\tau > x)$ for the interval $\tau$ between successive executions of hidden orders for AZN, plotted on semi-logarithmic
scale and compared to an exponential distribution. As expected the distribution is truncated for values of $\tau$ close to $\theta_{\text{max}} = 100$, but from roughly $\tau = 10$ to $\tau = 80$ there is reasonable agreement with an exponential. This is consistent with the Poisson hypothesis, and suggests that the probability of artificially splitting an order is about 1%, consistent with our previous estimates. The mean value of $\theta$ for each stock is reported in Table I.

E. Tests of efficiency

In this section we present empirical tests of the markets efficiency with respect to the two order flow models of Section VI. We first demonstrate explicitly that both of these models strongly predict order flow, comparing there predictability, and we show that the market compensates for this through an asymmetric price response. The linear time series order flow model is a $K^{th}$ order autoregressive model $\hat{\epsilon}_t = \sum_{i=1}^{K} a_i \epsilon_{t-i}$. The order of the model is chosen using the Akaike information criterion. To construct a model for predicting the order flow based on the detailed model, we first classify all the hidden orders using the algorithm described above. A prediction algorithm for the probability of the next sign is then constructed by using Eq. 3 and averaging over all active hidden orders. When combined with Eq. 34 this gives

$$\hat{\epsilon}'_t = \frac{1}{2} + \sum_j \epsilon_j \pi_j \mathcal{P}(n_j) \approx \frac{1}{2} + \sum_j \epsilon_j \pi_j \left(\frac{n_j}{n_j + 1}\right)^\alpha.$$ (37)

As a measure of how well these two models predict order flow we measure the empirical probability that a prediction made at time $t$ predicts the correct sign at time $t + \tau$, i.e. $p_+(\tau) = P[\epsilon_{t+\tau} = \text{sign}(\hat{\epsilon}_t)]$. This can be done for either information sets by using either $\hat{\epsilon}$ or $\hat{\epsilon}'$ as the predictor. The results are shown in Figure (3)(a).
FIG. 3: Several different tests of efficiency for both the linear and the detailed model. (a) The accuracy of the order flow prediction $p^+(\tau)$ as a function of the forecast horizon $\tau$ measured in number of transactions. (b) The quantity $p^+/p^- - 1$ is compared to the $r^-/r^+ - 1$, plotted on double logarithmic scale. (c) The inefficiency $\Delta(\tau)$. (d) The cumulative inefficiency $\sum_{i=1}^{\tau} \Delta(i)$ measured in units of the average bid-ask spread. (e) The ratio $r^+/r^-$ for the linear model as a function of the predicted sign $\hat{\epsilon}(\tau)$, compared to the predicted relationship under the symmetry condition. This is plotted for both $\tau = 1$ and $\tau = 100$. (f) Same as (e) for the detailed model.
For $\tau = 1$ the linear model is more accurate than the detail model, predicting correctly 61% of the time, in contrast to 57% for the detailed model. In both cases the prediction accuracy decays to $p^+ / p^- = 1/2$ for large $\tau$, but the linear model decays much more quickly so that by $\tau = 3$ the detailed model is more accurate. To allow us to see the asymptotic scaling of these results more clearly, we also plot $p_+ / p_- \ln$ in double logarithmic scale as a function of the number of transactions $\tau$. For the market to be efficient it must satisfy $p^+ r^+ - p^- r^- = 0$, i.e. $p^+ / p^- = r^- / r^+$. We can get an idea for how quickly the market becomes efficient by comparing $r^- / r^+$ to $p^+ / p^-$, as is done in Fig. 3(a). The market becomes efficient with respect to the linear time series model at roughly $\tau = 20$. In contrast, for the detailed model efficiency happens much more slowly, and is not complete until roughly $\tau = 400$.

This is seen more clearly in Figure 3(b), where we plot the inefficiency $\Delta(\tau) = p^+(\tau) r^+(\tau) - p^-(\tau) r^-(\tau)$. The inefficiency decays quickly to zero for the linear model and much more slowly for the detailed model. To assess whether either of these inefficiencies are large enough to permit arbitrage, in Figure 3(c) we plot the cumulative inefficiency $\sum_{i=1}^{\tau} \Delta(i)$ in units of the average bid-ask spread. The cumulative inefficiency for the linear model peaks at about 0.2, in contrast to the detailed model, which peaks at about 1.3. Thus the linear model is clearly less than the spread, and so is efficient; the detailed model is inefficient, but this inefficiency is not large. The linear model is efficient for every stock in our sample, as can be verified from Table I. In contrast, the efficiency of the detailed model is much more variable. For three of the stocks (LLOY, RTO and VOD) the cumulative inefficiency is always less than the spread, whereas for BSY, a lightly traded stock, at its maximum it rises to 2.8 in units of the average spread.

To explicitly test our assertion that efficiency in the face of predictable order flow is achieved through asymmetric price response, in Figure 3(d) we test this for the linear model by plotting the ratio $r^+ / r^-$ as a function of the predicted sign $\hat{\epsilon}(\tau)$, where $\tau$ is the lag between when the prediction is made and when the ratio is observed. We do the same for the detailed model in Figure 3(e). This is compared to the predicted relationship derived in Eq. 6, and reproduced here:

$$\frac{r^+}{r^-} = \frac{1 - \hat{\epsilon}(\tau)}{1 + \hat{\epsilon}(\tau)}.$$  (38)

In both cases we observe that for small values of $\tau$ the ratio $r^+ / r^-$ is far from the prediction, but by $\tau = 100$ the observed data are very close to the prediction. This explicitly demonstrates that efficiency is achieved by an asymmetric price response, and for sufficiently large times this satisfies the efficiency condition.

### F. Tests of symmetry

In Sections VIA and VIB.2 we assumed symmetric price response when deriving market impact as a function of hidden order size. We did this because the empirical results we present below suggest this assumption is valid. These results should not be considered definite, however, because we have limited data and because the algorithm we use to determine hidden orders is not exact. As mentioned before, we derive the consequence for other assumptions in a companion paper (Farmer, Gerig, Lillo, and Waelbroeck (unpublished)).
1. Tests of \( r^+ \) and \( r^- \)

As shown in Section V B, the symmetry assumption uniquely requires that (1) the response of \( r^+ \) and \( r^- \) to the predictors \( \hat{\epsilon} \) and \( \hat{\epsilon}' \) is linear, and (2) that the exact relationships are, \( r^+ \propto 1 - \hat{\epsilon} \) and \( r^- \propto 1 + \hat{\epsilon} \) for the linear model and \( r^+ \propto 1 - \hat{\epsilon}' \) and \( r^- \propto 1 + \hat{\epsilon}' \) for the detailed model. In Figure 4 we plot \( r^+ \) and \( r^- \) versus \( \hat{\epsilon} \) and \( \hat{\epsilon}' \) for AZN. Determining the exact relationship between \( r^+ \), \( r^- \) and \( \hat{\epsilon} \), \( \hat{\epsilon}' \) from these plots is difficult for two reasons. First, it is difficult to distinguish linear from nonlinear relationships for values close to the origin and this region is precisely the region where most of the empirical data is located. Second, as shown in Figure 3(a), returns are not immediately efficient, but become so after a certain length of time. This means the response of \( r^+ \) and \( r^- \) to \( \hat{\epsilon} \) and \( \hat{\epsilon}' \) as shown in Figure 4 is not sufficient to make market returns efficient – only when using lagged instead of immediate values of \( \hat{\epsilon} \) and \( \hat{\epsilon}' \) is the market efficient. Using longer lags for these variables is not feasible, however, as the the data becomes noisy very quickly. In spite of these cavets, when grouped together with the further tests of symmetry shown in the following sections, we interpret the plots in Figure 4 as being consistent with the symmetry assumption. For all stocks, the response of \( r^+ \) and \( r^- \) to \( \hat{\epsilon} \) and \( \hat{\epsilon}' \) is consistent with linearity and the slopes of these lines for \( r^+ \) and \( r^- \) are within a factor of 2.5. This is reported in Table I.

FIG. 4: Symmetry plots for AZN and VOD. In the top figures, the sign predictor is calculated using the detailed model (AZN is on the left and VOD is on the right). In the bottom figure, the sign predictor is calculated using the linear model (AZN is on the left and VOD is on the right).
2. Tests of $\nu(n)$

As reported in Section V B, the symmetry assumption predicts that the volatility-like variable $\nu(n)$ should remain constant. We show in Figure 5 $\nu(n)$ for each stock. As seen in the figure, the data is consistent with the theory.

FIG. 5: Volatility Plot.
G. Tests of market impact

If we accept that the symmetry assumption is valid, then the equations for market impact derived in Sections VIA and VIB2 can be tested. As reported in Section VIA, the linear time series model predicts that market impact should scale with hidden order size $N$ as a power law with exponent $1 - \phi$. As reported in Section VIB2, the detailed model of hidden order executions predicts that market impact should scale with hidden order size $N$ as a logarithm (this assumes the distribution of hidden order size is a power law). Here we test which model better fits the empirical data. In Figure 6(a), we plot the scaled impact of all hidden orders as a function of order size $N$ and compare this to the curve predicted by the theory. The empirical data does not correspond to the theory, and although we do not show it here, even when allowing for a free scale parameter, the data still does not conform to the theory. In Figure 6(b) we plot the scaled impact of all hidden orders as a function of order size $N$ and compare this to the curve predicted by the theory. Now, the empirical data matches nicely with theory. This suggests that hidden order impact scales logarithmically with size, and that the detailed model is much more accurate than the linear model in describing the market’s response to hidden orders. In Figures 6(c,d), we repeat these measurements of scaled hidden order impact for both models, but now show it as a function of $n$. Again, the data suggests the detailed model is more accurate than the linear model.

FIG. 6: Panel of Impact Plots.
H. Return distribution

In Figure 7 we plot the distribution of returns for all hidden orders for the stock AZN. The return distribution appears to be exponential, consistent with the assumption of a power law distribution of order size and logarithmic impact.

\[ P(R > x) = \begin{cases} \text{exponential} & \text{if } x > 0 \\ 1 & \text{if } x = 0 \end{cases} \]

FIG. 7: Return Distribution for AZN.

VIII. ADVANTAGES AND DISADVANTAGES OF THE PRESENT THEORY

A. Full accounting of impact under detailed model?

The perceptive reader may at this point want to make sure we have fully accounted for all possible sources of price changes. In the detailed model we imagine that observers know exactly when each increment of a hidden order arrives, as if a light goes off when each increment arrives. We have implicitly assumed so far that the price response to all orders of a given sign at a given point in time is the same whether or not the order originates from hidden order \( j \). I.e., we have defined \( r_t^+ = E_{t-1}[r_t | \epsilon_t = \epsilon_j] \). Letting \( H \) be a Boolean variable that is true if the order at time \( t \) comes from hidden order \( j \), this is not necessarily the same as \( \tilde{r}_t^+ = E_{t-1}[r_t | \epsilon_t = \epsilon_j & H] \). Does allowing for the possibility that \( r_t^+ \neq \tilde{r}_t^+ \) alter our conclusions? This is important because preliminary results suggest they are not equal (See Figure 8).

To understand this subtle point it helps to closely examine all the possible sources of price change. We defined the return \( r_t \) in terms of the logarithm of the midprice change from just after a transaction until just after the next transaction. From a microscopic perspective, during this time there can be an arbitrary number of quote-driven price changes followed by a final transaction-driven price change. The expected size of the final transaction-driven price change is determined by liquidity providers, who collectively determine the depth at the bid and offer and the distance to next bid or offer, as well as by liquidity takers, who determine the probability that they will consume the bid or offer, and thus move the
midprice. The quote-driven price changes and the component of the final transaction-driven price change due to liquidity providers cannot depend on whether or not the final transaction is a hidden order, since this is not known at the time they make their decisions. Thus the only component of $r_t$ that can depend on whether or not it is a hidden order is the portion of the final transaction driven price change that is determined by the liquidity taker executing the hidden order.

If indeed $r_t^+ \neq \tilde{r}_t^+$, how will that affect our results? As we will now show, this is equivalent to an inefficiency, which allows us to put an upper bound on its possible effect when we perform our empirical analysis. Assume that the \textit{ex ante} expectations formed based on the efficiency conditions in Eq. 4 use the \textit{ex ante} expected returns $r_t^+$ and $r_t^-$, which are conditioned only on the sign of the next order and not conditioned on whether or not it is from hidden order $j$. If $r_t^+ \neq \tilde{r}_t^+$, then when the hidden order arrives the \textit{ex post} return of Eq. 7 is modified to be

$$E_t[r_t] = \epsilon_j \left[ \pi \tilde{r}_t^+ + \frac{1 - \pi}{2} (r_t^+ - r_t^-) \right].$$

If we write $\tilde{r}_t^+ = r_t^+ + \delta$, then it is as if we had added $\pi \delta$ to the \textit{ex ante} expected return. In the efficiency condition $p^+ r^+ - p^- r^- = \Delta$, from an empirical point of view it is as if the inefficiency has changed to $\tilde{\Delta} = \Delta + \pi \delta$. Thus from an empirical point of view we can place an upper bound on such an effect, which is effectively included with the measured inefficiency of the market. Of course we can also measure this directly by simply measuring $r_t^+$ and comparing it to $\tilde{r}_t^+$. Both of these show that this component of impact can be neglected.

**B. Incentives for information revelation**

One of the most interesting aspects of these results concern information revelation. In particular, at large times the impact is lower under the detailed model, which assumes private information about order flow, than under the linear time series model, which only relies on public information. Thus it may be the to advantage of liquidity takers to reveal information about $n$, the number of hidden orders they have transacted so far. (Go on to
present the calculations about the optimal information revelation strategy under various hypotheses about market impact).

IX. CONCLUSIONS

We should stress that there are different kinds of market impact, which have different behavior and must be treated differently. We have developed a theory for revealed impacts in an anonymous context, both individually and in aggregate[10]. That is, one can consider a series of transaction without reference to what hidden orders they originate from. When we do this we find quite different behavior than we have reported here. The aggregate impact from summing together $N$ anonymous orders is not a power law, and is not a logarithm, but is a more complicated function for which we have developed a theory using a generalization of the central limit theorem[10]. These results will be reported elsewhere.

To conclude, by finding a theory that predicts the correct scale and shape of the price elasticity of hidden orders, we have partially solved a long-standing problem in economics. The fact that the theory gives good predictions for several other quantities, such as clustered volatility and the return distribution, adds to its power. It appears that participants are able to infer fairly strong information about the actions of other market participants. The crux of the behavior observed here depends on the power law distribution of trading volume, which we have not attempted to explain here. Our results should increase interest in finding a first principles explanation, which would make this theory more fundamental. Our approach shows that by making assumptions about underlying structure (e.g. the power law of volume and the mechanics of hidden order execution) one can make theories about social phenomena that are quantitatively accurate, without resorting to ad hoc assumptions about quantities such as utility.


