An Assessment of the Validity of Linear Propagator Models for Modelling Market Impact

by

William Goldberg (CID: 00931083)

Department of Mathematics
Imperial College London
London SW7 2AZ
United Kingdom

Thesis submitted as part of the requirements for the award of the MSc in Mathematics and Finance, Imperial College London, 2017-2018

Declaration

The work contained in this thesis is my own work unless otherwise stated.

Signature and date:

Acknowledgements

I would like to thank Pietro Siorpaes for his advice and Mikko Pakkanen for providing the NASDAQ limit order book data. Special thanks go to Christian Morgenstern, Stefan Curtress and James Kelly for proposing this project and for providing guidance and many thoughtful discussions.

Abstract

Market impact is a crucial area of study in financial markets. We explore a class of so-called 'propagator models' which consist of two main types. The Transient Impact Model (TIM) models returns as a convolution of order signs and a propagator kernel. We also explore the History Dependent Impact Model (HDIM) which states that the deviation between the realised order sign and it's expected value has a permanent and linear effect on the price. Both models are generalised to two events types: price changing and non-price changing. We show that these models perform well for small tick stocks but not for large tick stocks. Additionally, some perform well for futures data, provided the asset is traded actively enough in any given day.

Contents

List of Figures

1 Introduction

Financial markets are constantly evolving and adapting to new technologies. The last few decades have seen the rise of algorithmic trading. Algorithmic trading involves executing trading decisions via computer programs with electronic market access. They were initially introduced at the start of the 1990s as a tool for institutional investors but a number of factors including evolving exchanges, reduction in fees and creation of new markets aside from the traditional LSE or NASDAQ exchange, have led to an explosion in algorithmic trading. Now a typical sell side broker will trade trillions of dollars per year using algorithms [7, slide 2]. But what is the main advantage of algorithmic trading? Perhaps the most appealing feature it offers is not necessarily a way to maximise profits, but a way to minimise the execution costs of a trade. Execution costs are the difference between the cost of an ideal trade and the value at which it is actually traded. They consist of direct, predictable costs such as broker commissions, taxes and exchange fees, as well as indirect costs [8, page 1]. Prices are volatile and moments before a trade is placed, a piece of new information could be released causing the price to jump up. This leads to an increased cost of the trade and this is an example of *market risk*, an indirect cost. There is also the issue of *liquidity risk*. This is the extra cost incurred when there is insufficient liquidity to meet the required amount for a trade. More often than not small volumes can be executed close to the listed price so liquidity risk is insignificant for smaller traders, however when large volumes are under consideration these costs become significant and understanding the nature of the costs is highly nontrivial.

The scarce nature of the liquidity causes traders to split their orders into chunks (child orders) and execute them over a certain time period according to a pre-determined trading strategy (see section 2.4). Each time a child order is executed there is a price pressure in the same direction of the child order making subsequent child orders more expensive. This is known as *market impact*, another indirect cost. Additionally, if another market participant anticipates that another agent is going to make a large order, they can trade faster than the other agent and push the price in their favour. This is a strategy used by *High Frequency Trading* (HFT) firms. Understanding market impact is important for three reasons:

1. It is fundamental to understanding the nature of fair prices. Market impact and and price changes are inextricably linked and form a basis for determining a fair price for an asset.

2. A large proportion of execution costs can be from market impact. Understanding how a trade will impact the price is crucial for a trading firms profits, particularly when the order flow is large.

3. It aids financial regulation. A better understanding of market impact will prompt new market microstructure regulations which should promote competition. It also reveals connections

between systemic risk and the market setup [1, page 8, section 1.5].

The primary aim of this study is to investigate a specific class of market impact models known as 'propagator models', but before we continue any further we need to cover some market microstructure definitions and terminology.

1.1 Terminology

Traders execute their trades by submitting an *order* which specifies the quantity V, order sign $\epsilon = \pm 1$ (+1 for buy, -1 for sell), and some additional conditions which must be met to execute the trade. When a trade counter-party is found the order is *filled*. Orders can also be partially or entirely cancelled. The two main types of order are *market orders* and *limit orders* [9, pages 1-3].

Definition 1.1. A market order is an order where the price is not specified and the trade is executed at the best available price at the time of the order.

Definition 1.2. A limit order is a buy or sell order which is quoted with its corresponding *bid* or *ask* price: the bid for a buy order and the ask for a sell order.

There are two types of market: quote-driven and order-driven markets. In quote-driven markets the broker sets the bid and ask prices and provides liquidity using a centralised system of buy and sell orders. Order-driven markets are electronic markets where all unfilled limit orders are collected into a *limit order book* such that market orders are filled at the best available price. In recent years there has been a migration from quote-driven markets to order-driven markets which has led to substantial increases in market data. This has prompted numerous studies into order book dynamics.

The *best-bid* price $p_B(t)$ is the highest outstanding limit buy order in the limit order book and the *best-ask* price $p_A(t)$ is the lowest outstanding limit sell order in the limit order book. Orders remain in the limit order book until they are either filled or cancelled. We can formulate the limit order book mathematically [10, pages 4-6, section 2.1, 2.2] by considering a price grid $\{1, \ldots, n\}$ where each point is a multiple of a price tick. A price tick is the minimum increment at which a price quote can be submitted. We consider the number of outstanding limit orders $|X_p(t)|$, $1 \leq p \leq n$. If $X_p(t)$ is negative then there are $-X_p(t)$ bid orders at price p and if it is positive then there are $X_p(t)$ ask orders at price p. The order book follows the continuous time[1] process $X(t) := (X_1(t), \ldots, X_n(t))_{t \geq 0}$.

Definition 1.3. The best bid price $p_B(t)$ at time t is defined by

$$p_B(t) := \inf\{p = 1, \ldots, n, X_p(t) > 0\} \wedge (n+1). \tag{1.1}$$

[1] on an intra-day time scale we usually consider this as discrete time.

Definition 1.4. The best ask price $p_A(t)$ is defined by

$$p_A(t) := \sup\{p = 1, \dots, n, X_p(t) < 0\} \vee (n+1). \tag{1.2}$$

The difference between the best-bid p_B and the best-ask p_A is the *bid-ask spread* $s(t)$ and the *mid-point price* is the average of the best bid and best asking price

$$m(t) := \frac{p_A(t) + p_B(t)}{2}, \quad s(t) := p_A(t) - p_B(t). \tag{1.3}$$

Given most market orders are filled close to the best bid/ask a useful quantity is the number of outstanding limit orders $Q_i(t)$ at a given distance i from the best bid/ask

$$Q_i^B(t) = \begin{cases} X_{p_A(t)-i}(t) & 0 < i < p_A(t) \\ 0 & p_A(t) \leq i < n. \end{cases} \tag{1.4}$$

$$Q_i^A(t) = \begin{cases} X_{p_B(t)+i}(t) & 0 < i < n - p_B(t) \\ 0 & n - p_B(t) \leq i < n. \end{cases} \tag{1.5}$$

Essentially, the $Q_i(t)$ is the quantity available for *aggressive* orders[2]. If we consider a state of the order book $x \in \mathbb{Z}^n$ we define the new state of the order book following a limit order at price p

$$x_{p+1} := x \pm (0, \dots 1, \dots 0),$$

where the vector on the far right is a vector of zeros with a 1 at the p-th component. Table 1 shows the effect on the order book state x when an order of unit size arrives.

Order Type	Sign	Price	Quantity	Effect
Limit	Buy	$p < p_A(t)$	Increase	$x \to x_{p-1}$
	Sell	$p > p_B(t)$	Increase	$x \to x_{p+1}$
Market	Buy	-	Decrease	$x \to x_{p_A(t)-1}$
	Sell	-	Decrease	$x \to x_{p_B(t)+1}$
Cancellation	Buy	$p < p_A(t)$	Decrease	$x \to x_{p-1}$
	Sell	$p > p_B(t)$	Decrease	$x \to x_{p-1}$

Table 1: The impact on the order book at state x of an order of unit size at price p [10, pages 4-6, section 2.1, 2.2].

The dynamics of the order book are thus governed by market orders, limit orders and cancellations submitted by traders. There are two main types of traders in financial markets: market makers and informed traders.

[2] An aggressive order is an order that removes liquidity from the book and usually is executed at a non-favourable price

Definition 1.5. A market maker is an agent which opposes a trade and provides the liquidity for another market participant who wishes to buy or sell a financial instrument. They do not take a position in the market and their main source of profit is a fraction of the bid-ask spread. We will assume that market makers primarily submit limit orders.

Definition 1.6. An informed trader, or liquidity taker, is an agent who triggers trades by placing market orders. They hope to gain from some piece of information and aim to make profits from correct predictions in movements of the price [11, page 3].

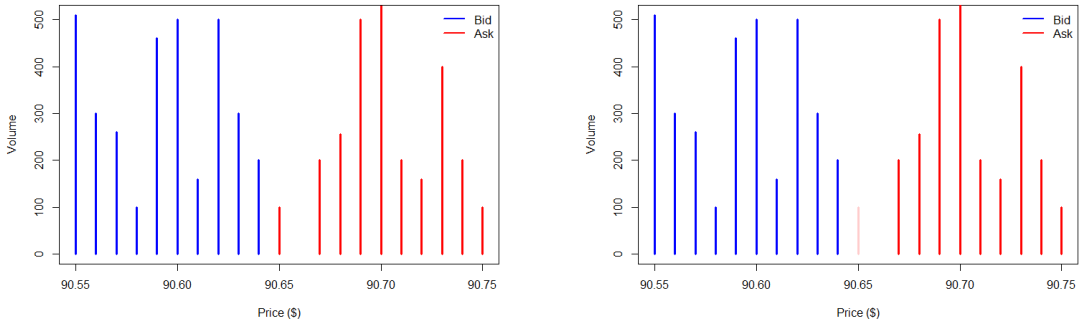
We can further characterise an order book event π into price changing and non-price changing events [12, table I, page 4]. Price changing events can be:

1. A market order with volume V greater than the outstanding volume at the best bid/ask.
2. A complete cancellation of the best bid/ask.
3. A limit order inside the bid-ask spread.

Non-price changing events consist of:

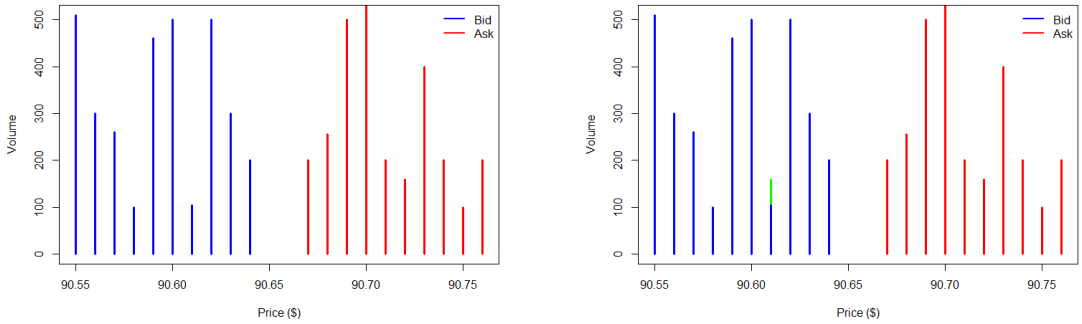
1. A market order with volume less than the outstanding volume V at the best bid/ask.
2. A partial cancellation of the bid/ask queue.
3. A limit order at or outside the bid-ask spread.

Figure 1 shows an example of a price changing event for Microsoft (MSFT). A market order of 100 shares is submitted and filled at \$90.65, the best-ask. The best-level consists of 100 shares so the level is completely used up and the best-ask moves to the next level at \$90.67, hence the mid-price increases from \$90.645 to \$90.655. Figure 2 shows a non-price changing event. A buy limit order is placed at the fourth level resulting in no change in the mid-price. Figure 3 shows another non-price changing event, a complete cancellation of a limit order at the best-bid. The cancellation is not large enough to deplete the entire best-bid volume



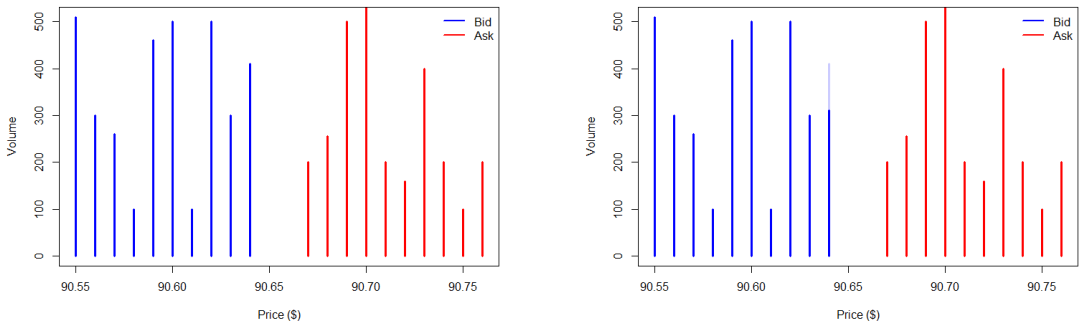
(a) Before (b) After

Figure 1: A buy market order of 100 shares (Microsoft) is filled at the best ask ($90.65) and the best-ask shifts two ticks to the next best price level ($90.67).



(a) Before (b) After

Figure 2: A non-price changing event. A buy limit order of 55 shares (Microsoft) is placed at the fourth level ($90.61), highlighted in green.



(a) Before (b) After

Figure 3: A complete cancellation of a limit order of 100 shares (Microsoft) at the best-bid ($90.64).

1.2 Motivation

We have already explained the importance of understanding market impact but before we continue we should explain the motivation behind propagator models. This stems from the origin of the random walk nature of prices. There are two different hypotheses which propose contradictory origins for the random walk nature of prices. The first of these is the *Efficient Market Hypothesis* [13, page 5] which posits that the price of a financial instrument reflects all of the available information and are thus efficient. Prices are by definition unpredictable and can only be influenced by news events which are not anticipated. As a result, prices are predicted to be random walks. Schiller et. al [14, page 14] noted that the measured volatility of stock prices was too high to be attributed to new information. A more concrete criticism of this hypothesis observes that trades on liquid stocks occur at fractions of a second, whereas the time increment between relevant news events is longer. The conclusion is that the assumption of perfectly informed and rational agents is flawed.

The second hypothesis, at the other extreme, assumes that market participants are completely irrational and take completely random decisions on whether to buy or sell a particular asset and their choices are interpreted by all other market participants as potentially containing some information [15, page 3] [16, page 1]. A trade decision will potentially change the bid or the ask price and thus the corresponding mid-price. This new price is adopted immediately by other market agents as the price around which new decisions are made. This hypothesis also predicts a random walk nature of prices and is popular due to its analytical tractability.

Clearly, the true case must lie somewhere between these two extremes. Bouchaud et. al [11, page 3] proposed that the random walk nature of prices originates from the delicate interplay between market makers and informed traders.

To demonstrate this fine-tuned competition consider the following. Suppose the informed trader believes the price will go up and so they want to buy. An increased number of buy orders are placed in the order book. The market maker is now incentivised to increase their asking price since these buy orders might be from well informed traders taking advantage of price inefficiencies. Should the price go up, the market maker will likely lose money when they have to close their position. To combat this, the informed trader will place a *meta-order*, whereby they split their total order into smaller orders. This leads to temporal correlations in the *signs* of the trades.

If the market orders do not contain any relevant information then the prices will mean-revert over time, rather than go up. The market makers drive this phenomena. The mean reversion has to be slow, otherwise the informed trader can take advantage of the lower prices. The action of these two types of market agents removes linear correlations in the price changes, prevents statistical arbitrage and leads to the random walk nature of prices.

Bouchaud et. al [11, section 3, page 13] proposed the class of propagator models to explain this origin of the random walk nature of prices. They based their models on empirical observations and provide numerous papers outlining their performance. Their focus is mainly on stocks, although they do occasionally consider futures. In this study we will implement a number of propagator models and assess their performance for futures data in more detail, as well as for common stocks.

The outline of the paper is as follows: first we present some background material on market impact models and optimal order execution. Then we outline the data and tools used for this study before presenting some empirical stylised facts found in the data. Following this we define each propagator model before implementing and then we present our findings in section 6. We offer a critique for each model before drawing conclusions on the effectiveness of each model for each dataset.

2 Background of Impact Models

2.1 Price Definition

First we consider the definition of price of a financial instrument, since defining the price becomes more challenging when considering the dynamics of the limit order book. We can define the price in a number of ways [1, page 6, section 1.3]:

1. Price of the transaction. The price of the asset is quoted as the price of the most recent transaction.

2. The average of the transaction prices. The average can be taken over time or volume and produces the time-weighted average price (TWAP) and volume weighted average price (VWAP) respectively.

3. The mid price m_t.

4. The *micro-price* P_t^{micro} (a generalistaion of the mid-price, see appendix).

Each of these price definitions has its own drawbacks, for example transaction prices can only be quoted on a discrete time setting (i.e. at every transaction time). For the remainder of this study, we will use the mid price to define the fair price of the asset under consideration. Extensions to the micro-price are left as an area for further study. Figure 4 shows an excerpt of MSFT's mid price and the corresponding bid-ask prices. This shows the dynamics of price changes at a microstructure level.

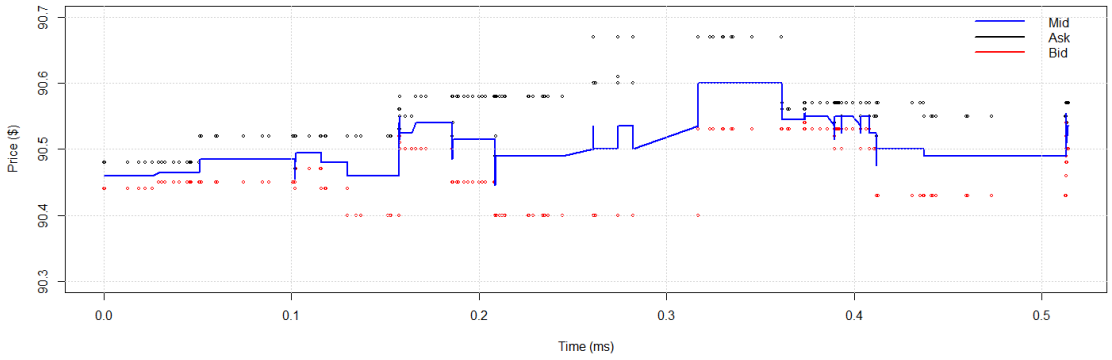


Figure 4: An excerpt from MSFT showing the best bid, ask and the mid-price. This indicates at a microstructure level how prices change.

2.2 Market Impact

Market impact can be thought of as the correlation between price changes and an incoming order. A buy trade will on average, depending on a number of factors, push the price up. Regarding order book dynamics, this occurs when a market order uses up all of the available volume at the first level. The best-bid moves to the second level which is higher, causing an increase in the mid-price. Price impact is generally considered a cost since a market agent's second buy trade will likely be more expensive than the original. The control of market impact is particular active area of research since order flow can total in the billions of dollars for some firms. Savings of even fractions of basis points can be worth a substantial amount of money. The primary considerations for industry practitioners will be the volume dependence of the market impact and the nature of the impact: is it permanent or temporary?

It would be reasonable to consider a transaction as a fair deal between the buyer and seller. This raises questions as to the origin of market impact, since a fair transaction should eliminate market impact. Bouchaud [17, pages 1-2] proposed three possible origins:

1. Market participants are able to predict temporal price movements and adjust their decisions accordingly. Whilst the corresponding trades may have no direct impact on the price, they do cause correlations between trades and future prices. For example, if a trader predicts a price rise he will likely buy immediately.

2. Trades reveal hidden information. A price change can be triggered by an news event or any form of incoming information. Since trades are anonymous it makes informed and non-informed traders indistinguishable. All trades will thus impact the market since other market participants will be aware that a proportion of these trades will be well informed decisions.

3. Market impact is a statistical effect caused by order flow variations. Differences in supply and demand can be totally random, yet an additional buy order, *ceteris paribus*, will on average increase the price. These fluctuations occur irrespective of information.

The first two points can be thought of as orders not directly impacting price but acting as predictors for future prices. If the third point is correct, then this indicates that order flow and price change correlations are essentially the same thing and prices only move due to trades, irrespective of the number of well-informed traders (see section 2.3.3).

2.3 Market Impact Models

Here we will cover simplistic market impact models. The main scope of this study is propagator models which are covered in section 5. These model market impact based on a number of empirical observations as outlined in section 4. Firstly we will consider the most simplistic of impact models: the Kyle Model.

2.3.1 The Kyle Model

Kyle [18] considered a scenario whereby a single informed trader will decide their order size such as to maximise their profits (this can also be thought of as minimising costs). He proposed the parameter λ which is generally interpreted as the inverse of the available liquidity in the market, or a measure of impact [17, section 2, page 3]. Thus the price change can be represented as

$$dp = \lambda \epsilon V, \tag{2.1}$$

where ϵV is the signed volume of the trade. The total price change between time $t = 0$ and $t = T = Ndt$ is therefore

$$p_T = p_0 + \sum_{n=0}^{N-1} dp_n = p_0 + \lambda \sum_{n=0}^{N-1} \epsilon_n V_n, \tag{2.2}$$

i.e. the price impact is permanent until time T and a linear function of the traded volume V. One assumption of this model is that the signs ϵ must be uncorrelated, otherwise the price would not

resemble a random walk and would have an element of predictability. We will show later in section 4 that this is not the case.

2.3.2 Hasbrouck's Vector Autoregressive Model

Hasbrouck [19] formulated a model under the assumption that the size of the price impact of a given trade is a function of the proportion of informed traders, the probability that said trader is actually well informed and the precision of the private information they possess. These factors can be considered as to what extent does information asymmetry exist. Information asymmetry comes with two predictions: firstly that the asymmetry is positively correlated with the spread and secondly that it is also positively correlated with price impact of a trade. Hasbrouck's classic paper investigates the second prediction.

Hasbrouck's VAR model posits that the trades from an econometric perspective form a system governed by auto and cross-correlations. He defines the information impact as the price impact caused by the unexpected component of the trade. Vector autoregression is the technique of choice since basic computations are linear and whilst the model is analytically tractable, it can also be extended to incorporate nonlinear effects. It is worth noting that although this model was initially designed for quote-driven markets it can be extended to order-driven markets. Hasbrouck's VAR model consists of the following bivariate vector autoregression model:

$$r_t = \sum_{j>0} B_{rr}(j) r_{t-j} + \sum_{j\geq 0} B_{xr}(j) x_{t-j} + v_{r,t} \tag{2.3}$$

$$x_t = \sum_{j>0} B_{rx}(j) r_{t-j} + \sum_{j>0} B_{xx}(j) x_{t-j} + v_{x,t}, \tag{2.4}$$

where r_t are the returns, $x_t := V_t \epsilon_t$ is the signed volume of the trade, the $B(j)$ are regression coefficients, $v_{1,t}$ is the disturbance reflecting public information and $v_{t,2}$ is the unanticipated component of the trade. This model can in theory be of infinite order but in practise has to be truncated at some maximum lag.

2.3.3 Random Order Flow and Latent Order Books

Random order flow is a widely studied topic in order flow dynamics since it provides simple, analytically tractable models. Random order flow assumes that there are no informed traders and decisions are made based on 'zero intelligence' [20] [21] [22]. This is analogous to point 3 in section 2.2 above. These models consider the order book as a price grid, as described in section 1.1. The limit order flow and market order flow are both described by a Poisson process with rates λ and μ per unit event and time respectively.

Donier et. al [23] extended these random order flow models to the *latent order book*. This model is an extension of propagator models and is primarily aimed at capturing nonlinear effects. The

latent order book is relevant when considering a mesoscopic time-scale. The visible limit order book hides the intentions of high-latency traders since the limit order book mainly displays the activity of low-latency traders. Therefore, the limit order book does not reflect the true supply and demand in the market [1, page 10, section 1.5.3]. The latent order book aims to describe the true supply and demand. It is a fictitious order book which records intentions of market participants.We consider the price $m(t)$ at some time $t \in [t, t + dt)$. A high-latency trader will decide to put a *latent order* of size V at price $m(t) + \Delta$ with probability λ, i.e. the order only appears when the price reaches around $m(t) + \Delta$. A schematic of a latent order book is shown in Figure 5.

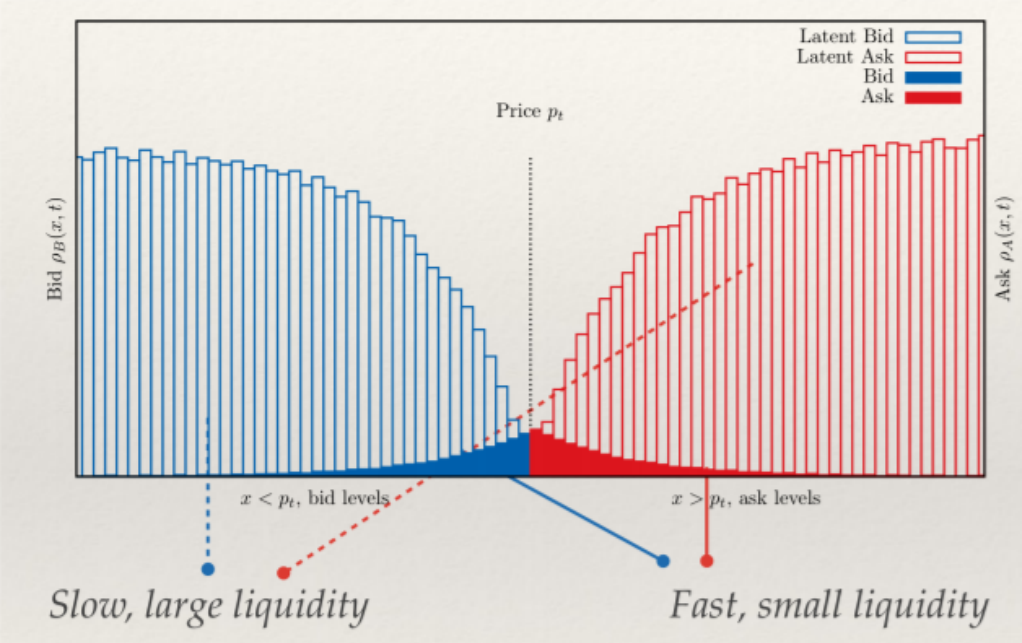


Figure 5: A diagram of the latent order book. The 'slow' liquidity largely exceeds the 'fast' liquidity [1, page 11, figure 1.4].

Donier et al. named $x(t) := m(t) + \Delta$ the reservation price. They set the following set of partial differential equations for the volume density $\rho(t, x)$:

$$\frac{\partial \rho_B(t,x)}{\partial t} = -v_t \frac{\partial \rho_B(t,x)}{\partial x} + D\frac{\partial^2 \rho_B(t,x)}{\partial x^2} - \nu \rho_B(t,x) + \lambda \Theta(p_t - x) - \kappa R_{AB}(x,t) \tag{2.5}$$

$$\frac{\partial \rho_A(t,x)}{\partial t} = -v_t \frac{\partial \rho_A(t,x)}{\partial x} + D\frac{\partial^2 \rho_A(t,x)}{\partial x^2} - \nu \rho_A(t,x) + \lambda \Theta(p_t - x) - \kappa R_{AB}(x,t), \tag{2.6}$$

where ρ_A is the density of sell orders, ρ_B is the density of buy orders and p_t is the price averaged over high frequency noise defined by $\rho_A(t, p_t) - \rho_B(t, p_t) = 0$. The other terms are explained below.

1. & 2. The first two terms represent the drift diffusion: the market participant revises their reservation price x due to, for example, new information. This consists of a random element, a diffusion characterised by coefficient D and an overall shift (caused by new public information for example) v_t which shifts the entire book.

3. The third term considers cancellations of the latent orders which occur at the ν^{-1}.

4. The fourth term considers deposition, i.e. the emergence of new intentions to buy/sell at an intensity λ, governed by the function $\Theta(u)$.

5. The final term represents the reaction. Two orders meet with reaction rate κ and $R_{AB}(t,x)$ is the average product of the buy/sell volume densities.

We can introduce a change of variables $\phi(t,x) := \rho_B(t,x) - \rho_A(t,x)$ and $y = x - \hat{p_t}$ where $\hat{p_t} = \int_0^t v_s ds$ is the un-impacted price. The above equations become

$$\frac{\partial \phi(t,y)}{\partial t} = -v_t \frac{\partial \phi(t,y)}{\partial y} + D\frac{\partial^2 \phi(t,x)}{\partial y^2} - \nu\phi(t,y) + \lambda \text{sign}(p_t - \hat{p_t} - y), \tag{2.7}$$

where p_t solves the equation $\phi(p_t, t) = 0$. For a meta order that starts at $t = 0$ the latent order book can be characterised by the exact solution

$$\phi(y,t) = -\mathcal{L} + \int_0^t \frac{dsm_s}{\sqrt{4\pi D(t-s)}} \exp\left\{-\frac{(y_t - y_s)^2}{4D(t-s)}\right\}, \tag{2.8}$$

where $\mathcal{L}$ is the *latent liquidity*. Hence we can represent the translated reservation price as an integral equation

$$y_t = \int_0^t \frac{dsm_s}{\sqrt{4\pi D(t-s)}} \exp\left\{-\frac{(y_t - y_s)^2}{4D(t-s)}\right\}, \tag{2.9}$$

where m_t is the signed trading intensity (positive for buy meta-order). This can be used to calculate the market impact of a meta-order in the latent order book. This model is consistent with empirical observation, as outlined in section 4, particularly the square-root impact of meta-orders which is not captured by the Kyle model.

The response function measures the average price change conditional to a buy order at time 0 (or sell order with the opposite effect) at time l. This quantity is *not* the market impact to a single trade [11, section 2.3, page 7]. When we consider different types of order book events we need the conditioned response function.

2.4 Optimal Order Scheduling

Understanding market impact is beneficial for reasons outlined earlier. However more is required to manage execution costs. In order for market participants to control their execution costs they must utilise their knowledge of market impact to devise optimal order scheduling methods. Controlling the execution costs using optimal scheduling involves a trade off between market impact and market risk. Reducing market impact requires more passive trading whilst reducing market risk demands more aggressive trading. The most widely known works in optimal order scheduling are the Bertsimas-Lo model, Almgren-Chriss model and the Obizhaeva-Wang model [24, page 1]. We will briefly cover them here.

2.4.1 Bertsimas-Lo Model

Bertsimas and Lo [25, section 2, pages 4-9] provided a definition of best execution by minimising the expected trade cost using stochastic dynamic programming. Their work was notable since it was the first time dynamic programming had been used for best execution. The problems arise from the observation that trading takes time and affects the price dynamics as well as the current price. This influence over price dynamics affects future trading costs.

The problem is set up as follows: we want to buy a large number of units X of an asset between $t = 0$ and $t = T$, where $t \in \{0, 1, \ldots, T\}$. We denote X_t as the number of outstanding units and we want to choose the quantities $V_t := -(X_{t+1} - X_t)$ and Δt to minimise the expected impact cost of the trade. We will consider the case of permanent market impact. The asset dynamics are modelled according to the following:

$$S_t = S_{t-1} + \theta \Delta V_{t-1} + \sigma S_0 \Delta W_{t-1}, \tag{2.10}$$

where W_t is a Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. We want to choose $\{\Delta V\}_0^{T-1}$ such that expected cost C is minimised:

$$C(X_0, S_0) = \min_{\Delta V} E_0 \left[\sum_{t=0}^{T-1} S_{t+1} \Delta V_t \right]. \tag{2.11}$$

Bertsimas and Lo constructed a dynamic programming algorithm the based on the observation that an optimal solution $\{S_0^*, S_1^*, \ldots, S_T^*\}$ must be optimal for the remaining program for every t. This property is summarised by the *Bellman equation*:

$$C(X_t, S_t) = \min_{\Delta V} E_t[S_{t+1} \Delta V_t + C(X_{t+1}, S_{t+1})]. \tag{2.12}$$

This equation can be solved backwards by noting that the execution must be completed by time T so that $\Delta V_{T-1} = X_{T-1}$. Using induction it can be shown that

$$\Delta V_{T-i} = \frac{X_0}{T}, \quad \text{for } i = 1, \ldots, T, \tag{2.13}$$

i.e. it is optimal to split the orders equally across time. By substituting this into equation (2.11) and with a little algebra we get the total cost

$$C(X_0, S_0) = X_0 S_0 + \frac{\theta X_0^2}{2} \left(1 + \frac{1}{T}\right), \tag{2.14}$$

where the second term represents the cost due to market impact.

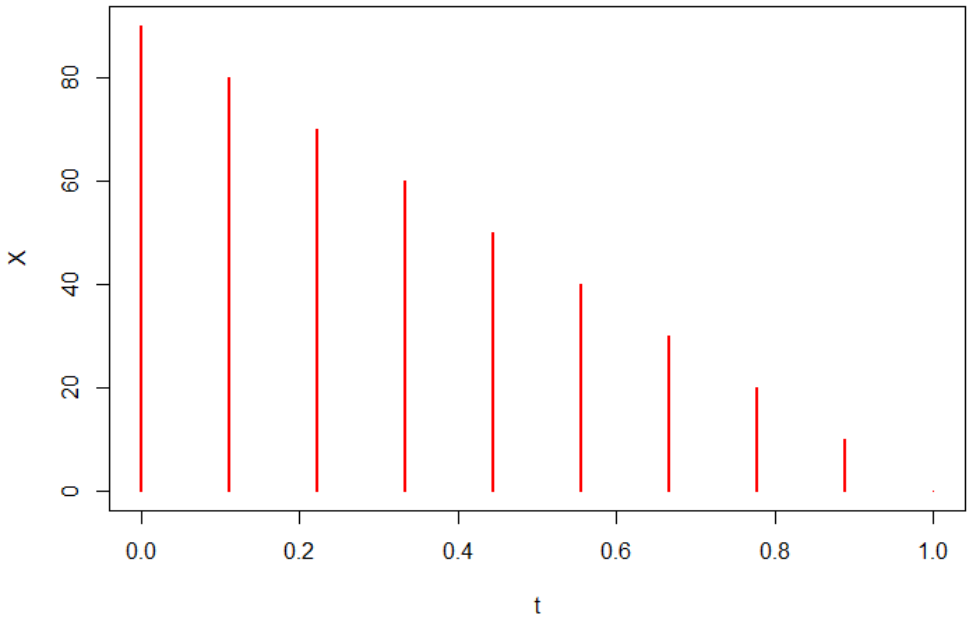


Figure 6: The execution profile in discrete time of the Bertsimas-Lo model for a parent order of 100 shares to be executed over the time period $T = 1$.

2.4.2 Almgren-Chriss Model

Bertsimas and Lo provided a definition of best execution, however their approach ignored the volatility of execution costs for different trading strategies. Almgren and Chriss [26, pages 3-15] worked in the framework of minimising the cost of trading with a penalty for cost uncertainty. The issue of cost uncertainty can be understood as follows: say we have an illiquid, highly volatile asset we want to trade. One could either trade all of the asset now at the expense of a high execution cost, or trade the asset in fixed amounts over a finite duration, with the added uncertainty of volatile prices. We want to optimise the trading strategy given a level of uncertainty, or appetite for risk.

We adopt a similar set up as in the Bertsimas-Lo model, except now we consider t continuous and the price dynamics are governed by

$$S_t = \tilde{S}_t + \eta v_t + \gamma(x_0 - x_t), \tag{2.15}$$

where $v_t := -\dot{x}_t$ is the rate of trading, η and γ are temporary and permanent respectively, and $\tilde{S}_t$ is the unaffected mid price given by a driftless arithmetic Brownian motion

$$d\tilde{S}_t = \sigma \tilde{S}_0 dW_t. \tag{2.16}$$

The execution cost is now

$$C = \int_0^T S_t dx_t, \tag{2.17}$$

so if we include the variance of the trading cost, our optimisation problem becomes the following:

$$\min_v E[C] + \lambda \text{Var}[C], \tag{2.18}$$

where λ represents the risk aversion of the trader. Following a little algebra the optimisation problem can be reduced to

$$\begin{aligned} &\min_v \eta \int_0^T \dot{x}_t^2 dt + \lambda \sigma^2 \tilde{S}_0^2 \int_0^T x_t^2 dt \\ &\text{subject to } x_0 = X, \quad x_T = 0. \end{aligned}$$

This optimisation problem can be solved using calculus of variations and has solutions

$$x_t = X \frac{\sinh[k(T-t)]}{\sinh(kT)}, \quad k = \sqrt{\frac{\lambda \sigma^2 \tilde{S}_0^2}{\eta}} \tag{2.19}$$

$$v_t = k x_t \coth[k(T-t)] \tag{2.20}$$

We can see from figure 7 that the execution profile depends on k, which is a function of the risk aversion parameter λ. A higher risk aversion means the trader wants to execute quickly since they are more concerned about market risk rather than market impact. This is evident in the figure since increasing λ leads to increasing k and earlier execution.

2.4.3 Obizhaeva-Wang Model

Both the Bertsimas-Lo and Almgren-Chriss models have one particular shortfall: they fail to describe the level of interactions with the order books. Obizhaeva and Wang [2, pages 1-5, 17-19] proposed a framework which describes the evolution of the limit order book. They showed that the optimal strategy depends on the interaction between the limit order book and a sequence of trades.

Obizhaeva and Wang noted that following the execution of a trade the order book does not revert to its previous state immediately, particularly in more illiquid markets. They modelled a trade as having a transient impact with an exponential decay with decay rate ρ. Therefore, the price has the following dynamics:

$$S_t = S_0 + \eta \int_0^t v_s e^{-\rho(t-s)} ds + \int_0^t \sigma_s dW_s \tag{2.21}$$

The optimisation problem then becomes

$$\min_v E_0 \left[\int_0^T S_t v_t dt \right] = S_0 x_0 + \min_v \eta \int_0^T v_t \left(\int_0^t v_s e^{-\rho(t-s)} ds \right) dt, \tag{2.22}$$

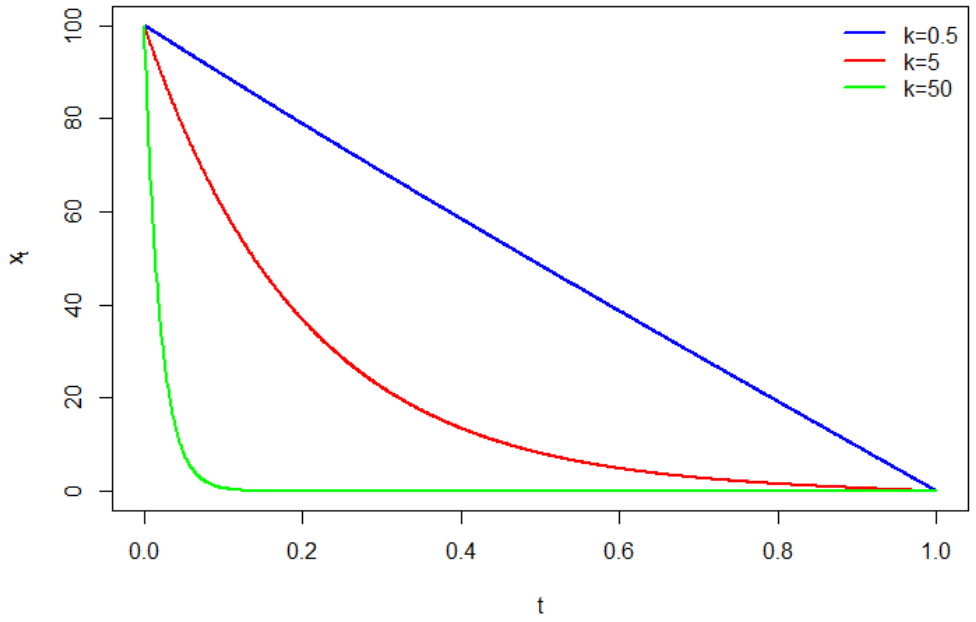


Figure 7: The execution profile of the Almgren-Chriss model for a parent order of 100 shares to be executed over time period $T = 1$. Execution profile depends on k which depends on risk aversion λ.

which has solution

$$v_s = \frac{x_0}{2 + \rho T} \left[\delta(s) + \rho + \delta(s - T) \right], \tag{2.23}$$

where $\delta(x)$ is the Dirac-delta function. The optimal execution strategy thus consists of two block trades at time $t = 0$ and $t = T$, both of size $x_0/(2 + \rho T)$. In between, the trading rate is constant $v_t = \rho x_0/(2 + \rho T)$. The total execution cost is

$$C^*(0, T) = S_0 x_0 + \eta \frac{x_0^2}{2 + \rho t} \tag{2.24}$$

The initial block trade serves the purpose of perturbing the limit order book from a steady state such as to attract new limit orders at better prices. The subsequent continuous trading is aimed at matching these limit orders and the final block trade matches any outstanding orders at the end of the time period. This different approach utilitses discrete and continuous trading methods and the cost saving depends on the dynamical properties of the order book, such as the time evolution following the trade. More strikingly, Obizhaeva and Wang showed that the optimal strategy is not sensitive to the instantaneous price-impact function, a feature not predicted by the other two models.

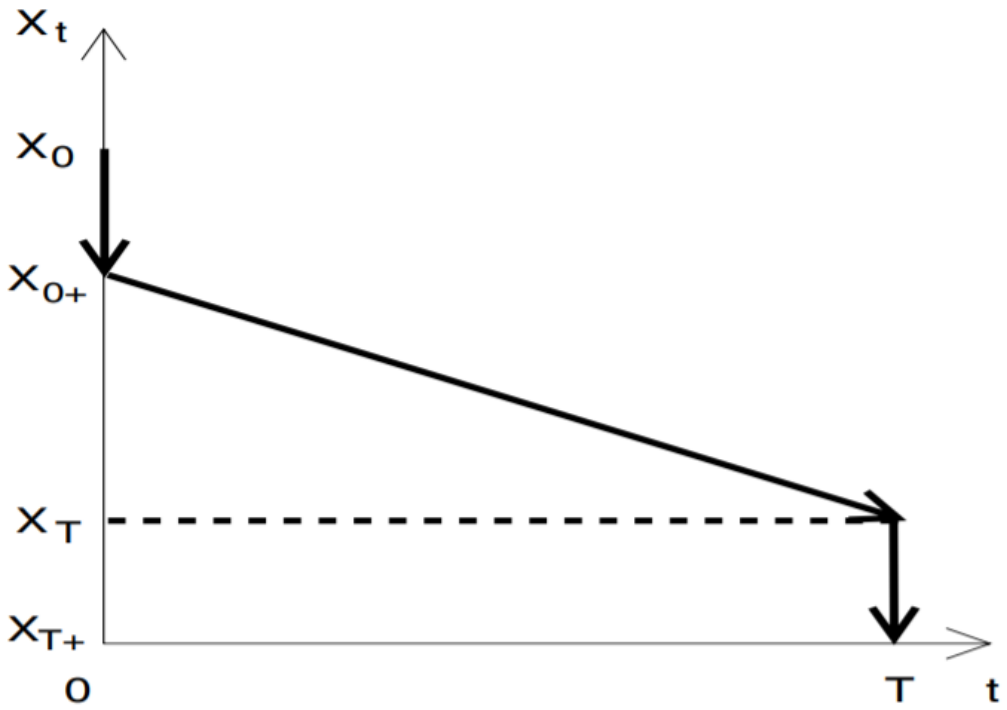


Figure 8: The execution profile of the Obizhaeva-Wang model for a parent order of X_0 shares to be executed over time period T. Optimal strategy is an initial block trade at $t = 0$ followe by a period of continuous trading until $t = T$, where another block trade of the same size as the initial trade is executed [2, figure 3, page 20].

3 Dataset and Tools

Here we outline the data and tools used to produce results in subsequent sections.

3.1 Discussion of Data

For each dataset, we collect the data per day and run our calculations on a daily basis before averaging.

3.1.1 Cash equities (common stock)

For this study, we selected a number of liquid stocks traded on the NASDAQ exchange between 02/04/2018 - 30/04/2018, a total of 21 trading days. Stocks from the NASDAQ change are convenient since the NASDAQ uses a constant tick size of \$0.01[3]. For this study we consider two types of stock: large tick and small tick stock. Large tick stocks tend to have a bid-ask spread of 1-2 ticks, whereas small tick stocks tend to have spreads of several ticks. As a rule of thumb, large tick stocks tend to have a lower price and small tick stocks tend to have higher prices. When we present our results, we will use EBAY as representative of large tick stocks and Amazon (AMZN) as representative of small tick stocks. The behaviour of these types of stock is different, as we will see later.

NASDAQ data was obtained from LOBSTER, a company which reconstructs limit order book data for all NASDAQ stocks[4]. The data from LOBSTER comes in two files: a message file and an order book file. Table 2 shows an example of the message file.

Time (s)	Event Type	Order ID	Size	Price	Direction
:	:	:	:	:	:
34567.254323	1	9086849	100	120050	-1
34567.334534	4	9086849	10	120050	1
:	:	:	:	:	:

Table 2: Sample of the message file

Time is the time of the order book event in seconds, accurate up to microseconds. Event type denotes what event has occured, for example 1 for a new limit order or 4 for an execution of a visible order (incoming market order). Order ID is the unique order reference number and Size is the number of shares. Price is given in dollars multiplied by 10^4 and Direction is the sign of the event: +1 for buy, -1 for sell. An example of a two level order book file is shown in table 3

[3]Stocks under a dollar use a separate tick size but these are not relevant for this study

[4]LOBs have to be reconstructed since only limited information is published by the exchange

A Price 1	A Size 1	B Price 1	B Size 1	A Price 2	A Size 2	B Price 2	B Size 2
:	:	:	:	:	:	:	:
1257700	1100	1257900	100	1258000	1000	1257800	2000
1257700	1000	1257900	100	1258000	1000	1257800	2000
:	:	:	:	:	:	:	:

Table 3: Sample of a 2 level order book file.

The A/B Price 1 is the price of the ask/bid at the first level (i.e. the best bid/ask). The A/B Size 1 is the corresponding liquidity at this level. A/B Price i where $i = 1, \dots n^{\text{levels}}$ is the similar for the next i levels. Each row in the message file corresponds to the same row in the order book file.

The data cleaning process consisted of selecting the required events for the required time-frames. The market operates between 9:30 and 16:00, however the first and last half an hour of trading per day was discarded. We also discarded trading days with shorter trading hours and checked for trading halts and volatility auctions. This study is concerned only with trades, hence all other order book events were discarded. We decided not to use the order signs provided by the exchange and defined a buy market order $\epsilon = +1$ as a trade executed above the mid-price, and vice-versa for sell orders. Trades at the mid-price were discarded. Returns were calculated as the log difference in mid-price between two trades: i.e.

$$r_t := \log m_{t+1} - \log m_t$$

, where m_t is the mid-price.

3.1.2 Futures

We also run analysis for two futures. We consider the S&P 500 E-mini futures contract for the same dates: the ESM8 [5]. S&P 500 futures are very actively traded so this provided a vast dataset for the specified days. We also consider the 10 year U.S Treasury Note Futures [6]. The 10Y T-Note futures dataset was collected for April 2018 and the contract valid at this time was TYM8.

The futures dataset was collected from Bloomberg and the data cleaning process was slightly different to the cash equities, since futures are traded over a longer period. They are not traded on the weekend. We discard any events which were not trades and calculated order signs as done for the equities.

[5] Futures contracts are only valid for a certain time period, after which a new one becomes valid. During April 2018 the contract was ESM8. It is currently ESU8.

[6] The 10 Y U.S. T-Note is a debt-obligation issue by the U.S government with a maturity of 10 years.

Figure 9: A screen shot from a bloomberg terminal for ESM8. The order book events we require are TSUM (Trade Summary).

3.2 Tools

All of the data cleaning and analysis was carried out in R. Some of the plots were produced in Python since the plotting tool was preferred in some cases. The following packages were used:

- R: zoo, latex2exp, stats, base, pracma, Rfast, dplyr, rblpapi (Bloomberg API), sparr, reticulate,
- Python: pandas, numpy, scipy, matplotlib, scorr, priceprop

The scorr library was written by the author of [27] in Python. Some of the functionality in this library was vital for this study, so the package was rewritten in R.

4 Empirical Findings

4.1 Persistence of order signs

One stylised fact identified in a number of papers [28, page 16, section 2.6] [11, pages 11-13, section 2.5] [29, page 8, section 4.1] from intra-day order data is the long-term memory of the trade signs ϵ. Temporal price changes are largely uncorrelated, else this would make markets predictable. The signs of the trade, however, display slow decaying correlations. We define the auto-correlation of the signs as

$$C(l) = \langle \epsilon_{t+l} \epsilon_t \rangle \,, \tag{4.1}$$

where $\langle \cdot \rangle$ denotes the time averaged value and l is the lag. Figure 10 shows the auto-correlation of order signs for AMZN in April 2018. The auto-correlation of the signs decays as a power law $C(l) \sim l^{-\gamma}$, with decay parameter $\gamma = 0.66$ in this case. Continuing from the random trading hypothesis in section 1.2, if this were the case then the auto-correlation of the signs would decay rapidly. The long term memory in order signs indicates that trades are not random. There are two likely explanations for this empirical observation [30, page 1, section 1]. The first explanation is informed traders executing meta-orders to manage their execution costs. The second explanation is 'copy-cat' traders, who interpret the initial trade as a result of a well informed future price prediction.

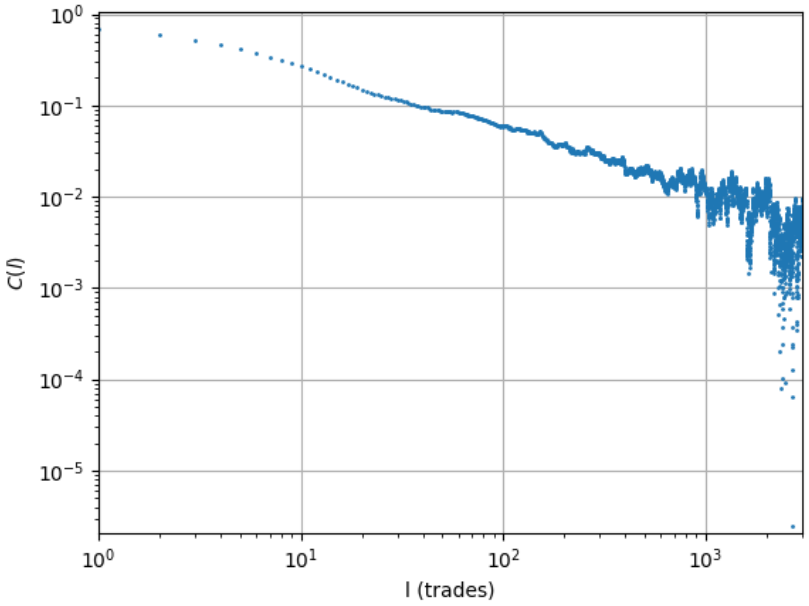


Figure 10: The autocorrelation of signs $C(l)$ for AMZN in April 2018. We see a long-ranged correlation with decay parameter $\gamma = 0.66$ which persists for several thousand trades.

4.2 Price pins

One of the more striking empirical findings identified by Patzelt et. al [31, page 6, section III.C] is that large order sign imbalances lead to price pina and are associated with minuscule returns. This is observed frequently on an intra-day timescale in both stocks and futures and one of these price pins is shown in figure 11. The excerpt was taken from the Microsoft stock price on 02/04/2018 and indicates the 50 trade log-return series as well as the average order sign over this 50 period. We see around trade 600 that the average order sign is $\bar{\epsilon} \approx 1$, indicating a series of sell trades which leads to a large bias in the order book. The corresponding return plot indicates a price pin for approximately the same period.

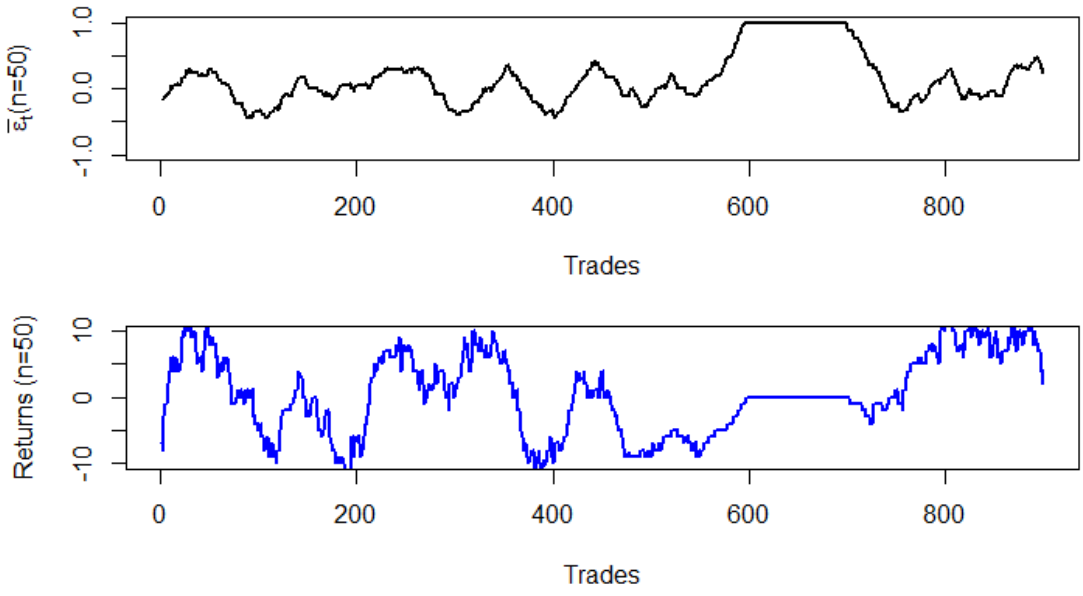


Figure 11: A small window of order flow for Microsoft with 50 day average order sign $\bar{\epsilon}_t$ (top) and the 50-trade returns (bottom). A price pin is observed at around trade 600 following a high order sign bias.

The price pin can be attributed to market makers reconsidering their limit orders and revising them to match the present order bias, as well as informed traders revising their volume of market orders to match the outstanding liquidity. If there is a high order book bias towards the buy side, then a continuous string of sell market orders will likely not completely deplete the available liquidity, hence the price will not change.

4.3 Nonlinear Effects

A detailed investigation into the dynamics of meta-orders is challenging since it requires a proprietary trading database where each trading decision can be mapped to the individual trades. Since

most order book data from major exchanges is anonymous we need another quantity to investigate aggregate price impacts [31, pages 1-2, section I].

Definition 4.1. The *aggregate-volume impact* over N trades $\mathcal{R}_N$ is defined as the average price return conditional on the total volume imbalance $\mathcal{Q}_N$

$$\mathcal{R}_N(\mathcal{Q}) = \left\langle \log m_{t+N} - \log m_t \mid \mathcal{Q} = \sum_{i=0}^{N-1} q_{t+i} \right\rangle, \quad q_i := \epsilon_i V_i, \tag{4.2}$$

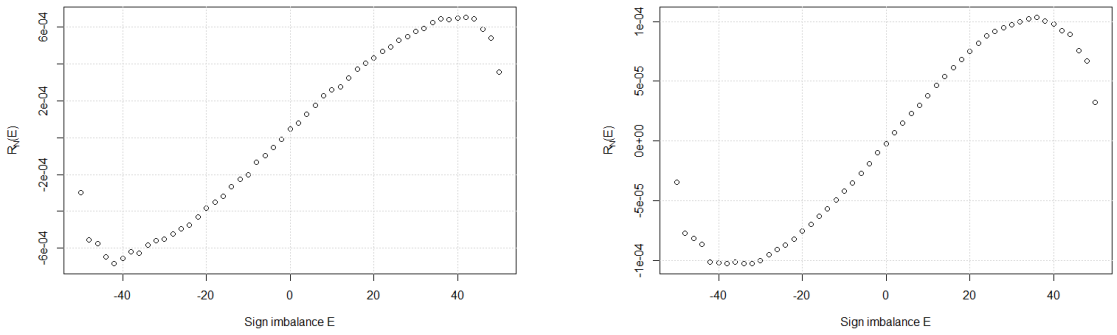
where m_t is the mid-price, q_t is the signed volume of the trade, ϵ_t is the sign of the trade and V_t is the volume of the trade.

Patzelt et al. [31, page 3, section III.A] showed that for $N \gtrsim 10$, the rescaled aggregate impact function is follows a non-linear sigmoidal shape with which is approximately independent of N and the asset under consideration. This is not the only quantity which exhibits nonlinear effects.

Definition 4.2. The *aggregate-sign impact* over N trades $\mathcal{R}_N(\mathcal{E})$ is defined as

$$\mathcal{R}_N(\mathcal{E}) = \left\langle \log m_{t+N} - \log m_t \mid \mathcal{E} = \sum_{i=0}^{N-1} \epsilon_{t+i} \right\rangle, \tag{4.3}$$

Since our study is mainly concerned with order signs rather than volume we will display results for the aggregate sign only. Figure 12 shows the aggregate sign impact for AMZN and the ESM8 with bin sizes of $N = 50$.



(a) AMZN (b) ESM8

Figure 12: The aggregate sign impact $\mathcal{R}_N(\mathcal{E})$ for AMZN and ESM8 Index for $N = 50$. Both display sinusoidal behaviour and the curves are very similar.

We see that both curves follow a sinusoidal curve. These curves are remarkably stable and, when $N \gtrsim 10$, are also independent of N. In fact, Patzelt et al. [31, page 3, section III.A] proposed the scaling ansatz

$$\mathcal{R}_N(\mathcal{E}) \sim R_N \mathscr{F}\left(\frac{\mathcal{E}}{E_N}\right), \tag{4.4}$$

where E_N and R_N both follow power-law scaling

$$E_N \sim E_1 N^{\xi}$$
$$R_N \sim R_1 N^{\phi}.$$

In this case we have $\mathscr{F}(x) = \sin x$. Figure 12 also corroborates the price pins seen in figure 11. When the sign imbalance reaches 100%, the aggregate sign impact reverts to zero, which results in small returns.

4.3.1 Concave volume dependence

A commonly observed feature in market micro-structure when studying market impact is the nonlinear dependence on volume. This is in direct contradiction of the Kyle model which predicts a linear dependence on volume. To see this volume dependence we need a volume conditional response function.

Definition 4.3. The volume conditional response function over a time period T is defined as

$$\mathcal{R}(T, V) = \langle (m_{t+T} - m_t) \cdot \epsilon_t | V_t = V \rangle , \tag{4.5}$$

where V is the volume of the trade and m_t is the mid-price.

Proposition 4.4. *The volume dependence of the price impact is sublinear and follows a power-law [17, section 4.1, page 5]:*

$$\mathcal{R}(T = \Delta t, V) \sim V^{\psi(\Delta t)}, \qquad \psi(\Delta t) \leq 1, \tag{4.6}$$

where Δt is the time scale under consideration, and the exponent ψ varies depending on the time scale.

The power law has exponent ψ which is a function of the time scale Δt. The time scale can range from milliseconds to days and typically takes values of $\psi \approx 0.1 - 0.3$ on a trade by trade time scale but tends to $\psi = 1$ when considering several thousands of trades. Figure 13 shows the volume conditioned response function of AMZN for April 2018 with $T = 50$. We see that the volume conditional response function follows a power law with exponent $\psi = 0.124$.

The small value of the exponent ψ for trade-by-trade time scales implies a week volume dependence. This can be attributed to large market orders which are only submitted if there is a sufficient quantity at the best price level, hence the impact is limited. It has been suggested that price changes are more strongly correlated with the number of trades rather than the volume [32].

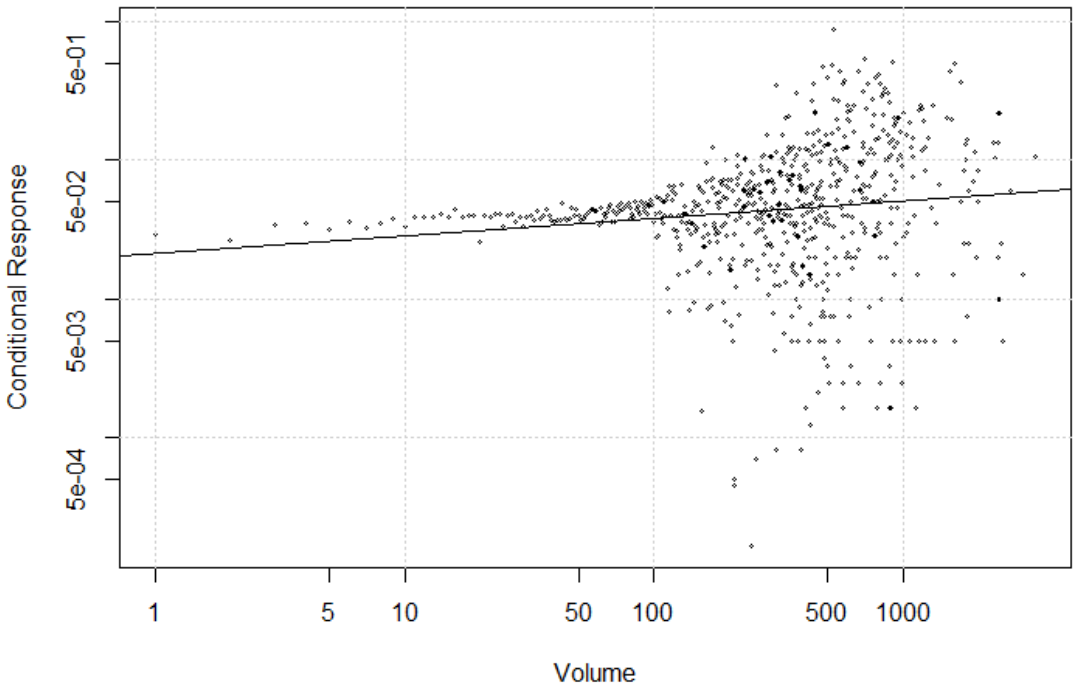


Figure 13: Plot of $\mathcal{R}(T = 50, V)$ against V for AMZN in April 2018. The conditional response function is a power law with exponent $\psi = 0.124$.

4.3.2 The square root law of a Meta-order

An empirical observation identified in numerous studies is the empirical square root law of a meta-order. This law considers the *peak impact* of a meta order. Consider the inventory $I(t)$. A meta order has total volume

$$Q = \int_{t_-}^{t_+} dt M(t), \tag{4.7}$$

where $M(t)$ is the trading rate

$$M(t) = \frac{\partial I(t)}{\partial t}, \tag{4.8}$$

and t_+ and t_- are the starting and ending times respectively.

Definition 4.5. Let X be an arbitrary set and $f : X \to \mathbb{R}$ be a real valued function. The *set-theoretic support* of f $\text{supp}(f)$ is the set of points in X where f is non-zero

$$\text{supp}(f) = \{x \in X | f(x) \neq 0\}.$$

Definition 4.6. The starting and ending times of a meta-order are defined as

$$t_- = \inf \operatorname{supp}(m) \tag{4.9}$$

$$t_+ = \sup \operatorname{supp}(m) \tag{4.10}$$

Definition 4.7. The peak-impact of a meta-order is defined as

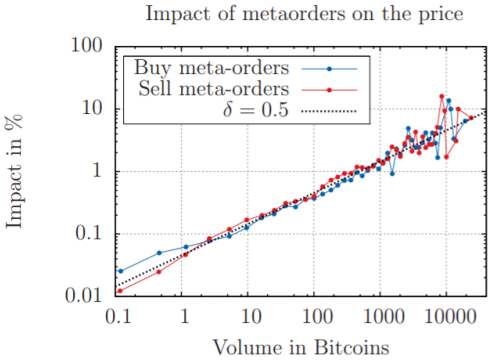
$$\mathcal{I}(T) = \mathbb{E}[p(t_+) - p(t_-)] \tag{4.11}$$

Proposition 4.8. *The peak impact of a meta order follows the relation*

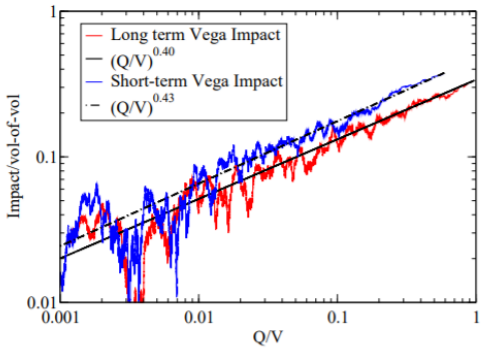
$$\mathcal{I}(T) = Y\sigma\sqrt{\frac{Q}{V}}, \tag{4.12}$$

where $T := t_+ - t_-$ is the length of the meta-order, V is the average daily volume of the security traded, σ is the daily volatility of the security and Y is a homogenisation constant of order 1 [33, slides 21-27] .

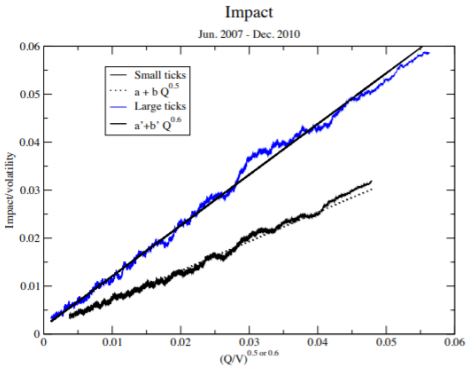
What is remarkable about this empirical law is the near universality it displays. It has been observed in stocks [4], futures, options [34] and even Bitcoin markets [3]. It does not appear to have any dependence on geographical location, time period nor the maturity of the market (S&P 500 vs. Bitcoin) [23, section I, page 1]. This also contradicts the predictions of the Kyle model. Since a propriety trading database is not available for this study, we refer to results from Toth, Lehalle and Donier. Figure 14 shows the square root law for a number of asset classes. The law holds, to a reasonable approximation, for all asset classes. The law predicts no dependence on execution path, an assumption which has been shown to be incorrect. Hence the square root law acts as a good benchmark for market impact models but can only been seen as a first order approximation [1, section 1.5.1, pages 8-9].



(a) Bitcoin



(b) LT and ST options



(c) CFM large tick and small tick stocks

Figure 14: The square root law for different asset classes [3, figure 5, page 9] [4, figure 1, page 2] [5, slide 8]

5 Propagator Models

We now come to the main focus of this study: propagator models. The motivation of these models stems from the autocorrelation of order signs as well as the origin of the random nature of price as discussed in section 1.2. A mathematical framework is required which describes the slowly decaying correlations in the trade signs as well as the (almost) diffusive nature of prices. Bouchaud et al. [11, pages 13-14] proposed a class of so-called 'propagator' models. These models posit that the mid-point price m_t at time t just before the trade can be expressed as a linear superposition of the impact of all past trades plus a noise term:

$$m_t = \sum_{t'<t} [G(t-t')\epsilon_{t'} + \eta_{t'}] + m_{-\infty}, \tag{5.1}$$

where $\epsilon_{t'}$ is the sign of the trade (+1 for a buy order and -1 for a sell order), $\eta_{t'}$ is a noise term which models price changes not induced by trades (e.g. jumps due to news). The function $G(t-t')$ is known as the 'propagator' and describes the time decay of the impact of a single trade. The propagator must decay with time in a precise way to counteract the autocorrelation of the sign of the trades. In the absence of decay, the sign of the the trades would be proportional to returns, leading to strong autocorrelations in time and high predictability which is clearly unphysical. On the other hand, a sudden rapid decay creates a price oscillation over a short period and the resulting long-term volatility would tend to zero [12, page 2, section 2].

As mentioned in section 4.1, the correlation of order signs $C(l) := \langle \epsilon_t \epsilon_{t+l} \rangle$ at large lags l decays as a power law with exponent $\gamma < 1$. This is to account for the effect of meta-orders. To account for this behaviour, $G(t-t')$ must also decay as a power law.

The above framework is clearly an over-simplification and is missing three key features [30, page 2]:

1. Equation (5.1) assumes that G only depends on $t-t'$ rather than t and t' individually. This implies that all market orders have identical impact and neglects any fluctuations in the impact which may arise.

2. There are other order book events which can change the price, such as limit orders within the bid ask spread and cancellations. These effects are only considered implicitly, since limit orders oppose market orders. These need to be accounted for explicitly.

3. This is a linear model and neglects any nonlinear effect, like those outlined in section 4.

Linear propagator models aim to address issues 1 and 2 by generalising equation (5.1) such that price-changing and non price-changing market orders are considered separately. We will consider two types of linear propagator models: the Transient Impact Model (TIM) and the History Dependent Impact Model (HDIM).

5.1 The one-event propagator model

5.1.1 The Transient Impact Model (TIM1)

Before we generalise the above model into price-changing and non price-changing events we consider the one-event transient impact model (TIM1). The '1' refers to the single propagator function. If we consider the price process as above then we can derive a formula for the TIM1 return process $r_t := m_{t+1} - m_t$ by taking the price difference:

$$r_t := m_{t+1} - m_t = \sum_{t'<t+1} [G(t-t'+1)\epsilon_{t'} + \eta_{t'}] - \sum_{t'<t} [G(t-t')\epsilon_{t'} + \eta_{t'}] \tag{5.2}$$

$$= \sum_{t'<t} [G(t-t'+1) - G(t-t')]\epsilon_{t'} + \eta_{t'}] + G(1)\epsilon_t \tag{5.3}$$

$$= \sum_{j>0} [G(j+1) - G(j)]\epsilon_{t-j}] + G(1)\epsilon_t \tag{5.4}$$

$$= \sum_{j\geq 0} \mathcal{G}(j)\epsilon_{t-j}, \tag{5.5}$$

where we have made the substitution $j = t - t'$ and we have neglected the noise term. We have also defined the differential kernel $\mathcal{G} := G(j+1) - G(j)$. Note $G(j \leq 0) \equiv 0$ by definition. The TIM1 is similar to equation (5.1), except the kernel G is now written as a differential kernel $\mathcal{G}$. The returns r_t are written as the convolution between the order signs ϵ_t and the kernel $\mathcal{G}$:

5.1.2 The History Dependent Impact Model (HDIM1)

A different interpretation of the TIM1 model is the History Dependent impact model HDIM1 [30, pages 4-5, section 2.3] which states that the difference of the sign ϵ_t from its expected level $\hat{\epsilon_t}$ has a permanent linear impact on the returns:

$$r_t = G(1)(\epsilon - \hat{\epsilon_t}) \tag{5.6}$$

If we consider $\hat{\epsilon}_t$ as the best predictor of ϵ_t then by construction the price process is a martingale. When $\hat{\epsilon_t}$ is linear then the the TIM1 and the HDIM1 are equivalent:

$$\hat{\epsilon_t} = -\sum_{j\geq 0} \frac{\mathcal{G}(j)}{G(1)}\epsilon_{t-j} \tag{5.7}$$

This best predictor is linear when the string of signs is generated by a *Discrete Autoregressive* (DAR) process. We consider the sign at time t as having been influenced by a sign at time $t-l$, where the distance is characterised by a discrete probability distribution λ_l such that

$$\sum_{l=1}^{\infty} \lambda_l = 1$$

When the model is truncated at $l = p$ the model becomes DAR(p). In this case we have:

$$\epsilon(t) = \begin{cases} \epsilon_{t-l} & \text{with probability } \rho \\ -\epsilon_{t-l} & \text{with probability } 1-\rho. \end{cases}$$

Lemma 5.1. *If the process is in a stationary state, the signs are probable with equal probability and the sign autocorrelation function $C(l)$ obeys the Yule-Walker equation*

$$C(l) = (2\rho - 1)\sum_{n=1}^{\infty} \lambda_n C(l-n) \tag{5.8}$$

Proof. Since ϵ_t is an auto-regressive process we have

$$\epsilon_t = \sum_{n=1}^{\infty} \lambda_n \epsilon_{t-n}.$$

If we multiply by ϵ_{t-l} and take the time averaged values we get

$$\begin{aligned}\langle \epsilon_t \epsilon_{t-l} \rangle &= \left\langle \sum_{n=1}^{\infty} \lambda_n \epsilon_{t-n} \epsilon_{t-l} \right\rangle \\ &= (2\rho - 1)\sum_{n=1}^{\infty} \lambda_n \left\langle \epsilon_{t-n} \epsilon_{t-l} \right\rangle \\ &= (2\rho - 1)\sum_{n=1}^{\infty} \lambda_n \left\langle \epsilon_t \epsilon_{t-(l-n)} \right\rangle,\end{aligned}$$

where in the final line, we have used stationarity of the ϵ_t. □

As we have shown empirically the correlation function $C(l)$ decays as power law $l^{-\gamma}$ with $\gamma < 1$. This then leads to $\lambda_l \sim l^{(\gamma-3)/2}$ and$\rho \to 1^-$. By construction we have

$$\hat{\epsilon}_t = (2\rho - 1)\sum_{l=1}^{\infty} \lambda_l \epsilon_{t-l},$$

so we can consider the HDIM1 as a TIM1 where

$$\mathcal{G}(l) = -(2\rho - 1)G(1)\lambda_l. \tag{5.9}$$

Since we can consider the two as equivalent, will not investigate the HDIM1 for the remainder of this study.

5.2 The generalised propagator model

Larger market orders will have a different impact to smaller market orders, since large market orders are likely to use up entire levels in the order book. To account for this phenomena, we introduce different event types π_t and extend our definition of the TIM1 and the HDIM1 models. We consider price-changing and non price-changing events:

$$\pi_t = \begin{cases} \text{n} & \text{if } r_t = 0 \\ \text{c} & \text{if } r_t \neq 0. \end{cases}$$

We now generalise equation(5.1) to include price changing and non-price changing events $\pi_t = \{\text{n, c}\}$ so now we incur an extra summation

$$m_t = \sum_{t'<t} \left[\sum_{\pi'} G_{\pi'}(t-t')\epsilon_{t'} \mathbf{1}_{\{\pi_{t'}=\pi'\}} + \eta_{t'} \right] + m_{-\infty}. \tag{5.10}$$

5.2.1 The Transient Impact Model (TIM2)

If we take equation (5.10) and calculate the returns $r_t := m_{t+1} - m_t$ we get

$$r_t = \sum_{t'<t+1} \left[\sum_{\pi'} G_{\pi'}(t - t' + 1)\epsilon_{t'} \mathbf{1}_{\{\pi_{t'}=\pi'\}} + \eta_{t'} \right] - \sum_{t'<t} \left[\sum_{\pi'} G_{\pi'}(t - t')\epsilon_{t'} \mathbf{1}_{\{\pi_{t'}=\pi'\}} + \eta_{t'} \right] \quad (5.11)$$

$$= \sum_{\pi} G(1)\mathbf{1}_{\{\pi_t=\pi\}}\epsilon_t + \sum_{j>0} \sum_{\pi'} \mathcal{G}_{\pi'}(j)\mathbf{1}_{\{\pi_{t-j}=\pi'\}}\epsilon_{t-j} \quad (5.12)$$

$$= \sum_{\pi'} \sum_{j\geq 0} \mathcal{G}_{\pi'}(j)\mathbf{1}_{\{\pi_{t-j}=\pi'\}}\epsilon_{t-j}, \quad (5.13)$$

where again we have defined the differential kernel $\mathcal{G}_{\pi'} := G_{\pi'}(j+1) - G_{\pi'}(j)$. This is the TIM2 model, where the '2' represents the two different propagator functions. Note that past events, through the decay of the kernel, aways influence the the price in the TIM2 model. This leads to the unphysical prediction that non-zero returns are possible when $\pi_t = n$.

5.2.2 The History Dependent Impact Model (HDIM2)

A modification of the TIM2 model adds a dependency on present events as well as past events. This is the HDIM2 model:

$$r_t = \sum_{\pi''} \mathbf{1}_{\{\pi_t=\pi''\}} \sum_{\pi'} \sum_{j\geq 0} \kappa_{\pi',\pi''}(j)\mathbf{1}_{\{\pi_{t-j}=\pi'\}}\epsilon_{t-j}, \quad (5.14)$$

where $\kappa_{\pi',\pi}(l)$ is know as the 'influence kernel' which is dependent on past and present events. The HDIM2 utilises four kernels through and 'influence matrix' $\kappa_{\pi',\pi''}$ which is dependent on the label of the most recent event. The HDIM2 and the TIM2 become equivalent when the kernel κ loses its dependency on the present event type π. In this case we have:

$$\kappa_{\pi',\pi''}(j) = \mathcal{G}_{\pi'}(j). \quad (5.15)$$

5.3 The Constant Impact Model (CIM)

The above models all assume some time dependence for the propagators G and κ. To verify this assumption it is worth reversing the logic and test how well time-independent models perform. We consider the *Constant Impact Model*, which has no such time-dependence. Consider the exact formula for the mid-price m_t

$$m_{t+l} = m_t + \sum_{t\leq t'<t+l} \epsilon_{t'}\Delta_{\pi'_t,\epsilon_{t'},t'}, \quad (5.16)$$

where $\Delta_{\pi'_t,\epsilon_{t'},t'}$ is the price change at time t' is event π happens. Since only price changing events affect the price this becomes

$$m_{t+l} = m_t + \sum_{t\leq t'<t+l} \epsilon_{t'}\Delta_{\epsilon_{t'},t'}\mathbf{1}_{\{\pi'_t=c\}}.$$

We remove the time dependency by averaging the Δ s over time; this is the *average realised gap*

$$\Delta_c^R = \left\langle \Delta_{\epsilon_{t'}} | \pi_t = c \right\rangle . \tag{5.17}$$

If we rewrite this in terms of returns we arrive at the CIM2 model

$$r_t^{\text{CIM2}} := \Delta_c^R \mathbf{1}_{\{\pi(t)=c\}} \epsilon_t \tag{5.18}$$

$$\Delta_c^R = \left\langle |r_t| \mid \pi_t = c \right\rangle . \tag{5.19}$$

5.4 Measuring Market Impact

Before moving any further we should consider how we should measure market impact. There are a number of ways we can measure the impact. One could consider the correlation $\rho(T)$ between the price change from 0 to T and the total signed volume within this time frame:

$$\rho(l) = \frac{\left\langle (m_{t+l} - m_t) \cdot \sum_{n=0}^{N-1} \epsilon_n V_n \right\rangle}{\sqrt{\langle (m_{t+l} - m_t) \rangle^2 \left\langle (\sum_{n=0}^{N-1} \epsilon_n V_n)^2 \right\rangle}}, \tag{5.20}$$

where $\langle \cdot \rangle$ denotes the time averaged value, ϵ_t is the sign of the trade, V is the volume of the trade and m_t is the mid-price. This study is primarily concerned with using *response function*, which is a measure of market impact used to calibrate the propagator models.

Definition 5.2. The empirical response function is defined as the covariance of the mid-price change m_t between two points in time separated by l, and the sign of the trade, or precisely:

$$\mathcal{R}(l) := \langle (m_{t+l} - m_t) \cdot \epsilon_t \rangle \tag{5.21}$$

Definition 5.3. The empirical conditioned differential response function $\mathcal{R}_{\pi,\pi'}(l)$ is defined as the average price behaviour after an order book event, or average impact function:

$$\mathcal{R}_{\pi,\pi'}(l) := \langle r(t) \cdot \epsilon(t-l) | \pi(t) = \pi', \pi(t-l) = \pi \rangle \tag{5.22}$$

$$= \left\langle r(t) \cdot \epsilon(t-l) \mathbf{1}_{\{\pi(t-l)=\pi\}} \mathbf{1}_{\{\pi(t)=\pi'\}} \right\rangle . \tag{5.23}$$

Note we can sum over events π' and π respectively to get the integrated response functions:

$$\mathcal{R}_\pi(l) := \sum_{\pi'} \mathcal{R}_{\pi,\pi'}(l), \qquad \mathcal{R}(l) := \sum_{\pi} \mathcal{R}_\pi(l). \tag{5.24}$$

$\mathcal{R}(l)$ is the expected one-trade return r at time $t+l$ given a trade at time t.

5.5 Model Calibration

To calibrate the models, we measure the empirical response function and use it to form a system of linear equations, which we solve for the kernel [30, page 3, section 2.1] [27, pages 9-10, appendix A].

Firstly we consider the HDIM2, which is the most challenging to calibrate. The TIM1 and TIM2 calibration methods are specific instances of this calibration process. First we introduce an error term defined as:

$$\nu_t := r_t - r_t^{\text{HDIM2}} \tag{5.25}$$

If we combine this with the predicted returns from the HDIM2 [equation (5.14)], we see

$$\begin{aligned}
\mathcal{S}_\pi(l) &= \left\langle (r_t^{\text{HDIM2}} + \nu_t)\mathbf{1}_{\{\pi_{t-l}=\pi\}}\epsilon_{t-l} \right\rangle \\
&\approx \left\langle r_t^{\text{HDIM2}}\mathbf{1}_{\{\pi_{t-l}=\pi\}}\epsilon_{t-l} \right\rangle \\
&= \sum_{\pi'',\pi',j\geq 0} \left[\kappa_{\pi',\pi''}(j) \left\langle \mathbf{1}_{\{\pi_t=\pi''\}}\mathbf{1}_{\{\pi_{t-j}=\pi'\}}\epsilon_{t-j}\mathbf{1}_{\{\pi_{t-l}=\pi\}}\epsilon_{t-l} \right\rangle \right] \\
&= \sum_{\pi'',\pi',j\geq 0} \left[\kappa_{\pi',\pi''}(j)\mathcal{C}_{\pi\pi'\pi''}(l,j) \right],
\end{aligned}$$

where $\mathcal{C}_{\pi\pi'\pi''}(l,j)$ is the triple triple cross-correlation. A well calibrated, consistent model will have no correlations between the model error ν_t and its input, hence this error term will be negligible for well formed models.

If we set

$$\begin{aligned}
f &= \mathbf{1}_{\{\pi_t:=\pi''\}} \\
g &= \mathbf{1}_{\{\pi_{t-l}:=\pi\}}\epsilon_{t-l} \\
h &= \mathbf{1}_{\{\pi_{t-j}:=\pi'\}}\epsilon_{t-j},
\end{aligned}$$

then we can define $\mathcal{C}_{\pi\pi'\pi''}(l,j) := C_{fgh}(l,j)$. For convenience, we will write it in the form $\mathbf{S} = \mathbf{Ck}$ or

$$\begin{bmatrix} \mathbf{S}^{nc} \\ \mathbf{S}^{cc} \end{bmatrix} = \begin{bmatrix} \mathbf{C}^{nnc} & \mathbf{C}^{cnc} \\ \mathbf{C}^{ncc} & \mathbf{C}^{ccc} \end{bmatrix} \begin{bmatrix} \mathbf{k}^{nc} \\ \mathbf{k}^{cc} \end{bmatrix},$$

where

$$\mathbf{S}_l^{\pi,\pi'} = \mathcal{S}_{\pi,\pi'}(l), \quad \mathbf{C}_{lj}^{\pi\pi'\pi''} = \mathcal{C}_{\pi\pi'\pi''}(l,j), \quad \mathbf{k}_j^{\pi\pi'} = \kappa_{\pi\pi'}(j).$$

The TIM2 is easier to calibrate as is expressed in the form:

$$\begin{bmatrix} \mathbf{S}^{nc} \\ \mathbf{S}^{cc} \end{bmatrix} = \begin{bmatrix} \mathbf{C}^{nn} & \mathbf{C}^{cn} \\ \mathbf{C}^{nc} & \mathbf{C}^{cc} \end{bmatrix} \begin{bmatrix} \mathbf{g}^{nc} \\ \mathbf{g}^{cc} \end{bmatrix},$$

where

$$\mathbf{C}_{lj}^{\pi\pi'} = \mathcal{C}_{\pi\pi'}(l,j), \quad \mathbf{g}_j^{\pi'} = g_{\pi'}(j).$$

The TIM1 is a further simplification of these linear equations and can be expressed as $\mathbf{S} = \mathbf{Cg}$, where $\mathbf{S}_l = S(l)$, $\mathbf{g}_j = g(j)$ and $\mathbf{C}_{lj} = \langle r(t)\epsilon(t+l-j)\rangle$

Calculations were done per day then averaged over the month. Each day has equal weighting however for more active trading days the contribution of each individual event is lower.

5.5.1 The Triple Cross-Correlation

Calculation of the triple cross-correlation requires a novel technique proposed by Patzelt et. al [27, page 10, appendix B].

Definition 5.4. For a function f in Schwartz Space $\mathcal{S}(\mathbb{R})$, its Fourier Transform $\mathscr{F}[f](\xi)$ is defined as

$$\mathscr{F}[f](\xi) := \int_{\mathbb{R}^n} \mathrm{e}^{i\xi\cdot x} f(x)dx, \qquad \text{for any } \xi \in \mathbb{R}^n \tag{5.26}$$

Definition 5.5. For a Schwartz function $f \in \mathbb{R}^n$, its inverse Fourier transform is defined as

$$\mathscr{F}^{-1}[f](\xi) := (2\pi)^{-n}\mathscr{F}[f](-\xi) \tag{5.27}$$

Definition 5.6. Let $f, g : \mathbb{R} \to \mathbb{R}$ then the convolution of f and g, written $f * g$ is defined as

$$(f * g)(t) := \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau \tag{5.28}$$

Theorem 5.7. *let f, g be in $L^1(\mathbb{R}^n)$. Then the convolution theorem states that:*

$$(f * g) = \mathscr{F}^{-1}\left\{\mathscr{F}\{f\}\mathscr{F}\{g\}\right\} \tag{5.29}$$

Proof. See Appendix □

Definition 5.8. Let $F := \mathscr{F}[f]$, $G := \mathscr{F}[g]$ and $H := \mathscr{H}[h]$. We define the cross-spectrum (or periodogram) between the functions f, g and h as

$$B_{fgh}(x, y) = \bar{F}(x + y)G(x)H(y) \tag{5.30}$$

We can calculate the cross-correlation $C_{fg}(l)$ between two functions using the convolution theorem

$$C_{fg}(l) := \int_{-\infty}^{+\infty} \bar{f}(t)g(t + l)dt = \mathscr{F}^{-1}\left[\bar{F}G\right](l), \tag{5.31}$$

where l is the lag, and $\bar{f}$ is the complex conjugate of f.

Proposition 5.9. *The three-point cross-correlation $C_{fgh}(l, j)$ between three time domain functions f, g and h is given by the inverse Fourier transform of the cross-bispectrum of the corresponding functions, namely*

$$C_{fgh}(l, j) = \mathscr{F}^{-1}_{\nu'\nu''}\left[B_{fgh}(\nu', \nu'')\right](l, j). \tag{5.32}$$

Proof. We can extend equation (5.31) to a three-point cross-correlation between the functions

$f(t), g(t+l)$ and $h(t+j)$.

$$\begin{aligned}
C_{fgh}(l,j) &:= \int_{-\infty}^{+\infty} \bar{f}(t)g(t+l)h(t+j)dt \\
&= \int_{-\infty}^{+\infty} dt \left(\int_{-\infty}^{+\infty} d\nu \bar{F}(\nu) e^{2\pi i \nu t} \int_{-\infty}^{+\infty} d\nu' G(\nu') e^{-2\pi i \nu'(t+l)} \int_{-\infty}^{+\infty} d\nu'' H(\nu'') e^{2\pi i \nu'(t+j)} \right) \\
&= \iiint_{-\infty}^{+\infty} d\nu d\nu' d\nu'' \left(\bar{F}(\nu) G(\nu') H(\nu'') \int_{-\infty}^{+\infty} dt e^{2\pi i t(\nu'+\nu''-\nu)} e^{2\pi i (l\nu' + j\nu'')} \right) \\
&= \iiint_{-\infty}^{+\infty} d\nu d\nu' d\nu'' \left(\bar{F}(\nu) G(\nu') H(\nu'') \delta(\nu'+\nu''-\nu) e^{2\pi i (l\nu' + j\nu'')} \right) \\
&= \iint_{-\infty}^{+\infty} d\nu' d\nu'' \left(\bar{F}(\nu'+\nu'') G(\nu') H(\nu'') e^{2\pi i (l\nu' + j\nu'')} \right) \\
&= \mathscr{F}^{-1}_{\nu'\nu''} \left[B_{fgh}(\nu', \nu'') \right](l,j)
\end{aligned}$$

where we have used the Dirac delta in the third line. □

5.5.2 Processing finite discrete-time signals

The autocorrelation of signs $C(l)$ cannot be calculated for an infinite number of lags in practise, hence we estimate for the truncated series $l = -T+1, \ldots, T-1$ [27, appendix B.2, page 11]. Hence for two process correlations we calculate

$$\hat{C}_{fg}(l) := \frac{1}{T-|l|} \sum_{t=\sup\{0,-l\}}^{T-1-\sup\{0,l\}} \bar{f}(t)g(t+l) \tag{5.33}$$

$$= \mathrm{iFFT}_{\nu}[F(\nu)G(\nu)](l), \tag{5.34}$$

where $\hat{C} = \langle C \rangle$ if f and g are jointly stationary, F and G are the zero-padded, Fourier transformed signals of f and g. The normalisation has a dependence on l since the number of summands decreases with l. Signals are padded with T zeros since the Fourier transformed signals contain frequencies for $T/2$ positive and negatice frequencies respectively. For the triple cross correlations we have $B_{fgh}(\nu, \nu')$ and $C_{fgh}(l,j)$ which are $T \times T$ matrices. We will apply a variation of Welch's method to reduce the number of calculations.

Welch's method [6] is used to estimate power spectral densities from their periodogram. We will use a variation to determine the periodogram. Before discussing the variation, we will introduce Welch's method. Let $X(j), j = 0, 1, \ldots, N-1$ be a sample taken from a second order stationary stochastic sequence with first zero first moment. The sample has spectral density $B(f)$ with $|f| \leq 1/2$. First we take overlapping segments of length L with starting points D steps apart, as shown in figure 15.

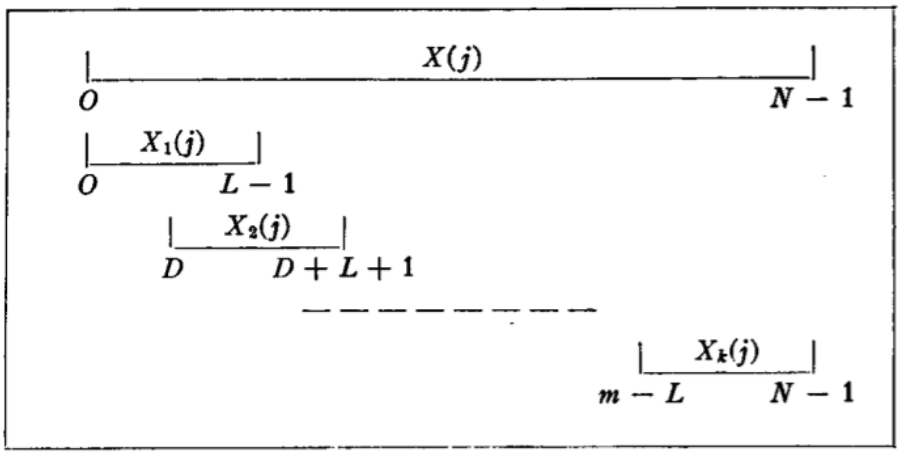


Figure 15: The segmentation of the signal for Welch's method. Source: [6, page 18]

The first segment is

$$X_1(j) = X(j), \qquad j = 0, 1, \ldots, L-1.$$

The next segments are

$$X_2(j) = X(j+D), \qquad j = 0, 1, \ldots, L-1,$$
$$X_k(j) = X(j+(K-1)D), \qquad j = 0, 1, \ldots, L-1.$$

The spectral density is estimated by calculating a modified periodogram, i.e. we take a window $W(j), j = 0, 1, \ldots L-1$ and construct sequences $X_1(j)W(j), \ldots X_k(j)W(j)$, of which we take the fast-Fourier transform:

$$A_k(n) = \frac{1}{L}\sum_{j=0}^{L-1} X_k(j)W(j)e^{-2\pi \mathrm{i} jn/L}. \tag{5.35}$$

The K modified periodograms a given by

$$I_k(f_n) = \frac{L}{U}|A_k(n)|^2, \qquad k = 1, 2, \ldots, K, \tag{5.36}$$

where $f_n = n/L$, $n = 0, 1, \ldots, L/2$ and

$$U = \frac{1}{L}\sum_{j=0}^{L-1} W^2(j). \tag{5.37}$$

The spectral density is then estimated using the average of the periodogams, or more precisely:

$$\hat{P}(f_n) = \frac{1}{K}\sum_{k=1}^{K} I_k(f_n). \tag{5.38}$$

The variation used is similar except equation (5.35) is not used. Instead, for each time segment we calculate

$$C_{fgh}(l,j) = \frac{1}{T - \sup(|l|,|j|}\mathscr{F}^{-1}_{\nu'\nu''}\left[B_{fgh}(\nu',\nu'')\right](l,j), \tag{5.39}$$

where B is calculated using zero padded fast-Fourier transforms to calculate F, G and H.

5.6 Link to Hasbrouck's VAR Model

It is worth considering how the propagator framework relates to the econometric perspective of Hasbrouck's VAR model outlined in section 2.3.2, since it has become a benchmark for microstructure studies [29, section2.4, pages 5-6]. Equation (5.1) can be seen as a special case of equation (2.3) provided:

1. The signed volume becomes the signs: $x_t \to \epsilon_t$
2. $B_{rr}(j)$ are zero
3. $G(l) = \sum_{0 \leq j' < j} B_{xr}(j')$, since equation (5.1) models price rather than returns.
4. The dynamical model for ϵ_t is non-specified.

Essentially, the main difference between these two types of model lies in the interpretation. For example, there is no interpretation for the B_{rr} coefficients since past returns cannot influence the present price alone.

5.7 Link to Latent Order Books

It is worth mentioning the link between propagator models and the model outlined in section 2.3.3. Consider the price impact equation (2.9). If the impact is small, i.e. for all t, s $|y_s - y_t| << D(t-s)$ then (2.9) becomes

$$y_t = \int_0^t \frac{dsm_s}{\sqrt{4\pi D(t-s)}}, \tag{5.40}$$

which is just a linear propagator model with square-root impact.

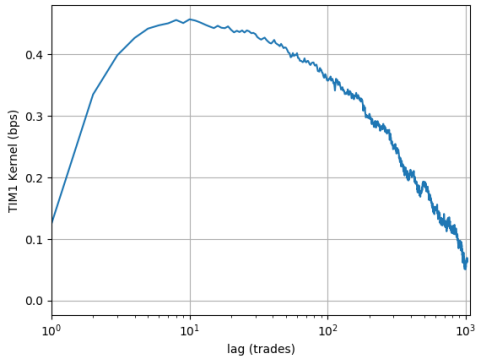
6 Results

6.1 Kernels

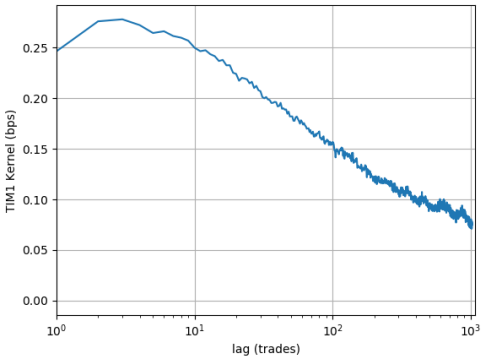
Here we present the calibrated kernels for the propagator models. Some terminology is interchanged: an order book event is taken to be a market order, i.e. a trade. We also present two cash equities (common stocks) for comparison: EBAY as representative of large tick stocks and AMZN as a representative of small tick stocks. Results were similar for all small tick stocks and large tick stocks respectively (see figure 32 in appendix for two more stocks). We compare the cash equities to an S&P 500 E-mini futures contract and 10-Year U.S. Treasury Note futures. Note, the maximum lag used for the TIM1 and TIM2 for the equities are up to 1024. Maximum lags for the other assets were capped at 256 due to data limitations.

6.1.1 TIM1

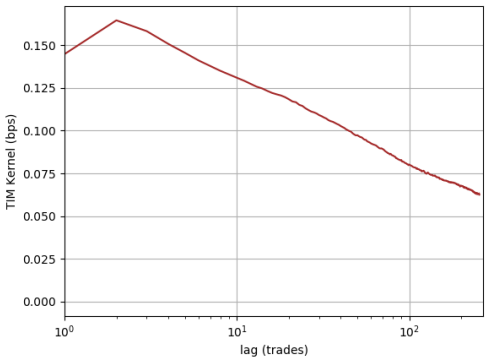
The TIM1 kernels were calibrated to their respective response functions and the results are shown in figure 16. The top two panels indicate kernels for cash equities and the bottom two panels are the kernels for futures.



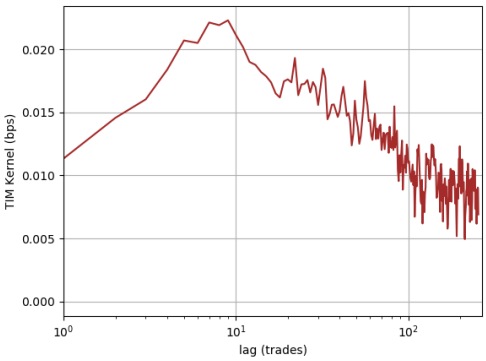
(a) EBAY (large tick)



(b) AMZN (small tick)



(c) ESM8



(d) TYM8

Figure 16: The TIM1 kernels. All kernels increase before following power law decay.

We see for all assets the kernel increases for the first few lags, resulting in a positive differential kernel $\mathcal{G}(l)$ at small lags, an effect which is much more prominent in large tick stocks and TYM8. This is not as expected since empirical observations suggest G should be a decreasing function, thus $\mathcal{G}(l > 0)$ should be a negative function. The physical interpretation of this is a market order will have a lower market impact provided it follows a sequence of trades with the same sign rather than the opposite sign. Lillo and Farmer referred to this as the 'asymmetric liquidity' mechanism [30, section 2, page 3] whereby the market impact of a market order is inversely proportional to the probability of occurrence. Suggested reasons for this include market makers placing their limit orders such as to oppose a trend in market orders, as well as informed traders adjusting their

request for liquidity to the available volume, thus reducing their market impact.

Knowing this, the increasing integrated kernel G does not behave as expected for small lags. Since $G(1) > 0$ market impact will be increased by a sequence of orders of the same sign as the initial market order, which will violate market efficiency over shorter time scales [30, section 2, page 3]. Following a positive price change, the probability that the next order will have the opposite sign is increased, which promotes price efficiency. This is a feature which the TIM1 fails to capture. This effect clearly increases with tick size.

For larger lags, the integrated kernels G follow a slow monotone decay, larger so than the noise observed at longer lags. The decay of the EBAY kernel exceeds that of the AMZN kernel, a feature which is common in large tick stocks.

The ESM8 TIM1 kernel appears very similar in shape to the AMZN kernel. The ESM8 kernel is much less noisy and follows an almost perfect power law decay. This reflects the increase in the number of events per trading day, as the ESM8 is very frequently traded. The TYM8 appears more noisy but the vertical axis is an order of magnitude smaller than the ESM8 axis hence the noise appears worse than it is. The TYM8 has a very slow decaying kernel.

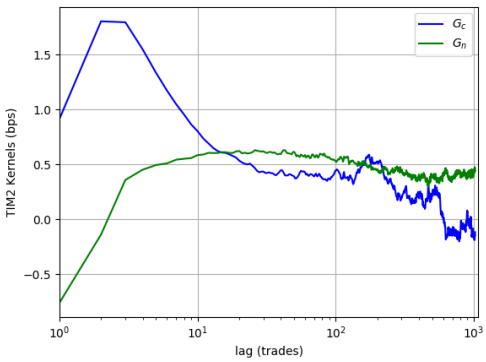
6.1.2 TIM2

The TIM2 kernels were calibrated using conditional response functions and are shown in figure 17. It is immediately obvious that there are some calibration difficulties, particularly for large tick stocks, as shown in the top left panel. The price changing event kernel G_c follows a noisy power law decay for $L \gtrsim 10$. The kernel is too noisy to observe any monotone decay and the noise exceeds the decay rate at some points. The non-price changing kernel G_n is increasing for small lags and plateaus for larger lags. Small tick stocks exhibited similar behaviour, except the non-price changing kernel G_n increases before decreasing as a power law, which means that non price changing events has less of an effect on the price at larger lags. An additional discrepancy is the relative positions of the kernels. The large tick stock kernels even interweave. Notwithstanding, the distance of the price changing kernel G_c above the non-price changing kernel G_n is larger for small tick stocks, implying that price changing events have, relatively speaking, more impact on the price than non-price changing events for small tick stocks.

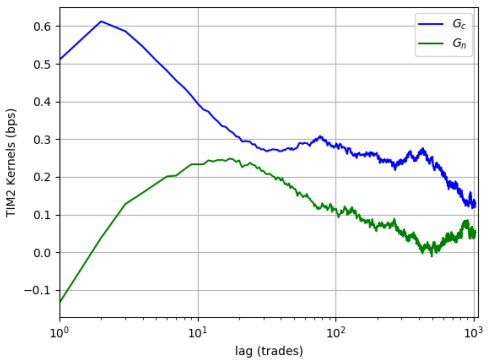
The ESM8 has a different shaped G_c kernel to the cash equities. Initially it rises and then decays, as occurs in the other G_c kernels, except the decay rate is much faster and there is a turning point at around $l = 10$, after which the kernel increases and then plateaus. This implies that price changing events have more of a permanent impact rather than a temporary impact. The non-price changing kernel behaves similarly to AMZN's: increasing initially before decaying to

zero. The TYM8 has a non-price changing kernel approximately constant zero, meaning most of the impact comes from the price changing events. This kernel G_c also increases before decaying as a power-law.

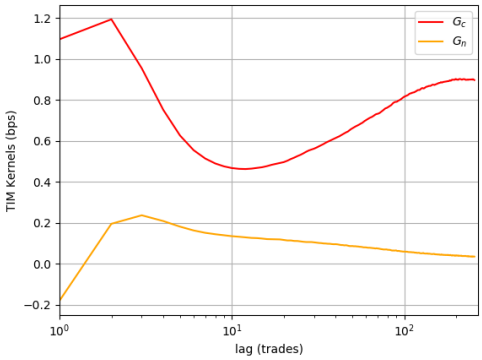
There is an additional inconsistency regarding $G_n(1)$ which is present for all asset types considered. By construction, a non-price changing event must have $G_n(l = 1) < 0$. We see here that $G_n(l = 1) \neq 0$, hence the outputs of the model disrespect the labels of $\{n, c\}$. This is an inherent inconsistency within the TIM2 model, the effects of which we will see later. The fact that $G_n(l = 1)$ is negative means that the market impact of a trade will be larger if it is immediately preceded by a non-price changing trade on the opposite side of the book, a feature which does not have any logical explanation.



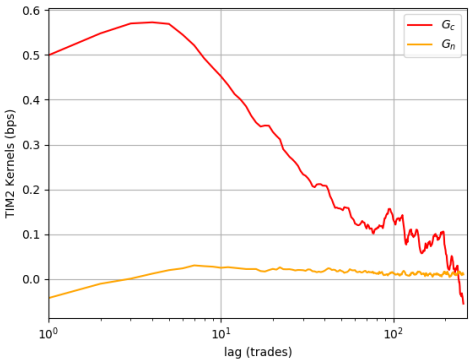
(a) EBAY (large tick)



(b) AMZN (small tick)



(c) ESM8



(d) TYM8

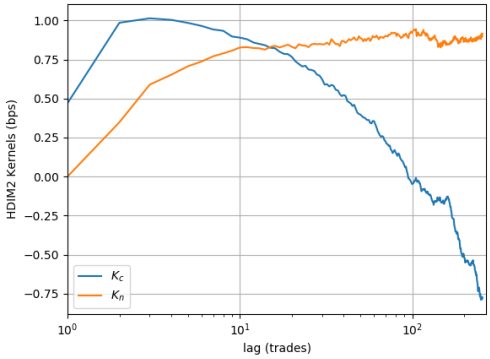
Figure 17: The TIM2 kernels.

6.1.3 HDIM2

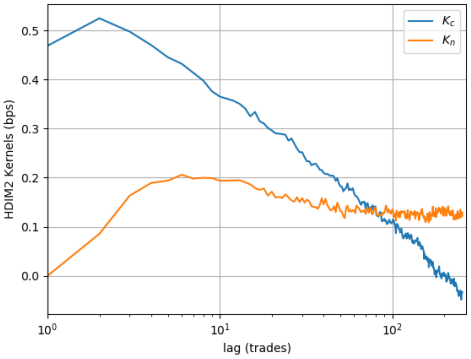
The HDIM2 kernels were calibrated using conditional response functions and are shown in figure 18. Note, these are not the $\kappa_{\pi,\pi'}$ influence kernels. Instead, we plot

$$K_\pi(l) \equiv K_{\pi c}(l) := \sum_{l'=0}^{l} \kappa_{\pi c}(l'), \tag{6.1}$$

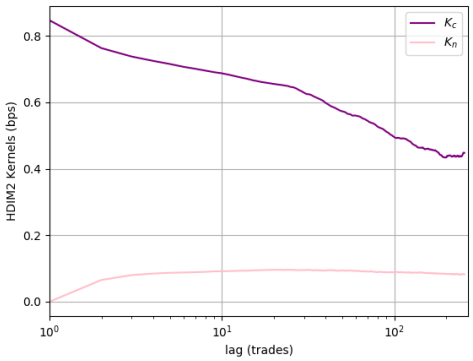
so the kernels now represent the influence of past order signs up until a price changing trade c at lag $l = 0$. The kernels are all only shown up to lag $l = 256$ since the data had to be grouped by date, else the kernels diverged to very large values; the cause of this was undetermined. Since the calculations were done in groups, these groups were too small to calibrate for lags up to $l = 1024$.



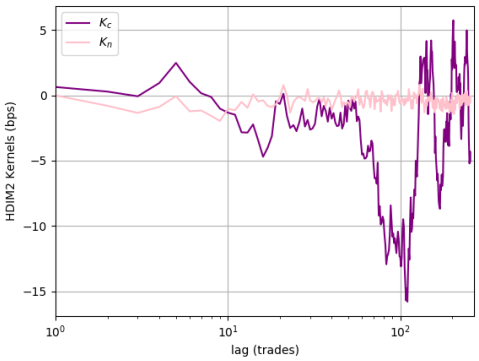
(a) EBAY (large tick)



(b) AMZN (small tick)



(c) ESM8



(d) TYM8

Figure 18: The HDIM kernels

The inconsistency encountered in the TIM2 kernels is no longer present because $K_\pi(l = 1) = 0$ by definition. For both stocks, the decay rate of K_c is particularly high, so much so that it drops below zero for large enough lags; we will denote this lag l^*. For $l < l^*$, the price changing kernel K_c is positive, which means a sequence of price changing market orders of the same sign as the

final price changing event c will have a greater market impact that a sequence of price changing signs of the opposite sign. However, for lags $l > l^*$, the converse is true since K_c is now negative. This is evidence of the asymmetric liquidity discussed above. The optimal trading strategy would incorporate price changing orders of the same sign until l^* trades before the final trade c, after which price changing events of the same sign will have lower impact.

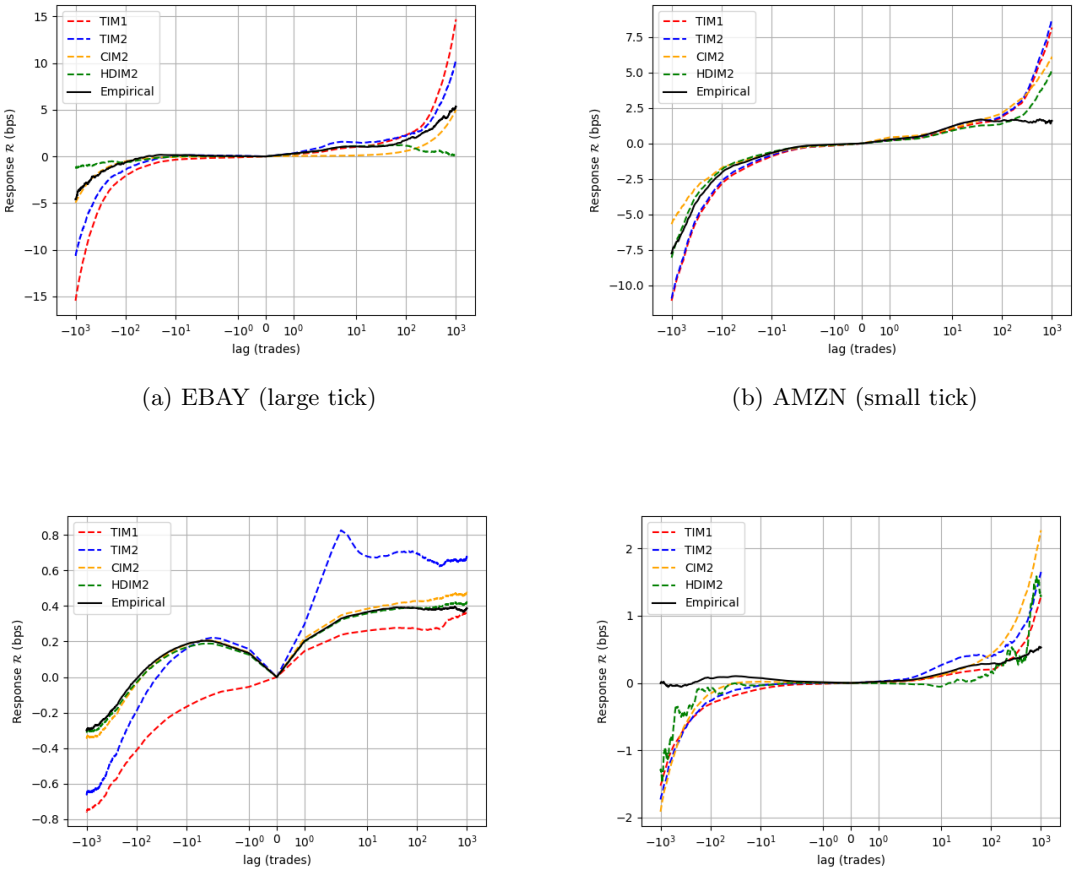
The main differences between small and large tick stocks is in the non-price changing kernel K_n. It increases before plateauing for EBAY whereas it increases then decays as a power law for AMZN. Additionally, a sequence non-price changing events before a price changing event has a greater market impact for larger tick stocks. This is intuitive since a sequence of non-price changing orders on the same side of the book as c will deplete the best bid/ask until the price changing event c which will move the price by at least a tick. Returns in basis points will be higher for larger tick stocks. The decay of K_n for small tick stocks implies that non-price changing events before c put more pressure on the price in the same direction as c.

The price changing kernel K_c for the ESM8 shows a steady power law decay. The kernel does not drop below zero as seen in the equity kernels, hence market impact will be reduced with a sequence of market orders on the other side of the book of c before c occurs. The non-price changing kernel K_n increases for the first lag and then remains remarkably constant at around 0.1bps. Whether this is a very slow decay or constant is not distinguishable from the figure but it does imply that non-price changing events to have a slow decaying impact on the price. The TYM8 contract price changing kernel has a completely different shape to the other kernels. It does not appear to follow a power law, nor any linear relation. This is likely because there are not enough trades per day or price changing events per day to allow a good calibration. The price changing kernel K_n remains approximately constant at zero, although it is very noisy.

6.2 Response Functions

We now assess the performance of the models by doing an out of sample analysis. We use data from first days trading in May 2018 to determine the performance of the models. Subsequent results are calculated with the kernels up to a max lag of $l = 256$, since this is the maximum available for the HDIM2 and using TIM1 and TIM2 kernels up to $l = 1024$ will unfairly penalise the HDIM2. Empirical response functions and response functions generated by the models are shown in figure 19 for positive and negative lags l (negative lag is response following market order). For both small and large tick stocks, we see the response function is generally well replicated up until lags $l \sim \pm 10$ for the TIM2/TIM2 and $l \sim \pm 100$ for the HDIM2/CIM2. The overall divergence between predicted response from the models and empirically measured response function could be amplified by the truncated maximum lag $l = 256$. Evidence for this is contained in the appendix in figure

31.



(a) EBAY (large tick)

(b) AMZN (small tick)

(c) ESM8

(d) TYM8

Figure 19: The response functions

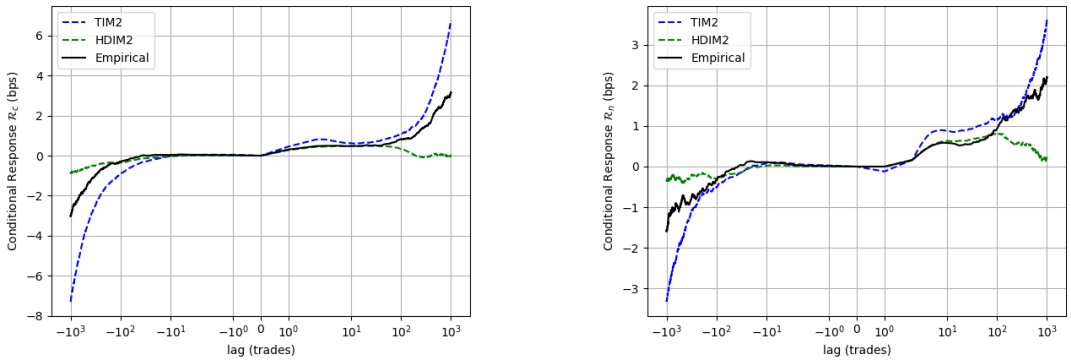
6.2.1 Large tick stocks

The TIM1 response proves the worst fit to empirical data at negative lags for EBAY. The measured response is larger than the predicted response for the TIM1, particularly at larger lags; this is mainly due to the assumption of rigid order flow which does not depend on past price changes. A consequence of this assumption is there must exist a negative correlation between future order flow and past returns, a feature that the TIM1 does not capture [30, section 4.2, page 9]. At positive lags, the TIM1 response function replicates the empirical observation slightly better for EBAY but still diverges for lags greater than a few hundred. The TIM1 therefore performs the worst out of the four models, as expected.

The TIM2 response function shows a slight improvement on the TIM1 for negative lags but the empirical response it still higher. This is due to the additional negative correlation between past

returns and future order flow as seen in the TIM1 response, although albeit lower meaning that some of this behaviour is captured by the model. At smaller positive lags we see there is a slight overshoot. We will observe this overshoot later when we consider the returns plots. There is also a divergence from the empirical response function at larger lags, so the TIM2 performs particularly badly for EBAY. This can be attributed to calibration bias which arises from inconsistent interpretation of the event labels $\{c, n\}$. This will be discussed in detail in section 6.6.

Figure 20 shows the conditional response functions $\mathcal{R}_\pi$. This allows us to see which type of event contributes the most to the overshoot. As expected, the non-price changing event component contributes the most to the deviation for positive lags, as shown in figure 20b. We see the TIM2 predicts a negative response for $l = 1$, which disrespects the non-price changing event label. For negative lags, however, the non-price changing response function appears to have a better fit to empirical data than it's counterpart. The negative correlation between past returns and future order flow is present in both price changing and non price changing TIM2 kernels, although is is less pronounced in the former. The empirical curve appears initially flat for positive lags, implying that non-price changing events do not have an immediate effect on future returns, which seems intuitive.



(a) Price changing

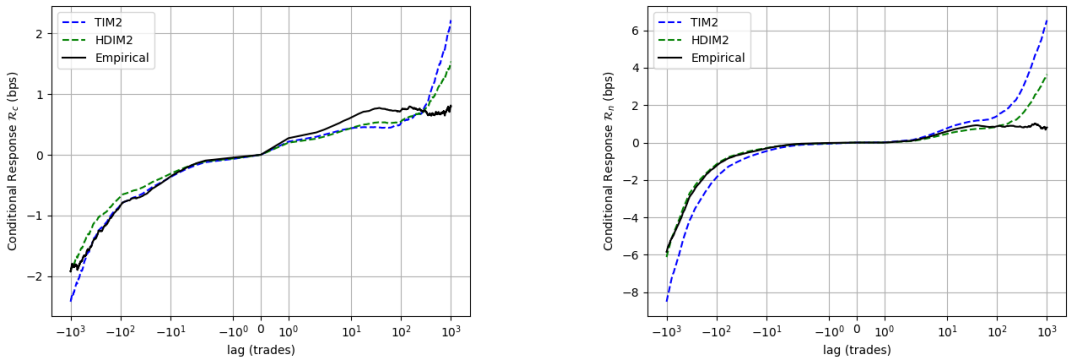
(b) Non-price changing

Figure 20: The conditional response functions for EBAY

Finally we consider the HDIM2 model. For both positive and negative lags, this model matches the empirical response function very well up until lag $l \sim \pm 100$, after which the response function diverges from the true response. Since the HDIM2 has the best theoretical grounding it is expected that this model will reproduce the best results, and up to certain lags it does. The divergence from the true response could be attributed to the truncated lag of the kernels $K_{\pi,\pi'}$. If we now turn out attention to the price changing and non price changing response functions $\mathcal{R}_\pi$, we see that both response functions for the HDIM2 replicate the empirical data up to $l \sim \pm 100$.

6.2.2 Small tick stocks

For AMZN, we see a substantial improvement in the TIM1 and TIM2 response functions at negative lags. The response is still lower than the empirical data but less so than for EBAY, which implies the negative correlation between between future order flow and past returns is smaller. This is common across all smaller tick stocks. The empirical response function increases and then plateaus at larger positive lags and this behaviour is not captured by the TIM1 nor the TIM2, event when calculated using longer kernels as in figure 31b.



(a) Price changing

(b) Non-price changing

Figure 21: The conditional response functions for AMZN

For the TIM2 model, we can check the contributions from price changing or non-price changing events. Figure 21 shows the conditional response functions for the TIM2 and HDIM2. For the TIM2 we see at negative lags, the price changing response function matches the data well, indicating that the negative correlation is captured by the model. However, at positive lags the fit is quite poor. For non-price changing events, the negative correlation is not captured since the TIM2 predicted response is too low for negative lags. The divergence of this response function at positive lags is also present here since the event label inconsistency is still present.

The CIM2 model does not perform as well for small tick stocks as it did for large tick stocks. This is as expected since the TIM2 kernels cannot be interpreted as constant. The plateau of the response function at large positive lags is not captured by the CIM2 either.

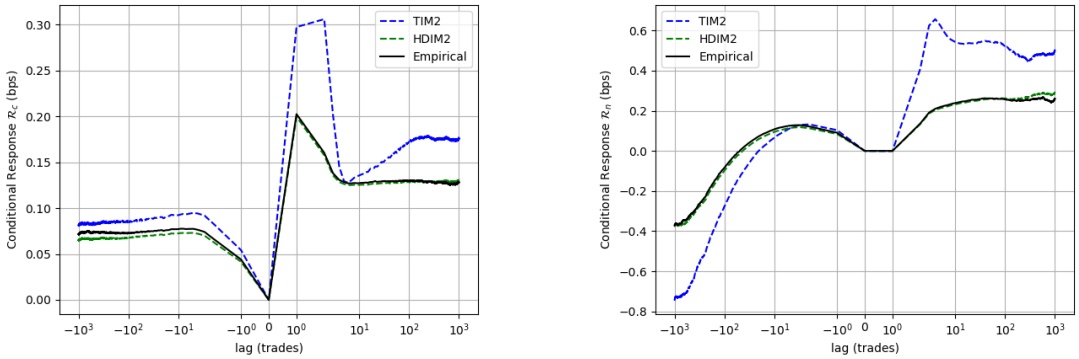
The HDIM2 replicates the response function of AMZN well for negative lags. At positive lags it reproduces the data well until around $l \sim 100$, after which is diverges like the other models. It is not clear whether the cause of this is the truncation of the kernel to $l = 256$ or if this is an issue with the model itself. When considering the price changing kernel, the HDIM2 does reasonable well at negative lags but is relatively poor at reproducing the data at positive lags. There are relatively

few differences between the TIM2 and HDIM2 for the price changing kernel, which is intuitive since the models both receive the same inputs. The difference lies in the non-price changing kernels.

6.2.3 Futures

The response function for the ESM8 Index is different in shape to the equity response functions. Firstly, it spans a much lower range. Secondly, at negative lags the response function increases before decreasing at around lag $l \sim -100$. This suggests that, for lags approaching this point from zero, on average there exists a negative correlation between the present order sign and past returns [30, section 4.3.1, page 16]. At positive lags the response function appears similar to AMZN's: increasing and then plateauing.

TIM1 model does not capture this feature at all and is, as a whole, the worst model for replicating the empirical response function as it consistently under-estimates the response. The TIM2 offers a better fit at negative lags but diverges after the first few lags. For positive lags we still see the overshoot present in equities for the TIM2. The HDIM2 and CIM2 both replicate the empirical response function very well, with the CIM2 only being slightly worse than the HDIM2.



(a) Price changing

(b) Non-price changing

Figure 22: The conditional response functions for the ESM8 Index

If we turn out attention to the conditional response functions, as shown in figure 22, we see that these are very different to the conditional response functions produced by the equities. The non-price changing kernel G_n at negative lags increases initially which indicates that if a non-price changing event occurs, then for smaller lags there is a negative correlation between the current order sign and past returns. The price changing kernel G_c also increases for negative lags which means this negative correlation also exists for price changing events. These both contribute to the overall negative correlation seen in figure 19c. At positive lags, the price changing kernel starts from the value of the spread in basis points and subsequently decreases, which means the general

reaction of the market is to mean revert the price following a price-changing order [30, section 4.3.1, page 16].

A remarkable feature of figure 19c is the performance of the CIM2 model. The CIM2 was introduced as a time independent model such that the time dependent models could be evaluated against some simplified reference model. The surprising success of the CIM2 model in this case can be traced back to the TIM2 kernels G_π. At lags $l \gtrsim 10$, both kernels are non-decaying. This hints at permanenent impact, rather than transient. If the kernel G_n was constant at zero, the price process could be considered as the sum of non-zero price changes, of which market impact is permanent [30, section 4.3.1, page 16]. The price could then be modelled by the sequence of random variables $\{(\epsilon_t, \pi_t)\}_{t\in\mathbb{N}}$, and their correlation structure. Since G_n is non-zero constant, we can still apply this with a vertical shift. This implies that the a constant impact model will work well for ESM8.

The HDIM2 model replicates both price changing and non-price changing kernels well, but the TIM2 is a relatively poor fit, especially for the non-price changing kernel.

The response function for TYM8 is similar in shape to the ESM8. All models are relatively poor at replicating this response function. We still observe the overshoot for the TIM2 and all models begin to diverge after the first few lags. We have omitted the conditional response functions since they does not provide any insightful information.

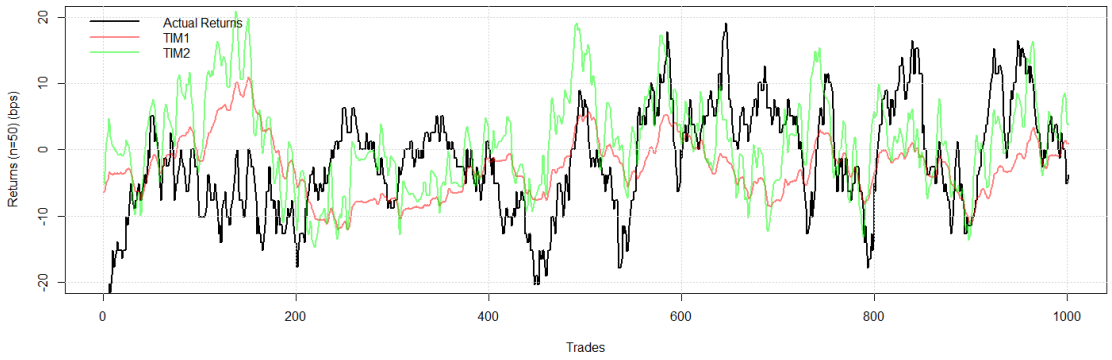
6.3 Returns

We now turn our attention to the main assessment for the performance of the models: the predicted price returns. We want to assess how well each model's predicted returns match the observed returns in the market. We use 50 trade log returns and display excerpts of the returns for each dataset for visualisation. Out of sample analysis is also used here on the first days trading in May 2018.

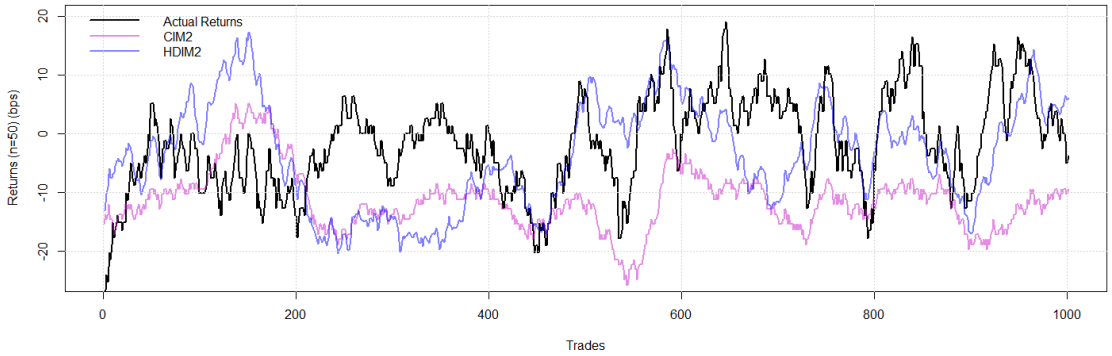
6.3.1 Large tick stocks

An excerpt of the EBAY 50 day log returns as well as the predicted returns for the models is shown in figure 23. The top panel includes predicted returns for the TIM1 and TIM2 models. Both models do not appear to replicate the true returns accurately. A number of times when the returns peak, like at trade 250 and trade 350, the models do not capture this. The TIM1 frequently underestimates the returns and the TIM2 frequently overshoots when the returns peak. This is to be expected since both models have their drawbacks: the TIM1 requires rigid order flow and the TIM2 is inherently inconsistent. Additionally, the calibration data used consisted of only 21 days. In the literature, calibration of multiple years data was used but this was not a viable option

given the resources available. Notwithstanding, since the returns are so poorly replicated by the TIM1 and TIM2 models there is no guarantee that calibration over more data would substantially improve the results.



(a) TIM1/TIM2



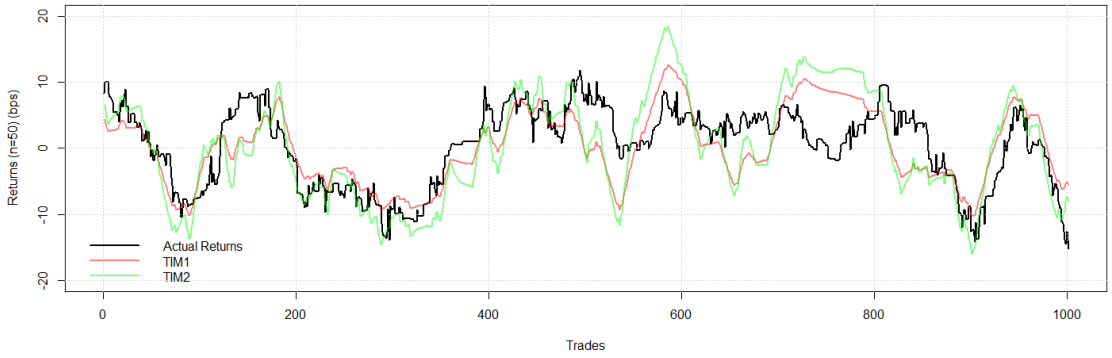
(b) CIM2/HDIM2

Figure 23: An excerpt of 50 trade returns from EBAY with the corresponding predicted returns from the models. Two figures were used for clarity.

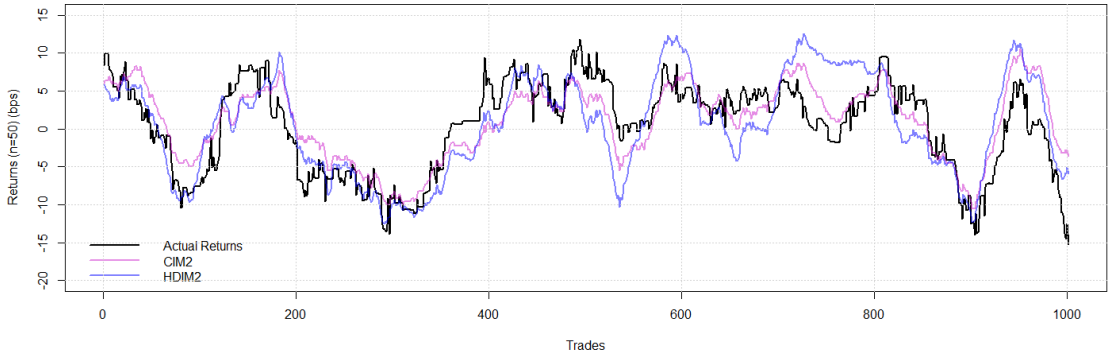
The CIM2, on inspection, appears to be the worst fit of all. This is as expected since the model is very simple. The general direction of the returns is reasonably well reproduced, i.e. when true returns increase the CIM2 predicted returns will likely increase too. For the most part, the predicted returns underestimate the true returns. The HDIM2 predicted returns produce the best fit to the data but its performance is still relatively poor. It represents a slight improvement on the TIM2 model, but still has similar issues. For example, the two peaks a trade 250 and 350 are not captured by the model. There are also overshoots in the predicted returns but not as pronounced as for the TIM2. The poor performance of all the models when reproducing returns is expected since their corresponding predicted response functions were poor.

6.3.2 Small tick stocks

The model predicted returns for AMZN are substantially closer to the true returns than for EBAY. Figure 24 shows an excerpt of the true 50 trade log returns and their corresponding predictions from the models. The TIM1 and TIM2 returns shown in the top panel exhibit a remarkable improvement over the EBAY returns. The rigid order flow assumption has less of an influence on small tick stocks which is the primary reason for the increase in performance for the TIM1 and TIM2 returns. There are still overshoots from both models and occasionally the model will fail to capture a peak or trough, as shown around trade 550.



(a) TIM1/TIM2



(b) CIM2/HDIM2

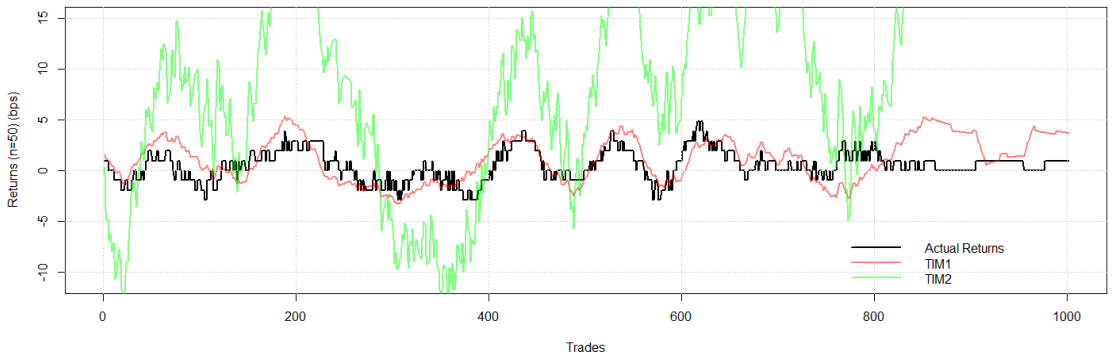
Figure 24: An excerpt of 50 trade returns from AMZN with the corresponding predicted returns from the models. Two figures were used for clarity.

The excerpt also shows a remarkable improvement in the CIM2 predicted returns. If considering this sample alone, the CIM2 appears to be the best performing model of them all which seems counterintuitive since the CIM2 is expected to perform better for large tick stocks. We will see later that this excerpt is misleading and actually the CIM2 has the worst fit of all the models. The

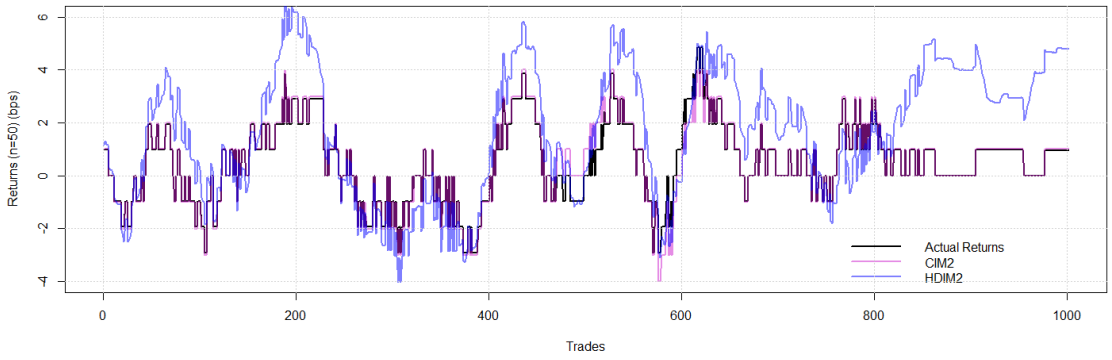
HDIM2 model also provides a reasonably good fit to data which is as expected since the HDIM2 has the best theoretical grounding and the models are expected to perform better for small tick stocks.

6.3.3 Futures

The first thing that is immediately obvious from figure 25 is the very poor performance of the TIM2 model. The returns are out by almost an order of magnitude. A poor fit for the TIM2 is to be expected since the response function is not replicated well, however the true extent of the discrepancy is alarming. A possible explanation for this lies in the non-price changing events. The non-price changing kernel is hugely over estimated and the futures data consists mainly of non-price changing events. The TIM1 performs substantially better than the TIM2, although it fails to capture the (approximate) price pin in the final 200 trades of the excerpt.



(a) TIM1/TIM2



(b) CIM2/HDIM2

Figure 25: An excerpt of 50 trade log returns from ESM8 with the corresponding predicted returns from the models. Two figures were used for clarity.

The HDIM2 performs reasonably well although there are prominent overshoots whenever the returns peak and the model also fails to capture the price pin in the final segment. The CIM2, however, exhibits the best fit of the models. The returns match so well that it can be hard to distinguish the CIM2 returns in the figure since they hug the true returns so well.

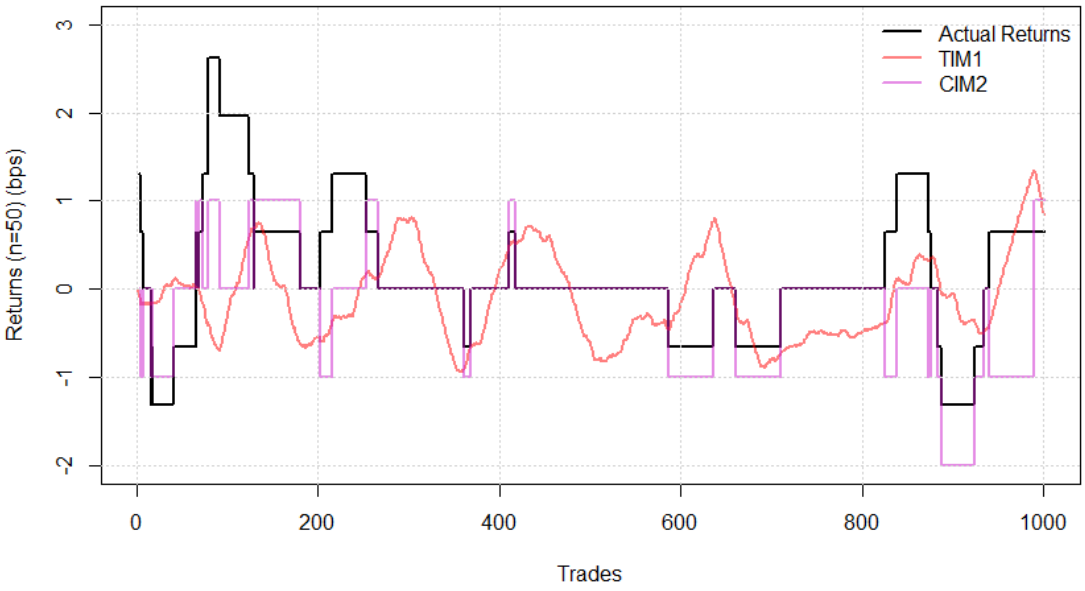


Figure 26: The 50 trade log returns and the model predictions for the TIM1 and CIM2.

For the TYM8, none of the models appear to accurately reproduce the returns. Based on the model's ability to replicate the response function, this is as expected. The CIM2 offers the best fit and the TIM2 and HDIM2 produce returns an order of magnitude higher than expected, hence they are not displayed. This is expected of the TIM2 but is rather surprising for the HDIM. The strange shape of the kernel implies that the HDIM2 is not a suitable model for this dataset.

6.3.4 Quantitative comparison

Since we have only considered small excerpts of 1000 trades; it would not be fair to draw any conclusions from this. A more quantitative description is required which encompasses the entire dataset. Here we offer a simple quantitative comparison which we will use to conclude which combination of model and dataset fits the best.

We consider a second order comparison of the predicted 50 day log returns and the true returns.

Model	EBAY	AMZN	ESM8	TYM8
TIM1	9.499	5.447	1.842	4.866
TIM2	11.023	6.167	12.944	9.482
HDIM2	10.860	5.120	1.632	21.442
CIM2	14.256	7.110	0.493	4.009

Table 4: The root mean square error $\mathcal{E}^2_{\mathrm{m}}$ of the returns for each model.

We will consider the root mean-squared error $\mathcal{E}$, defined as

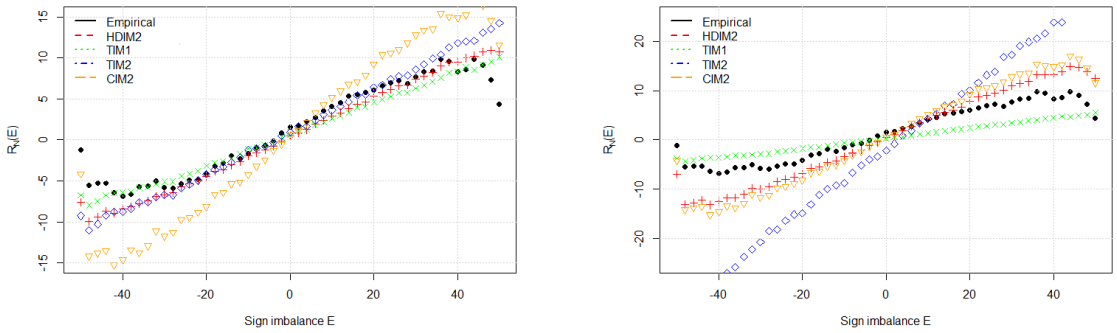
$$\mathcal{E}^2_{\mathrm{m}} := \frac{1}{N}\sum_{i=1}^{N}(r_t - r_t^{\mathrm{m}})^2\,, \tag{6.2}$$

where N is the number of data points, r_t is the true out-of-sample returns, r_t^{m} is the predicted returns of each model and $\mathrm{m} \in \{\mathrm{TIM1}, \mathrm{TIM2}, \mathrm{HDIM2}, \mathrm{CIM2}\}$.
The results are shown in table 4. We see there is a substantial increase in the performance in results for AMZN over EBAY. One surprising result is the TIM1 and TIM2 outperform the CIM2, something that was not evident in the returns plots. The HDIM2 proves the best fit for small tick stocks. For futures, we see a substantial increase in performance for the ESM8 over the TYM8, as anticipated. The CIM2 model for ESM8 yields the best results, we a root mean square error of just 0.493.

6.4 Replication of Empirical Observations

Here we assess the performance of the models at replicating some of the findings displayed in 4. The models account for the autocorrelation of signs by construction. Additionally, since this study is mainly concerned with order signs, we will not verify the concave volume dependence. We will mainly consider price pins and aggregate sign impact. The price pins are not replicated well by all models apart from the CIM2 for futures. Figure 25 shows an example of this. There is an approximate price pin for the final 100 trades and this is not captured by any of the models except the CIM2. This was observed frequently in all of the datasets. Each of the models except the CIM2 largely over estimated returns around price pins.

Figure 27 shows the aggregate sign impact $\mathcal{R}_N(\mathcal{E})$ for $N = 50$ predicted by all of the models.
For AMZN, most models are able to approximately reproduce the sinusoidal shape, but all models overestimate the aggregate sign impact when the sign imbalance is 100%. This is as expected since the models were not able to capture the price pins. The CIM2 produces the best sinusoidal shape but still deviates from the true values the most. The TIM1 and TIM2 aggregate sign impact functions are both linear, which does not reflect the empirical findings. The ESM8 aggregate sign



(a) AMZN (b) ESM8 Index

Figure 27: Aggregate sign impact $\mathcal{R}_{50}(\mathcal{E})$ for the models.

impact results are not replicated well by the models. All models, except the TIM1 overestimate the aggregate impact and the TIM2 proves the worst fit again for reasons discussed above.

6.5 Calibration to Special Days

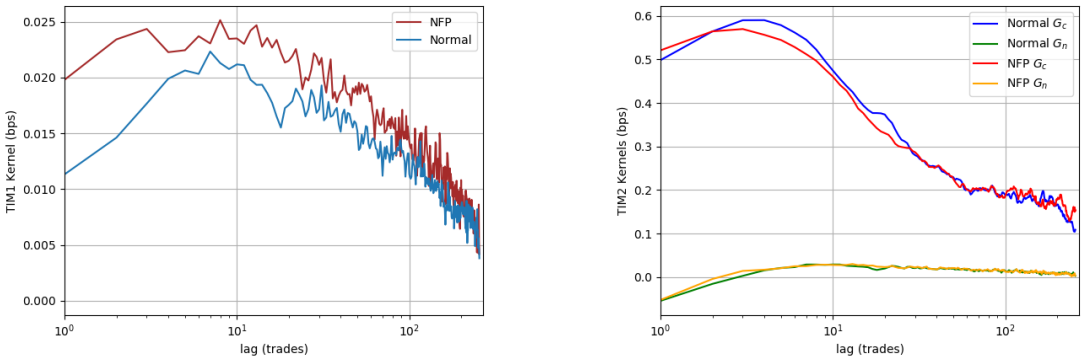
As an extension, we offer a comparison for the 10Y U.S. T-Note treasury futures calibrated on a 'special day'. A special trading day refers to any trading day where trading is different to usual. Examples include month-end, contract expiry date or contract rollover dates. The special day we will consider is the *Non-Farm Payroll* date. Non-farm payroll (NFP) refers to any job in the U.S except farming, self-employment, non-profit employment and military employment. The U.S Bureau of Labor Statistics releases the change in the total number of non-farm payrolls on the first Friday of each month and it is generally seen as an indicator of economic health. The release of this statistic leads to different trading behaviour on these days. The days are generally more active, especially if the figure comes as a surprise. The difference in trading activity will be more pronounced for long-maturity governmnent debt securities, hence why we choose the 10Y U.S. T-Note futures. We want to see if there is a discrepancy between the kernels and associated returns when calibrated to these NFP days only. Since Bloomberg data only goes back a few months, only 5 days data was available for the 10 Y U.S T-Note futures. Since we don't want to penalise either dataset, we will recalibrate the 10Y U.S T-Note futures using only 5 days data[7] and we will compare with kernels calibrated to NFP days only.

Figure 28 shows the calibrated TIM1 and TIM2 kernels for the 10-Year U.S. Treasury Note futures kernel. The calibration process for the HDIM2 failed. This is expected since it also failed when calibrated to a whole months data. This is likely due to insufficient trades per day. We will omit

[7]The days were arbitrarily chosen as the days before NFP Dates.

the HDIM2 from subsequent results for this asset.

The TIM1 kernels are both noisy which is as expected since the dataset used to calibrate is much smaller. The kernel is also an order of magnitude smaller in range than the ESM8 kernel but still exhibits a slow power law decay. The kernel calibrated to the NFP date dataset is larger than its regular counterpart, and this is intuitive since these trading days are more active and volatile, hence returns are likely to be higher.



(a) TIM1 (b) TIM2

Figure 28: Calibrated kernels for the 10Y U.S. T-Note futures

The 10-Year U.S. Treasury Note non-price changing kernel is approximately constant at zero, whilst the price changing kernel follows an approximate power law decay for lags greater than 5. There does not appear to be a significant discrepancy between the NFP dates and non-NFP dates when calibrating the kernels. The G_c kernel is slightly higher for NFP dates up until around $l = 50$, for similar reasons as outlined above.

Figure 29 shows the response functions predicted by the TIM1 and TIM2 models. The CIM2 is omitted since it does not change when calibrated NFP dates. There do not appear to be any large discrepancies between the TIM models calibrated to NFP dates and those calibrated to regular days. This is as expected since the kernels were very similar. The responses from NFP dates are slightly higher than normal days, which is also as expected from the kernels above.

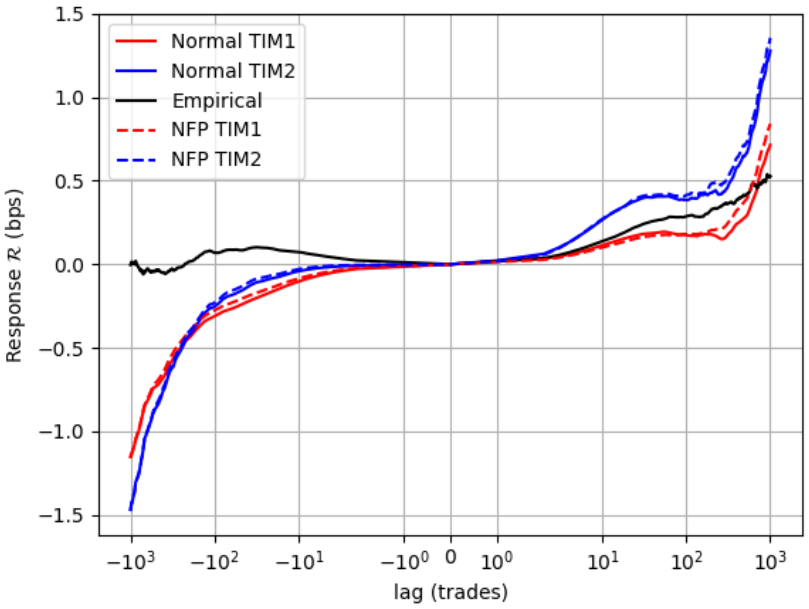


Figure 29: The response functions for the TIM1 and TIM2 models calibrated for normal and NFP dates. The curves appear very similar.

Figure 30 shows the predicted returns of the TIM1 and CIM2 models for the 10Y U.S. T-Note futures. The TIM2 model is not included since the returns were an order of magnitude larger than the predicted returns. This is primarily due to the high number of non-price changing events, since the TIM2 is inherently inconsistent for non-price changing events. The TIM1 performs poorly, regardless of NFP date or not. Predicted returns are slightly higher at peaks, which represents a slight improvement of the models when calibrated to NFP days. We consider this an improvement since TIM1 returns usually underestimate the returns. Nonetheless, there is an argument that this is not an improvement since the overall fit of the TIM1 is so poor. Given the available data, this is not surprising. The CIM2 model provides the best fit for returns in this excerpt which is intuitive since it does not have to be calibrated in the same way as the other models.

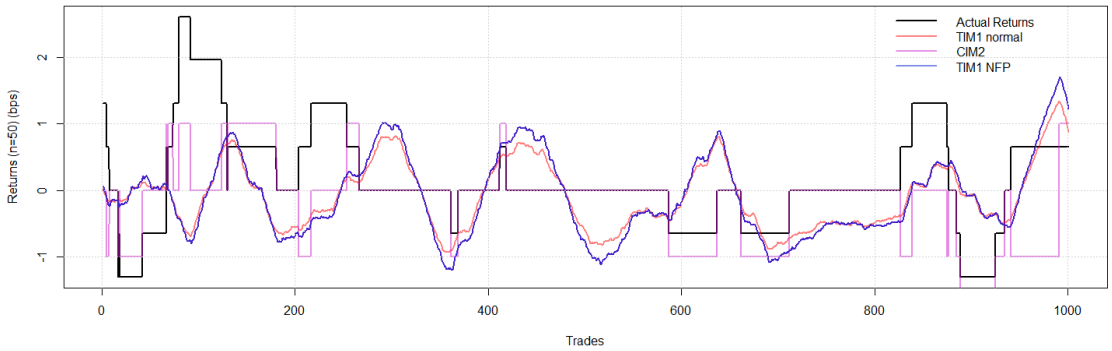


Figure 30: An excerpt of returns for 10Y U.S. T-Note futures. The TIM2 is omitted since returns were too large and greatly exceed that of the true returns. The CIM2 returns match the true values the best.

6.6 Discussion

We have explored each propagator model for a number of different datasets, each performing differently depending on the dataset. We will put the results in section 6.5 aside for now, since they are only calibrated to 5 days data which is insufficient to fully calibrate the models. Otherwise, the HDIM2 has the best overall fit if the entire data range is considered. This is as expected since the HDIM2 has the best theoretical grounding. That said, performance is dependent on the dataset of choice. For small tick stocks, the models out-perform those calibrated to large tick stocks. One of the key assumptions used in propagator models is rigid order flow, meaning that the order book does not explicitly react to past event hence it can be described as a whole by its correlations [30, section 5, page 17]. In this sense, the structure of order flow can be considered as exogenous and does not depend on the short term dynamics of prices. This assumption is mostly valid for small tick stocks where price changes are more common. For large tick stocks, however, price changes are rare and this assumption is no longer valid.

The TIM2 is poor at replicating the response functions for positive and negative lags. The returns also corroborate this observation, since the returns over shoot before reverting. This is somewhat counterintuitive since the TIM2 received the same input as the HDIM2. The reason lies in the calibration bias. The calibration bias, defined in equation (5.25) only vanishes if there is are no correlations between the the prediction error and the inputs. The TIM2 contains one substantial inconsistency which will likely lead to correlations between the prediction error and the inputs to the model: the incorrect interpretation of event label. This leads to negative correlations between past returns and future order signs. As seen in the results above, the discrepancy between TIM2 predicted returns and the true returns is magnified significantly when there are a fewer proportion

of price changing events in the dataset.

The TIM1 as a whole also performs particularly badly. Since the model is so simplistic and only depends on order signs, this is hardly surprising. One interesting observation is the slight improvement in returns when calibrating to NFP dates. For small tick stocks the TIM1 consistently underestimates returns so calibrating to NFP dates will likely train the model to predict larger returns and provide a better fit. This will be an interesting area for further study, provided more NFP day data can be acquired.

The CIM2 performs poorly for both small tick and large tick stocks, however for futures it represents the best fitting model. The returns for the futures appear more discrete than for the stocks. This benefits the CIM2 the most since non-price changing events are assumed to have zero impact. Since the CIM2 fits so well, this implies that for futures a non-price changing event will have negligible effects on returns. The main conclusion of this is the fit of the models depends on the dataset. The CIM2 predicts constant, permanent impact for returns rather than transient impact which is much better suited to futures than for stocks.

The dependency of the validity of the models on the dataset seems intuitive. Each dataset has a different number of price changing events and different daily activity. The poor performance of the propagator models for the TYM8 could be attributed to insufficient trades per day, or it could be down to the number of price changing events. In proportional terms, the ESM8 has around 5 times the amount of price changing events as the TYM8.

Before concluding we will divert a little to consider the applications and impact of these findings. As mentioned in the introduction, understanding market impact is crucial for controlling execution costs and for setting regulation. Whilst some of our models have not been able to successfully reproduce empirical findings, in some cases we have managed to find a good fit, to a reasonable degree of accuracy. We are able to perform out-of-sample analysis to predict different scenarios based only on two inputs: the signs of the trades and whether it is price changing or not. We have shown for futures that returns can be predicted. For a sell side business, this allows them to predict their execution costs in advance[8], therefore they can adjust their execution strategy accordingly. Nonetheless, methods of predicting whether an event will be price changing or not are required to firm up the predictive power of these models, hence these models are better utilised as a way of understanding market impact, rather than predicting it.

[8]The predictions will not be far into the future since they will have to know whether their child orders will be price changing or not. In the short term you can predict this, for example if there is a high order book imbalance to one side, you know a small order will likely not be price changing.

7 Conclusion

The aim of this study was to investigate linear propagator models as a class of models for modelling market impact. We have found a large dependency of the performance of the models on the dataset under consideration. All models perform badly for large tick stocks but exhibit a substantial improvement when tested on small tick stocks. This is due to the assumption of rigid order flow. We find that the TIM2 is the weakest model due to inconsistent interpretation of the price changing label. When applied to futures, provided the futures contracts are traded actively enough, we see strong results for the HDIM2 and particularly the CIM2 which is particularly good for modelling returns which appear more discrete. When calibrating the models to NFP days, we show there is a slight improvement in the TIM1 results but the HDIM2 and TIM2 failed to exhibit any improvement, mostly due to the calibration difficulties with the asset under consideration.

We have left a number of problems for future studies. The main drawback of these models is we are not able to simulate artificial order-flows with them. As discussed, a sell-side company could anticipate a price changing event, but there is only a small window in time at which this prediction is viable. Further studies could include predictions as to whether an event will be price changing or not. Furthermore, we have only considered trades for this study. Additional studies could look in more detail into the impact of limit orders, market orders and cancellations (see [29] and [12]).

We have touched on the idea of calibrating these models to so-called 'special days', such as NFP dates. Since the dataset used was limited; a more in-depth study into these special days could yield some interesting results. This could include contract roll dates or month-end. Finally, other assets could be considered. We have noted the improvement of the models when considering futures over stocks (particularly large tick stocks). An investigation into the performance of these models for other asset classes may uncover further improvements.

A Appendix

A.1 The Micro Price

The micro-price is was introduced by Stoikov [35, page 4, section 2] to address a number of issues with using the mid price as the fair price. Firstly, it has been observed that changes in the mid price can be correlated, an effect known as *bid-ask bounce*. This means that the mid price is not a martingale. Secondly, the price signal is low frequency since price changes occur fairly infrequently compared to the quote updates. Lastly, it does not consider the volume at the best bid/ask prices. The set up of the micro price is as follows: consider the prediction of the i-th mid-price

$$P_t^i = \mathbb{E}[m_{\tau_i}|\mathcal{F}_t],$$

where τ_i, $i = 1, \dots, n$ are stopping times when the mid price m_t changes defined by

$$\tau_1 = \inf\{u > t | m_u - m_{u-} \neq 0$$
$$\tau_{i+1} = \inf\{u > \tau_i | m_u - m_{u-} \neq 0\},$$

and $\mathcal{F}_t$ is the information contained in the order book up until time t. The price processes P_t^i are martingales by construction.

Definition A.1. The micro price is defined as

$$P_t^{micro} = \lim_{i \to \infty} P_t^i \tag{A.1}$$

A.2 Proof of convolution theorem

Proof. Define $h(z) := (f * g)(z)$. First we need to check that h is in L^1. Note that

$$\int\int |f(x)g(z-x)|dzdx = \int |f(x)| \int |g(z-x)|dzdx = \int |f(x)| \|g\|_1 dx = \|f\|_1 \|g\|_1,$$

so by Fubini's theorem we have that $h \in L^1(\mathbb{R}^n)$

$$H(\xi) := \mathscr{F}[h] = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x)g(z-x)dx e^{i\xi \cdot z} dz$$

Note that $|f(x)g(z-x)e^{i\xi \cdot z}| = |f(x)g(z-x)|$ so we can apply Fubini's theorem again

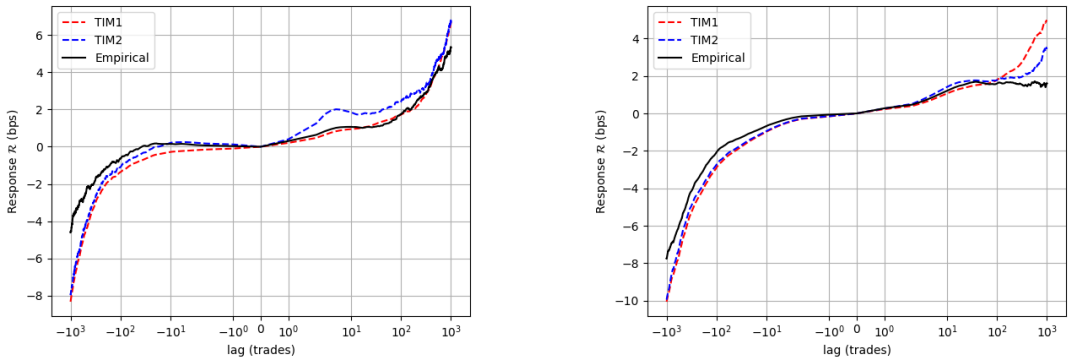
$$H(\xi) = \int_{\mathbb{R}^n} f(x) \left(\int_{\mathbb{R}^n} g(z-x) e^{i\xi \cdot z} dz \right) dx.$$

By substituting $y = z - x$ we have

$$\begin{aligned} H(\xi) &= \int_{\mathbb{R}^n} f(x) \left(\int_{\mathbb{R}^n} g(y) e^{i\xi\cdot(y+x)} dy \right) dx \\ &= \int_{\mathbb{R}^n} f(x) e^{i\xi\cdot x} \left(\int_{\mathbb{R}^n} g(y) e^{i\xi\cdot y} dy \right) dx \\ &= \int_{\mathbb{R}^n} f(x) \left(\int_{\mathbb{R}^n} g(y) e^{i\xi\cdot(y+x)} dy \right) dx \\ &= \int_{\mathbb{R}^n} f(x) e^{i\xi\cdot x} dx \int_{\mathbb{R}^n} g(y) e^{i\xi\cdot y} dy \\ &= F(\xi) G(\xi). \end{aligned}$$

Applying the inverse Fourier transform to both sides yields the required result. □

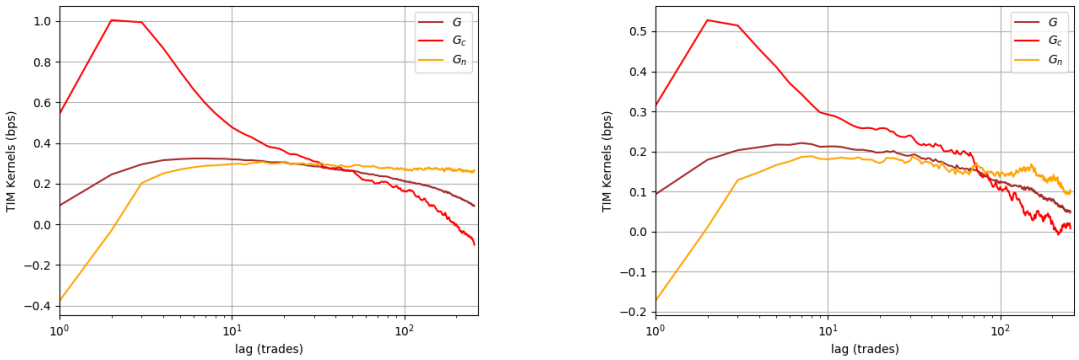
A.3 Improved response functions



(a) EBAY (b) AMZN

Figure 31: The TIM1 and TIM2 response functions calculated using kernels up to lag $l = 1000$. We see a substantial improvement from figures 19a and 19b for these two models, implying this would be the case for the HDIM2 too.

A.4 Kernels for other Stocks



(a) INTC (large tick)

(b) MSFT (small tick)

Figure 32: The TIM kernels for MSFT and INTC. Both exhibit the same behaviour as AMZN and EBAY respectively.

References

[1] I. Lemhadri, "Market impact in a latent order book," 2017.

[2] A. A. Obizhaeva and J. Wang, "Optimal trading strategy and supply/demand dynamics," *Journal of Financial Markets*, vol. 16, no. 1, pp. 1–32, 2013.

[3] J. Donier and J. Bonart, "A million metaorder analysis of market impact on the bitcoin," *Market Microstructure and Liquidity*, vol. 1, no. 02, p. 1550008, 2015.

[4] B. Tth, Z. Eisler, and J. Bouchaud, "The squareroot impace law also holds for option markets," *Wilmott*, vol. 2016, no. 85, pp. 70–73, 2016.

[5] B. Toth and J.-P. Bouchaud, "Price impact of metaorders: the latent order book," 2011.

[6] P. Welch, "The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on audio and electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[7] M. Avellaneda, "Algorithmic and high-frequency trading: an overview," in *New York University & Finance Concepts LLC Quant Congress USA 2011*, 2011.

[8] R. Almgren, "Execution costs," *Encyclopedia of quantitative finance*, 2010.

[9] A. B. Schmidt, *Financial Markets and Trading: An Introduction to Market Microstructure and Trading Strategies*. Wiley, 2011.

[10] R. Cont, S. Stoikov, and R. Talreja, "A stochastic model for order book dynamics," *Operations research*, vol. 58, no. 3, pp. 549–563, 2010.

[11] J. . Bouchaud, Y. Gefen, M. Potters, , and M. Wyart, "Fluctuations and response in financial markets: the subtle nature of random price changes," *Quantitative Finance*, vol. 4, no. 2, p. 176, 2004.

[12] Z. Eisler, J.-P. Bouchaud, and J. Kockelkoren, "Models for the impact of all order book events," *arXiv preprint arXiv:1107.3364*, 2011.

[13] S. Harder, *The Efficient Market Hypothesis and Its Application to Stock Markets*. GRIN Verlag, 2010.

[14] R. J. Shiller *Do stock prices move too much to be justified by subsequent changes in dividends?*, 1980.

[15] P. Bak, M. Paczuski, and M. Shubik, "Price variations in a stock market with many agents," *Physica A: Statistical Mechanics and its Applications*, vol. 246, no. 3-4, pp. 430–453, 1997.

[16] S. Maslov, "Simple model of a limit order-driven market," *Physica A: Statistical Mechanics and its Applications*, vol. 278, no. 3-4, pp. 571–578, 2000.

[17] J.-P. Bouchaud, "Price impact," *arXiv preprint arXiv:0903.2428*, 2009.

[18] A. S. Kyle, "Continuous auctions and insider trading," *Econometrica: Journal of the Econometric Society*, pp. 1315–1335, 1985.

[19] J. Hasbrouck, "Measuring the information content of stock trades," *The Journal of Finance*, vol. 46, no. 1, pp. 179–207, 1991.

[20] R. Cont and A. De Larrard, "Price dynamics in a markovian limit order market," *SIAM Journal on Financial Mathematics*, vol. 4, no. 1, pp. 1–25, 2013.

[21] E. Smith, J. D. Farmer, L. s. Gillemot, and S. Krishnamurthy, "Statistical theory of the continuous double auction," *Quantitative finance*, vol. 3, no. 6, pp. 481–514, 2003.

[22] M. G. Daniels, J. D. Farmer, L. Gillemot, G. Iori, and E. Smith, "Quantitative model of price diffusion and market friction based on trading as a mechanistic random process," *Physical Review Letters*, vol. 90, no. 10, p. 108102, 2003.

[23] J. Donier, J. Bonart, I. Mastromatteo, and J.-P. Bouchaud, "A fully consistent, minimal model for non-linear market impact," *Quantitative finance*, vol. 15, no. 7, pp. 1109–1121, 2015.

[24] O. Guant, C.-A. Lehalle, and J. Fernandez-Tapia, "Optimal portfolio liquidation with limit orders," *SIAM Journal on Financial Mathematics*, vol. 3, no. 1, pp. 740–764, 2012.

[25] D. Bertsimas and A. W. Lo, "Optimal control of execution costs," *Journal of Financial Markets*, vol. 1, no. 1, pp. 1–50, 1998.

[26] R. Almgren and N. Chriss, "Optimal execution of portfolio transactions," *Journal of Risk*, vol. 3, pp. 5–40, 2001.

[27] F. Patzelt and J.-P. Bouchaud, "Nonlinear price impact from linear models," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2017, no. 12, p. 123404, 2017.

[28] C. Hopman, "Are supply and demand driving stock prices," *Quantitative Finance*, 2002.

[29] Z. Eisler, J.-P. Bouchaud, and J. Kockelkoren, "The price impact of order book events: market orders, limit orders and cancellations," *Quantitative Finance*, vol. 12, no. 9, pp. 1395–1419, 2012.

[30] D. E. Taranto, G. Bormetti, J.-P. Bouchaud, F. Lillo, and B. Toth, "Linear models for the impact of order flow on prices i. propagators: Transient vs. history dependent impact," *arXiv preprint arXiv:1602.02735*, 2016.

[31] F. Patzelt and J.-P. Bouchaud, "Universal scaling and nonlinearity of aggregate price impact in financial markets," *Physical Review E*, vol. 97, no. 1, p. 012304, 2018.

[32] M. D. Evans and R. K. Lyons, "Order flow and exchange rate dynamics," *Journal of political economy*, vol. 110, no. 1, pp. 170–180, 2002.

[33] J. Bonart, "The price impact of trades," 2015.

[34] E. Bacry, A. Iuga, M. Lasnier, and C.-A. Lehalle, "Market impacts and the life cycle of investors orders," *Market Microstructure and Liquidity*, vol. 1, no. 02, p. 1550009, 2015.

[35] S. Stoikov, "The micro-price: A high frequency estimator of future prices," 2017.