Relating Market Impact to Aggregate Order Flow: The Role of Supply and Demand in Explaining Concavity and Order Flow Dynamics

Christopher R. Stephens
C_3 - Centro de Ciencias de la Complejidad and
Instituto de Ciencias Nucleares
Universidad Nacional Autónoma de México
A. Postal 70-543 México D.F. 04510, MEXICO

Adaptive Technologies Inc.
3534 East Shea Boulevard
Phoenix, AZ 85434, USA

Henri Waelbroeck
Pipeline Trading LLC
60 East 42nd Street, Suite 624
New York, NY 10165, USA

Alejandro Mendoza
C_3 - Centro de Ciencias de la Complejidad and
Instituto de Ingeniería
Universidad Nacional Autónoma de México
A. Postal 70-543 México D.F. 04510, MEXICO

Adaptive Technologies Inc.
3534 East Shea Boulevard
Phoenix, AZ 85434

October 26, 2009

Abstract

We investigate how the dynamics of supply and demand affect the relationship between aggregated order flow and returns/market impact. We classify order flow according to the different event types that comprise it: Limit orders, cancelations and market orders executed against displayed or hidden liquidity, and summarize it using three imbalance variables that are aggregate measures of net supply and demand across these different order types. We postulate state equations that link the imbalances among themselves, and, subsequently, to returns, and use empirical market data to understand their functional relationship. By examining aggregated order flow across different time intervals we establish that there is a characteristic time scale over which the market can reliably infer the existence of an imbalance in supply/demand and adapt to it, thus reducing the imbalance. We show that although components of order flow are highly self-correlated, the expected value of an imbalance decays over time. We argue that this is due to a "bounce", whereby a strong imbalance of a given sign has a higher than expected probability to lead to an imbalance of the opposite sign in the future. We further argue that the origin of this reversal is the withdrawal or termination of large institutional orders. We show that the associated conditional probability distributions for future imbalances and returns are bimodal, with a large past imbalance/return of a given sign leading to either a large imbalance/return of the same sign in the future, or a large imbalance/return of the opposite sign. This bimodal distribution explains why market impact estimates have such high variance. We show that the concavity of the relation between market impact and order flow depends sensitively on the relative contributions of the three contemporaneous imbalance components, as well as on their past values.

1 Introduction

Understanding how the dynamics of order flow is related to the formation of price is a fundamental goal of finance and, especially, of microstructure research. From a more macro perspective it has often been thought that order flow is only one determinant of price formation. However, at the most microscopic level, where every quote and transaction is considered, it becomes clear that price is completely determined by order flow, and that the real problem lies in understanding the dynamics of order flow itself. Of course, a better understanding of the relation between price and order flow also has important practical implications, as the cost impact of trading very much affects a trade's profitability. Thus, the study of market impact is, and will continue to be, a subject of intense scrutiny.

An important, empirically well established fact is that returns, or market impact, are a concave function of some *single* variable that is used to describe order flow. However, a proper empirical characterization and theoretical understanding of market impact is still lacking [1]. Studies that show concavity are based on two types of data - publically available data, such as trades and quotes data, and proprietary trading data [2, 3]. Usually, a single variable, V, is used as a measure of order flow. An equation of state relating price changes and order flow, $\Delta P = F(V)$, is then considered. Although there are variations in what constitutes the variable V, and also how price changes are measured (mid-point changes, transaction prices etc.), all robustly show that $F(V)$ is a concave function of V. In other words, the marginal cost of trading becomes less the larger the trade. The robustness of this result impressively stretches across a wide variety of situations - from a trade by trade analysis, where V is the size of a single order [4, 5], through to different aggregation levels, where V is a coarse grained representation of order flow [6, 7, 8].

Although there is ample consensus on the concavity of $F(V)$, there is less agreement as to its specific functional form. The most widely accepted one is a power law, $F(V) \sim V^{\gamma}$, where $\gamma < 1$. There are theoretical arguments [8] that argue that $\gamma = 1/2$. This particular square root impact function can also be derived [9] using an arbitrage argument, if one assumes that the distribution of hidden order sizes has a Pareto exponent of 1.5 and that institutions break even on average after reversion. However, if the break even condition is replaced by a symmetry assumption, the same arbitrage argument leads to a logarithmic impact function [10]. Most empirical models take market impact to also scale as a power of trade size with an exponent close to 0.5. There are also empirical studies, however, that show that a logarithmic function offers a more appropriate measure [11]. Although these differences may sound somewhat academic, this is not the case, these differences being particularly acute for very large trade sizes, the logarithmic function suggesting much lower impact costs for giant trades. This is relevant to the fund scaling problem, i.e., to determine how large a fund should be allowed to become before its trading costs grow beyond the manager's ability to generate alpha. Unfortunately, the data on impact of giant trades is almost by definition sparse, and is also subject to selection bias created by limit prices and cancelations.

More important than the specific functional form however, is the deeper underlying questionsof whether the full richness and complexity of order flow can be successfully characterized by a single variable V and, consequently, whether the relation between returns and order flow can be effectively described by means of a single "universal" function $F(V)$. Do we really expect order flow to be characterizable by a single variable, or the relationship between order flow and returns to be so simple? For instance, in the case of a single trade analysis, Farmer et al. showed that the degree of concavity also depended on the market capitalization of the associated stock. Such a result begs the question: Does the degree of concavity depend on variables other than V, and if so, how so? This is a very important question, because by determining the degree to which the concavity of the relationship between price and order flow can depend on other variables we may better understand why concavity exists in the first place. This latter question in itself is highly non-trivial and must involve interesting dynamical effects, as as been emphasized in [12]. The observation there is that the price change or market impact, that would accrue due to a single market order of a given size, would be determined by the orders already present in the order book. The structure of the order book has been well studied and it is known that, generally, it is a *convex* function of depth, i.e., the density of limit orders or shares available at a given price level is a decreasing function of distance from the best bid/offer. This has an important consequence - that a large order, placed randomly in time, that sweeps up the book must result in a convex impact or price change, contrary to what has been observed. In the case of single trades, the logical inference is that such trades are therefore not carried out in a random fashion, otherwise the aforementioned convexity would be observed.

An important associated problem is that when one talks of the impact of a "trade" one faces the question of how the trade was done. A buy can be achieved through many different mechanisms, such as a market order that executes against displayed liquidity, a market order that executes against dark liquidity, a buy limit order, a cross in a dark pool etc. All of these different mechanisms can be expected to leave different imprints as far as market impact is concerned. Moreover, one would expect that this imprint also depends on what other traders are doing: A buy market order having substantially different effects depending on what liquidity is available at the offer. An associated question then is: Are there variables that permit us to understand how different trading styles can impinge on the degree of market impact?

This points us, once again, in the direction of asking whether it is sufficient to understand the relationship between price and order flow via a single variable V. Usually, this V is taken to be identified as a market order imbalance, i.e., the volume of buy market orders minus the number of sell market orders. Of course, without access to the proprietary information as to the nature of an order and the associated transaction, whether or not it was a market order of a particular sign is something that must be inferred. This is usually done using the Lee-Ready tick rule [13].

If order flow dictates price, then it is obviously important to understand the endogeneous dynamics of order flow itself. The most basic determinant of this dynamics is just that of supply and demand, i.e., the provision and removal of liquidity. For instance, if there is an excess demand at the offer from buy market orders then this is likely to cause the offer price to increase, and therefore induce the arrival of limit orders (supply) at those new, improved (for sellers) prices. As we shall see later on, such elementary notions of supply and demand do, indeed, govern the dynamics of price and are also responsible for the concavity of returns.

In this paper we will present an empirical study and analysis of the dynamics of order flow, and its relationship with price returns/market impact, using trade and quote data from all stocks on all US exchanges over the period April 2008 to April 2009. We will be particularly concerned with a characterization of order flow that goes beyond that of just using market orders, considering other components, such as limit orders and cancelations, and a further subdivision of market orders into those executed against displayed limit orders versus those executed against hidden limit orders (iceberg orders). We will consider the empirical properties of an equation of state that goes beyond the simple one variable model considered previously, characterizing order flow through a set of imbalances that summarize net supply and demand across the different modalities of buying/selling - limit orders, cancelations, and market orders executed against displayed or hidden liquidity. We will consider not only how these different components of order flow influence price, but also how they are causally related among themselves. By doing so we will gain much more insight into what is the right dynamical model for order flow, how the concavity of market impact is related to the non-linear predictability of order flow and how market impact increases the risk of short-term volatility.

In section 2 we will discuss the general problem of relating order flow to price, arguing that, at the microscopic level, price changes are completely determined by individual order flow events. However, as these individual events are only known *ex post* this does not help us to predict, nor have a deeper understanding of how more macroscopic properties of order flow affect price. We therefore consider an analysis whereby order flow is characterized in terms of a small set of coarse grained variables that take into account the net supply and demand - imbalances - across different order types - limit orders, cancelations and market orders - and these variables and returns are aggregated across different time intervals. An equation of state that relates functions of returns to these imbalances is then posited. In section 3 we discuss our data set and then in sections 4 and 5 our principal results. Finally, in section 6 we draw attention to the most important take home message from these results.

2 The Problem of Relating Returns to Order Flow

We will consider returns and market impact in the context of continuous double auction markets, such as the NASDAQ, NYSE etc., as these represent the principle type in US equity trading and are also by far the most common mechanism in other countries. In such markets, traders place either limit orders or market orders, the former specifying a specific price, while the latter are executed immediately at the best available price in the market (the best offer for buy market orders and the best bid for sell market orders). Trades only take place when a market order (or executable limit order) executes against one of more corresponding limit orders.

The most basic understanding of the relationship between order flow and price comes from simple considerations of supply and demand. At the most microscopic level, demand - the desire to buy a stock - manifests itself in the form of individual buy limit orders or individual buy market orders. The latter remove liquidity at the best price level in the order book, leaving the best offer at a higher price if the market order is for more volume than is available. Thus, demand tends to lead to a higher price. The case of a buy limit order is a little bit more subtle. There, price increases are associated with the need to price a limit order more aggressively in order to get a fill when there is competition from other orders at the bid. The net effect however, is the same - a tendency for the price to rise in the presence of demand. A similar argument holds for the supply side, where sell market orders or sell limit orders lead to a decreasing price.

Contrary to what might be imagined, at the microstructure level, the relation between price and order flow is completely deterministic. If we define price from the order book, via the transaction price or the midpoint price, then it too changes deterministically. Of course, some microscopic events do not cause a change in price, even though though they do change the state of the book. For instance, adding volume to the best bid/ask does not immediately change the midpoint price, but it changes the book and therefore may impact how price changes following a subsequent event. The net change due to a sequence of events is a deterministic function of the specific events sequence. The relation between price and individual order flow events then is purely "mechanical".

Of course this is an *ex post* point of view: Returns are predictable only after order flow events have occurred, at least if we have the complete set of them. Still, it emphasizes the fact that if one can predict order flow then one can predict returns and, further, formally at least, one can imagine trying to predict order flow at the microscopic level.

The above considerations also apply to market impact. The only difference to the above is that market impact is associated with a particular subset of orders associated, say, with a particular trader, or a particular meta-order, as distinct from the order flow of the whole market. In reality, it is important to further distinguish the order flow from the rest of the market into order flow that would have existed in absence of the execution process, and the order flow that was generated as a result of or in response to the execution.

Thus, to understand market impact one must consider how different order flow events are related, or how price feeds back into the order flow, thus affecting future events. Distinct to the case of price, where an individual event leads to a deterministic change, the relation between one order flow event and another is probablistic. Thus, the conditional probability for an event $P(e_i^j(t+1)|E(t),P(t),I(t))$ is a function of the set of past events up to the moment t, $E(t) = \{e_{i'}^{j'}(t')\}$, the set of past price points up to time t, $P(t) = \{p(t')\}$, and any other information, $I(t)$, that influences the decision to submit an order that is not accounted for in the past order flow and/or price. In principle, this event probability can be associated with the full set of market participants, or a particular one. Thus,

$$P(e_i^j(t+1)|E(t),P(t),I(t)) = F(E(t),P(t),I(t)) \tag{1}$$

An important question is to what extent there are generic, or "universal", features of F in determining a future order flow event versus it depending on the microscopic details of $E(t)$, $P(t)$ and $I(t)$. That there should exist some generic features can be surmised by contemplating some potential scenarios. For instance, if we imagine that in $E(t)$ there was a predominance of buy/sell market order events in the recent past, then one would expect to see a corresponding enhanced probability for a sell/buy limit order. Similarly, if there is a predominance of buy/sell limit order events in the recent past, then one would expect to see a corresponding enhanced probability for a sell/buy market order. These are simple considerations of supply and demand.

However, more complex scenarios can be easily envisioned. For instance, in the case of hidden liquidity not displayed on the order book, an alert trader who wants to buy/sell using market orders might note that there are consistent executions of such orders without a change in the price of the best offer/bid and therefore infer the existence of hidden liquidity. This in its turn would increase the probability that the trader keeps putting in market orders, in order to take advantage of the discovered liquidity. Another realistic potential scenario is that an institutional trader is executing a large block, or "meta-order", by dividing it up into smaller orders. In this case a sequence of orders in the recent past would be associated with an increased probability to have similar orders in the future.

The point of this discusion is to emphasize that in order to understand the relation between order flow and price at other than a purely mechanical level we must understand the dynamical correlations inherent in the order flow itself. This, in its turn, requires some classification of such events. Order type is certainly an important classification - limit order versus market order. These are dynamically related through demand/supply considerations and also have different impacts on price. However, the above discussion leads us to conclude that there are other degrees of freedom that are important for characterizing order flow. For instance, if a sequence of trades is associated with a particular large block, or if a sequence of market orders is executed against hidden as opposed to displayed liquidity. In these different scenarios however, important information about the nature of the order flow is hidden within a sequence of orders, not just a single one. This information is then subsequently inferred by market participants who adapt their trading strategies and react accordingly. An obviously important classification variable, though one that is jealously guarded, is to which market participant does a particular order or sequence of orders belong. A generalization of this would be the association of a subset of order flow with a particular trading strategy, such as momentum trading or value trading etc.

The point is that patterns and correlations exist in order flow at various aggregation levels, and that many of these are due to the trading activities of market participants who by the trading strategies that they adopt leave imprints. Market participants are constantly trying to decipher these imprints in order to glean information about what is happening, and what will happen, to future order flow and price. As the statistical inference problem is acute, market participants not having access to each other's trading strategies, we will here consider a classification of individual order flow events into four basic types - addition and cancelation of limit orders, market orders executed against displayed liquidity and market orders executed against hidden liquidity. As each of these is associated with a given side we have eight basic event types. Furthermore, in order to avoid the complication of dealing with the whole limit order book we will restrict attention here to orders at the best bid/offer.

As statistical inferences are more reliably made from observations of sequences of trades, it is necessary to consider an aggregation of individual events by performing a coarse graining. We will do this in the simplest way possible by considering net aggregate order flow in a given time interval. We will however, consider different interval lengths - 5, 15, 30, 65, 130 and 195 minutes, corresponding to simple rational fractions of a trading day: 1/78, 1/26, 1/13, 1/6, 1/3, 1/2.

For each interval (t, α), where α denotes the interval type - 5, 15, etc. - we consider the following order types: limit order buys/sells (LOB/LOS); market order buys/sells against displayed limit orders (MOB/MOS), market order buys/sells against hidden limit orders (DOB/DOS), cancels on the best offer/bid (CO/CB). Associated with each of these is a corresponding volume $V_i(t, \alpha)$, where $i = LOB$, LOS, MOB, MOS, DOB, DOS, CO, CB. We will also make a further simplification by considering the net change in orders of a given type. Thus, the excess liquidity at the best bid relative to the best offer for the interval (t, α) due to the net arrivals of limit orders in that interval can be denoted $LCO(t, \alpha) = V_{LOB}(t, \alpha) - V_{LOS}(t, \alpha) + V_{CO}(t, \alpha) - V_{CB}(t, \alpha)$. Similarly, the difference in volume between buy and sell market orders in the interval can be denoted by $MO(t, \alpha) = V_{MOB}(t, \alpha) - V_{MOS}(t, \alpha)$ and $DO(t, \alpha) = V_{DOB}(t, \alpha) - V_{DOS}(t, \alpha)$ for market orders executed against displayed or hidden liquidity respectively. These three components $MO(t, \alpha)$, $DO(t, \alpha)$ and $LCO(t, \alpha)$ can be thought of as forming the components of a three-dimensional imbalance vector, $\mathbf{I}(t, \alpha)$. Thus, we coarse grain the order flow over an interval of time so that it is summarized by these three variables.

2.1 Equations of State for Returns and Order Flow

With the above descriptors of order flow in hand we can posit a state equation that relates them to returns, or indeed, any function of returns, $G(r(t, \alpha))$. Of course, an immediate question is as to whether $r(t, \alpha)$ depends only on the imbalances in the contemporaneous interval t or can depend on the order flow imbalances in past intervals t'. We will see shortly evidence that, indeed, returns are also sensitive to past order flow. Thus we posit an equation of state of the form

$$G(r(t, \alpha)) = F_G(\mathbf{I}(t, \alpha), \{\mathbf{I}(t', \alpha)\}) \tag{2}$$

where $\{\mathbf{I}(t', \alpha)\}$ represents the set of imbalances over different past intervals t'. For the case of past imbalances we consider differences between the different contemporaneous intervals in order to have non-overlapping segments in the past. Thus, we consider past intervals $0 - 5$ mins, $5 - 15$ mins, $15 - 30$ mins, $30 - 65$ mins, $65 - 130$ mins and $130 - 195$ mins. For instance, $G(r(\alpha, \beta)) =< r(\alpha, \beta) >$ could be the average return in the interval t as a function of the contemporaneous and past imbalances. Another function of interest would be $G(r(t, \alpha)) = P(r(t, \alpha)|\mathbf{I}(t, \alpha), \{\mathbf{I}(t', \alpha)\})$, the conditional probability for a given return as a function of past and present imbalances.

We will mainly be concerned with trying to empirically establish the relationship (2) for some $G(r(t,\alpha))$ of particular interest. Although (2) is associated with a substantial coarse graining, whereby the full microscopic order flow has been subsumed into a three dimensional vector taken over a set of discrete time intervals, it is still associated with potentially many variables. However, we can imagine considering a reduced equation of state as a function of a smaller set of variables. For instance,

$$G(r(t,\alpha)) = F_G(MO(t,\alpha), DO(t,\alpha), LCO(t,\alpha)) \tag{3}$$

describes the relation between contemporaneous returns and contemporaneous order flow, as proxied by the three-dimensional imbalance vector. Similarly,

$$G(r(t,\alpha)) = F_G(MO(t,\alpha), LCO(t,\alpha)) \tag{4}$$

would describe how contemporaneous returns depend only on the net liquidity, as described by $LCO(t,\alpha)$, and the market order imbalance against displayed liquidity, $MO(t,\alpha)$. If we take $G(r(t,\alpha))$ to be average returns and the order imbalance to be described by the single variable $V(t,\alpha) = MO(t,\alpha) + DO(t,\alpha)$ then the associated one-variable equation of state would be analogous to that considered in many previous studies.

Of course, one can imagine also constructing an analogous equation that describes the relation between a particular order flow component at t and others at t and t'. For instance,

$$G(MO(t,\alpha)) = F_G(LCO(t,\alpha), DO(t,\alpha), \{\mathbf{I}(t',\alpha)\}, \{\mathbf{r}(t',\alpha)\}) \tag{5}$$

would describe the relation between a given function of the contemporaneous imbalance of market orders against displayed liquidity to the contemporaneous imbalances of dark orders and net liquidity, as well as past values of all three imbalance vector components. Additionally, it is natural to surmise that contemporaneous imbalances should also be sensitive to past returns, $\{\mathbf{r}(t',\alpha)\}$. One might imagine that (5) could be determined by inverting the equation (2). However, we here take the point of view that, unlike a normal thermodynamic equation of state, given that we have argued that present price is causally determined by present order flow rather than vice versa, then it is inappropriate to include contemporaneous price in F_G in (5). Once again, as for returns, corresponding equations on a reduced state space can be considered. One of particular interest will be

$$< MO(t,\alpha) >= F_G(LCO(t,\alpha), MO(t-1,\alpha), LCO(t-1,\alpha)) \tag{6}$$

i.e., the expected contemporaneous market order imbalance as a function of the contemporaneous net liquidity imbalance and also the values of these imbalances in the previous interval. In this way we can begin to understand the adaptive dynamics of the relationship between supply and demand.

As mentioned, for the task of constructing the above functions F_G, rather than try and construct them theoretically or by positing a highly biased model, such as a linear regression, we will rather observe them directly from the data we consider.

3 The Data

Trade and quote data from all US exchanges was collected over the period April 2008 to April 2009 for all (approximately 7000) listed stocks. To understand the relationships between the different components of order flow and returns it is important to first construct accurate aggregate metrics from market data. Unfortunately, the complexity of US market structure makes this a daunting task. Trades take place in dozens of distinct market destinations; some of these represent orders that are required to be displayed publicly and made available to all. To make such a fragmented market structure work, the National Market System regulation (reg NMS) requires a broker to execute an order at the best displayed price across all "fast" markets, which generally means all markets that handle orders electronically without human intervention. Technology vendors provide smart order routers that keep track of the quotes and response times at every market and automatically route an order to execute in compliance with reg NMS. In this way, the multitude of displayed markets becomes effectively equivalent to a single aggregate order book, for all purposes except time priority.

Regulation also requires markets to submit their market data (quotes and prints) to a data aggregator. Unfortunately, the aggregator introduces time delays and does not report the actual timestamp of the event on the exchange, setting its own timestamp instead. Making matters worse, market data on different symbols is handled through different systems, each of which is subject to its own delays; these data synchronicity issues complicate the task of matching prints to quoted prices.

This paper circumvents these issues by using data collected at Pipeline using direct feeds from each significant exchange, and using the aggregated feed only for secondary exchanges or other execution venues (such as dark pools) that do not sell direct data feeds. A matching algorithm was developed to identify the origin of each print as follows: If a print could be matched to a reduction in the displayed quote size at the same price and at the same exchange, it was identified as a displayed market order execution. Exchanges typically execute orders against a multitude of small quotes, resulting in a burst of several prints at the same price followed by a single quote update. To correctly identify these executions as displayed-market executions we considered aggregates of multiple prints that could not be matched individually. We aggregated unmatched prints from the same exchange and at the same price when they occurred within less than 1 second of each other. If the aggregation of contiguous print reports matches a quote size reduction at the same exchange the aggregate trade was recorded as a displayed-market execution. Trades at the best bid or offer that could not be matched to a quote size reduction were identified as dark buyer- or seller-initiated transactions according to the Lee and Ready tick rule. Quote size reductions at the best bid or offer that could not be matched to a print were identified as cancellation events and quote size increases at the current best bid or offer prices or making a new best bid or offer price were identified as limit order arrivals.

3.1 Variable Aggregation

The perspective we take is a "data mining" one, where the idea is to consider a large set of potential descriptors, or "features", of the state of the world as it affects a variable of interest, such as returns. We can intuitively divide such features into two types: low frequency and high frequency. The former involve relatively static stock dependent features such as volatility, expected earnings, firm value etc., while the latter take into account dynamic factors, principally aggregate order flow as discussed above. In this paper we will be concerned only with aggregate order flow as described by past and present values of the imbalance vector $I(t, \alpha)$, leaving a more complete analysis that includes stock dependent factors for another paper.

Another innovation we wish to introduce here is that, instead of necessarily considering returns and order flow components as continuous variables, we will discretize them. By doing so, we can then consider the construction of the equation of state from a "classification" point of view. Of course, such discretization is necessary from a statistical point of view in the first place anyway, as the probability for any combination of data values when the variables are continuous is essentially zero or one; i.e., the data point with those values exists or it doesn't. It also obviates the need to construct a continuous probability distribution with which to fit the data. You just let the data speak for themselves.

Explicitly, to discretize: For any metric observable, such as returns, shortfall, order flow of a given type etc. we will rank the observations from "biggest" to "smallest" then divide up the list into a fixed number of categories. We will use various numbers of bins - 10, 20, 30 depending on the specific variable of interest and the size of the data sample. There are various criteria by which membership of a given category may be defined. Here we will do so by requiring that the categories form bins that contain equal numbers of observations. Noting the values of the observable at the boundaries of the bins allows for any new data point to be classified as a member of a specified bin.

Of particular interest will be those bins that contain the intervals associated with the largest positive values of a variable (10th decile) and the largest negative values (1st decile). The variables are naturally divided up into dependent and independent - which in our classification point of view means classes and features of that class. So, an important object of interest is then the conditional probability, $P(C|\mathbf{X})$, where C is the class of interest and the $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ form a vector of features that are taken to be the variables of interest that affect class membership. Similarly, we can consider, for example, $< r_{X_1 X_2 \ldots X_N}(t) >$ the average return in the bin associated with the discrete values of the features X_i.

An important question is how data will be coarse grained by aggregating across time. Many studies have analyzed market data on a trade by trade basis. An important drawback of this approach is that it does not allow one to see how market impact manifests itself within a trading strategy that evolves temporally, such as dividing up a large block order into smaller orders. More importantly however, it does not allow one to see how the market as a whole develops relating only market impact to the trade size of the individual trade. In other words market context is lost in gauging the impact of the trade. In this study, as mentioned above, we considered aggregation across several distinct time interval. An interval is defined to be the greater of one minute or 5 prints on the tape. For liquid stocks an interval is equivalent to a minute, while for less liquid stocks an interval may span many minutes. An interval may also span more than one trading day. A second aggregation level comes from combining intervals together. We consider 5, 15, 30, 65, 130 and 195 intervals. For liquid stocks, where an interval is a minute, this is equivalent to dividing the trading day into 2, 3, 6, 13, 26 and 78 equal parts.

This time aggregation aspect influences both the class and the features we will consider. For instance, for returns we have $r(t, \alpha) = \ln(p_f(t, \alpha)/p_i(t, \alpha))$, where t refers to the tth interval, α to the interval type, i.e., 5, 15, etc. and $p_f(t, \alpha)$ and $p_i(t, \alpha)$ refer to the last and first trade prices in that interval. We will consider return classes $C_{r_i}(t, \alpha)$, where i refers to the returns decile, 1-10. Similarly, we will consider $X_{ij}(t, \alpha)$, where j refers to the corresponding bin associated with the coarse graining of the variable X_i.

4 The Dynamics of Supply and Demand

To understand the relation between returns and order flow it is first important to understand the endogeneous dynamics of the latter. To this end we will consider the relation between the different components of the imbalance vector as representing different effective degrees of freedom of the aggregated order flow.

4.1 Contemporaneous Supply and Demand

First we will consider the relationship between the net supply and demand of liquidity in the same time interval, as determined by the net liquidity imbalance, $LCO(t)$, and the market order imbalance, $MO(t)$. This corresponds to the state equation (6), where we have marginalized over the past imbalances and the contemporaneous dark imbalance considering $< MO(t, \alpha) >= F(LCO(t, \alpha))$. In Figures 1, 4, 2 and 3 we see the results for $\alpha = 5$, 15, 30 and 65 minute intervals respectively. For 5 minute intervals, the most notable aspect of Figure 1 is that the slope is negative. Given that $LCO(t)$ is positive when there is more liquidity on the offer, we see that this is associated with a negative market order imbalance. So, over 5 minute intervals, selling/buying via market orders is highly correlated with selling/buying via limit orders. To understand why this is possible without creating or destroying shares, it is important to remember that the definition of $LCO(t)$ only takes into account order arrivals and cancelations at the best quoted price. When an aggressive buyer exhausts all liquidity available at the current best offer, a new offer price is revealed and shares that were entered prior are now available at this new best offer price. Since these limit orders were not at the best offer at the time of their placement they were not counted in $LCO(t)$. The negative slope in Figure 1 shows that as limit order buyers arrive on the market, they add shares to the bid and make new bid prices, while limit order sellers are supplying insufficient liquidity on the offer to meet the demand at the current best offer thus inducing a positive market order imbalance.

Note however, that the relationship is highly nonlinear, showing a large degree of concavity for large limit order imbalances. This is an indication that, for large net liquidity imbalances, the associated marginal increase in market order imbalance on the opposite side of the spread is small. As we will discuss in more detail later, this is the beginning of an adaptive reaction whereby more liquidity at the offer/bid is associated with a preponderance of market orders on the same side of the spread.

In Figures 2 and 3 we see the analogous graphs for contemporaneous intervals of 30 and 65 minutes. What is apparent from these two graphs is that, in distinction to the situation for 5 minute intervals, now the slope is positive, thus indicating that over these time scales selling/buying by limit orders is highly correlated with buying/selling using market orders. So, for 5 minute intervals, demand is manifest across the different modalities of buying: Buying with limit orders is correlated with buying by market orders. In other words liquidity givers are preferentially on the opposite side to liquidity takers. However, over time intervals of 30 minutes or more liquidity now preferentially appears on the same side as the liquidity takers. Clearly, the implication is that something is happening between 5 and 30 minutes in that the relation between the limit order and market order imbalances reverses. To see this further, consider in Figure 4 the relation between $LCO(t)$ and $MO(t)$ for 15 minute intervals. Interestingly, we now see that there is very little correlation between the market order and limit order imbalances. This is due to the fact that it is precisely at this time scale that there is a change in relation between $MO(t)$ and $LCO(t)$ such that market orders change from being on the opposite side of the spread to the net liquidity imbalance to the same side. We can clearly see that this change depends on the “strength” of the net liquidity signal, i.e., the largest net liquidity imbalances at the 15 minute timescale are clearly associated with market order imbalances on the same side of the spread, whereas the correlation is weaker for smaller liquidity imbalances.

One possible way to interpret these results is to bring in Hasbrouck’s observation [15] that only the unpredictable component of order flow can cause market impact: After 5 minutes, the excess demand becomes, in part, predictable and is therefore associated with a greater supply from liquidity providers. This is the basic mechanism for co-adaptation between liquidity providers and liquidity takers on the market. Information arbitrage theory [9] formalizes this argument and shows that it explains the *average* concavity of market impact.

In practice, the timescale for detecting the presence of a potentially large hidden order (together with the corresponding prediction of future order flow imbalances) depends on the stock’s liquidity on the market and on the speed at which the hidden order is executed, 5 minutes being the correct value for execution speeds of 20-25% typically preferred by buyside traders. Understanding the timescale at which the market reaction develops is key to understanding how this theory becomes applicable in optimal trade execution [16].

The story from these figures is fairly clear: At timescales of the order of 5 minutes buy/sell pressure manifests itself on both sides of the spread. At this timescale events are typically dominated by high frequency traders, who today represent the vast majority of order flow on the markets. As such traders operate on extremely thin per-share margins, in the US markets they have an incentive to trade on the passive side of the spread as passive executions are awarded a liquidity provider rebate. At 0.2 cents per share this can easily exceed the net margin of a high frequency operation. These traders are therefore trying to take advantage of the spread and earn the rebate when possible, i.e., they will buy/sell if possible using limit orders. As buy/sell pressure develops, bids/offers are less likely to get filled and traders need to become liquidity takers in anticipation of a price move. Thus, an excess of limit orders on the bid is in the short term associated with an excess of market buy orders. Thus, over shorter time scales it is possible to have significant imbalances in supply or demand on both the corresponding sides of the spread.

On a larger timescale, the activity of large institutional hidden orders on the market leads to a persistent imbalance which becomes increasingly predictable. This gives a logic as to the backreaction of the market that acts so as to reduce the imbalance. In order to do this the market itself must be able to detect these imbalances. This is a statistical inference problem for traders, who must determine whether an imbalance is due to the existence of a true "signal", e.g., the presence of a large institutional hidden order, or is just a random fluctuation. This signal has two important dimensions - size and duration. The larger the size of the imbalance and the longer it has lasted the more easily it is detected. We see sensitivity to both these dimensions manifest in the figures.

When their presence becomes detectable one can imagine that markets begin to operate in a different mode, where knowing the potential size of the institutional hidden order becomes more important than responding to the news stream and ordinary market events. Thus, expectations of positive/negative prices changes cause more liquidity to become available at the offer/bid. This has the effect that the market order imbalance now changes sign, demand is met by supply at prices that reflect market expectations about the size of the hidden order. In this regime, the market is dominated by a cat and mouse game where the institution's interest is to guide expectations of their order size down while arbitrageurs are using statistical models and tactics such as dark pool pinging to get an edge in more accurately estimating the size of a hidden order.

4.2 Relation between Past and Present Supply and Demand

It is well known that there is a high degree of autocorrelation in supply or demand in time [11, 14]. To illustrate this with the present data we see in Figure 5 the relation between $MO(t)$ and $MO(t-1)$ for 5 minute intervals, while in Figure 6 we see the analogous relation for $LCO(t)$ and $LCO(t-1)$. In both these graphs, the extreme degree of linearity inherent in the time dependence of both market order and net limit order imbalances is manifest. However, note that the expected imbalance at t is substantially less than the corresponding imbalance at $t-1$. For market order imbalances the suppression is such that the imbalance at t is about 30% of the imbalance at $t-1$, whereas for $LCO(t)$ it is about only 15% of its value at $t-1$. This is an important fact, given that much has been made of the extreme degree of auto-correlation in sign of the market order imbalance. What it shows is that, even though there is a large degree of auto-correlation in sign, this is not so useful for computing magnitude, given that, if the predicted imbalance at the next time step is only a fraction of that in the previous one, then the expected magnitude will be exponentially suppressed as a function of time. The reason why the magnitude decays we will explain in more detail shortly.

We might also ask why there is an asymmetry in the magnitude of the decay between liquidity takers and liquidity providers? We believe that this is due to the fact that liquidity provision is more discretionary than liquidity taking. In other words, institutional buying/selling with a strong immediacy has to be carried out using market orders. On the other hand with liquidity provision the urgency of trading is less and therefore there will be more tendency to wait for optimal trading conditions.

As well as the auto-correlation of a given flow component, we can also consider the cross-correlations between different flow components. In particular, that between $LCO(t)/MO(t)$ and $MO(t-1)/LCO(t-1)$. We begin with the relation between $LCO(t-1)$ and $MO(t)$, as seen in Figure 7, where we see that more liquidity on the offer/bid at $t-1$ is associated with more buying/selling via market orders at t. However, the signal is not very strong, with a suppression of about an order of magnitude of $MO(t)$ for a given $LCO(t-1)$ relative to the value of $MO(t)$ for a given contemporaneous net liquidity imbalance, $LCO(t)$. However, one can expect this to be a supply-demand adaptation. Traders wishing to buy/sell note the presence of excess liquidity on the offer/bid and enter the market to conform with the presence of that liquidity. We believe this to be the selection bias associated with individual trades, but now at an aggregated time scale. i.e., traders preferentially trade when there is more liquidity.

Although a linear correlation analysis focused on average values is interesting it obscures a much more interesting picture that emerges when one considers the full probability distributions. Thus, a much deeper insight into the relationship between different components of order flow in different intervals becomes apparent in Figures 8 and 9. Reviewing Figure 8 we see that a decile 10 (excess buying) market order imbalance at $t-1$ leads to a decile 10 market order imbalance at t with a probability of 35%, i.e., 3.5 times higher than chance. Significantly however, it is decile 1 of $MO(t-1)$ that has the next highest contribution to the probability of a decile 10 value of $MO(t)$, with a value of about 17% - 70% higher than chance! Thus, we see that the probability distribution, $P(MO(t) = 10|I_i(t-1) = n)$, over the different deciles, n, of the different imbalance vector components at $t-1$ is bimodal as a function of n. Note also that the probability of there being a decile 10 market order imbalance at t when the $t-1$ imbalance was in deciles 2-8 is less than random, i.e., there is a negative correlation, with the probability being up to 50% less than the 10% random distribution level, which we could take as a null hypothesis. Note that the greatest asymmetry is for MO. With $LCO(t-1)$ the distribution is almost symmetric with decile 10 and decile 1 net liquidity imbalances almost having the same probability to lead to a large positive market order imbalance at t. Finally, we can see that price is the least "predictive" variable, with values that are much closer to the 10% random benchmark than any of the flow variables.

So, there is a substantial probability that the order flow imbalance flips sign in the next time step. In other words, if it was high and positive/negative, then there is a much higher probability than one would naively expect that in the next time step it is high and negative/positive. We believe that the reason why this happens can once again be associated with large institutional trades: The market becomes habituated to the continuing presence of order flow of a given sign, especially flow of market orders, leading to the arrival of more liquidity on the opposite side. However, eventually the institutional order finishes. In this circumstance, the market must infer that the pressure from this auto-correlated order flow has finished. As emphasized above, in the discussion of the relation between comtemporaneous supply and demand - this takes time, the exact amount depending on the strength of the imbalance signal. While the market gathers data to infer the end of the institutional trade it meanwhile is "pushing" back in the opposite direction. This reverse order flow now causes an opposite imbalance. Thus, if the institutional trade continues, one will obtain a high order imbalance of a given sign, while, if it ends, one can get a high imbalance of the opposite sign. As institutional trades will have durations of more than 5 minutes, it is more likely that the trade will continue than it will end, thus this "bounce" in order flow is asymmetric and leads to a partial suppression of the magnitude.

Another important observation with respect to the probability distributions of the dynamics of order flow, is that the bimodality implies a very large variance. In other words, if we calculate, approximately, the expected value of a particular imbalance component, $< I_i(t) >$

$$< I_i(t) >= \sum_{k=1}^{10} < I_i = k > P(I_i(t) = k | I_j(t-1) = 10) \tag{7}$$

as a weighted sum over the probabilities to be in the different deciles conditioned on the fact that a particular imbalance I_j component at $t-1$ was in decile 10, then the fact that there is a "bounce" implies that both $P(I_i(t) = 10 | I_j(t-1) = 10)$ and $P(I_i(t) = 1 | I_j(t-1) = 10)$ contribute substantially to $< I_i(t) >$. As $< I_i = 1 > \approx - < I_i = 10 >$, where $< I_i = n >$ is the expectation value of the imbalance component I_i in the nth decile, then we see that $< I_i(t) >$ will be very much less than $< I_i = 10 >$. This is, indeed, the source of the suppression of the magnitude of an order flow imbalance in spite of the fact that there is a great deal of auto-correlation. Similar considerations hold for the dependence of the probability distribution of decile 10 values of the net liquidity imbalance,

$P(LCO(t) = 10|I_i(t-1) = n)$, as a function of the imbalance vector components at $t-1$. Once again, the distribution is bimodal.

5 Relation between Returns/Market Impact and Order Flow

We now turn our attention to how both contemporaneous and past order flow affects returns.

5.1 Contemporaneous Returns/Market Impact Depend on Available Liquidity

We argued in section 4 that returns and therefore, by implication, market impact should depend on the relative balance between supply and demand and that this manifested itself in several different components of which we considered: Limit orders, cancelations, and market orders against displayed and hidden liquidity. To investigate this further, we will consider the average return, $DP(t)$, as a function of the market order imbalance against displayed liquidity, $MO(t)$, conditioned on the net liquidity imbalance, $LCO(t)$ as measured by its decile, decile 10/1 representing the strongest imbalance on the offer/bid versus the bid/offer. In Figure 10 we see the relation for 5 minute intervals. Similar results hold true for other interval lengths. There are two important features of this graph to emphasize: Firstly, the returns for a given $MO(t)$ depend on the liquidity imbalance; and, secondly, the concavity of the relation between $DP(t)$ and $MO(t)$ also depends on the liquidity imbalance.

Certainly we can see that a given imbalance in market orders can lead to both positive and negative returns depending on the sign of the net liquidity imbalance. If there is a large imbalance on the offer relative to the bid, then buy market orders can lead to a negative return as opposed to a positive one. However, when the liquidity imbalance is fairly neutral - deciles 5 and 6 - we see that the sign of $DP(t)$ is essentially given by the sign of the corresponding market order imbalance. Note that there are non-trivial returns/market impact, even in the absence of a liquidity imbalance. This is clear in the intercepts of the curves at $MO(t) = 0$, with the highest magnitude imbalances - deciles 10 and 1 - being associated with price changes of about 40bps.

The second important feature of Figure 10 is that the degree of concavity of the returns as a function of market order imbalance also depends on the liquidity imbalance. If we take $LCO(t)$ decile 10, that corresponds to the largest imbalance in the offer versus the bid, we see that the marginal increase in returns for positive $MO(t)$ decreases as a function of $MO(t)$. On the contrary, if we compare this to decile 1 of $LCO(t)$ we see that, there, the slope of the curve is greater for a given $MO(t)$. In other words, the marginal return is higher for a given market order imbalance when the liquidity imbalance is against you. In terms of market impact, this means that you pay more when there is not much liquidity in your favor.

If we consider negative market order imbalances and deciles 10 and 1 of the net liquidity imbalance, then we see that analogous conclusions hold. The asymmetries in the curves for large liquidity imbalances reflect the fact that the marginal costs of market orders are less, irrespective of whether you're buying or selling, when the liquidity imbalance is on your side.

Passing next to the dependence of returns on contemporaneous net liquidity we see a complementary picture to that of Figure 10. In this case the intercepts for $LCO(t) = 0$ are less for a given decile of $MO(t)$ compared to intercepts for $MO(t) = 0$ for a given decile of $LCO(t)$. This indicates that a given imbalance of $MO(t)$ has less impact than a given, similar, in the sense of decile, imbalance of $LCO(t)$, being about $\pm$20bp for deciles 10 and 1 of $MO(t)$. An analogous asymmetry to that found in Figure 10 is present here: The marginal price increase for a given liquidity imbalance is higher the higher the market order imbalance on the same side. Thus, if there is more liquidity on the offer, then for a negative market order imbalance, i.e., more sells than buys, this leads to a higher marginal impact than if the market order imbalance is positive.

5.1.1 Contemporaneous Returns/Market Impact Depend on All Imbalance Components

A corollary to the above is how returns/market impact depend on the different components of the imbalance vector. In Figure 12 we see a graph that shows the probability of different deciles of the return $DP(t)$ given different states of the imbalance vector $\mathbf{I}(t) = (MO(t), DO(t), LCO(t))$. Decile 10 on the x-axis signifies the top 10% of most positive returns, while decile 1 is the top 10% of most negative returns. Take as example, $\mathbf{I}(t) = (1, 1, 10)$. This corresponds to the market order imbalance being in the bottom decile - most excess of sells over buys; the dark order imbalance in the bottom decile, once again the most excess of sells over buys; and the net liquidity imbalance being in the top decile, i.e., most imbalance on the offer versus the bid. As the latter implies more selling than buying by limit orders, then $(1, 1, 10)$ implies all three imbalance components are aligned, all giving rise to an excess sell pressure. The imbalance vector $(1, 1, *)$ on the other hand, corresponds to the market order and dark order imbalances being in the top decile but without conditioning information on the net limit order imbalance. In this case, two of the three imbalance vector components are aligned. Similarly, $(1, *, 10)$ is associated with the market order and net liquidity imbalances being aligned. The imbalance vector $(1, *, *)$ is then associated with the market order imbalance being in the bottom decile, but without conditioning information on the other components.

What is clearly visible is the effect that the higher the degree of alignment between the different order flow components, the higher the probability for a large negative return. With complete alignment, the probability is 44%, i.e., nearly four and a half times bigger than random chance. With two components aligned this figure drops to about 36%, while for only one component it is about 26%, i.e., about 43% less than the corresponding figure for total alignment. If we study further these curves, associated with varying degrees of sell pressure across the different imbalance vector components, we see that instead of these curves being monotonic, with successively smaller probabilities for being in higher return deciles, on the contrary, as with the relation between order flow components, there is a significant "bounce", meaning that the probability distribution for returns is bimodal, with the probability for a decile 10 return being higher than the corresponding probabilities of deciles 4 through 9. The explanation of why this ocurs is analogous to that presented for the case of order flow and, given that price is determined by order flow, follows from it. The bounce occurs because a correlated order flow, potentially due to an institutional trader, pushes the market which, in its turn, pushes back via an induced counterflow. When the institutional pressure is removed then there is a reversion in price due to the fact that there is a change in sign of the order flow imbalance. As mentioned for order flow, the fact that the distribution is bimodal will clearly be associated with a very high variance when compared to any standard unimodal distribution. This variance is, in fact, the reason why estimates of market impact are notoriously noisy. Put simply, with such a bimodal distribution, the average return or impact will be particularly unrepresentative.

Although the relation between returns and the contemporaneous imbalance vector components is bimodal, the distribution is asymmetric, with the degree of asymmetry depending on the relative alignment of the components - more alignment, more asymmetry. We can ask then: What is the relation between returns at time t and the imbalance vector components at $t-1$? The results are shown in Figure 13. The bimodal nature of the distribution is now even more manifest due to the almost complete symmetry of the curves. In plain speaking: If all order flow components are aligned at $t-1$, the probability to have a return in the opposite direction to that of the order flow is almost the same as that as when they are in the same direction. In other words, a large buy asymmetry at $t-1$ is as likely to lead to a large negative return as a large positive one. This is a manifestation of market efficiency at work. Note - this is not just a volatility effect. The probability for a non-extreme return, i.e., not decile 10 or 1, is relatively very low. Basically, the harder you hit the market at time $t-1$ the less likely you are to have a non-extreme event at t.

5.2 Contemporaneous Returns/Market Impact Depend on Past Trading Behavior

Another important question is to what extent contemporaneous returns depend on past trading behavior? Given the strong autocorrelation in order flow, it would be expected that this would have some bearing on returns. To investigate this question we considered the relationship between returns, $r(t)$, and the contemporaneous market order imbalance, $MO(t)$, but now considering different conditioning information on past order flow. Here we will show what happens to the relation between returns/market impact and order flow when conditioned on the presence in the past of one of the following three scenarios: A strong positive - excess buying - market order imbalance - decile 10; a strong negative - excess selling, decile 1 - imbalance; and, finally, in the presence of an imbalance that was close to zero - decile 5.

In particular, we wished to study the dependence on the presence of imbalances that had different durations. To this end we may denote the past imbalances as a vector $\mathbf{M} = (MO_5, MO_{15}, MO_{30}, MO_{65}, MO_{130}, MO_{195})$, where MO_i denotes the market order imbalance, in terms of decile, in the ith past interval. Remember that our past intervals are non-overlapping so that MO_5 is the market order imbalance in the last 5 mins before the beginning of the current interval. MO_{15} is then the imbalance decile in the interval beginning 5 minutes ago and terminating 15 minutes in the past. Similarly, MO_{30} represents the imbalance in the interval that started 15 minutes before the present interval and terminated 30 minutes ago. We can consider marginal conditioning information by representing a given imbalance component by the symbol $*$, which represents that there is no conditioning on the corresponding variable. Thus, $\mathbf{M} = (MO_5, MO_{15}, *, *, *, *)$ represents a conditioning on the imbalances in the first past 5 minutes before the current interval and the interval between $t = -5$ and $t = -15$ but without further conditioning information on the other intervals.

In Figures 14, 15 and 16 we see the results for contemporaneous intervals of 15 mins. For example, in Figure 14 we see what happens in the presence of a large buy market order imbalance in the past. The first curves to note are *buy015* and *sell015*, that correspond to the past imbalance vector $\mathbf{M}(t) = (*, *, *, *, *, *)$, i.e., the relation between returns and positive/negative market order imbalances without any past conditioning information. These curves correspond to the standard ones that demonstrate the concavity of returns as a function of market order imbalance. Note that the curves are symmetric in that a buy or sell market order imbalance of a given magnitude leads to a corresponding positive/negative return of the same magnitude. Turning now to the curves with conditioning information, there are two important features to note. Firstly, for a given positive contemporaneous market order imbalance, the corresponding price impact is less if there has been a positive imbalance in the past and even less according to the duration of this past imbalance. Thus, for instance, at $MO(t) \sim 0.025$, which corresponds to an imbalance of about 65% of ADV, without conditioning on the presence of a large past positive imbalance the average impact is about 53bp, whereas when there has been a strong positive imbalance for the last 30 mins (curve *buy015-10-10-10*) the impact for the same contemporaneous imbalance is only about 25bp, i.e., about 50% less! Even more striking is the fact that in the presence of a strong past imbalance, the relation between impact and imbalance is not even monotonic, i.e., that a larger imbalance can lead to less impact, corresponding to a negative marginal cost.

Considering now what happens to returns when the contemporaneous market imbalance was negative, i.e., excess selling, we see that the situation is the complete opposite of that found for a large positive imbalance. Now, for a given imbalance, the price impact increases according to the presence and duration of a past large, positive imbalance. Without conditioning information (curve *sell015*), the impact for a negative imbalance of about 65% of ADV is about 57bp, results that are very similar to that for a positive buy imbalance of the same magnitude. However, conditioned on the presence of a large positive past imbalance (curve *sell015-10-10-10*), we see that the corresponding impact is about 75bp, i.e., about 30% more. Note though, that despite the great asymmetry between the impact of buying or selling in the presence of a large amount of excess buying in the past, there is still a great degree of concavity in the curves that correspond to negative contemporaneous market order imbalances. We will return to this point momentarily.

In Figure 15 we see an analogous graph to that of Figure 14 but now in the presence of a past market order imbalance that was in decile 1, i.e., there was a large excess of sellers versus buyers. The results are essentially the mirror image of those found for decile 10 past imbalances. Finally, in Figure 16 we see the relation between returns and contemporaneous market order imbalance conditioned on the presence of a neutral - decile 5 - market order imbalance in the past. In contrast to the cases of large positive or negative imbalances, we see here that the curves are completely symmetric with respect to the magnitude of the contemporaneous market order imbalance. In other words, the presence of a neutral past imbalance does not affect differently the relative impact of a buy versus a sell. However, this does not mean that there is no effect. Comparing the curves *buy15* and *sell15* with *buy15-05* and *sell15-05*, we see that the impact when there has been neutrality in the past market order imbalance is actually greater than when this conditioning information is not accounted for, the extra impact for an imbalance of about 15% of ADV being about 10bp.

6 Conclusions

In this paper we have presented a largely empirical study of the relation between order flow and returns/market impact. We initially argued, at the microscopic level, where all order flow events were considered, that price was a deterministic function of order flow. Thus, the true underlying task to try and understand the dynamics of price was to try and understand the dynamics of order flow itself. We argued that the right framework for this was just the standard one of understanding the relation between supply and demand. In distinction to many other studies we argued that these could not be adequately described using a single variable, such as a Lee-Ready tick rule determined market order imbalance.

Although, it is not clear what is the optimal set of variables with which to describe universal properties of order flow, we argued that a useful set consisted of eight basic event types: limit order, cancel, market order executed against displayed liquidity and market order executed against dark liquidity, each with its side - buy or sell. From these we argued that much could be understood by considering the three components of an imbalance vector - associated with the net liquidity - signed limit order volume minus signed cancel volume - and the imbalance in the two different types of market order.

With the different characterizations of order flow we postulated an associated equation of state, relating functions of returns/market impact to both the contemporaneous and past values of our phenomenological imbalance vector. Our goal was then to study this equation of state using empirical data consisting of all trades and quotes at the best offer/bid across all US exchanges over the period April 2008 to April 2009 and across all traded stocks. The data was aggregated over different time intervals - 5, 15, 30, 65, 130 and 195 mins.

We first studied the relationship between contemporaneous values of the net liquidity, $LCO(t)$, and the market order imbalance, $MO(t)$. We found over short time intervals, about 5 mins, that there was a substantial correlation between buying through limit orders and buying through market orders. However, over longer time intervals, greater than 15 mins, we saw that this relation was reversed, with now an anti-correlation - more buying/selling through market orders being associated with more liquidity on the offer/bid. We argued that this change was due to the market adapting to the presence of correlated order flow of a given type, such as might be associated with large institutional orders, the time scale for the market to detect such orders being typically of the order of 15 mins. The actual statistical inference has two important dimensions: The size of the imbalance caused by the order and its duration. Both increase the signal to noise ratio in the general order flow of the market and lead to an easier detection problem.

We showed that, as is well known, order flow is highly self-correlated. However, in this study we considered two novel elements: First, the change in magnitude of a particular imbalance component, showing that there was a substantial suppression from one time interval to another, that was more acute for market orders than limit orders; and, second, the relation between different imbalance components. For the latter, we showed that excess liquidity at the offer/bid was associated with an excess of market order buys/sells in the next interval. We also explained the suppression in the magnitude of the auto-correlated imbalances by noting that there was a substantial probability of a market "bounce", so that high positive/negative imbalances in the past could lead to high negative/positive imbalances in the next interval. We argued that this was due to the removal of an institutional order flow stream associated with a large meta-order. Such a stream is detected by the market, which in its turn, according to the laws of supply and demand, induces a compensating back reaction, e.g., liquidity provision at the offer in the case of an excess of buy market orders. Thus, the market pushes back against this institutional order flow. When the latter is removed, the order flow imbalance changes sign causing, generically, a mean reversion in price.

We studied this bounce in more depth by considering the full probability distribution for a particular imbalance given different conditioning information in past time intervals. We showed that this distribution is highly bimodal and thereby associated with a high variance, which is an important source of the difficulties in predicting future order flow and returns. Especially the latter, where we showed that the associated bimodal probability distribution relating past order flow to future returns was almost completely symmetric. In other words, for example, a high positive market order imbalance in the past was just as likely to lead to a large negative price movement as a positive one. This is a manifestation of market efficiency.

We showed that returns/market impact depended sensitively on all imbalance components, the returns being greater the more the different components were "aligned", e.g., that there was an excess of buys over sells for both types of market order and also limit orders. We showed that the effect of a market order imbalance was very different depending on the corresponding net liquidity imbalance. The more liquidity available, the lower the impact for a given market order imbalance. We also showed that returns/market impact also depended on past trading behavior. The impact of a given contemporaneous imbalance was less in the presence of a continuous corresponding past imbalance of the same sign than of the opposite sign. In other words, buying after the market is habituated to the corresponding order flow has less impact than if the order flow was in the opposite direction. We also saw that impact was relatively greater if the market had been directionless and then is suddenly moved by a large imbalance.

By these observations we saw that there are many more aspects of order flow that affect returns in a "universal" way than just a market order imbalance, thus showing the inadequacy of simple one-variable equations of state.

Acknowledgements

CRS and HW are grateful to Bruce Lehmann and Larry Glosten for helpful discusions. CRS is grateful to Enrique Martínez for help with the figures. This work was partially supported by DGAPA grant IN120509 and a CONACyT grant to the C_3 - Centro de Ciencias de la Complejidad.

References

[1] J.P. Bouchaud, J. D. Farmer, and F. Lillo (2009) How markets slowly digest changes in supply and demand, in Handbook of Financial Markets: Dynamics and Evolution, T. Hens and K.R. Schenk-Hoppe (Editors), 57 (North-Holland, 2009).

[2] N. Torre (1997) BARRA Market Impact Model Handbook. BARRA Inc., Berkeley.

[3] Almgren, A., Chee Thum, Emmanuel Hauptmann, and Hong Li (2005) Direct Estimation of equity market impact, Risk, July 2005.

[4] F. Lillo, J. D. Farmer and R. N. Mantegna (2003) Nature 421 129-130.

[5] M. Potters, J.P. Bouchaud (2003) Physica A, 324, 133.

[6] A. Kempf, O. Korn (1999) Journal of Financial Markets 2, 29.

[7] X. Gabaix, P. Gopikrishnan, V. Plerou, and H.E. Stanley (2003) Nature 423, 267.

[8] H. Eugene Stanley, Vasiliki Plerou, Xavier Gabaix, A statistical physics view of financial fluctuations: Evidence for scaling and universality, Physica A 387 (2008) 3967-3981.

[9] Farmer, J.D., Gerig, A., Lillo, F. and Waelbroeck, H.: Fair Pricing and the Market Impact of Large Institutional Trades, Pipeline Preprint 2009-10-002.

[10] Gerig, A.,"A Theory for Market Impact: How Order Flow Affects Stock Price." 2007, PhD Dissertation, University of Illinois.

[11] Bouchaud, J.P., Gefen Y., Potters, M., and Wyart, M. Fluctuations and response in financial markets: The subtle nature of "random" price changes, Quantitative Finance, 4(2), 176-190, April 2004

[12] P. Weber and B. Rosenow, Order book approach to price impact, Quantitative Finance 5, 357 (2005).

[13] Lee, Ch. and M. Ready (1991) Inferring Trade Direction from Intraday Data, Journal of Finance 46, 733-746.

[14] F.Lillo and J.D. Farmer (2004) The long memory of efficient markets. Studies in Nonlinear Dynamics and Econometrics 8, 1.

[15] Hasbrouck, J. (1991) Measuring the Information Content of Stock Trades, J. Finance 46, 179-207.

[16] Criscuolo, A. and Waelbroeck, H.: Optimal Execution of Portfolio Transactions: The Effect of Information Leakage, Pipeline Preprint 2009-10-004

Figures

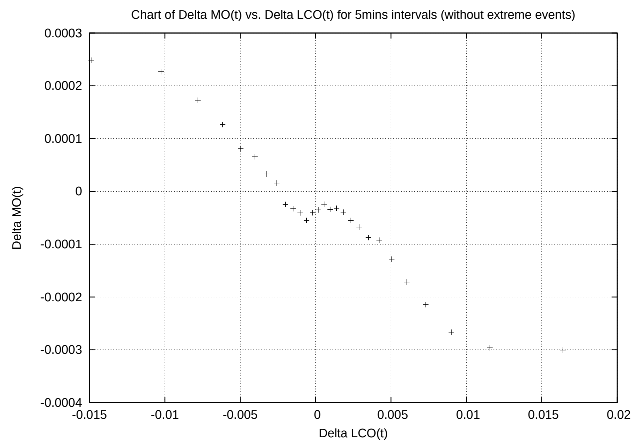

Figure 1: Graph of average contemporaneous market order imbalance for 5 minute intervals conditioned on the contemporaneous net limit order imbalance.

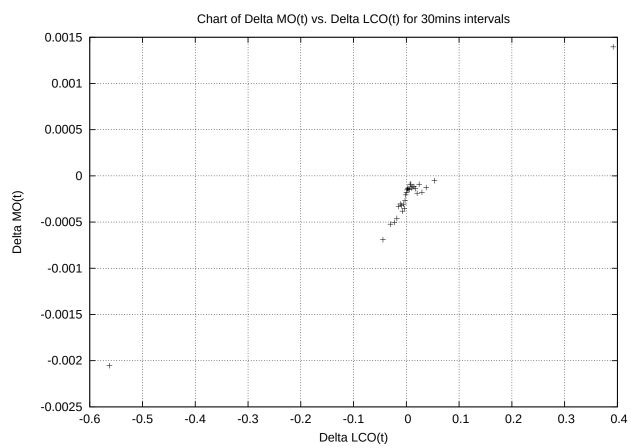

Figure 2: Graph of average contemporaneous market order imbalance conditioned on the contemporaneous net liquidity imbalance for 30 min intervals.

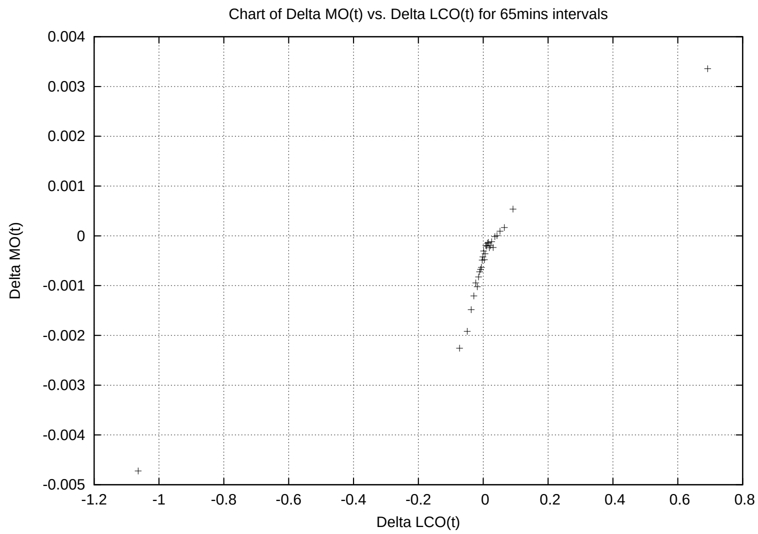

Figure 3: Graph of average contemporaneous market order imbalance conditioned on the contemporaneous net liquidity imbalance for 65 min intervals.

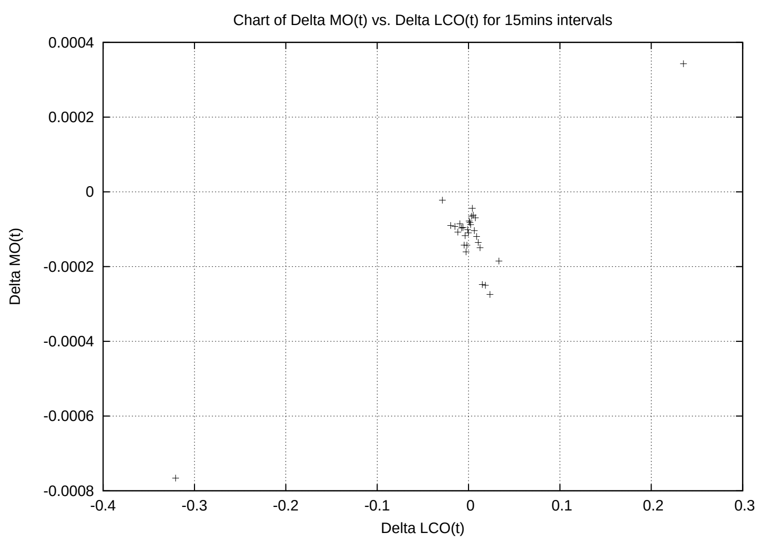

Figure 4: Graph of average contemporaneous market order imbalance conditioned on the contemporaneous net liquidity imbalance for 15 min intervals.

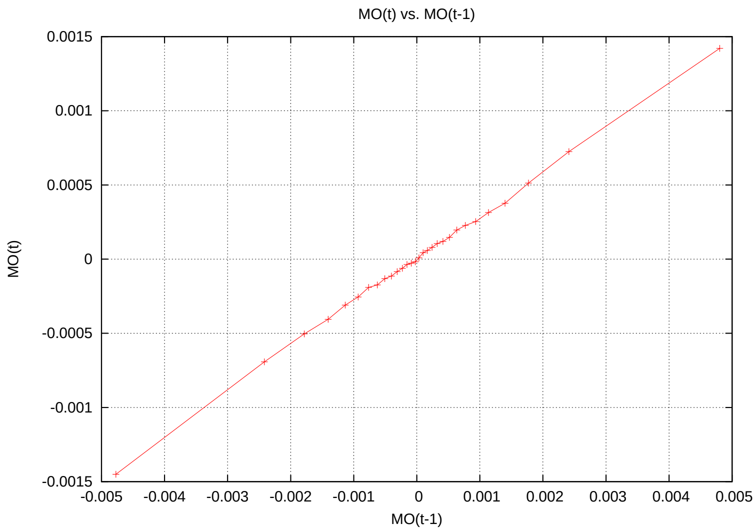

Figure 5: Graph of $MO(t)$ vs. $MO(t-1)$ for 5 minute intervals.

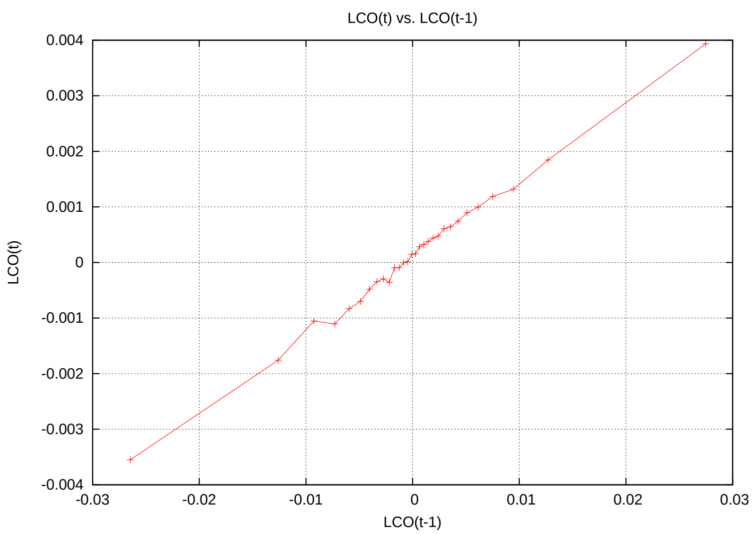

Figure 6: Graph of $LCO(t)$ vs. $LCO(t-1)$ for 5 minute intervals.

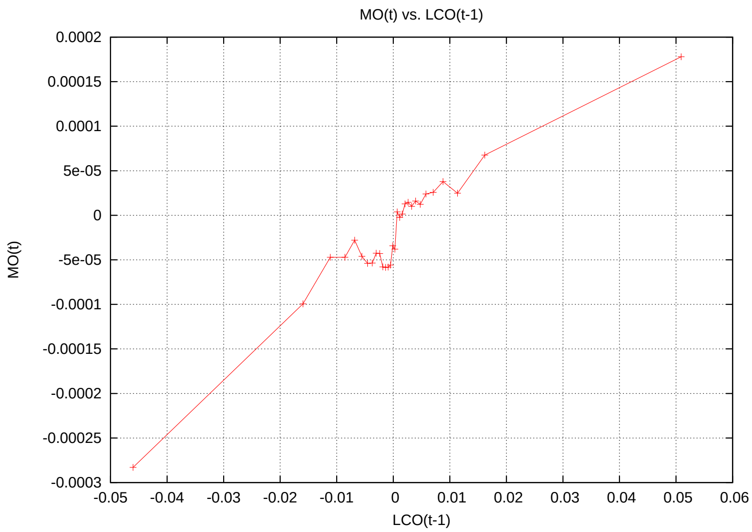

Figure 7: Graph of $MO(t)$ vs. $LCO(t-1)$ for 5 minute intervals.

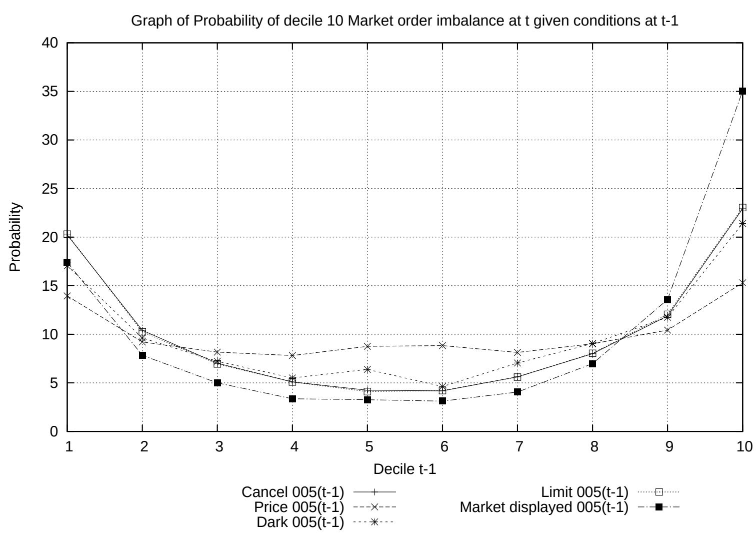

Figure 8: Graph of probability of decile 10 value of $MO(t)$ given conditions at $t-1$.

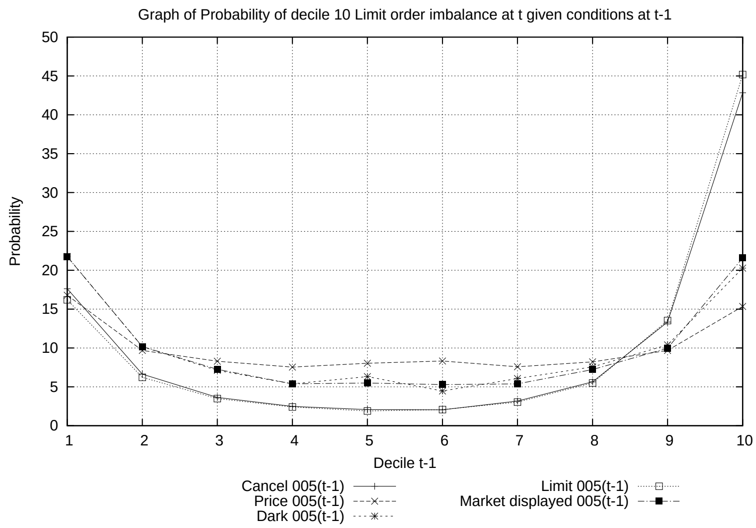

Figure 9: Graph of probability of decile 10 value of $LCO(t)$ given conditions at $t-1$.

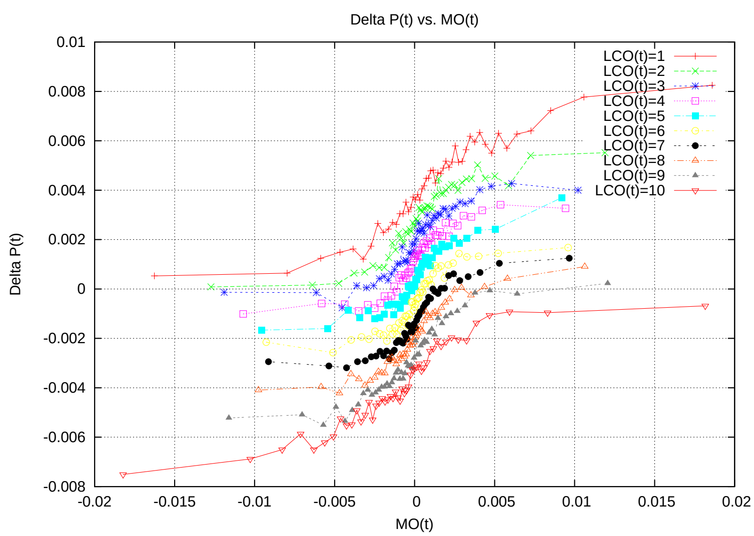

Figure 10: Graph of Returns versus contemporaneous market order imbalance for aggregated 5 min intervals

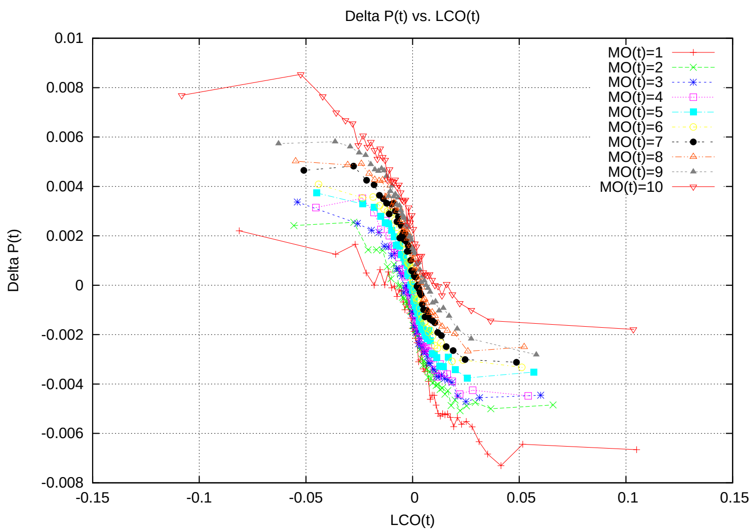

Figure 11: Graph of Returns versus contemporaneous net liquidity (limit order - cancel) imbalance for aggregated 5 min intervals

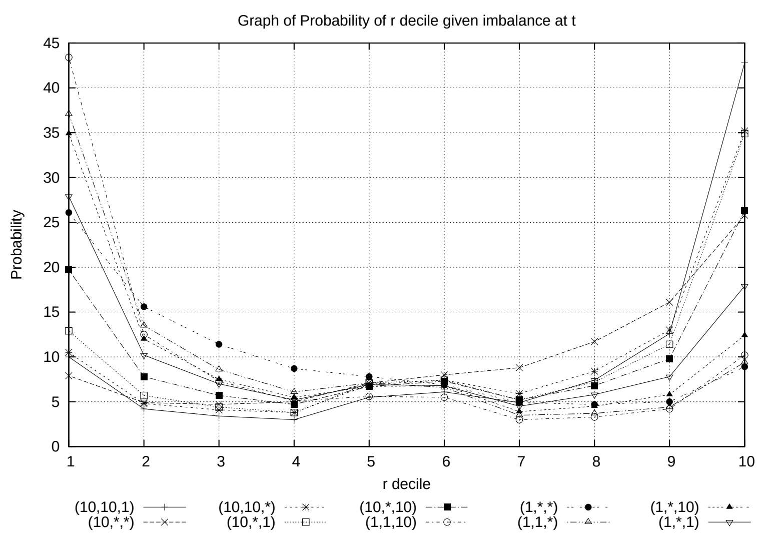

Figure 12: Graph of Returns versus different states of the contemporaneous imbalance vector aggregated over 5 min intervals

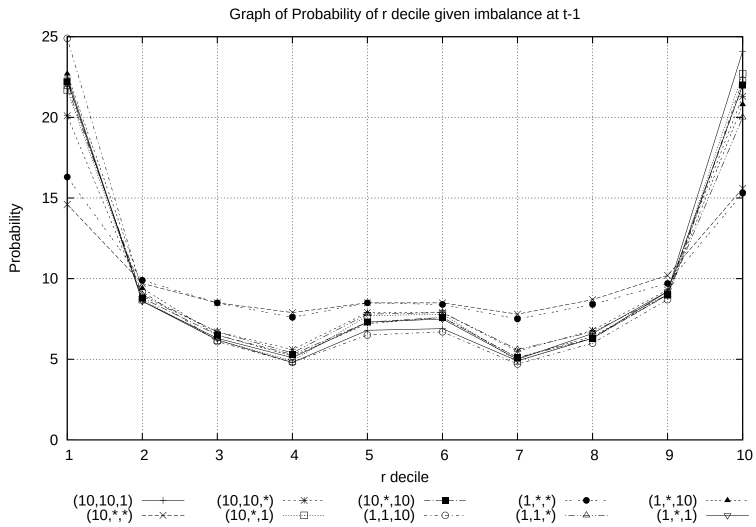

Figure 13: Graph of Returns versus different states of the imbalance vector at $t-1$ aggregated over 5 min intervals

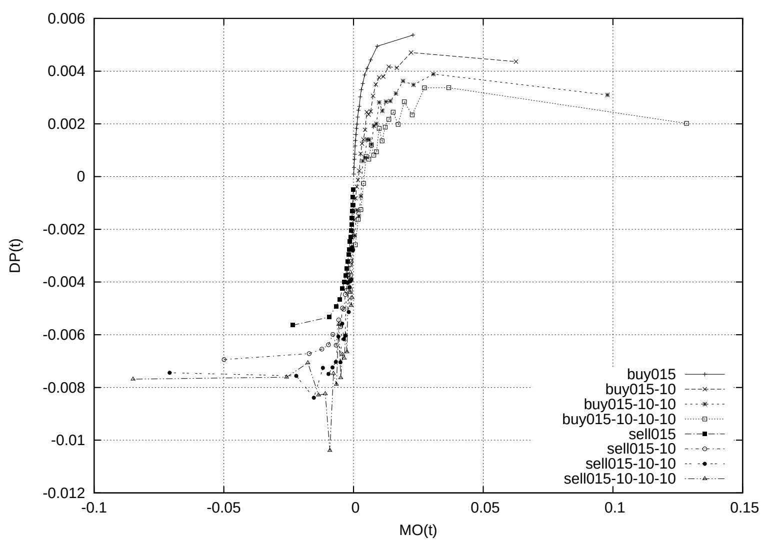

Figure 14: Graph of Returns versus contemporaneous market order imbalance conditioned on the presence of strong past positive market imbalances

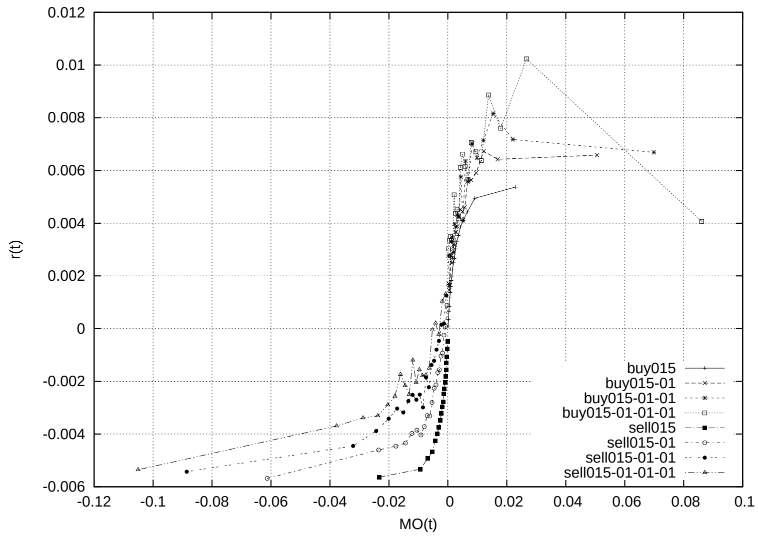

Figure 15: Graph of Returns versus contemporaneous market order imbalance conditioned on strong past negative market order imbalances

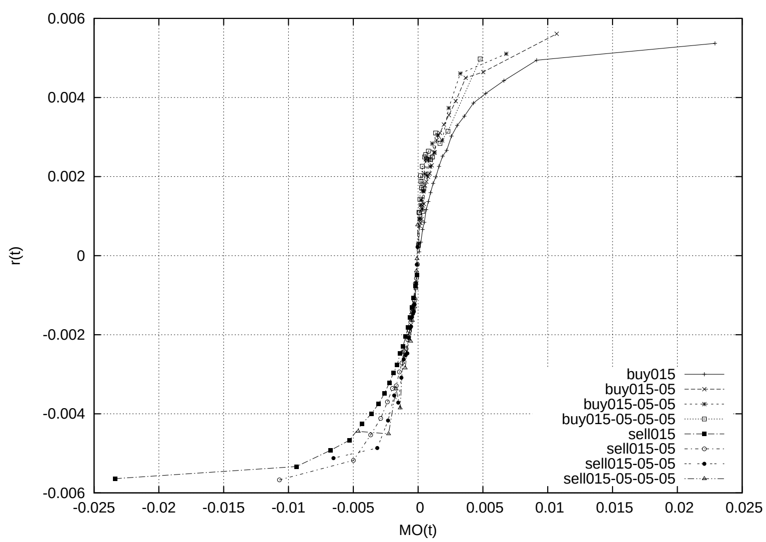

Figure 16: Graph of Returns versus contemporaneous market order imbalance conditioned on past neutral market order imbalances