Do price trajectory data increase the efficiency of market impact estimation?

FENGPEI LI ⊂*,†, VITALII IHNATIUK‡§, YU CHEN†, JIAHE LIN†, RYAN J. KINNEAR¶, ANDERSON SCHNEIDER†, YURIY NEVMYVAKA† and HENRY LAM∥

†Machine Learning Research, Morgan Stanley, New York, NY, USA
‡Quantitative Research, Morgan Stanley, New York, NY, USA
§Department of Economics, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
¶Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada
∥Department of IEOR, Columbia University, New York, NY, USA

(Received 27 May 2022; accepted 30 April 2024; published online 29 May 2024)

Market impact is an important problem faced by large institutional investors and active market participants. In this paper, we rigorously investigate whether price trajectory data from the metaorder increases the efficiency of estimation, from the view of the Fisher information, which is directly related to the asymptotic efficiency of statistical estimation. We show that, for popular market impact models, estimation methods based on partial price trajectory data, especially those containing early trade prices, can outperform established estimation methods (e.g. VWAP-based) asymptotically. We discuss theoretical and empirical implications of such phenomenon, and how they could be readily incorporated into practice.

Keywords: Market microstructure; Market impact model; Parameter estimation

JEL Classification: C9

1. Introduction

Market impact is a crucial feature of market microstructure for large traders, manifesting as adverse effects on the price of the underlying asset due to order execution. In other words, upon completion of the trade, aside from direct costs (i.e. commissions/fees), slippage from effective bid-ask spread or delay/timing risk, investors are also subject to the transaction cost generated from the price impact of their own actions (Robert et al. 2012). For example, a trader who needs to liquidate a large number of shares will take liquidity from the Limit-Order Book (LOB) and push the price down, resulting in a implementation shortfall (see Perold 1988) or liquidation cost (see Gatheral and Schied 2013), which is the difference between the realized revenue and the initial asset value. Besides the short-term correlation between price changes and trades, or the statistical effect of order flow fluctuations (see Bouchaud 2010), one notable explanation for this dynamics of market impact relates to the reveal of new, private infor-

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.
forms/parameters admit different types of price manipulation strategies (see Huberman and Stanzl 2004, Alfonsi et al. 2012, Gatheral and Schied 2013) which potentially lead to risky/unstable trading behavior and mathematically preclude the existence of unique optimal execution strategies (Gatheral et al. 2012). Second, the estimation of model parameters has important implications in explaining and understanding various stylized facts/empirical findings including the square-root impact law (Grinold and Kahn 2000, Bucci et al. 2019), the logarithmic dependence in market impact surface (Zarinelli et al. 2015), the power law decay (Bouchaud et al. 2003, Gatheral 2010) and so on. Finally, the design of trading strategies that minimize execution cost, as well as pre-trade analytic software that delivers a reliable pre-trade estimate of the expected trade cost, relies crucially on an accurate and efficient estimation of the model parameters. Indeed, following the prevalence of automated trading algorithms, both of the above have become standard considerations for active market participants, especially large institutional investors.

Despite its theoretical importance in market microstructure literature and its practical importance among institutional traders, there are relatively few studies (both empirically and, to a greater extent, theoretically) of the parameter estimation problem for market impact models, and we aim to fill the gap in this paper. One main reason is the limited access to the metaorder data for academics and practitioners, as these data are typically proprietary data of brokerage firms or funds. Consequently, most empirical studies of the market impact estimation are based on publicly available datasets, which collectively suffer from certain inherent drawbacks (e.g. limited information on trades being initiated by buyer vs seller, an unknown number of metaorders, unknown trading duration Almgren et al. 2005) and can only provide a ‘partial view’ of the market (Zarinelli et al. 2015). Another possible reason is the limited consensus on the appropriate model for price impact (e.g. linear vs non-linear, permanent vs temporary, transient vs instantaneous, see Bouchaud 2010, Cont et al. 2014). Notably, a few exceptions include (Almgren et al. 2005, Moro et al. 2009, Zarinelli et al. 2015) which conducted empirical investigations with large metaorder datasets, but the model fitting procedures typically relied only on some summary statistics during the execution. For example, in Almgren et al. (2005), the authors tried to determine the exponent of the power law functional form in price impact, for which a nonlinear least square regression is carried out with two statistics: the realized impact and the permanent impact (which we explore in detail later). In Zarinelli et al. (2015), in order to fit the temporary market impact surface, the market impact measured at the moment when the metaorder is completed was regressed on the metaorder duration and the metaorder participation rate. Moreover, it was suggested in Curato et al. (2017) that, as ‘one of the major attractions of the propagator model to practitioners’, the historical data of the cost of VWAP executions† (Volume Weighted Average Price, can be seen as another summary statistics, similar to the realized impact in Almgren et al. 2005) can be easily used to estimate the functional forms/parameters of the instantaneous market impact function.

While these summary statistics contain important information about the price trajectory during the order (e.g. cost of VWAP involves averaging of price along the trajectory), discarding a significant portion of price trajectory data during model fitting seems uneconomical, especially since the trajectory—beyond merely the trade cost—can offer additional insights. In particular, considering how the trajectory, driven by the price movement, arrives at the cost (shape of the ‘master curve’ Lillo et al. 2003) potentially reveals extra information (see also Briere et al. 2020). For example, it has been empirically observed that the market impact of metaorders is concave with respect to the order size (Tóth et al. 2011, Zarinelli et al. 2015), from which one might conjecture the price movements around the early stages of the trade can be especially informative for predicting the total order cost or the market impact shape. Meanwhile, for the owner of metaorder data (i.e. asset managers or brokerage firms), compared with modeling approaches based on LOB, modeling approaches based on the price dynamics would also be feasible, as the additional collection and storage of these extra data points during the life of order should generally not come at a high cost. On the other hand, it is not unusual to see financial data intentionally discarded (due to structural noise or data corruption) for more accurate estimates. For example, it is common practice (see Zhang et al. 2005, Aït-Sahalia and Mykland 2009) when estimating the volatility of an asset return process to throw away a large fraction of the high-frequency data as a way to avoid the contamination of market microstructure noise (e.g. bid-ask spread). In particular, the realized volatility is typically computed by the sum of less frequently sampled squared returns, i.e. selected on time intervals much larger (i.e. 5 or 10 min) than when originally collected (i.e. every a couple of seconds or less), thus effectively discarding a substantial portion of a dataset. Although the bid-ask spread should not have a substantial effect on the market impact model, as trades within a metaorder mostly have the same sign (i.e. a large buy program usually does not contain any sell orders) and most market impact models in the literature do not include a bid-ask spread (see detailed discussion in Alfonsi and Schied 2010 regarding a two-sided versus one-sided LOB model), it remains largely unclear/undiscussed whether the extra price trajectory information should benefit the estimation of market impact models.

In this paper, using the Fisher information as a measure of asymptotic efficiency in statistical estimations, we investigate in a principled manner whether incorporating extra price trajectory data enhances the efficiency of estimating market impact. We compare the Fisher information matrix (FI) of different statistical experiments constructed from the same underlying stochastic process, and quantify the relative efficiency of their respective maximum likelihood estimators (MLE). The validity of this approach in assessing the optimality of experimental design is rooted in the asymptotic optimality of the MLE estimator in regular parametric models among the class of regular estimators (see, e.g. Van der Vaart 2000 or sections below). Specifically, among the popular existing market impact models, we compare different estimators by their respective Fisher information matrix

†To be specified in section 2.
and observe when one information matrix would dominate another, as this implies an asymptotically smaller variance for estimating any quantity of interest under that parametric model, e.g. the impact of metaorder or the cost of execution. To ensure the broad applicability of our findings, we separately investigate two broad categories of market impact models: the Almgren-Chriss model and the propagator models, which cover a large portion of the parametric models that are currently adopted or studied. Whether the price trajectory data could increase the efficiency of estimation is directly related to whether the current statistic is sufficient. Perhaps surprisingly (or even puzzlingly), we observe that, even when one does not have access to the full price trajectory data, it does not take many price points at all to achieve a more efficient estimation than well-established (also highly intuitive) methods, e.g. VWAP-based estimation method. For example, we show that in the Almgren-Chriss model, even substituting the realized impact data (the terminology from Almgren et al. (2005), equivalent to the cost of VWAP) with just two price points, one in a sufficiently early stage (within the first quarter of trade duration) of the order and the other one at the end of the order, we would get a strictly more asymptotically efficient estimation. One possible, intuitive explanation could be the concavity of the market impact function (see, e.g. Zarinelli et al. 2015), where two carefully chosen points can leverage information more efficiently than VWAP. Similar results can also be observed for the family of propagator models, except more price trajectories are typically more useful. We will explore our findings in detail in sections 3 and 4.

1.1. Related works and motivation

1.1.1. Market impact modeling. As one of the central themes in quantitative finance and market microstructure literature, the modeling of market impact is of great interest to practitioners and academic researchers. There is a wide range of literature on market impact modeling for which we only give a partial review here (for a more comprehensive review, see e.g. Gatheral and Schied 2013, Zarinelli et al. 2015). As one of the best-known and widely adopted models of market impact, the Almgren-Chriss model in the influential papers of Almgren and Chriss (2001b) and Almgren and Chriss (2001a) quantifies two distinctive components of price impact: a temporary impact induced by and affecting only the ongoing trade, and a permanent impact affecting all current and future trades equally. Under the Almgren-Chriss model, Almgren (2003) extended the analysis in Grinold and Kahn (2000) and Loeb (1983) and solved the optimal execution problem by framing it as a mean-variance optimization between the expected execution cost and its variance (representing the uncertainty/liquidity risk during execution). The optimal execution problem was also investigated, from an optimal control perspective, by Bertsimas and Lo (1998), Forsyth et al. (2012), Forsyth (2011), and Gatheral and Schied (2011) under the geometric Brownian motion assumption for the unaffected stock price process, rather than the arithmetic Brownian motion (ABM) assumption in the Almgren-Chriss model. The discrete and continuous time variants of these models (Bank and Baum 2004, Brunnermeier and Pedersen 2005, Almgren and Lorenz 2007, Carlin et al. 2007, Cetin et al. 2010) are collectively termed by Curato et al. (2017) as ‘first-generation’ market impact models, as opposed to the ‘second-generation’ models (Bouchaud et al. 2003, 2006, Alfonsi et al. 2010, Gatheral 2010, Gatheral et al. 2012, Obizhaeva and Wang 2013).

The ‘second-generation’ models, among which the propagator model is perhaps the most notable representative, postulate that the price impact should be neither permanent nor temporary, but transient, as it decays over time (Moro et al. 2009, Lehalle and Dang 2010). As one of the pioneering ‘second-generation’ models, Obizhaeva and Wang (2013) proposed a model with linear transient price impact and exponential decay, by modeling an exponential resilience of LOB. The model based on dynamics of the LOB was further developed by Alfonsi et al. (2010), Alfonsi et al. (2008), Gatheral et al. (2012), Curato et al. (2017), Almgren and Stoikov (2008), Bayraktar and Ludkovski (2014), Cont et al. (2014), and Guéant et al. (2012), which include non-linear price impact, as well as LOB with general shape function and time-dependent resilience. On the other hand, instead of modeling the dynamics of LOB, the discrete-time and continuous-time propagator models developed by Bouchaud et al. (2003) and Gatheral (2010) directly model the dynamics of price, using decay kernels to reflect the transient nature of the market impact. Detailed discussions on the connection and comparison between these two approaches, as well as further generalizations of propagator models, can be found in Bacry et al. (2015), Gatheral et al. (2012), Alfonsi and Schied (2010), Donier et al. (2015), Tóth et al. (2011), and Curato et al. (2017). Finally, aside from the aforementioned approaches focusing on the interactions between large trades and price changes, other approaches from alternative perspectives also provided many valuable insights on the price impact dynamics. For example, Cont et al. (2014) investigate how price changes are driven by order flow imbalance in the order book events (e.g. quote events); Cardaliaguet and Lehalle (2018) also investigates the optimal execution problem in a mean-field game setting, where the trader strategically deals with the uncertainty in price and behavior of other market participants.

1.1.2. Stochastic process and asymptotic statistical inference. In this section, we briefly review some of the basic concepts from asymptotic statistical inference in the context of a continuous-time stochastic process. To see their connections with the market impact literature, note that it is typical to first assume the the underlying process is driven by a continuous-time stochastic differential equations (SDE). For example, in Gatheral et al. (2012) and Almgren and Chriss (2001a), the ‘unaffected’ stock price follows an arithmetic Brownian motion (ABM) (The term unaffected price process, as used in Gatheral et al. (2012), refers to the price as determined by market participants excluding the executing trader, highlighting diffusion without the drift component) while Bertsimas and Lo (1998) adopts a geometric Brownian motion (GBM) price dynamics. As mentioned in Almgren and Chriss (2001a), while it could be beneficial to consider geometric Brownian motion (GBM) or correlation in price
movements for longer investment horizons, ABM remains a suitable model for the unaffected stock price over the short-term horizon, and the improvement gained by incorporating short-term serial correlation in price movement is also small (Almgren and Chriss 2001a). In fact, it was also investigated in Gatheral and Schied (2011) that, for the Almgren-Chriss model, the cost-risk efficient frontier under GBM and ABM assumption is ‘almost identical’.

In fact, as spelled out in Merton and Samuelson (1992), most theoretical models in finance use a continuous-time diffusion process (or general continuous-time Markov process) driven by SDE. However, as documented early in Grenander (1950), the extensive literature on stochastic process ‘rarely touched upon’ questions of inference. Over the years, some notable exceptions include Jacod and Protter (2011), Kutoyants (2013), Liptser and Shiryaev (2013), Liptser and Shiryaev (2014), and Bishwal (2007). In Kutoyants (2013) and Bishwal (2007), the maximum likelihood estimator (MLE) and its asymptotic property are investigated based on the likelihood function (the Radon-Nikodym derivative) of the realized continuous sample path or certain path integral. In addition, Kutoyants (2013) discussed inference based on other estimation methods such as minimum distance estimation, trajectory fitting estimation or method of moments, although Kutoyants (2013) note that these methods are ‘generally not asymptotically efficient’ compared to MLE. In Jacod and Protter (2011), such inference is conducted via central limit theorems for functional related to the SDE. Finally, in the comprehensive monographs of Liptser and Shiryaev (2013, 2014), parameter estimation and testing of statistical hypothesis are discussed, for both discrete and continuous diffusion-type processes, where the concept of sufficient statistic is closely related to our methods. In practice, the observed data can only possibly be discrete (up to a resolution beyond which point the price dynamics can no longer be appropriately modeled as SDE). Thus, the parameter estimation and inference problem for discretely, or sometimes non-synchronously/randomly observed diffusion processes (e.g., high-frequency trading) are of much more practical interest. These empirical estimation methods typically require discrete observations (e.g., see Alt-Sahalia and Mykland 2004, Alt-Sahalia 2008).

Based on discrete observations of a diffusion processes, a specific parameter estimation problem regarding the drift coefficients has been studied by many authors (see Bishwal 2007, Liptser and Shiryaev 2013 for a review). Due to the Markovian nature of diffusion processes, one can readily calculate the log-likelihood function of discretely sampled observations simply as the sum of successive pairs of log-transition function and the maximum likelihood estimation (MLE) can be a natural choice for such cases. Although a large portion of such studies indeed utilized (quasi-) likelihood-based estimation/inference (see Bibby et al. 2009 for review), the key difficulty is that the transition densities themselves from one time point to another generally do not have analytic forms, except in some special cases. In order to implement the efficient MLE-based methods, many attempts have been made to approximate the likelihood function, which leads to approximate maximum likelihood estimators (AMLE). Notably, the ground-breaking series of works in Alt-Sahalia (2020), Alt-Sahalia and Mykland (2004), and Alt-Sahalia (2008) proposed asymptotic expansions of the transition density for the diffusion process which could be used for approximation. Following this line of work by Alt-Sahalia, various subsequent analyses and noteworthy applications have been developed (see Chang and Chen 2011, Li 2013 for a review), as well as other numerical, simulation-based approaches to approximate likelihood (see, e.g., Pedersen 1995).

Fortunately, in market impact models, the ABM assumption for the price process allows one to evade the technical difficulty of transition density approximation for likelihood-based estimation of impact (i.e., drift term) because the joint distribution of discrete price observations reduces to a multivariate Gaussian. Within this canonical statistical model, a wide range of basic results in asymptotic inference, especially estimation for regular parametric models, are readily available. Precise definitions of regular experiments or regular parametric models can be found in Bickel et al. (1993), which we also specify in the appendix. As termed in Bickel et al. (1993), a regular parametric statistical experiment has a ‘nice’ parameter space in \( \Theta \subseteq \mathbb{R}^k \) and a density ‘smooth’ in \( \theta \). Most importantly, a regular statistical experiment possesses a non-singular Fisher information matrix at each point \( \theta \in \Theta \). As we shall see, the relationship between MLE and Fisher information matrix plays a crucial role in our arguments of asymptotic optimality.

1.1.3. Asymptotic optimality and experimental design. MLE is a ubiquitous method in statistical inference that has many desirable properties in terms of efficiency, feasibility, and generality. In fact, it has been argued that MLE attains asymptotically optimality among the classes of regular estimators (precise definitions on regular estimators can be seen in Van der Vaart (2000), which we specify in the appendix for completeness. Intuitively, a regular estimator admits a considerable regularity where a small change in parameters does not change the distribution of the estimator too much.). For example, for results stated below), in this regime, the local asymptotic normality (LAN) and Lipschitzness of log-likelihood can be used to establish the \( \sqrt{n} \) - convergence of the MLE estimator to the true parameter under a Gaussian distribution with the inverse of Fisher information matrix as its covariance matrix. This limiting property attained by the MLE, as shown in the Hájek-LeCam convolution theorem and its variant, is the ‘best’ limiting distribution asymptotically for any regular estimator, in the sense that, it is (1) locally asymptotically minimax for any bowl-shaped loss function, i.e. non-negative function with level sets convex and symmetric around the origin, (2) achieves the lowest possible variance (i.e. a quadratic form based on the inverse of the Fisher information matrix) for any asymptotically regular sequence of the estimator and (3) any improvement over this limit distribution can only be made on a Lebesgue null set of parameters. The asymptotic efficiency of MLE has also been discussed in the sense of Bahadur’s asymptotic efficiency or C.R.Rao’s efficiency (see Ibragimov and Khasminskii 2013). A better-known result, regarded as a simpler version of the Hájek - LeCam convolution theorem, is the Cramér-Fréchet-Rao information lower bound, which also establishes the asymptotic variance lower
bound as the inverse of Fisher information under unbiasedness. Notice the various prerequisites one must declare before one claims the asymptotic optimality of MLE. This is not a mere technicality because, aside from the fact the optimality criterion is not singular in nature, various counter-examples exist outside the confines of such conditions. For example, it is well-known that James-Stein’s shrinkage estimator (Lehmann and Casella 2006) achieves strictly smaller risk for estimating any mean of a $K \geq 3$-dimension multivariate Gaussian with identity covariance matrix under quadratic loss when compared to the MLE (i.e. sample mean). However, the James-Stein estimator is not regular and the improvement over MLE here is for finite sample scenarios, not asymptotic ones. For a counter-example with asymptotic improvement on a Lebesgue null set, check the famous Hodges’ estimator (Van der Vaart 2000).

The discussion above does not aim to debate whether one should necessarily use MLE for the estimation of market impact models. Rather, one can make the observation that, the asymptotic variance attained by MLE, being the ‘best’ (or ‘lowest’) possible as the inverse of the Fisher information matrix, quantifies an upper limit on how efficiently one can learn the parameter from a given statistical experiment. As a result, one naturally questions whether one can, by designing statistical experiments that maximize Fisher information in some sense (more about it below), reduces uncertainty in parameter estimation. Indeed, this line of work is pursued extensively in experimental design literature, where the Fisher information matrix has been used to measure the amount of information gained and to design optimal experiments. For example, recently Durant et al. (2021) uses Fisher information to optimize the experimental design in neutron reflectometry. In this paper, we investigate whether the statistical experiments based on price trajectory are more efficient than the ones based on certain summary statistics. In the experimental design literature (see, e.g. Whittle 1973, Chaloner and Verdinelli 1995, Wolkenhauer et al. 2008, Fedorov 2010), a unifying, single optimality-criteria for designing experiments has been studied. Traditional methods include maximizing expected trace, minimal eigenvalue, or determinant of Fisher information, corresponding to so-called A-optimality, E-optimality, or D-optimality. However, we note that in the case of estimating multiple parameters, there is an inherent difficulty in estimating all of them accurately, as an optimal way to estimate one particular parameter may not be optimal for the other ones, especially in the context of market impact models where the scales of parameters (or their variances) are vastly different. As a consequence, the meaning of the traditional criteria becomes less clear, unless the Fisher information from one experiment strictly dominates the other (i.e. their difference matrix is positive semi-definite), which we focus on showing in this paper.

1.2. Roadmap and outline

In this section we clarify the roadmap and main contributions of this paper. As mentioned in the previous paragraph, although we discuss and quantify the asymptotic efficiency of MLE (based on which numerical simulation is conducted), the main discourse is concentrated on comparing the Fisher information of different statistical experiments or designs. The MLE serves as natural candidate for numerical verification and its asymptotic optimality is quantified by the Fisher information; this also justifies the comparison among statistical experiments based on dominance of Fisher information. To this end, we propose an efficient method—that outperforms those based on common summary statistics—for sampling price trajectory data from the underlying continuous-time stochastic process assumed in the market impact models, where efficiency is measured by the Fisher information. The sampling scheme and the analysis revolving around it do not rely on any discretization of the continuous process, although we do consider, for theoretical purpose, discretized price trajectory with fixed increments approaching 0 (similar to the canonical model in Aït-Sahalia (2020), with equal spaced discretization). In particular, in the discussion about maximizing the Fisher information, we borrow insight from the concept of sufficiency in an idealized discretization of price trajectory. The flexibility regarding sampling scheme is an important practical consideration because, although experiment design with samples along a single trajectory approaching infinity could theoretically improve the Fisher information, in reality, the price trajectory at a scale finer than a certain threshold would not behave as a continuous-time diffusion process and one would practically want to design robust experiments based on fewer representative trajectory datapoints (e.g. 3 or 4 price points). We focus our discussion in this spirit.

The remainder of the paper is organized as follows. In section 2, we present the basic market impact estimation framework given price trajectory data, including technical lemmas about conditions. In section 3, we investigate two popular market impact models: Almgren-Chriss and the family of propagator models. The main result for the Almgren-Chriss model, connecting asymptotic efficiency and sampling of three trajectory points, is presented in theorem 3.3. The main result for propagator models is presented in theorem 4.1. In sharp contrast with the Almgren-Chriss result, it states that the only sufficient statistic is the full trajectory data when considering general instantaneous and kernel functions. Section 3 also shows a numerical study comparing different sampling strategies against VWAP-based estimation methods where the importance of early price data is explored. The last section concludes with discussions on some limitations of the theorems, specifically on both model misspecification and model selection. Simulation and empirical results are placed within each section. Reviews of basic concepts, proofs, and technical conditions are left in the appendix.

2. Basic setup and framework

2.1. Background and model

Throughout this paper, we assume the metaorder is a buy program so that we do not need to specify the sign of a trade. The case for a sell program can be derived analogously. We first focus our discussion on the VWAP (Volume Weighted Average Price) execution strategy, where the trading rate $\dot{s}_t = v$ is
constant in volume time, where, as a standard assumption in market impact literature, time units are measured by traded volume or volume time instead of physical time. Typically volume time is scaled to adjust for different levels of trading activity during the day, but for this paper, we do not actively distinguish the two times. Consequently, a VWAP execution strategy aims to trade equally/evenly in volume time, which implies trading at a constant proportion against the current traded volume of the stock. In this setting, VWAP strategies can be identified as strategies with a constant trading rate. See Almgren et al. (2005) and Gatheral and Schied (2013). Note that this is not a restriction on the order types, since we are considering the estimation rather than the optimal execution problem. In particular, different execution strategies can be approximated by sequences of interval VWAP strategies with different trading rates (see Remark 3.1 of Curato et al. 2017) and a provably reliable estimation procedure for VWAP execution provides insight for non-VWAP orders as well. Thus, the discussion for non-VWAP strategies shall be deferred to section 3.3.

We consider a continuous-time model for the evolution of the underlying stock price during the execution of a metaorder during $0 \leq t \leq T$:

$$S_t = S_0 + \mu_\theta(t,v) + \sigma \int_0^t dW_s,$$  \hspace{1cm} (1)

where $t$ represents time. The trade duration $T$ and trading rate $v$ are given beforehand, i.e. the SDE in (1) is conditional on $(v,T)$. For a buy program, the drift term $\mu_\theta(t,v)$ in (1) is initialized with $\mu_\theta(0,0) = 0$ and is typically a concave function in $t$ (Lillo et al. 2003, Nadtochiy 2022) (various forms of $\mu_\theta(t,v)$ will be discussed, see also Almgren et al. 2005, Gatheral 2010) representing the price impact generated from trading at rate $v$ for a period of $t$. Moreover, $\{\mu_\theta\}_{\theta \in \Theta}$ is a family of impact functions parameterized by $\theta \in \Theta \subseteq \mathbb{R}^K$ as the parameter space. Given fixed $(v,T)$ and $\theta \in \Theta$, $\{S_t\}_{t \in \mathbb{T}}$ is defined on some filtered probability space $(\Omega, \{\mathcal{F}_t\}_{t \in \mathbb{T}}, \mathbb{P})$ where $\mathbb{W}$ is a standard Brownian motion, $\sigma$ is the volatility and $\mathbb{T}$ is the time span (usually $\mathbb{T} = [0,T]$ for some $T$). As it is typically the case for market impact models (and in practice), we only aim to calibrate function $\mu_\theta(t,v)$ while we assume $\sigma$ is fixed or estimated separately. An execution strategy is represented by a continuous function of time $\{x_t\}_{t \in \mathbb{T}}$ (here we write $x_t$ instead of $X_t$ but in general $x_t$ can also be random and adaptive w.r.t $\mathcal{F}_t$) with $x_0 = 0$ and $x_T = X$ indicating the units of shares bought by time $t$, where $X$ is the metaorder size and $T$ is order duration measured by volume time. In general, impact function $\mu$ depends on the entire trajectory of trading rate $\{x_t\}_{t \in \mathbb{T}}$, but for VWAP strategy with constant $x_t = v$, we can simplify this dependence into only one variable $v$. As in Zarinelli et al. (2015), we characterize two metaorder features (or hyper-parameters): participation rate $\nu$ and duration $T$. In particular, if $V_D$ is the daily traded volume and $V_M$ is the volume traded by the whole market during the order execution period, then we define

$$T = V_M/V_D, \hspace{1cm} v = X/V_M,$$  \hspace{1cm} (2)

where both quantities are unitless and are in $[0,1]$. Notice the definition in (2) would suggest that $T \cdot v = X/V_D$, making the order size $X$ effectively scaled by a factor of $1/V_D$. For ease of notation, henceforth we assume $V_D = 1$ and treat $X$ as has been scaled. Moreover, as we do not actively distinguish between physical time and volume time so that we retain $X = T v$. Below are our two canonical classes models from (1).

**Example 1** As a first example, in the Almgren-Chriss model (Almgren et al. 2005), price dynamics follow as:

$$S_t = S_0 + \int_0^t (g(v) t + h(v)) + S_0 \sigma g(v) dt,$$

where, in view of (1), one can see $\mu(v,t) = \int_0^t g(v) dt$ and $\sigma = \sigma_0$ scaled by $S_0$. As we shall discuss in detail, $g$ represents the permanent impact and $h$ the temporary impact. As $S_0$ can be viewed as fixed and execution cost is typically expressed in basis points (bps), an affine transformation of $P_t = S_t/S_0$ is carried out in Almgren et al. (2005) which reduces the dynamics of $P_t$ to the canonical form (1) with $\mu(t,v) = g(v) t + h(v)$ and volatility $\sigma$.

**Example 2** A second example under the framework of (1), is the continuous-time propagator model (see, e.g. Bouchaud et al. 2003, 2009, Gatheral et al. 2012, Curato et al. 2017) developed to quantitatively reflect the transient nature of price impact:

$$S_t = S_0 + \int_0^t f(\xi_s) G(t-s) ds + \sigma \int_0^t dW_s,$$  \hspace{1cm} (3)

where $f(\cdot)$ is referred to as the instantaneous market impact function and $G(\cdot)$ as the decay kernel (Gatheral 2010). For a VWAP strategy, (3) reduces to (1) with $\mu(t,v) = f(v) \int_0^t G(s) ds$. We shall discuss (3) in detail in section 3.2.

**2.2. Statistical experiment and asymptotic inference**

In the context of this paper, rather than a statistical model, we speak of a statistical experiment constructed from discrete observations or statistics sampled from (1), as we cannot observe the whole path under the SDE. We briefly review some basic concepts in the statistical experiment and asymptotic inference, with the technical proofs and additional related material provided in the appendix. We first lay out some assumptions.

**Assumption 2.1** (Model Specification and Identifiability) $\Theta$ is an open subset of $\mathbb{R}^K$ and there exists a unique $\theta^* \in \Theta$ as the true parameter of the SDE in (1) for all $(v,T)$ almost surely (a.s.).

**Assumption 2.2** (Metaorder Characteristics) The distribution of $(v,T)$ in metaorder data does not depend on $\theta$ and follows some exogenous distribution function $G_{\text{order}}$. We write it as $(v,T) \sim G_{\text{order}}$ with pdf $(v,T) \sim g_{\text{order}}(v,T)dvdt$, i.e. $G_{\text{order}}(dv,dt) = g_{\text{order}}(v,T) dv dt$. We further assume $(v,T) \sim G_{\text{order}}$ satisfies

$$v_L \leq v \leq v_U \hspace{1cm} \text{and} \hspace{1cm} T_L \leq T \leq T_U$$
for some constants $0 < v_1 \leq v_U < \infty$ and $0 < T_L \leq T_U < \infty$ a.s.

**Remark 1** We make a couple of remarks on the assumptions above. We first note that assumption 2.1 states (1) is well-specified within the parametric family. The discussion for model misspecification is deferred to section 5. The uniqueness of $\theta^*$ implicitly requires certain identifiability conditions within the parametric family. This is relatively easy to satisfy when the support of $G_{\text{order}}$ is not too narrow, because, as long as one parametrizes $\mu_0$ carefully, it would then be difficult to find $\theta_1$ and $\theta_2$ such that $\mu_0(t, v) = \mu_0(t, v)$ for all $0 \leq t \leq T$ and all $(v, T)$ pair almost surely. We shall also see later that such “sufficient variability” condition on $G_{\text{order}}$ is crucial in our discussion regarding the nonsingular Fisher information matrix. Such condition on $G_{\text{order}}$ is also not restrictive as the metaorder data typically contains various execution styles and order characteristics reflecting the demands or specifications of clients, resulting in a broad range of values for $(v, T)$ in practice (see the figures on the empirical distribution of $(v, T)$ from real metaorder data in Zarinelli et al. 2015). For assumption 2.2, the lower and upper bounds on $(v, T)$ are reasonable, as one typically does not trade too slowly or too fast. On the other hand, assumption 2.2 of the exogenosity of $G_{\text{order}}$ may not be as straightforward. Indeed, although the distribution $G_{\text{order}}$ undeniably depends on many exogenous factors such as requests of clients or trading styles, it is not immediately clear whether one can claim $G_{\text{order}}$ has no dependence on $\theta$. For example, it is conceivable that the traders would, over time, estimate the parameter up to some accuracy and adapt their trading strategy accordingly for all the exogenous requirements, to satisfy certain optimal trading schedules (e.g. Almgren 2003), resulting in a “shift” of $G_{\text{order}}$ towards their acquired knowledge of $\theta$. In this paper, we assume such dependence is negligible, but we do note that assumption 2.2 could be a source of bias and should be a subject of future deliberation.

### 2.2.1. Statistical experiment & regularity conditions.

We are now ready to discuss the experiment design derived from (1). In this paper, we primarily consider statistical experiments consisting of discrete observations or summary statistics of the following form: given $N, M \in \mathbb{N}^+$, let $S = \{S_i\}_{i \in [N]}$ and $J = \{\int_{t_i}^{t_f} S_i \, dt\}_{i \in [M]}$, the statistical experiment on $\mathbb{R}^{2+N+M}$ has the triplet,

$$
(X, \mathcal{B}^{2+N+M}, \{\mathbb{P}^S_{\theta,J} \}_{\theta \in \Theta})
$$

where $X = \{(v, T) \cup S \cup J\}$. Here $S$ are $N$ price trajectory data (i.e. observations from a single trajectory) with $t_i \equiv t_i/T$ fixed; $J$ are $M$ summary statistics proportional to the average cost (or price) along a certain time window with $t_i^j \equiv t_i^j/T$. Such selection of $0 \leq t \leq T$ based on fixed ratio $\tau$ avoids inconsistencies in choosing $t$ when $T$ is random. Here $\mathcal{B}^{2+N+M}$ is the Borel-sigma algebra on $\mathbb{R}^{2+N+M}$ and $\mathbb{P}^S_{\theta,J}$ is the product measure of $G_{\text{order}}$ and the probability measure induced from the SDE in (1) conditional on $(v, T)$ when restricted to $S \cup J$. For ease of notation, we omit the upper script in $\mathbb{P}_{\theta}^{S_{\cup J}}$ and simply write $\mathbb{P}_{\theta}$ (and density function $p_{\theta}$ as well) when there is no ambiguity about the sample space in question. One can consider a sample $X$ under (4) generated as: first sample $(v, T) \sim G_{\text{order}}$, then generate a sample path based on SDE (1), and finally record $S$ and $J$.

Based on (1), it is straightforward to see that, conditional on $(v, T)$, the sample $S \cup J$ follows a $N + M$-dimensional multivariate Gaussian distribution

$$
\mathcal{N}(\mu(\theta, T, v), \Sigma(T)),
$$

with a mean function $\mu(\theta, T, v)$ given by

$$
\mathbb{E}S_i = \mu_0(t, v) \text{ and } \mathbb{E}\int_t^{t_f} S_i \, dt = \int_t^{t_f} \mu_0(t, v) \, dt,
$$

and a covariance matrix $\Sigma(T)$ given by

$$
\text{Cov}[(S_i, S_j)] = \sigma^2 \text{Cov}[W_{t_i}, W_{t_j}],
$$

$$
\text{Cov}\left[\left(\int_{t_i}^{t_f} S_i \, dt, \int_{t_i}^{t_f} S_j \, dt\right)\right] = \sigma^2 \text{Cov}\left[\left(\int_{t_i}^{t_f} W_i \, dt, \int_{t_i}^{t_f} W_j \, dt\right)\right],
$$

all of which can be readily computed using elementary Itô calculus on standard Brownian motion (e.g. $\text{Cov}[W_{t_i}, W_{t_j}] = \min(s, t)$). A notable consequences is that $\Sigma(T)$ has no dependence on $\theta$ or $v$.

We assume there is no linear dependence among elements in $S \cup J$ so that $\Sigma(T) > 0$ and $\Sigma^{-1}(T)$ exists. Thus, the likelihood function of a sample from (4) can be written as

$$
p_{\theta}(X) = \frac{g_{\text{order}}(v, T)}{\sqrt{2\pi|\Sigma(T)|}} \exp\left(-\frac{(\mu(\theta, T, v) - (S, J))^T \Sigma^{-1}(T)}{2}\right),
$$

and log-likelihood

$$
\ell(\theta|X) = \log p_{\theta}(X).
$$

Let $X_1, X_2, \ldots, X_n$ be $n$ i.i.d. samples from experiment (4). Let $\hat{\theta}$ be the maximum likelihood estimator such that

$$
\hat{\theta} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \ell(\theta|X_i).
$$

**Example 3** Typically one has price trajectory data $S = \{S_i\}_{i \in [N]}$ and $J = \{\int_{t_i}^{t_f} S_i \, dt\}_{i \in [M]}$ from metaorders, in the forms of (Almgren et al. 2005, Zarinelli et al. 2015):
(i) The cost of execution: \( \int S \, \mathrm{d}t - X S_0 \) (divide by \( X S_0 \) for bps),
(ii) The peak impact \( S_T \),
(iii) The ‘permanent’ impact \( S_{T_{\text{pos}}} \) for \( T_{\text{pos}} > T \) (e.g., 30 mins after trade Almgren et al. 2005),
(iv) The quantification of price trajectory for \( \mu_t(t, v) \) for some \( t < T \).

For example, conditional \((v, T)\), the cost of VWAP execution \( C_{\text{VWAP}}(v) \) is given as in (5)
\[
C_{\text{VWAP}} \sim N(\mu_{\text{VWAP}}(\theta, T, v), \sigma_{\text{VWAP}}^2(T, v)),
\]
where the mean \( \mu_{\text{VWAP}}(\theta, T, v) = V \int_0^T \mu_0(t, v) \, \mathrm{d}t \) and variance \( \sigma_{\text{VWAP}}^2(T, v) = \int_0^T \sigma^2(T, v) \, \mathrm{d}t \) (derivation left in appendix). Another example, given a single sample corresponding to price trajectory data \( X = (v, T) \cup [S_{ij}]_{i \in [N]} \), the log-likelihood follows (derived using independent increments of \( S_j \)):
\[
\ell(X | \theta) \equiv \log(p_\theta(X)) = \log g_{\text{order}}(v, T) - \frac{N}{2} \log(2\pi \sigma^2) + \sum_{t \in [N]} \frac{1}{2} \log(\tau_T - \tau_{t-1}) - \frac{1}{2} \tau^2(\tau_{t-1})T \\
\quad \times \left( (S_{\tau_T} - S_{\tau_{t-1}}) - (\mu_0(\tau_T, v) - \mu_0(\tau_{t-1}, v)) \right)^2.
\]

Then, given \( n \) total orders samples \( X_j \), executed by VWAP strategy under \((T_j, v_j)\), with \([S_{ij}]_{i \in [N]} \) being the \( j \)-th price trajectory data \( S_j = [S_{i,j}]_{i \in [N]} \), the MLE is the estimator that maximizes the log-likelihood ratio:
\[
\hat{\theta}_n = \arg \max_{\theta \in \Theta} \sum_{j=1}^n \ell(X_j | \theta)
\]
\[
= \arg \min_{\theta \in \Theta} \sum_{j=1}^M \sum_{i \in [N]} \frac{1}{t_j - t_{i-1}} \\
\quad \times \left( (S_{t_j} - S_{t_{i-1}}) - (\mu_0(t_j, v_j) - \mu_0(t_{i-1}, v_j)) \right)^2,
\]
where the second equality follows from (10).

**Assumption 2.3 (Gaussian Experiment)** For the statistical experiment in (4), we assume
(i) \( P_\theta(\ell|X) < \infty, \forall \theta \in \Theta \).
(ii) For the multivariate Gaussian in (5), there exists some \( \epsilon > 0 \) such that \( \Sigma(T) \geq \sigma^2 I \forall T \) a.s.
(iii) For the multivariate Gaussian in (5), \( \mu(\theta, T, v) \) is continuously differentiable in \( \theta \forall (v, T) \) a.s.
(iv) For the multivariate Gaussian in (5), there exists a neighborhood around every \( \theta \) such that \( \forall \theta' \) in this neighborhood, \( |\mu(\theta', T, v) - \mu(\theta, T, v)| \leq L|\theta' - \theta|^{\alpha} \) for some \( B, L, \forall \theta, \alpha \in \text{this neighborhood and } V(v, T) \) a.s.
(v) For the multivariate Gaussian in (5), \( \theta^* \) is the unique true parameter.

**Remark 2** We make a couple of remarks on assumption 2.3. Assumption (i) is standard. Assumption (ii) is also standard and is related to the design of the Gaussian experiment, where one must choose \( S \cup J \) so that they are not linearly dependent (no observation can be linearly replicated by others, otherwise one can simply delete this redundancy). This condition requires that \( \Sigma(T) > 0 \). The uniform lower bound \( \epsilon \) on its eigenvalue for all \( T \), as we shall see, hinges on the lower and upper bound of \( T \) in assumption 2.2. For assumption (iii), one simply can check whether \( \mu_0(t, v) \) in (1) is continuously differentiable in \( (v, T) \) within the bounded support of \( G_{\text{order}} \). Assumption (iv) ensures Lipschitz continuity of \( \mu_0(t, v) \) in \( (v, T) \) within the bounded support of \( G_{\text{order}} \). The identifiability condition in Assumption (v) is not quite the same as assumption 2.1 because \( P_\theta \) is only a projection of the probability measure induced from SDE (1) onto a finite-dimensional space (4). However, this condition is needed to ensure the parameters of interest can be recovered by finite dimensional observations. Intuitively, the higher the dimension for \( \theta \), the higher the dimension of \( X \) we need for the model identifiability in (4).

**Definition 1** With assumption 3.3 in place, we write the derivative of \( \mu(\theta, T, v) \) w.r.t \( \theta \) as a Jacobian matrix
\[
\mathcal{J}(\theta, T, v) = \begin{bmatrix}
\frac{\partial \mu_1}{\partial \theta_1} & \frac{\partial \mu_1}{\partial \theta_2} & \cdots & \frac{\partial \mu_1}{\partial \theta_k} \\
\frac{\partial \mu_2}{\partial \theta_1} & \frac{\partial \mu_2}{\partial \theta_2} & \cdots & \frac{\partial \mu_2}{\partial \theta_k} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial \mu_{M}}{\partial \theta_1} & \frac{\partial \mu_{M}}{\partial \theta_2} & \cdots & \frac{\partial \mu_{M}}{\partial \theta_k} 
\end{bmatrix}.
\]

Consequently, given \( \theta_0 \in \Theta \), we define the Fisher information matrix in \( \mathbb{R}^{k \times k} \) for experiment (4):
\[
\mathcal{I}_X(\theta_0) = E_{\theta_0} \left[ \left( \frac{\partial \ell(\theta|X)}{\partial \theta} \right)^T \right],
\]
where the subscript in \( E_{\theta_0} \) denotes expectation under \( P_{\theta_0} \) and the subscript in \( \mathcal{I}_X \) denotes the experiment design in (4) based on sample \( X \).

Before we discuss the relevance of the Fisher information and its importance, we first provide a known result which allows convenient computation for it in a Gaussian experiment (5).

**Lemma 2.4** Let \( X \sim N(\alpha(\theta), \Sigma) \) be the \( N \)-dimensional multivariate Gaussian distribution with known \( \Sigma \) where \( \theta \in \mathbb{R}^k \). Then, let \( D(\theta) \in \mathbb{R}^{N \times k} \) be Jacobian matrix where
\[
D(\theta) = \begin{bmatrix}
\frac{\partial \mu_1}{\partial \theta_1} & \frac{\partial \mu_1}{\partial \theta_2} & \cdots & \frac{\partial \mu_1}{\partial \theta_k} \\
\frac{\partial \mu_2}{\partial \theta_1} & \frac{\partial \mu_2}{\partial \theta_2} & \cdots & \frac{\partial \mu_2}{\partial \theta_k} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial \mu_M}{\partial \theta_1} & \frac{\partial \mu_M}{\partial \theta_2} & \cdots & \frac{\partial \mu_M}{\partial \theta_k} 
\end{bmatrix},
\]
we have \( \mathcal{I}_X(\theta) = D(\theta)^T \Sigma^{-1} D(\theta) \).

**Proof** See example 7.7 in Van der Vaart (2000).

Based on lemma 2.4, we have the following Proposition.

**Proposition 2.5** Under assumption 2.2 and 2.3, the statistical experiment (4) has its Fisher information in (13).
as
\[ I_X(\theta) = E_{(v,T)} G_{\text{order}} \left[ J^T(\theta, T, v) \Sigma^{-1}(T) J(\theta, T, v) \right]. \]  
(14)

**Proof** Due to assumption 2.2, we can infer from (6) that
\[
l(\theta|X) = \log g_{\text{order}}(v, T) + \log p_0(X|(v, T))
\]
\[
= \log g_{\text{order}}(v, T) + \log p_0(S \cup J|(v, T))
\]
Since \( g_{\text{order}}(v, T) \) does not depend on \( \theta \) by assumption 2.2 and \( S \cup J \) is multivariate Gaussian conditional on \( (v, T) \), we can evoke lemma 2.4 and the tower property of conditional expectation to write
\[
I_X(\theta) = E_{(v,T)} G_{\text{order}} \left[ E \left[ \frac{\partial l(\theta|X)}{\partial \theta} \left( \frac{\partial l(\theta|X)}{\partial \theta} \right)^T |(v, T) \right] \right]
\]
\[
= E_{(v,T)} G_{\text{order}} \left[ I_{(v,T)}(\theta) \right]
\]
\[
= E_{(v,T)} G_{\text{order}} \left[ J^T(\theta, T, v) \Sigma^{-1}(T) J(\theta, T, v) \right]
\]
where the \( I_X(\theta) \) is computed from conditional likelihood, treating \( (v, T) \) as fixed. □

Lastly, we present our final assumption.

**Assumption 2.6** (Nonsingular Fisher Information) We assume the Fisher information \( I_X(\theta) \) in (14) is well-defined, non-singular and continuous in \( \theta \).

As we shall see, assumption 2.6 is related to \( G_{\text{order}} \). Simply put, it would be hard to determine a high-dimensional \( \theta \) in the market impact model if trading style \( (v, T) \) is too singular, as one might struggle to separate different impact effects from \( v \) or \( T \). One would need to check assumption 2.6 on a case-by-case basis. For example, a point mass distribution of \( G_{\text{order}} \) (i.e. one pair of \( (v, T) \)) is almost never capable of fitting a good market impact model.

### 2.2.2. Asymptotic optimality and Fisher information.

Now we present some consequences of assumptions 2.1–2.6. The results discussed here are related to basic concepts in asymptotic inference, such as *regular* parametric model, local asymptotic normality (LAN), and *regular* estimators. We do not dive into the details of these established results and we defer both the proof and the reference for these concepts to the appendix. The purpose of the following proposition is to establish that, under fairly reasonable criteria and a fairly broad class of estimators, the MLE achieves asymptotic optimality given the aforementioned conditions and that optimality is quantified by the Fisher information.

**Proposition 2.7** Let \( X_1, X_2, \ldots, X_n \) be \( n \) i.i.d. samples from experiment (4). Let \( \hat{\theta}_n \) be the maximum likelihood estimator such that \( \hat{\theta}_n = \arg \max_{\theta \in \Theta} \sum_{i=1}^n l(\theta|X_i) \). Then, under assumptions 2.1–2.6, the MLE is consistent for \( \theta^* \), i.e. \( \hat{\theta}_n \to \theta^* \) in probability and \( \sqrt{n}\)-asymptotically normal:
\[
\sqrt{n}(\hat{\theta}_n - \theta^*) \overset{d}{\to} N\left(0, \left.I_X^{-1}(\theta^*)\right)\right)
\]  
(15)
for any \( \phi(\theta) \) differentiable in \( \theta \). Moreover, for any bowl-shaped loss function \( l \) (i.e. \( x : l(x) \leq c \) is convex and symmetric around the origin) and any estimator sequence \( \{T_n\}_{n \geq 0} \),
\[
\sup_{\theta \in \Theta} \lim_{n \to \infty} \inf_{h \to 0} E_{\theta + \frac{h}{\sqrt{n}}} \left[ \left( \sqrt{n}(T_n - \phi(\theta^* + h + \frac{h}{\sqrt{n}})) \right) \right] 
\]
\[
\geq l(\theta^*)
\]  
(16)
where \( l \) is taken over all finite subset of \( \mathbb{R}^k \) and \( Z_\theta \sim N(0, \nabla \phi(\theta^*)^T I_X^{-1}(\theta^*) \nabla \phi(\theta^*)) \).

The proof is presented in the appendix. The argument follows from section 8.7 of Van der Vaart (2000) where the asymptotic optimality of the MLE and its Fisher information is argued from the local asymptotic minimax perspective. Many other optimality criteria exist and are discussed in appendix as well. The main takeaway is that, if (15) is taken as the basis for asymptotic optimality, one should aim to design an experiment that maximizes the Fisher information. To see why, consider two statistical experiments \( \chi \) based on different designs \( X \) and \( Y \) (e.g. price trajectory versus total cost) and one wants to predict \( c_{\text{vwap}}(\theta) \) = \( \int c_{\text{vwap}}(\theta^*, T, v) G_{\text{order}}(dv, dT) \) based on \( \hat{\theta}_n \) or \( \hat{\theta}_n \).

If \( c_{\text{vwap}}(\theta) \) is differentiable w.r.t \( \theta \) with gradient \( \nabla \phi c_{\text{vwap}} \), one can show that, based on (15) and the delta method (Van der Vaart 2000), the asymptotic variance of the plug-in estimator satisfies
\[
\sqrt{n}(c_{\text{vwap}}(\hat{\theta}_n) - c_{\text{vwap}}(\theta^*)) \overset{d}{\to} N(0, \nabla \phi c_{\text{vwap}}(\theta^*)^T I_X^{-1}(\theta^*) \nabla \phi c_{\text{vwap}}(\theta^*))
\]
and
\[
\nabla \phi c_{\text{vwap}}(\theta^*)^T I_X^{-1}(\theta^*) \nabla \phi c_{\text{vwap}}(\theta^*)
\]
\[
\leq \nabla \phi c_{\text{vwap}}(\theta^*)^T I_Y^{-1}(\theta^*) \nabla \phi c_{\text{vwap}}(\theta^*)
\]
(17)
which implies \( \hat{\theta}_n \) has lower asymptotic variance for estimating \( c_{\text{vwap}}(\theta^*) \), or any other differentiable function of \( \theta \), simply based on its greater Fisher information.

### 2.2.3. Sufficiency and other lemmas.

In this section, we introduce several technical lemmas we shall use later. Following the previous discussion, one might want to maximize the Fisher information of an experiment (4). An important related concept is sufficient statistics. Formally, given \( X \) in some statistical experiment (4) and a function \( \phi \), one can create another statistical experiment using the statistics \( \phi(X) \). This typically incurs a loss of information in terms of the Fisher information unless the statistic is sufficient. Here, a statistic \( \phi(\cdot) \) is sufficient for \( \{P_\theta\}_{\theta \in \Theta} \) if the conditional distribution of \( S \) given \( \phi(S) \) is free of \( \theta \), under \( P_\theta \) for any \( \theta \in \Theta \). We also give a formal characterization here which we shall use later.

**Lemma 2.8** (Neyman-Fisher Factorization Theorem) Consider the statistical experiment (4) and its sample \( X \). A
statistic $\phi(X)$ is sufficient if and only if the likelihood function in (4) has the factorization $p(X|\theta) = h(X)g(\phi(X), \theta)$, for some non-negative $h(\cdot), g(\cdot)$.

The following lemmas relate a sufficient statistic with the Fisher information.

**Lemma 2.9 (Data Processing Inequality Zamir 1998)** Consider the statistical experiment (4) and its sample $X$. Let $\phi(X)$ be a statistic of data, then the statistical experiment based on $\phi(X)$ satisfies, for $\theta \in \Theta$,

$$I_X(\theta) \geq I_{\phi(X)}(\theta),$$

with the equality obtained if the $\phi(X)$ is a sufficient statistic for the original statistical experiment.

**Lemma 2.10 (Data Refinement Inequality)** Consider the statistical experiment (4) and its sample $X$. Let $U$ and $V$ be two statistics of the experiments, respective; then the statistical experiment based on them satisfies, for $\theta \in \Theta$, that $I_U(\theta) \geq I_V(\theta)$. Moreover, if $V = \phi(U)$ for some $\phi(\cdot)$ which is a bijective mapping, then $I_U(\theta) = I_V(\theta)$.

The proofs for lemmas 2.9 and 2.10 are in the appendix for completeness. In other words, the Fisher information of the experiment using $\phi(X)$ is generally smaller than the one using $X$, unless statistic $\phi(X)$ is sufficient, in which case $X$ provides no extra information over $\phi(X)$ for estimating $\theta$ (thus the equality in (18)). Note the original sample $X$ is always, trivially, a sufficient statistic. However, the whole data $X$ is not always required.

**Example 4** Consider a simple model under (1) which is a random walk with drift expressed by $S_t = S_0 + \theta g(v) t + \sigma W_t$ for some $f(\cdot)$. Let $X = (v, T) \cup S$ be the statistical experiment in (4) where $S = \{S_t\}_{t \in [1]}$. It is straightforward to check that the log-likelihood in (10) is given by

$$l(\theta|X) = \log g_{\text{order}}(v, T) - \frac{1}{2\sigma^2} \sum_{i=1}^{N} \left( (S_i - S_{i-1}) - \theta f(\cdot)(t_i - t_{i-1}) \right)^2$$

which can be reduced to $h(S, T, v) + g(S_T - S_0, T, v, \theta)$ for some $h, g$ (i.e. isolate the term with $\theta$). Using theorem 2.8, we can see that $(S_T - S_0, v, T)$ is sufficient. Thus, for this model, price trajectory $S$ is not needed for the estimation of $\theta$, the difference between the end price and start price would suffice (in fact $S_T$ would suffice since we assume $S_0$ is known).

In the next two sections, we examine sampling schemes from price trajectory in two popular market impact models: namely the Almgren Chriss Model and the Propagator Model. In particular, given lemma 2.9, which establishes the connection between sufficient statistics and the Fisher information, we are able to find some insight in designing efficient statistical experiments.

3. The Almgren-Chriss model

The Almgren-Chriss model remains one of the most popular and widely adopted market impact models since its introduction (Almgren and Chriss 2001a) and the paper (Almgren et al. 2005) is dedicated to parameter estimation problems for the model:

$$S_t = S_0 + S_0 g(v) T + S_0 \int_0^T dW_t, \quad \text{when} \ t \leq T$$

In Almgren-Chriss model, $g$ is the permanent impact function and $h$ is the temporary impact function. In Almgren et al. (2005), $h$ and $g$ are parameterized by power law:

$$g(v; \theta) = \gamma v^\alpha \quad \text{and} \quad h(v; \theta) = \eta v^\beta,$$

for $\theta = (\gamma, \eta, \alpha, \beta)$. Although $\alpha$ can generally vary between $[0, 1]$, it was argued in Almgren et al. (2005) and Huberman and Stanzl (2004) that the linearity of permanent impact function $g$ (i.e. $\alpha = 1$) is desirable for a model to preclude price manipulation strategy. Moreover, $\beta = 0.5$ corresponds to a plausible square-root temporary impact function for $h$. However, during the fitting process, these are not taken as restrictions. There are several practical modifications in the model estimation section of Almgren et al. (2005):

(i) In order to facilitate the estimation of the permanent impact in the Almgren-Chriss model, the price trajectory $S$ includes an additional price point $S_{T_{\text{post}}}$ where $T_{T_{\text{post}}} > T$; this can also be seen from the ‘piece-wise’ dynamic in (19). Almgren et al. (2005) indicated that the choice of $T_{T_{\text{post}}}$ being 30 min after $T$ works well. In this way, in the experiment design (4), we can fix a constant $t_{\text{delay}}$ so that $T_{T_{\text{post}}} = T + t_{\text{delay}}$, for $t_{\text{delay}} = 0.077$ (i.e. 30 min delay in a day with 6.5 trading hours). In such manner, we remove the randomness of choosing $T_{T_{\text{post}}}$ given $T$, similar to the way we use fixed $\tau_i$ to choose $t_i = \tau_i T$.

(ii) The impact functions $g$ and $h$ are further scaled by volatility $\sigma$. For a single stock, this can be subsumed in the coefficient $\gamma, \eta$ in (20).

(iii) A liquidity factor in the form of shares outstanding with power law exponent is inserted to characterize the fact that the market impact is not just based on $(T, v)$, but also on the stock’s liquidity condition. For a single stock, this can also be subsumed in $\gamma, \eta$ in (20).

We will again discuss these modifications in detail in the simulation section where we reproduce and compare with the simulation results in Almgren et al. (2005). In summary, these modifications facilitate a cross-sectional description of market impact, although we focus on the single stock analysis in this paper. Finally, for our discussion on (19), aside from assumptions in section 2.2, we do not put restrictions on the exact form of $h(\cdot; \theta), g(\cdot; \theta)$, i.e. we allow them to take in forms other than the power law parametric form (20). With
the affine transformation $P_t = \frac{S_t - S_0}{S_0}$, Almgren et al. (2005) estimated $\theta$ using two quantities termed as \textit{permanent impact} $I$ and \textit{realized impact} $J$:

$$I = \frac{S_{T_{\text{post}} - S_0}}{S_0} = P_{T_{\text{post}}}, \quad \text{and}$$

$$J = \frac{\int_0^T S_t dt / T - S_0}{S_0} = \frac{1}{T} \int_0^T P_t dt,$$

(21)

where, under the model (19), conditional on $(v, T)$ the joint distribution of $I$ and $J$ follow a Gaussian one (5):

$$\left( \begin{array}{c} I \\ J - I/2 \end{array} \right) \sim N\left( \begin{array}{c} \mu_{I,J}(\theta, T, v), \\ J(T) \end{array} \right),$$

where $\mu_{I,J}(\theta, T, v) = \left( \begin{array}{c} Tg(v; \theta) \\ h(\theta; v) \end{array} \right)$, (22)

$$\Sigma_{I,J}(T) \triangleq \sigma^2 \left[ \begin{array}{ll} \frac{T_{\text{post}}}{\mu_{\text{post}}} - T & -\frac{T_{\text{post}} - T}{4} \\ -\frac{T_{\text{post}} - T}{4} & \frac{T_{\text{post}}}{4} - \frac{T}{6} \end{array} \right].$$

A partial derivative of the above can be found in Almgren et al. (2005), but we also include a full derivation in the appendix for completeness. Then, in the parameter estimating phase (section 4.2 of Almgren et al. 2005), Almgren et al. (2005) uses the metaorder data to fit the normalized residuals of $(I, J - I/2)$ using Gauss-Newton optimization algorithm, since this is a non-linear least square problem for the $\mu$ given by (20). This fits exactly in our frame of (4), as the maximum likelihood estimation (8) for the Gaussian experiment based on $I, J$ in (5) is equivalent to solving the following non-linear least squares problem:

$$\max_{\theta \in \Theta} \sum_{i=1}^{n} l(\theta | I_i, J_i)$$

$$\Leftrightarrow \min_{\theta \in \Theta} \sum_{i=1}^{n} \left( \left[ I_i - I_i/2 \right] - \mu_{I,J}(\theta, T, v_i) \right)^T$$

$$\times \Sigma_{I,J}^{-1}(T) \left[ I_i - I_i/2 \right] - \mu_{I,J}(\theta, T, v_i) \right),$$

(23)

based on $n$ total orders samples $(I_i, J_i)_{i\in[n]}$, executed under $(T_i, v_i)_{i\in[n]}$. This equivalence allows us to connect the parameter estimation method in Almgren et al. (2005) with the asymptotic theory of statistical estimation in section 2.2. It follows from lemmas 2.10, 2.4 and proposition 2.5, if the Jacobian of $\mu_{I,J}(\theta, T, v) \ w.r.t \ \theta$ is $J_{I,J}(\theta, T, v)$, then the Fisher information for experiments based on $I, J$ is

$$J_{I,J}(\theta) = J_{I,J-1/2}(\theta)$$

$$= \text{Var}_{(v, T) \sim G_{\text{order}}} \left[ J_{I,J}^T(\theta, T, v) \cdot \Sigma_{I,J}^{-1}(T) \cdot J_{I,J}(\theta, T, v) \right]$$

$$= \int_{(v, T) \sim G_{\text{order}}} \left[ J_{I,J}^T(\theta, T, v) \cdot \sigma^{-2} \right.$$

$$\left. \begin{bmatrix} \frac{T_{\text{post}}}{\mu_{\text{post}}} - T/6 \\ \frac{T_{\text{post}}}{\mu_{\text{post}}} - T/4 \\ \frac{T_{\text{post}}}{\mu_{\text{post}}} - T/2 \\ \frac{T_{\text{post}}}{T_{\text{post}} - 3} - \frac{T_{\text{post}}}{T_{\text{post}} - 4} \end{bmatrix} J_{I,J}(\theta, T, v) \right).$$

(24)

\textbf{Example 5} For fitting the power law model (20) in Almgren-Chriss model with $\theta = (\gamma, \eta, \alpha, \beta)$, one can calculate

$$J_{I,J}(\theta, T, v) = \begin{bmatrix} T^\mu & 0 & T^\nu v^\sigma \ln(v) & 0 \\ 0 & v^\sigma & 0 & \eta^\nu \ln(v) \end{bmatrix},$$

which means that, based on (24), $J_{I,J}(\theta) \in \mathbb{R}^{n \times 4}$ does not have full rank (hence singular) if $(v, T) \sim \delta_{(v_0, T_0)}$ (i.e. a point mass). In such case, the optimization in (23) could also be under-determined. However, if one only tries to determine the exponent of the power law, then $\theta = (\alpha, \beta)$ and

$$J_{I,J}(\theta, T, v) = \begin{bmatrix} T^\nu v^\sigma \ln(v) & 0 \\ 0 & \eta^\nu \ln(v) \end{bmatrix}$$

for some given $(v_0, \eta_0)$. In summary, the non-singularity of $J_{I,J}(\theta)$ largely depends on the variability of $(v, T) \sim G_{\text{order}}$ for multi-dimensional $\theta$.

In a case-by-case fashion, we prove the following in the appendix, so that our results from previous sections can be applied.

\textbf{Lemma 3.1 (Regularity Check) Consider the statistical experiment (4) based on Almgren-Chriss model with power law (20) impact and $I, J$ in (22). Under assumptions 2.1, 2.2, 2.3 and 2.6 are satisfied provided that the marginal distribution of $G_{\text{order}}$ on $v$ has a support with infinite cardinality.}

\textbf{Lemma 3.1} provides a sufficient condition for assumptions 3 (3.5 in particular) and 4 to hold, although it is not necessary. Note that the condition on $G_{\text{order}}$ is satisfied for any distribution with a continuous density. By checking the proof, one can also establish the result for discrete distribution $G_{\text{order}}$ with carefully chosen support with cardinality larger than the dimension of $\theta$.

\textbf{3.1. Sufficient statistic from a discrete price trajectory}

From lemma 3.1, in view of section 2.2 and the regularity check of experiment based on (20), we want to see if we can design experiments with Fisher information larger than $J_{I,J}$. We first investigate this for the canonical price trajectory data of the form $S_{\text{full}} = \{ S_t \}_{t\in[N]}$ on a given grid \{3, 2, 2, 1\} on a grid \{$\Delta t, 2\Delta t, \ldots, N\Delta t = T, T_{\text{post}}$\} with $\Delta t = T/N$ and $N$ is a fixed number. Notably, this is also the setup for Almgren and Chriss (2001a), when they consider the optimal execution problem with $N$-trading period.

In this example, again taking the transform $P_t = \frac{S_t - S_0}{S_0}$, from (19), we can see that the likelihood for the experiment based on $S_{\text{full}}$, conditional on $(v, T)$, under Almgren-Chriss model is:

$$p_{\theta}(S_{\text{full}} | v, T) \propto \exp \left( -\frac{(P_{\text{harm}} - g(v; \theta) \Delta t - h(v; \theta))^2}{2\sigma^2 \Delta t} \right)$$

$$- \frac{N \sum_{i=0}^N (P_i - P_{i-1} - g(v; \theta) \Delta t)^2}{2\sigma^2 \Delta t}$$

$$- \frac{(P_{\text{post}} - P_T + h(v; \theta))^2}{2\sigma^2 (T_{\text{post}} - T)}$$

$$\times f_1(S_{\text{full}}, T, v)f_2 \left( g(v; \theta) P_T + h(v; \theta) \right)$$
for some $f_1, f_2$. Using lemma 2.8, we can see that $(P_{t_1}, P_T, P_{T_{post}})$ is a sufficient statistic for $\theta$. In fact, $(P_{t_1}, P_T, P_{T_{post}})$ is sufficient, but not minimal sufficient, as $(P_T, \Delta t - \frac{P_{T_{post}} - P_T}{\Delta t}, \theta)$, a non-invertible mapping of $(P_{t_1}, P_T, P_{T_{post}})$, is also sufficient. For the definition of minimal sufficient, see e.g. Keener (2010) or theorem 4.1 below. However, the observation $\Delta t - \frac{P_{T_{post}} - P_T}{\Delta t}$ is somewhat artificial to this example. Consequently, from lemma 2.9 we have:

$$I_{S|N}(\theta) = I_{P_{t_1}, P_T, P_{T_{post}}}(\theta).$$

(26)

Although the construction of $t_1$ is artificial and related to the choice of $N$. However, since the choice of $N$ can be arbitrarily large, it is reasonable to suspect that the insight from (26) can be translated to generic price trajectory data $S$. Moreover, since the post-order price $I = P_{T_{post}}$, we can actually then substitute the VWAP cost $J$ for two price points $P_t$ (for some $t$ in ‘early’ stages) and $P_T$ along the price trajectory data for a more asymptotically efficient estimate? Indeed, in the next section, we shall see these insights in several important ways.

### 3.2. Sampling strategy along the price trajectory

Based on the previous insight, given an experiment design based on partial trajectory data $S = \{P_t\}_{t \in [N]} \cup \{P_{T_{post}}\}$ with $t_N = T$, one might ask:

(i) Does sampling more than three points along the trajectory (i.e. $N > 2$) increase the Fisher information?

(ii) If sampling more than three points does not improve the asymptotic efficiency, is simply sampling two points $P_T$ and $P_{T_{post}}$ good enough?

(iii) For sampling three points, in addition to $P_T$ and $P_{T_{post}}$, how should we pick the extra point $P_t$ for some $t \in (0, T)$?

To answer these questions, suppose we have the following grid of observations

$$S = \{P_{t_1}, P_{t_2}, \ldots, P_{t_{N-1}}, P_T, P_{T_{post}}\}.$$

Then writing out the density $p_0(S|v, T)$ as in (25), we have

$$p_0(S|v, T) \propto f_1(S)f_2\left[\left(\frac{P_{t_1}}{t_1} - \frac{P_{T_{post}} - P_T}{T_{post-T}}\right), \theta\right]$$

(27)

for some $f_1, f_2$. Following the same reasoning as in the previous section, we see that

$$I_S(\theta) = I_{P_{t_1}, P_T, P_{T_{post}}}(\theta),$$

where $t_1 = \min\{t : P_t \in S\}$. Thus, we have the following corollary.

COROLLARY 3.2 For the Almgren-Chirss model and experiment based on $S = \{P_t\}_{t \in [N]} \cup \{P_{T_{post}}\}$, sampling more than three points on the trajectory does not increase the Fisher information, as:

$$I_S(\theta) = I_{P_{t_1}, P_T, P_{T_{post}}}(\theta).$$

(28)

This answers the first question. To answer the second question, one can calculate the (derivation is left in the appendix):

$$I_{P_{t_1}, P_T, P_{T_{post}}}(\theta) = \mathbb{E}_{(v, T)}G_{\hat{V}_v} \left[\frac{J_{IJ}^T(\theta, T, v)}{(J_{IJ}^T(\theta, T, v))^\gamma} \cdot \sigma^{-2} \right]$$

$$ \times \left[\left(\frac{7}{T_{post-T}}\right)^{\frac{7}{2}} \frac{1}{T_{post-T}} \right].$$

(29)

Comparing (29) with (24), we note that, in general, neither $I_{P_{t_1}, P_T, P_{T_{post}}}(\theta) \succ I_{IJ}(\theta)$ nor $I_{IJ}(\theta) \succ I_{P_{t_1}, P_T, P_{T_{post}}}(\theta)$ can be established. Thus, one cannot definitively say sampling two points is more efficient than using VWAP cost $J$ and post-order price $I$.

So far, we have shown that sampling two price points $P_T$ and $P_{T_{post}}$ is not optimal while sampling more than one point along price trajectory $\{P_t\}_{t \in (0, T)}$ is not necessary. This naturally brings us to the third question above. We summarize the answer as our first main theorem.

THEOREM 3.3 Under assumptions 2.1–2.6, we sample $(P_{t_1}, P_T, P_{T_{post}})$, as long as $t$ is chosen early enough so that $\frac{1}{T} \leq \frac{1}{4}$, then, under the Almgren-Chriss model (19), we have

$$I_{P_{t_1}, P_T, P_{T_{post}}}(\theta) \succ I_{IJ}(\theta).$$

(30)

Moreover, (30) holds for generic forms of $g$ and $h$ and all distributions $G_{\hat{V}_v}$ of $(v, T)$.

Proof The proof of theorem 3.3 is left in the appendix A.3.3.

REMARK 3 Notably, also shown in the proof, the ratio $\frac{1}{4}$ in theorem 3.3 is somewhat tight, in the sense that for $\frac{1}{T} > \frac{1}{4}$, there might exist some form of $h$ and $g$ for which (30) is no longer true. This could be of interest in its own right, as the significance of number $\frac{1}{4}$ is not immediately clear from observing (19).

Theorem 3.3 suggests that, at least for the Almgren-Chriss dynamics (19), as long as we sample $P_T$ early enough (e.g. any price sampled within the first 25 percent of the filled order), then MLE estimated using $(P_T, P_T, P_{T_{post}})$ would outperform the estimation method using $I, J$ in Almgren et al. (2005) asymptotically, for any distributions of $(v, T)$, rendering theorem 3.3 applicable for metaorders across different magnitudes and for generic form of $h$ and $g$ (specifically the original power law models in Almgren et al. 2005). Perhaps interestingly, this result seems to suggest earlier stages of an order are more informative for calibrating impact models. Finally, we use the data based on sections 4.1, 4.2 and table 3 of Almgren et al. (2005) to provide an example of the level of improvement based on theorem 3.3. More extensive simulation studies verifying theorem 3.3 are provided at the end of the section.
Example 6 Consider estimating the power law exponent of (20) in (19). Suppose the underlying true parameters are $\theta^* = ( \gamma^*, \eta^*, \alpha^*, \beta^*) = (0.314, 0.142, 0.891, 0.600)$. We set the hyper-parameters of the metamodels to be $(X, v, t, T_{\text{post}}, \sigma) = (0.1, 0.5, 0.2, 0.275, 0.0157)$. The data is from Almgren et al. (2005), which can be viewed as one specific instance from $G_{\text{order}}$, i.e. a point mass. We set $t = 0.1 \cdot T$ and $(\gamma^*, \eta^*) = (0.314, 0.142)$, so we only estimate $(\alpha, \beta)$. This avoids non-singular issues from $G_{\text{order}}$ being a point mass. Then, one can calculate $I_{P_{\text{f}}, P_{\text{r}}, P_{\text{post}}}(\alpha^*, \beta^*)$ and $I_{P_{\text{f}}, \alpha^*, \beta^*}$, which leads to $I_{P_{\text{f}}, P_{\text{r}}, P_{\text{post}}}(\alpha^*, \beta^*) = 10^4 \left[ \begin{smallmatrix} 3.504 \times 10^6 & -2.259 \times 10^7 \\ -2.259 \times 10^7 & 1.845 \times 10^9 \end{smallmatrix} \right]$. The closed-form expression of $I_{P_{\text{f}}, P_{\text{r}}, P_{\text{post}}}(\theta)$ is calculated in appendix A.3.3. In view of (15) and (17), using $(P_t, P_f, P_{\text{post}})$ over $(I, J)$ implies an asymptotic variance reduction (i.e. equivalent to relative efficiency, Van der Vaart 2000) of 21% for estimating $\alpha$ (calculated by $4.438 - 3.504$, or equivalently a 21% increase in sample-efficiency), 51% for estimating $\beta$, and 18.5% for estimating the average VWAP cost in bps, given by $c_{\text{VWAP}}(\alpha, \beta) = \frac{\gamma^* \eta^*}{\alpha^*} + \eta^* \beta^*$. The $\gamma^* \eta^* \log(\alpha), \eta^* \beta^* \log(\beta)$. On the other hand, also following remark 5, given $(\alpha^*, \beta^*) = (0.891, 0.600)$, suppose ones wants to estimate $(\gamma, \eta)$. Similarly one can show a 20.6% variance reduction for estimating $\gamma$, 51.5% for $\eta$. In the context of Schied and Schöneborn (2009), the form of the (20) with $\alpha^* = \beta^* = 1$ and a utility $u(R) = -\exp(-AR)$ for some $A > 0$ is assumed (e.g. we let $A = 5$ below). Then the optimal adaptive liquidation strategy (notice this is a sell-program so $x_0 = X$) is shown to be of $x_t = X \cdot \exp(-\sqrt{\frac{2A}{3}T})$. One can check if $\eta$ is estimated around $\eta^*$ with a 10% error, then the optimal liquidation at $T$ would be off-target by around 6.88%, whereas a 5% estimated error would that error to around 3.36%, an improvement close to 50%.

3.3. Non-VWAP executions

In this section, we use the framework from Almgren and Chriss (2001) to briefly discuss the estimation of non-VWAP orders under (19). In particular, given full grid observation $S_{\text{full}} = \{S_t \mid t \in [N]\}$ and a sequence of trading rate $|v_t| \in [N]$ so that a trading rate of $v_t$ is executed during interval $(t_{j-1}, t_j)$ and $\frac{1}{\sqrt{N}} \sum_{i=1}^N v_t = X$. We assume trading rates and trading period are all fixed and we only investigate the efficiency issue here. Thus, under (19) and $P_t = \sqrt{\frac{3}{2N}}$, we have

$$P_h = P_{h-1} + g(v_t; \theta) \Delta t + h(v_t; \theta) - h(v_{t-1}; \theta) + \sigma \int_{t_{j-1}}^{t_j} dW_s.$$  \hspace{1cm} (31)

Note that $|v_t| \in [N]$ need not have $N$ distinct values (i.e. $|\{v_t\}| < N$) as the same trading rate is allowed to prevail over several periods of $\Delta t$. To account for such strategy, suppose there are $N'$ distinct ($N' \leq N$) constant trading rates $|v'_t| \in [N']$ with $v' \neq v'_t$ for all $t \neq t'_j$ and each $v'_t$ corresponds to some union of trading intervals where the trading rate is $v'_t$. We represent such a union by a sequence of disjoint, non-adjacent intervals $I_t = \cup_{k}(I_{t_k}^{\text{f}}, I_{t_k}^{\text{r}})$. For example, if $v'_t = 0.5$ and the order is executed with a trading rate of 0.5 during $(t_0, t_1)$, $(t_1, t_2)$ and $(t_2, t_3)$, then $I_t = (t_0, t_2) \cup (t_2, t_3)$. Note that the disjoint and non-adjacent property makes sure that 1) the representation of $I_t$ is unique and 2) for any $t$ being $I_{t_k}^{\text{r}}$ or $I_{t_k}^{\text{f}}$ in the representation of $I_t$, then

(i) either $t = 0$ or $t = T$

(ii) or the interval $(t - \Delta t, t]$ and $(t, t + \Delta t]$ has two different trading rates.

In other words, if we view the time before $t = 0$ and after $t = T$ as having 0 trading rate, then any $t$ being $I_{t_k}^{\text{f}}$ or $I_{t_k}^{\text{r}}$ in the representation of $I_t$ implies $t$ connects two periods with different trading rates. In this setting, the likelihood can then be written as

$$p_t(S_{\text{full}}; V, T) \propto \exp \left( - \sum_{i=1}^N \frac{P_t(h_v; \theta) \Delta t}{2\sigma^2}(P_{\text{final}} - P_T + h(v_t; \theta) - h(v_{t-1}; \theta))^2 \right)$$

$$= \exp \left( - \sum_{t \in [N]} \sum_{v_t \in [N']} \frac{((P_h - P_{h-1}) - (g(v_t; \theta) \Delta t + h(v_t; \theta) - h(v_{t-1}; \theta)))^2}{2\sigma^2 \Delta t} \right)$$

$$= \left( \prod_{i \in [N]} f(h(v_t; \theta)) \right) \cdot \left( \prod_{i \in [N']} \int_{I_{t_k}^{\text{f}}} \left( \sum_{i \in [N']} (P_{t_k}^{\text{r}} - P_{t_k}^{\text{f}}) g(v_t; \theta) \right) \right)$$

$$\cdot \left( \prod_{i \in [N]} \int_{I_{t_k}^{\text{f}}} \left( \sum_{i \in [N']} (P_{t_k}^{\text{r}} - P_{t_k}^{\text{f}}) g(v_t; \theta) \right) \right)$$

$$+ \left( \prod_{i \in [N]} \int_{I_{t_k}^{\text{f}}} \left( \sum_{i \in [N']} (P_{t_k}^{\text{r}} - P_{t_k}^{\text{f}} + \Delta t \frac{\Delta_t}{T_{\text{post}} - T} I_{v_t} \cdot h(v_t; \theta) \right) \cdot f_z(\theta). \right)$$

(32)
3.4. Simulation

In this section, we verify the results from corollary 3.2 and theorem 3.3. For a fair comparison, we recover, to the best of our abilities, the IBM stock simulation result in Table 3 from Almgren et al. (2005). Several practical modifications in Almgren et al. (2005) are carried out, so impact equations in (7) and (8) of Almgren et al. (2005) become

\[
\frac{I}{\sigma} = \gamma_T \cdot \text{sgn}(X_A) \left| \frac{X_A}{V_A} \right|^\alpha \left( \frac{\Theta_A}{V_A} \right)^{\delta_A} + \text{(noise)}
\]

and

\[
\frac{1}{\sigma} \left( J - \frac{I}{2} \right) = \eta_A \cdot \text{sgn}(X) \left| \frac{X_A}{V_A} \right|^\beta + \text{(noise)}
\]

whereas our equation based on (22) gives

\[
I = \gamma T \cdot \nu^\alpha + \text{(noise)} \quad \text{and} \quad J - \frac{I}{2} = \eta \cdot \nu^\beta + \text{(noise)}.
\]

Here \( \Theta_A \) is called shares outstanding, \( V_A \) is average daily volume and \( \frac{X_A}{V_A} \) is called inverse turnover. These concepts are originally included in Almgren et al. (2005) for cross-sectional market impact model using the liquidity factor, which is not necessary for a single stock analysis simulation here. However, for the clarity of comparison, the parameters from Table 3 in Almgren et al. (2005) in different notations from ours are subscripted by \( A \). Their transformation to our model notations can be cast as

\[
v = \left| \frac{X_A}{V_A} \right|, \quad \gamma = \sigma \gamma_A \left( \frac{\Theta_A}{V_A} \right)^{\delta_A}, \quad \eta = \sigma \eta_A.
\]

Following Table 3 of Almgren et al. (2005), the initial hyperparameters for IBM stocks are given:

\[
V = 6561000, \quad \Theta = 1728000000, \quad \sigma = 0.0157, \\
T_{\text{post}} - T = 0.5/6.5 = 0.077 \quad \text{and} \quad T = 0.2.
\]

With these transformations, we can reproduce values in Table 3 in Almgren for IBM using equation (33) and our version of the parameters, up to a negligible error (i.e. no discrepancy up to the forth digit). The results are summarized in table 1 (figure 1).

For simulation, we let \( X_A/V_A \) follow a uniform distribution on an interval from 0.05 to 0.15 with \( T = 0.2 \) fixed. This would imply our parameter \( v = \frac{X_A}{V_A} \) follows a uniform distribution from 0.25 to 0.75. This specifies \( G_{\text{order}} \). The ground truth parameters are set following table 1 as \( \alpha^* = 1.0, \beta^* = 0.6, \gamma^* = \sigma \gamma_A (\frac{\Theta_A}{V_A})^{\delta_A} = 0.314 \cdot 0.0157 \cdot 4.0285 = 0.01986, \quad \eta^* = \sigma \eta_A = 0.142 \cdot 0.0157 = 0.002229 \). In the simulation, we fit all four parameters \( \theta = (\gamma, \eta, \alpha, \beta) \) in (20). Each path is generated by the Euler–Maruyama method with \( \Delta = 0.001T \) and the stochastic integral in \( J \) is approximated by the discrete sum. We consider five experiments to verify the findings in the previous sections:

(i) Using \((I, J)\), referred to as ‘Almgren approach \((I, J)\)’
(ii) Using \((P_T, P_{T_{\text{post}}})\), referred to as ‘Two points \((T, T_{\text{post}})\)’

(iii) Using \((P_{0.1T}, P_{T}, P_{T_{\text{post}}})\), referred to as ‘Three points \((0.1T, T, T_{\text{post}})\)’, with \( \frac{T}{t} = 0.1 \leq 0.25 \).
(iv) Using \((P_{0.5T}, P_{T}, P_{T_{\text{post}}})\), referred to as ‘Three points \((0.5T, T, T_{\text{post}})\)’, with \( \frac{T}{t} = 0.5 \leq 0.25 \).
(v) Using \((P_{0.1T}, P_{0.5T}, P_{T}, P_{T_{\text{post}}})\), referred to as ‘Four points \((0.1T, 0.5T, T, T_{\text{post}})\)’.

The results are presented in table 2. For each method, we report the experiment results for metaorder size \( n = 500, n = 1000 \) and \( n = 10000 \). For each sample size \( n \), we repeat the experiment 1000 times to report (1) average estimates of each parameter denoted by ‘Avg estimate \( \hat{\theta} \)’, across 1000 simulations, (2) the theoretical standard error of each parameter, based on empirical Fisher information, denoted by ‘Theoretical SE(\( \hat{\theta} \))’ and (3) the empirical standard deviation of each parameter, based on bootstrap from 1000 simulations, denoted by ‘Empirical SE(\( \hat{\theta} \))’. The best method, for each \( n \), is recorded in bold. We can see, consistent with the results in section 3.2:

(i) The ‘Three points method’ with \( t/T \leq 0.25 \) outperform the Almgren approach, especially in terms of standard deviation, as indicated by theorem 3.3.
(ii) The ‘Two points method’ and ‘Three points method’ with \( t/T \geq 0.25 \) do not visibly outperform the Almgren approach and
(iii) The ‘Three points method’ perform almost exactly the same as the ‘Four points method’, as indicated by corollary 3.2.

Finally, the Theoretical SE(\( \hat{\theta} \)) provides a good estimate for Empirical SE(\( \hat{\theta} \)), especially for large \( n \), which can be useful for inference.

4. The propagator models

As discussed in Zarinelli et al. (2015), the linear permanent market impact predicted by the Almgren-Chrisk model can deviate from the market impact trajectory for non-VWAP executions and fail to reproduce the concavity of the market impact curve (Zarinelli et al. 2015). The propagator models (Bouchaud et al. 2003, Gatheral and Schied 2013, Obizhaeva and Wang 2013), on the other hand, can consistently recover...
Do price trajectory data increase the efficiency of market impact estimation?

Table 2. Simulation results: true parameter $(\alpha^*, \beta^*, \gamma^*, \eta^*) = (1.0000, 0.6000, 0.01986, 0.002229)$.

| Method                          | Avg estimate $\hat{\theta}$ (average over 1000 simulations) | Theoretical $SE(\hat{\theta})$ (avg. of hessian implied SD) | Empirical $SE(\hat{\theta})$ (SD of estimate over 1000 sim) |
|--------------------------------|-------------------------------------------------------------|-------------------------------------------------------------|-------------------------------------------------------------|
|                                | $\alpha$  | $\beta$  | $\gamma$  | $\eta$  | $\alpha$  | $\beta$  | $\gamma$  | $\eta$  | $\alpha$  | $\beta$  | $\gamma$  | $\eta$  | $\alpha$  | $\beta$  | $\gamma$  | $\eta$  |
| $n = 500$                      |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
| Almgren approach $(I, J)$      | 1.0641   | 0.6209   | 0.022222 | 0.002285 | 0.7777   | 0.3327   | 0.010810 | 0.000520 | 0.7825    | 0.3337   | 0.011847 | 0.000533 |
| Two points: $(T, T_{post})$    | 1.0694   | 0.6334   | 0.022755 | 0.002336 | 0.7752   | 0.4935   | 0.012370 | 0.000784 | 0.8197    | 0.4814   | 0.023695 | 0.000781 |
| Three points: $(0.1T, T, T_{post})$ | **1.0464** | **0.6161** | **0.021577** | **0.002267** | **0.6835** | **0.2282** | **0.009126** | **0.000355** | **0.6891** | **0.2235** | **0.010070** | **0.000357** |
| Three points: $(0.5T, T, T_{post})$ | 1.0609   | 0.6248   | 0.022008 | 0.002230 | 0.7449   | 0.4164   | 0.010123 | 0.000656 | 0.7420    | 0.4160   | 0.010897 | 0.000669 |
| Four Point: $(0.1T, 0.5T, T, T_{post})$ | **1.0464** | **0.6161** | **0.021577** | **0.002267** | **0.6835** | **0.2282** | **0.009126** | **0.000355** | **0.6891** | **0.2235** | **0.010070** | **0.000357** |
| $n = 1000$                     |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
| Almgren approach $(I, J)$      | 1.0294   | 0.6215   | 0.020860 | 0.002273 | 0.5240   | 0.2324   | 0.006782 | 0.000362 | 0.5120    | 0.2316   | 0.006791 | 0.000367 |
| Two points: $(T, T_{post})$    | 1.0286   | 0.6257   | 0.020847 | 0.002312 | 0.5198   | 0.3401   | 0.006719 | 0.000537 | 0.5095    | 0.3329   | 0.006757 | 0.000537 |
| Three points: $(0.1T, T, T_{post})$ | **1.0372** | **0.6021** | **0.020831** | **0.002245** | **0.4628** | **0.1594** | **0.005990** | **0.000247** | **0.4676** | **0.1561** | **0.006152** | **0.000245** |
| Three points: $(0.5T, T, T_{post})$ | 1.0249   | 0.6262   | 0.020774 | 0.002296 | 0.4995   | 0.2891   | 0.006450 | 0.000454 | 0.4948    | 0.2844   | 0.006514 | 0.000457 |
| Four Point: $(0.1T, 0.5T, T, T_{post})$ | **1.0372** | **0.6021** | **0.020831** | **0.002245** | **0.4628** | **0.1594** | **0.005990** | **0.000247** | **0.4676** | **0.1561** | **0.006152** | **0.000245** |
| $n = 10000$                    |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
| Almgren approach $(I, J)$      | 1.0111   | 0.5994   | 0.020039 | 0.002224 | 0.1604   | 0.0728   | 0.002020 | 0.000112 | 0.1634    | 0.0734   | 0.002035 | 0.000114 |
| Two points: $(T, T_{post})$    | 1.0110   | 0.6006   | 0.020037 | 0.002234 | 0.1601   | 0.1063   | 0.002016 | 0.000164 | 0.1633    | 0.1052   | 0.002033 | 0.000164 |
| Three points: $(0.1T, T, T_{post})$ | **1.0102** | **0.6000** | **0.020007** | **0.002230** | **0.1421** | **0.0503** | **0.001787** | **0.000078** | **0.1431** | **0.0505** | **0.001773** | **0.000076** |
| Three points: $(0.5T, T, T_{post})$ | 1.0117   | 0.5996   | 0.020034 | 0.002232 | 0.1540   | 0.0903   | 0.001939 | 0.000139 | 0.1581    | 0.0919   | 0.001952 | 0.000142 |
| Four Point: $(0.1T, 0.5T, T, T_{post})$ | **1.0102** | **0.6000** | **0.020007** | **0.002230** | **0.1421** | **0.0503** | **0.001787** | **0.000078** | **0.1431** | **0.0505** | **0.001773** | **0.000076** |
market impact. The word transient of impact. In the framework of (1), we have† empirical findings on market efficiency and strong concavity assumption is consistent with real data and is in line with et al. (2003) and Busseti and Lillo (2012), such a transient decay kernel G(t) is later extended by Alfonsi et al. (2020) for generally-shaped LOB. The connection/ equivalence of price impact reversion and volume impact reversion is further discussed in Alfonsi and Schied (2010), Gatheral et al. (2011), and Webster (2023). (iii) logarithmic \( f(v) \propto \log(v/v_0) \) and \( G(s) \propto l_0(l_0 + s)^{-\gamma} \) or \( G(s) \propto (l_0^2 + s^2)^{-\gamma/2} \). Here \( \gamma \approx \frac{1}{\alpha} \) is related to the exponent of auto-correlation among trade signs studied in Bouchaud et al. (2003). The logarithmic impact has also been supported empirically in Zarinelli et al. (2015) to fit the data across different magnitudes better over the square root function.

There are many other forms of \( f \) and \( G \) (e.g. Gaussian kernel, trigonometric kernel in Gatheral et al. 2012 as examples of where one can solve Fredholm equation Gatheral et al. 2012) and propagator models have been extensively studied (e.g.Bouchaud et al. 2003, 2006, Alfonsi and Schied 2010, Gatheral 2010, Gatheral et al. 2011, 2012, Curato et al. 2017) to characterize conditions of the model which prevent price manipulation strategy or guarantee the stability of trades (Gatheral 2010) in optimal execution strategy. The various forms of \( f \) and \( G \) are typically modeled separately. As a result, their regularity conditions in assumption 2.1–2.6 need to be checked in a case-by-case basis, although they are satisfied for common propagator models.

### 4.1. Sufficient statistics for a discrete price trajectory

In this section, we first fix \((v, T)\) by letting \(G_{\text{order}} = \delta_{v,T}\) be a point mass to simplify the discussion. The general case follows by averaging over \((v, T) \sim G_{\text{order}}\). We can do this since we assumed \(G_{\text{order}}\) has no dependence on \(\theta\). In the same spirit as section 3.1, we first study the experiment based on \(S_{\text{full}} = \)

---

† The equation (34) is typically referred to as the continuous-time propagator model or the Gatheral (JG) model. The discrete-time propagator was first proposed by Bouchaud et al. (2003).

‡ Here \(\alpha\) indicates in proportional to.
in this setting we no longer include a $S_{\text{full}}$ with $T_{\text{post}} > T$ to measure the permanent impact, i.e. the height of plateau that peak impact relaxes to, as its determination might be difficult (Zarinelli et al. 2015). Formally, we have the following characterization of separability before we can state the theorem based on $S_{\text{full}}$, from a parameter estimation point of view as in section 3.1.

**Definition 2** The form of $f$ and $G$ in (34) is separable if they are modeled by separate parameters, i.e. there exist parameter spaces $\Theta_1$ and $\Theta_2$ such that $\Theta = \Theta_1 \times \Theta_2$ and $f(\cdot; \theta) = f(\cdot; \Theta_1, \cdot; \Theta_2)$ for all $\theta \in \Theta$.

**Theorem 4.1** Fix $(v, T)$ and consider (5). With a full-grid observations $S_{\text{full}} = \{S_r\}_{r \in [N]}$, for power-law kernel $G(s) \propto s^{-\gamma}$, $G(s) \propto \log(l_0 + s)^{-\gamma}$ where $\gamma \in \Gamma \subseteq (0, 1)$ contains some open set or exponential kernel $G(s) \propto e^{-\rho s}$ where $\rho \in \Gamma \subseteq \mathbb{R}^+$ contains some open set, assume the instantaneous impact $f$ is positive (i.e. $f(v; \theta) \neq 0$ for $v > 0$) and is separable from from the kernel $G$, then the unique sufficient statistic (up to a one-to-one transformation) for propagator model (34) is the full trajectory data $S_{\text{full}}$.

**Proof** A sufficient statistic is minimally sufficient if it can be represented as a function of every other sufficient statistic, i.e. $X$ is minimally sufficient if for every other sufficient statistic $Y$, there exists some function $f$ such that $X = f(Y)$ (a.e. for all $\theta$). Thus, if we can show $S_{\text{full}}$ is minimally sufficient, then for any sufficient statistic $X$, we know both $X$ is a function of $S_{\text{full}}$ (by definition of a statistic) and $S_{\text{full}}$ is a function of $X$ (by minimal sufficiency of $S_{\text{full}}$), thus proving $S_{\text{full}}$ is the unique sufficient statistic, up to a one-to-one mapping. Under (1), we have $\mu(t; v; \theta) = f(v; \theta) \int_0^t G(s; \theta) \, ds$. Then we can write, as in (25):

$$p_0(S_{\text{full}}|v, T) = \exp \left( - \sum_{i=1}^N \frac{\left( S_i - S_{t_{i-1}} - (\mu(t, v; \theta) - \mu(t_{i-1}, v; \theta)) \right)^2}{2\sigma_i^2 A_i} \right)$$

$$= \exp \left( - \sum_{i=1}^N \frac{\left( S_i - S_{t_{i-1}} - f(v; \theta) \int_{t_{i-1}}^t G(s; \theta) \, ds \right)^2}{2\sigma_i^2 A_i} \right)$$

$$= f_1(S_{\text{full}}) \exp \left( \sum_{i=1}^N \left( S_i - S_{t_{i-1}} \right) - f_2(v; \theta) \int_{t_{i-1}}^t G(s; \theta) \, ds \right).$$

(35)

for some $f_1, f_2$. It is known that (see e.g. Lehmann and Casella 2006 or theorem 3.19 Keener 2010) that, for an exponential family with density

$$p_0(x) = \exp(\eta(\theta) \cdot T(x) - B(\theta))b(x),$$

(36)

where the $\eta(\theta), T(x) \in \mathbb{R}^k$ and $\eta(\theta) \cdot T(x)$ denote their dot product, the statistic $T(x) \in \mathbb{R}^k$ is minimal sufficient if (1) $\{T_i(x)\}_{i \in \mathbb{N}}$ is linearly independent and (2) the dimension of the convex hull of $\{\eta(\theta)\}_{\theta \in \Theta}$ has dimension $k$. The convex hull of a set $S$ is $\{\sum_{i=1}^k t_i x_i : t_i \geq 0, x_i \in S, \forall i, \sum_{i=1}^k t_i = 1, n > 0\}$. An equivalent condition for (2) is: there exists $\{\theta_0, \theta_1, \ldots, \theta_k\} \subseteq \Theta$ such that $\{\eta(\theta) - \eta(\theta_0)\}_{\theta \in \Theta}$ is linearly independent. Now, compare (35) with (36), we let $T(S_{\text{full}}) \in \mathbb{R}^N$ with $T_i(S_{\text{full}}) = S_i - S_{t_{i-1}}$ for $i \in [N]$ and let $\eta(\theta) \in \mathbb{R}^N$ with $\eta_i(\theta) = \frac{\partial}{\partial \theta_i} \int_0^t G(s; \theta) \, ds$ for $i \in [N]$. Clearly $\{T_i(S_{\text{full}})\}_{i \in [N]}$ are linearly independent, as there are no linear constraints among the price increments $S_i - S_{t_{i-1}}$. Thus (1) is satisfied. For (2), since $f(v; \theta) \neq 0$ and is separable from $G$, it suffices to show the convex hull of $\{\eta_i(\theta)\}_{i \in \Theta} \subseteq \mathbb{R}^N$ has dimension $N$ for the various kernel $G$ mentioned above, where we redefine $\eta_i(\theta) = \int_0^T G(s; \theta) \, ds$ using $G$ only. This is left to the appendix A.3.4. Finally, the minimal sufficiency of $T(S_{\text{full}}) = \{S_i - S_{t_{i-1}}\}_{i \in [N]}$ is equivalent to the minimal sufficiency of $S_{\text{full}}$ since $S_0$ is given.

Theorem 4.1 shows, when fitting a general propagator model, only the entire price trajectory provides the whole information, not any summary statistics (e.g. cost of VWAP, or the three points as in theorem 3.3 for Almgren-Chriss model). To see more clearly why this should be the case, we can contemplate a toy example.

**Example 7** Suppose we are estimating a 1-d parameter $\theta \in \Theta \subseteq \mathbb{R}$ from $(1)$ and $(v, T)$ is fixed. We can estimate $\theta$ using either the average cost of VWAP (up to a constant factor $v$):

$$J = \frac{1}{T} \int_0^T S_i \, dt \sim N\left( \frac{1}{T} \int_0^T \mu_0(t, v) \, dt, \frac{\sigma^2 T^3}{3} \right),$$

(37)

or the entire path $S_{\text{full}}$. Using lemma 2.4, the Fisher information $I_j(\theta)$ is $I_j(\theta) = \frac{1}{\sigma^2 T} (\int_0^T \frac{\partial^2 \mu(t, v)}{\partial \theta dt^2} dt)^2$, whereas, using lemma 2.4 and the independence of $\{S_i - S_{t_{i-1}}\}_{i \in [N]}$ (multivariate Gaussian), the Fisher information $\mathcal{I}_{S_{\text{full}}}(\theta)$ is (let $\Delta t \downarrow 0$ and assume boundary condition $\frac{\partial \mu(t, v)}{\partial \theta dt}(0) = 0$):

$$\lim_{\Delta t \to 0} \mathcal{I}_{S_{\text{full}}}(\theta) = \frac{1}{\sigma^2 T} \int_0^T \left( \frac{\partial^2 \mu(t, v)}{\partial \theta dt} \right)^2 dt.$$ 

(38)

Let $h(t; \theta, v) \triangleq \frac{\partial \mu(t, v)}{\partial \theta dt}$, $h'(t; \theta, v) \triangleq \frac{d h(t; \theta, v)}{dt}$ and assume $h'(T; \theta, v) = 0$. Then $\lim_{\Delta t \to 0} \mathcal{I}_{S_{\text{full}}}(\theta) = I_j(\theta)$ then immediately follows from the inequality: $(\int_0^T h(t; \theta, v) \, dt)^2 \leq (\int_0^T \int_0^s h'(s; \theta, v) \, ds \, dt)^2 = (\int_0^T \int_0^s h'(s; \theta, v) \, ds \, dt)^2 \leq \frac{T^3}{3} (\int_0^T h'(s; \theta, v) \, ds)^2$, where the second equality follows from the Fubini-Tonelli theorem and the last inequality follows from Cauchy–Schwarz inequality.

### 4.2. Estimation of impact function $f$

In the previous section, theorem 4.1 addresses the joint estimation of $f$ and $G$, but the proof also works for the estimation of $G$ given $f$. However, theorem 4.1 can be adjusted for estimation of impact function $f$ given $G$, which typically can be calibrated from other methods (Bouchaud et al. 2003, Busseti and Lillo 2012). As discussed in section 2.1 of Curato et al. (2017), one of the ‘major attractions of the propagator models to practitioners’ is that given $G(\cdot)$ and a large data collection of ‘VWAP-like executions’, the expected cost of
Theorem 4.2 In the same setting as theorem 4.1, suppose a decay kernel \( G(\cdot) \) is fixed and the impact function \( f(v; \theta) \) is parameterized by some \( \theta \in \Gamma \subseteq [0,1] \) where \( \Gamma \) contains some open set. Then the sufficient statistics for estimating \( f \) in (34) is \( \phi(S_{\text{full}}) = \sum_{i=1}^{N} (S_i - S_{i-1}) \int_{t_{i-1}}^{t_i} G(s) \, ds \), which (trivially) implies
\[
I_{S_{\text{full}}} = I_{\phi(S_{\text{full}})}.
\] (40)

In the limit as \( \Delta t \to 0 \), we have
\[
I_{S_{\text{full}}} \to \frac{\mu^2}{2\sigma^2} \int_0^T G^2(t) \, dt,
\] (41)
and
\[
I_{\Delta \text{tiny}} \to I_{\Delta I} = \frac{\mu^2}{2\sigma^2} \cdot \frac{3}{T^3} \left( \int_0^T G(t)(T-t) \, dt \right)^2.
\] (42)

Finally, we have
\[
I_{S_{\text{full}}} \to I_{\Delta I} = \frac{\mu^2}{2\sigma^2} \cdot \sum_{i=1}^{N} \left( \int_{t_{i-1}}^{t_i} G(t) \, dt \right)^2.
\] (43)

Proof The proof is left in appendix A.3.5.

As a straightforward result of theorem 4.2, we can compare the Fisher information for calibrating \( f \) of using two points \((S_i, S_j)\) versus \((S_i, S_j)\).
4.3.1. Example 1: power law. In the first example, we take the power-law kernel \( G(s) = s^{-\gamma} \) with \( \gamma = 0.4 \) (Bouchaud et al. 2003, Busseti and Lillo 2012) and power-law impact \( f(v) = v^d \) with \( d = 0.6 \) (Almgren et al. 2005) (hence \( \mu(v, t; \theta) = \frac{\delta t^d}{1-t^d} \)). We shall compare the asymptotic variance for \( t \rightarrow \infty \). Moreover, we note that this depends on \( \delta \) under different estimation methods. In this example, we assume \( T = 0.1, v = 0.3, \sigma = 1 \) and \( \sigma^* = 0.4 \) and \( \delta^* = 0.6 \). Assuming we are only estimating \( \delta \) (i.e. all other parameters are given), then as computed in (38), as \( \Delta t \rightarrow 0 \), one can check the entry for \( \delta^* \) in Table 3. The range of values for \( \tau = \frac{t}{T} \) where \( \mathcal{J}_{S_t, \delta} \geq \mathcal{J}_f(\theta) \) is

| \( \gamma = 0.35 \) | \( \gamma = 0.45 \) | \( \gamma = 0.5 \) | \( \gamma = 0.55 \) | \( \gamma = 0.65 \) | \( \gamma = 0.75 \) |
|---|---|---|---|---|---|
| \( \tau = 0.369 \) | \( 9.41 \times 10^{-7} \) | \( \leq \tau \leq 0.252 \) | \( \tau \leq \frac{1}{4} \) | \( \tau \leq 0.257 \) | \( \tau \leq 0.279 \) |

We again verify our results in simulation, output is summarized in Table 5. The performance is evaluated using the squared error between the true cost and the estimated cost, and the second row shows the standard error of the mean error. \( J \) is the model based on the average cost of VWAP, others are based on the points along the price trajectory. \( t_1 = \frac{T}{3}, t_2 = \frac{T}{2}, t_3 = \frac{2T}{3} \) and \( t_4 = T \). In the simulation study, we use similar settings as in the above theoretical analysis with the power-law kernel \( G(s) = s^{-\gamma} \) and power-law impact \( f(v) = v^d \). We set \( T = 0.1, v = 0.3, \sigma = 0.2, \gamma = 0.4 \) and \( \delta = 0.6 \). The diffusion process is discretized into 1024 time bins. The data-generating model and the estimation model are in the same parametric format and only \( \delta \) is estimated. The first model is based on the average cost of VWAP as in (37). Other models rely on the price trajectory as discussed in section 3.2 with a subset or all points at \( t_1 = \frac{T}{3}, t_2 = \frac{T}{2}, t_3 = \frac{2T}{3} \) and \( t_4 = T \). Each estimation has 300 samples, and the simulation is repeated 40000 times. The performance is evaluated using the squared error of the cost with the estimated \( \hat{\delta} \). Bold numbers in Table 5 indicate the method has a better performance than using \( J \), after a standard two-sample t-test. The conclusion agrees with the theoretical analysis that the cost-based method is inferior to the price trajectory-based methods if earlier price trajectory points are selected (for example \( t_1 \) or \( t_2 \)). In addition, if more trajectory points are chosen or the earlier the points are selected, the performance will be better.

4.3.2. Example 1: exponential kernel. Next, we consider the setting of Obizhaeva and Wang (2013), where \( f(v) \propto v \) and \( G(s) \propto e^{-\rho s} \) so that \( \mu(t, v; \theta) = \frac{\delta t}{1-e^{-\rho t}} \) for some \( c \). For this example, we assume \( T = 0.1, v = 0.3, \sigma = 1 \) and \( \rho^* = 0.01 \) and \( \rho^* = 0.15 \). In this example, we estimate \( c \) and \( \rho \) jointly and we can compute \( \mathcal{I}_{S_t, \rho} \) now as (again we let \( \Delta t \rightarrow 0 \)): \( \mathcal{I}_{S_t, \rho} = \frac{1}{\sigma^2} \int_0^T \left( \frac{\delta t}{1-e^{-\rho t}} \right)^2 \left( \frac{\delta t}{1-e^{-\rho t}} \right) dt \). For joint estimation, we no longer simply compare the ratio of specific entries in \( \mathcal{I} \) against \( \mathcal{I}_{\text{full}} \) as in the last example. Instead, we compute the matrix 2-norm (i.e. the spectral norm) of \( \mathcal{I}_{S_t, \rho} \). To see why, note that \( \| \mathcal{I}_{S_t, \rho} \cdot \mathcal{I}_{X}^{-1} \cdot \mathcal{I}_{S_t, \rho} \|_2 \leq 1 + \epsilon \Rightarrow \epsilon \ll \mathcal{I}_{S_t, \rho} \cdot \mathcal{I}_{X}^{-1} \cdot \mathcal{I}_{S_t, \rho} \|_2 \leq 1 + \epsilon \| \mathcal{I}_{S_t, \rho} \cdot \mathcal{I}_{X}^{-1} \cdot \mathcal{I}_{S_t, \rho} \|_2 \leq 1 + \epsilon \| \mathcal{I}_{S_t, \rho} \cdot \mathcal{I}_{X}^{-1} \cdot \mathcal{I}_{S_t, \rho} \|_2 \leq 1 \). Moreover, same as the last example, we also compare the optimal estimation based on \( S_{\text{full}} \) has \( \| \mathcal{I}_{S_t, \rho} \cdot \mathcal{I}_{X}^{-1} \cdot \mathcal{I}_{S_t, \rho} \|_2 = 1 \).
sampling times \( t_1 = \frac{T}{3}, t_2 = \frac{T}{4}, t_3 = \frac{T}{5} \) in addition to \( T \). The results are summarized in table 6.

The numerical results from table 6 are in line with the findings from the first example, where adding more points always increases the accuracy of estimation. However, for this model, the inclusion of earlier points \( S_t \) does not improve the estimation as much as the inclusion of later points \( S_t \). This is due to the fact that, we are not just calibrating impact function \( f \), but also the decay kernel \( G \). As we can see, in this example the asymptotic variance of estimation (for any function of \( \theta \)) using four points method is within 12.2% of the optimal one. On the other hand, for calibrating the impact function \( f \), all sampling-based methods in the table are close to optimal (i.e. close to 1) and considerably outperform \( J \) (i.e. 75.4%).

**Remark 4** Although we have shown that the full price trajectory data \( S_{\text{full}} \) is most efficient and adding more price points \( S_t \) increases the efficiency, there seems to be, based on previous examples, a ‘diminishing return’ effect to adding price points, as the increase in efficiency tends to decrease in each addition. However, the quantification of such an effect or the characterization of efficiency for \( n \) given points requires further research.

### 4.4. Square-root law: a special case

A well-known, widely-used rule-of-thumb to produce a pre-trade estimate of cost is the **square-root law**. As stated in Tóth et al. (2011), the square-root law provides a remarkably good fit on empirical data (although it is argued in Zarinelli et al. (2015) that the logarithmic function provides a better fit across more orders of magnitude than the square root function). It is suggested that the square-root law remains a robust statistical phenomenon across a spectrum of traded instruments/markets and is roughly independent of trading period, order type, trade duration/rate, and stock capitalization (Zarinelli et al. 2015). In particular, under the framework (1), the square-root law can be stated in terms of the impact (Zarinelli et al. 2015):

\[
\mu(T, v) \propto (vT)^{\frac{1}{2}} = X^{\frac{1}{2}},
\]

or the average cost per-share (Gatheral 2010):

\[
c_{\text{vwap}} \triangleq \frac{\mathbb{E}[vT S_t - X S_0]}{X} \propto X^{\frac{1}{2}}.
\]

The original statement uses \( X/V_0 \) but we have scaled \( X \) by \( V_0 \) (i.e. \( V_0 = 1 \), see remarks following (1)). As noted in Zarinelli et al. (2015), under (46), the market impact no longer depends on the specific trading rate \( v \) or trade duration \( T \), but only on their product \( X = vT \). This phenomenon can be seen as a special case of propagator model under the power-law impact \( f(v) \propto v^\gamma \) and power-law decay \( G(s) = s^{-\delta} \), with the constraint that \( \delta + \gamma = 1 \) (see Gatheral 2010). Under such model, (46) is consistent with

\[
\mu(T, v; \theta) \propto \frac{v T^{1-\gamma}}{1-\gamma} = \frac{1}{\delta} (vT)^{\delta},
\]

and (47) is consistent with

\[
\frac{c_{\text{vwap}}}{X} = \frac{v}{T} \frac{\mathbb{E}[s_t - X S_0]}{X} \propto \frac{1}{\delta (1+\delta)} (vT)^{\delta},
\]

when \( \delta = \gamma = \frac{1}{2} \), as a special case of \( \gamma + \delta = 1 \). From an estimation point of view, if we want to calibrate such a propagator model, where the impact only depends on the product \( vT \), by estimating \( \delta \) under the constraint \( \gamma + \delta = 1 \), we can compare the estimation method using sampled points \( (S_t, S_T) \) based on (48) or using the cost of the VWAP \( J = v \int_0^T S_t \, dt - X S_0 \) based on (49). Leveraging techniques from sections before, we have the following Corollary. Again we observe the power of an early sample around 20%. Yet, just as in table 3, sampling too early is not sufficiently helpful.

**Corollary 4.4** Under the propagator model \( f(v) \propto v^\delta \) and \( G(s) \propto s^{-\gamma} \), if \( \delta \leq 0.8 \) and \( 0.005 \leq vT \leq 1 \), then the Fisher information matrix (one dimension, estimating \( \delta \) only) satisfies \( I_{\delta, S_t}(\delta) \geq I_{\delta}(\delta) \), whenever \( 1.2 \% \leq t/T \leq 22.2 \% \).

| \( [I_{\delta, S_t}]^{-1} \cdot [\delta_{\text{full}}]_{\delta, \delta} \) | 3.025 | 1.769 | 1.426 | 1.732 | 1.219 | 1.137 | 1.122 |
| \( [I_{\delta, S_t}]^{-1} \cdot [\delta_{\text{full}}]_{\delta, \delta} \) | 3.025 | 1.769 | 1.426 | 1.732 | 1.219 | 1.137 | 1.122 |

Table 4. Comparison of Fisher information for calibrating power-law impact.

| \( S_{\delta, S_T} \) | \( S_{\delta, S_T} \) | \( S_{\delta, S_T} \) | \( S_{\delta, S_T} \) | \( S_{\delta, S_T} \) | \( S_{\delta, S_T} \) | \( S_{\delta, S_T} \) | \( S_{\delta, S_T} \) | \( S_{\delta, S_T} \) |
| \( [I_{\delta, S_t}]^{-1} \cdot [\delta_{\text{full}}]_{\delta, \delta} \) | 1–1.258e−6 | 1–8.184e−6 | 1–5.590e−6 | 1–7.962e−6 | 1–3.376e−6 | 1–2.274e−6 | 1–2.051e−6 |

Table 6. Comparison of Fisher information for calibrating (Obizhaeva and Wang 2013).
5. Limitations

In this section, we discuss some limitations and their practical implications on the aforementioned theorems. First, theorems in section 3 rely on assumption 2.1, which states, the stochastic law (or the underlying ‘true’ structure) governing the process of market impact cannot be correctly specified within the parameterized model. It is natural to discuss, when this assumption fails, how the results of previous theorems hold and what types of adjustments need to be made regarding the quantification of estimation ‘efficiency’.

5.1. Model misspecification

Model misspecification occurs when the ‘true’ underlying process of market impact cannot be correctly specified by (1) (or a specific variant of (1), e.g. Almgren Chriss model or the propagator model). In other words, given \((T, v) \sim G_{\text{order}}\), the joint distribution of the discretized price trajectory \(S \in \mathbb{R}^N\) follows distribution \(F\); however, a \(\theta^* \in \Theta\) such that \(F(\theta) = F\) does not exist, where 
\(\theta^*\) is characterized by \(S_t = S_0 + \mu_\theta(t, v) + \sigma W_t\) for all \(t \in [0, T]\) and \((T, v) \sim G_{\text{order}}\). In this setting, under regularity conditions, it is shown in White (1982) and Bishwal (2007) that the MLE (or QMLE, quasi-maximum likelihood estimator in this setting White 1982) \(\hat{\theta}_{\text{MLE}} = \arg \max_{\theta \in \Theta} \sum_{t=0}^{T} l(S_t|\theta)\) estimates the parameter \(\theta_{\text{KL}}^*\), which minimizes the Kullback-Leibler (KL) divergence

\[ \theta_{\text{KL}}^* = \arg \min_{\theta \in \Theta} D_{\text{KL}}(F||F(\theta)) \]  

(50)

with consistency and asymptotic normality. As a widely used statistical distance measure, the KL divergence (also known as relative entropy) \(D_{\text{KL}}(P||Q)\) between two probability measure \(P\) and \(Q\) is defined as

\[ D_{\text{KL}}(P||Q) = \int \log \frac{dp}{dq} dp, \]

where \(\frac{dp}{dq}\) is the Radon-Nikodym derivative of \(P\) w.r.t \(Q\). In other words, under model misspecification, the MLE estimates the parameters whose corresponding price trajectory distribution is the closest to the true one, in terms of the statistical distance measured by KL-divergence. Since \(F\) is unknown, to gain some intuition on the property of \(\theta_{\text{KL}}^*\), we consider the idealized case, where we observe the entire continuous path \([S_t]_{0 \leq t \leq T}\).

EXAMPLE 10 Given \((v, T)\), let the true price process \([S_t]_{0 \leq t \leq T}\) be governed by a diffusion process:

\[ dS_t = \mu^*(S_t; t, v) dt + \sigma^*(S_t; t, v) dW_t, \quad 0 \leq t \leq T, \]  

(51)

for some \(\mu^*(\cdot), \sigma^*(\cdot)\), which cannot be described by (1) for any \(\theta \in \Theta\): \(dS_t = \frac{\partial \mu_\theta (t, v)}{\partial \theta} dt + \sigma dW_t, \quad 0 \leq t \leq T, \theta \in \Theta\). Then, using Girsanov theorem, McKeague (1984) showed the log-likelihood function of the process \([S_t]_{0 \leq t \leq T}\), as an element of \(C[0, T] \subset C[0, T]\) (\(C[0, T]\) represents the space of continuous function on \([0, T]\)), can be written as

\[ l(\theta|S_t)_{0 \leq t \leq T} = -\frac{1}{2\sigma^2} \int_0^T \left( \frac{\partial \mu_\theta (t, v)}{\partial \theta} - \mu^*(S_t; t, v) \right)^2 dt + \frac{1}{2\sigma^2} \int_0^T \left( \mu^*(S_t; t, v) \right)^2 dt \]

Given sufficient metaorder data, the sum of the log-likelihood function can be maximized to obtain an optimal value \(\theta_{\text{MLE}}\) (if exists), which can be seen as an estimator for

\[ \theta^* = \arg \max_{\theta \in \Theta} \mathbb{E} \left[ l((S_t)_{0 \leq t \leq T}|\theta) \right] \]

\[ = \arg \min_{\theta \in \Theta} \int_0^T \mathbb{E} \left[ \left( \frac{\partial \mu_\theta (t, v)}{\partial \theta} - \mu^*(S_t; t, v) \right)^2 \right] dt. \]

Thus, in this example, under model misspecification, the \(\theta^*\) corresponds to a model whose drift term (i.e. \(\frac{\partial \mu_\theta (t, v)}{\partial \theta}\)) or the gradient of impact function \(\mu_\theta (t, v)\) w.r.t \(t\), on average, best matches true drift term \(\mu^*(S_t; t, v)\) of \(S_t\) in mean squared error over \([0, T]\). This example suggests the MLE from (1) seeks to recover an impact function that best describes the shape of the impact function by matching gradients (or how the gradient diminishes, in presence of concavity of impact function Zarinelli et al. 2015), rather than the direct value of the impact function.

More importantly, how does model misspecification affect our discussion, largely based on the Fisher information, regarding the (asymptotic) ‘efficiency’ of MLE? When the model is correctly specified, as mentioned in (15), the asymptotic covariance matrix of \(\hat{\theta}_{\text{MLE}}\) scaling with \(\frac{1}{n}\) (i.e. the number of metaorders) is the Fisher information matrix evaluated at the true model parameter \(\theta^*\): \(I(\theta^*) = A(\theta^*) = \mathbb{E}(\frac{\partial \mu_\theta (t, v)}{\partial \theta} )^2\), which can be equivalently defined in Hessian form, known as the information matrix equivalence theorem (White 1982), as

\[ A(\theta^*) = B(\theta^*), \]  

(52)

where \([B]_{ij}(\theta^*) = -\mathbb{E}[\frac{\partial^2 \mu_\theta (t, v)}{\partial \theta_i \partial \theta_j}]\). Note both \(\mathbb{E}\) above refer to expectation taken w.r.t. the true probability measure \(\mathbb{E}_\theta\), when the model is correctly specified and is generally unknown under model misspecification). However, as shown in White (1982), under model misspecification, (52) no longer holds and the (scaled) asymptotic variance as \(\hat{\theta}_{\text{MLE}}\) approaches \(\theta_{\text{KL}}^*\), becomes

\[ C(\theta_{\text{KL}}^*) = B^{-1}(\theta_{\text{KL}}^*)A(\theta_{\text{KL}}^*)B^{-1}(\theta_{\text{KL}}^*). \]

Moreover, White (1982) shows the discrepancies between empirical estimates of \(A(\theta_{\text{KL}}^*)\) and \(B(\theta_{\text{KL}}^*)\) can be used to test model misspecification. However, this implies the conclusion drawn from our previous theorems no longer holds, as they rely on the comparison of the Fisher information of correctly specified models. In this case, it becomes less clear how to choose the appropriate models or consider estimation efficiency of MLE based on experiment design, as theoretical properties of \(A(\theta_{\text{KL}}^*)\) or \(B(\theta_{\text{KL}}^*)\) depends on the unknown, true distribution of the price trajectory.

Thus, the presence of model misspecification will pose a limitation to the results of previous theorems. To better discuss the extent of such limitations and possible remedies, we revisit relevant concepts on model selection from statistical learning/machine learning.
5.2. Simulation

In this section, we perform a simulation study for the model misspecification case. The estimator model follows the power-law kernel $G(s) = s^{-\gamma}$ and power-law impact $f(v) = v^\beta$ (so $\mu_0(t, v) = \frac{\beta^\beta v^\beta}{1-\beta}$), where $\delta$ is the only parameter that needs to be estimated. However, the sample-generating model does not have the same parametric format. The generating model has $G(s) = s^{-\gamma}$ and $f(v) = 1.5 \ln(1 + 0.5v^2 + 0.7v) + 1.5 \ln(1 + 0.5v^2 + 0.7v)$ (so $\mu^*(t, v) = 1.5 \ln(1 + 0.5v^2 + 0.7v)$). We set $\sigma = 0.2$, $\gamma = 0.4$. Each estimation has 300 samples, and each sample trajectory is assigned a random price trajectory length $T \sim \text{Uniform}(0.1, 0.15)$ and a random $v \sim \text{Uniform}(0.3, 0.4)$. The simulation is repeated 40,000 times for each scenario. The diffusion process is discretized into 1024 time bins. The first model is based on the average cost of VWAP as in (37). Other models rely on the price trajectory as discussed in section 3.2 with a subset or all points at $t_1 = \frac{T}{8}, t_2 = \frac{T}{4}, t_3 = \frac{3T}{8}$ and $t_4 = T$ relative to the corresponding path length $T$. The error of estimation is the mean squared error of the cost across all sample paths. Results are shown in Table 7. The conclusion is similar to the well-specification model in section 4.3 and the simulation study therein that the cost-based method is worse than the price trajectory-based methods if earlier price trajectory points are selected (for example $t_1$ or $t_2$). In addition, if more trajectory points are chosen or the earlier the points are selected, the performance will be better. The performance is evaluated using the squared error between the true cost and the estimated cost, and the second row shows the standard error of the mean error. $X = J$ is the model based on the average cost of VWAP, others are based on the points along the price trajectory, $t_1 = \frac{T}{8}, t_2 = \frac{T}{4}, t_3 = \frac{3T}{8}$ and $t_4 = T$. Bold numbers indicate the method has a statistically significant better performance than using $J$ based on two sample $t$-tests.

5.3. Model selection

As in (50), given $(T, v) \sim G_{\text{order}}$, we use $F$ to denote the true distribution of price trajectory $S$ and $F(\theta)$ the one characterized under $S = S_0 + \mu_0(t, v) + \sigma W_t$, for $S_t \in S$ and $(T, v) \sim G_{\text{order}}$. Then, as discussed in the previous section, if we measure the risk as

$$R(\theta) = D_{\text{KL}}(F||F(\theta)),$$

which quantifies the discrepancy between distributions fitted by the model and truth. Then, the model selection problem decomposes the risk as

$$R(\theta) = R(\theta) - \inf_{\theta \in \Theta} R(\theta) + \inf_{\theta \in \Theta} R(\theta),$$

where $R(\theta)$ is the estimation error and $R(\theta)$ the approximation error.

In a correctly specified model, $\theta^{*}_{\text{KL}} = \theta^*$ and the estimation error is 0. The approximation error is directly related to the efficiency for estimating $\theta^*$ and the results of our previous theorem, based on the Fisher information matrix, can be applied. When a model misspecification occurs, different factors/concerns can come into play and the landscape is less clear. For example,

- If one is concerned with an accurate estimation of the price trajectory for various values of $(T, v)$, then the estimation using the price trajectory would increase the sample size, compared to using summary statistics, e.g., VWAP, which could lead to a positive effect on reducing estimation error.
- If one is only concerned with an accurate estimation of summary statistics, e.g., VWAP. Then, this could lead to a reduced approximation error because KL divergence monotonically decreases under the non-invertible transformation (it is well known that KL divergence is invariant under invertible transformations, i.e., given random variable $X$, $Y$ and an invertible transformation $g()$, we have $D_{\text{KL}}(X \mid \mid Y) = D_{\text{KL}}(g(X) \mid \mid g(Y))$). However, it can be shown that $D_{\text{KL}}(X \mid \mid Y) \geq D_{\text{KL}}(g(X) \mid \mid g(Y))$ when $g$ is non-invertible (see Mena et al. 2018). The transformation from the price trajectory $S$ to summary statistics are typically non-invertible.). However, it could lead to an increase in approximation error due to reduced sample size or inefficiencies of estimation.

During practical implementation, we should take these issues into consideration. Typical approaches include cross-validation using different methods (or some hybrid of them) under the specified loss function or Bayesian inference based on carefully calibrated prior from previous data (or even online data, where one can update the posterior as the order is executed and adjust the execution). This is left for discussion and future research.

| $J$   | $S_t, S_T$ | $S_t, S_T$ | $S_t, S_T$ | $S_t, S_T$ | $S_t, S_T$ | $S_t, S_T$ | $S_t, S_T$ | $S_t, S_T$ |
|-------|------------|------------|------------|------------|------------|------------|------------|------------|
| Mean error $\times 10^{-7}$ | 1.079 | 1.049 | 1.077 | 1.144 | 1.038 | 1.040 | 1.073 | 1.034 |
| Standard error $\times 10^{-9}$ | 4.831 | 4.612 | 4.822 | 5.294 | 4.538 | 4.546 | 4.788 | 4.509 |

Table 7. Numerical comparison of models with misspecification.
6. Conclusion

In this study, we have methodically chosen the Fisher information as a measure of the asymptotic efficiency of statistical estimations and experimental designs. This approach serves as a gateway, not necessarily the optimal one but a principled one, to systematically determine how the inclusion of price trajectory data affects the estimation accuracy of market impact models. Through this principled examination, we compared, both theoretically and empirically, estimators derived from price trajectory data against those obtained via traditional VWAP-based methods under Almgren-Chriss models, propagator models, and scenarios involving model misspecification. Although the specific technical tweaks vary across different settings, our principal finding can be summarized as that incorporating price trajectory data, particularly the executed prices recorded at the early stages of an order, appears to enhance the accuracy of model parameter estimations in a statistically significant way. This observation does not serve as a literally prescriptive recommendation on the inclusion of price trajectory data for each and every specific model mentioned above, but it suggests that useful information carried in (early) price trajectory data could be a general phenomenon worthy of consideration across a variety of market impact modeling, including the ones not mentioned. Given this insight, incorporating this strategy into market impact model estimations could be advantageous.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Supplemental data

Supplemental data for this article can be accessed online at https://doi.org/10.1080/14697688.2024.2351457.

ORCID

Fengpei Li http://orcid.org/0000-0003-0904-4723

References

Adrian, T., Capponi, A., Fleming, M., Vogt, E. and Zhang, H., Intra-day market making with overnight inventory costs. J. Financ. Mark., 2020, 50, 100564.

Alt-Sahalia, Y., Closed-form likelihood expansions for multivariate diffusions. Ann. Stat., 2008, 36(2), 906–937.

Alt-Sahalia, Y., Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. Econometrica, 2020, 70(1), 223–262.

Alt-Sahalia, Y. and Mykland, P.A., Estimators of diffusions with randomly spaced discrete observations: A general theory. Ann. Stat., 2004, 32(5), 2186–2222.

Alt-Sahalia, Y. and Mykland, P., Estimating volatility in the presence of market microstructure noise: A review of the theory and practical considerations. In Handbook of Financial Time Series, pp. 577–598, 2009 (Springer: Berlin).

Alfonsi, A. and Schied, A., Optimal trade execution and absence of price manipulations in limit order book models. SIAM J. Financial Math., 2010, 1(1), 490–522.

Alfonsi, A., Fruth, A. and Schied, A., Constrained portfolio liquidation in a limit order book model. Banach Center Publ., 2008, 83, 9–25.

Alfonsi, A., Fruth, A. and Schied, A., Optimal execution strategies in limit order books with general shape functions. Quant. Finance, 2010, 10(2), 143–157.

Alfonsi, A., Schied, A. and Slynko, A., Order book resilience, price manipulation, and the positive portfolio problem. SIAM J. Financial Math., 2012, 3(1), 511–533.

Almgren, R., Optimal execution with nonlinear impact functions and trading-enhanced risk. Appl. Math. Finance, 2003, 10(1), 1–18.

Almgren, R. and Chriss, N., Optimal execution of portfolio transactions. J. Risk, 2001a, 3, 5–40.

Almgren, R. and Chriss, N., Optimal execution of portfolio transactions. Risk, 2001b, 12, 61–63.

Almgren, R. and Lorenz, J., Adaptive arrival price. Trading, 2007, 1, 59–66.

Almgren, R. and Stoikov, S., High-frequency trading in a limit order book. Quant. Finance, 2008, 8(3), 217–224.

Almgren, R., Thum, C., Hauptmann, E. and Li, H., Direct estimation of equity market impact. Risk, 2005, 18, 58–62.

Bacry, E., Iuga, A., Lasnier, M. and Lehalle, C.-A., Market impacts and the life cycle of investors orders. Market Microstruct. Liquid., 2015, 1(2), 1550009.

Bank, P. and Baum, D., Hedging and portfolio optimization in financial markets with a large trader. Math. Finance, 2004, 14(1), 1–18.

Bayraktar, E. and Ludkovski, M., Liquidation in limit order books with controlled intensity. Math. Finance, 2014, 24(4), 627–650.

Bertsimas, D. and Lo, A.W., Optimal control of execution costs. J. Financ. Mark., 1998, 1(1), 1–50.

Bibby, B.M., Jacobsen, M. and Sørensen, M., Estimating functions for discretely sampled diffusion-type models. In Handbook of Financial Econometrics: Tools and Techniques, pp. 203–268, 2009 (Elsevier: North-Holland).

Bickel, P.J., Klaassen, C.A., Bickel, P.J., Ritov, Y., Klaassen, J., Wellner, J.A. and Ritov, Y., Efficient and Adaptive Estimation for Semiparametric Models, Vol. 4, 1993 (Springer: New York, NY).

Bishwal, J.P., Parameter Estimation in Stochastic Differential Equations, 2007 (Springer: Berlin).

Bouchaud, J.-P., Price Impact. Encyclopedia of Quantitative Finance, 2010.

Bouchaud, J.-P., Gefen, Y., Potters, M. and Wyart, M., Fluctuations and response in financial markets: The subtle nature of random price changes. Quant. Finance, 2003, 3(2), 176.

Bouchaud, J.-P., Kockelkoren, J. and Potters, M., Random walks, liquidity molasses and critical response in financial markets. Quant. Finance, 2006, 6(2), 115–123.

Bouchaud, J.-P., Farmer, J.D. and Lillo, F., How markets slowly digest changes in supply and demand. In Handbook of Financial Econometrics: Dynamics and Evolution, pp. 57–160, 2009 (Elsevier: North-Holland).

Briere, M., Lehalle, C.-A., Nefedova, T. and Raboun, A., Modelling transaction costs when trades may be crowded: A bayesian network using partially observable orders imbalance. In Machine Learning for Asset Management: New Developments and Financial Applications, pp. 387–430, 2020 (Wiley-ISTE).

Brunnermeier, M. and Pedersen, L.H., Predatory trading. J. Finance, 2005, 60(4), 1825–1862.

Bucci, F., Benzaquen, M., Lillo, F. and Bouchaud, J.-P., Crossover from linear to square-root market impact. Phys. Rev. Lett., 2019, 122(10), 108302.

Busseti, E. and Lillo, F., Calibration of optimal execution of financial transactions in the presence of transient market impact. J. Stat. Mech. Theory Exp., 2012, 09, P09010.
