TRADING WITH CONCAVE PRICE IMPACT AND IMPACT DECAY – THEORY AND EVIDENCE

NATASCHA HEY[1], IACOPO MASTROMATTEO[2], JOHANNES MUHLE-KARBE[3], KEVIN WEBSTER[4]

[1]Chair of Econophysics and Complex Systems, École Polytechnique
[2]Capital Fund Management
[3]Department of Mathematics, Imperial College London
[4]Department of Mathematics, Columbia University

Abstract. We study statistical arbitrage problems accounting for the *nonlinear* and *transient* price impact of metaorders observed empirically. We show that simple explicit trading rules can be derived even for general nonparametric alpha and liquidity signals, and also discuss extensions to several impact decay timescales. These results are illustrated using a proprietary dataset of CFM metaorders, which allows us to calibrate the levels, concavity, and decay parameters of the price impact model and analyze their effects on optimal trading.

1. Introduction

Trading costs play a central role in designing and implementing quantitative trading strategies. Indeed, Loeb (1983) refers to them as the "critical link between investment information and results"; 40 years later Harvey et al. (2022) still write that "market impact costs are a crucial component of strategy performance – yet these costs are routinely ignored in most academic research studies."

For sizable funds, the crucial concern is their trades' adverse *price impact*.[1] It is well known that impact is concave in trade sizes, in that large trades have a smaller impact than predicted by a linear model and are instead better described by a "square-root law".[2] Price dislocations are also not static but gradually dissipate over time.[3]

E-mail address: natascha.hey@ladhyx.polytechnique.fr, iacopo.mastromatteo@cfm.com, jmuhleka@ic.ac.uk, kw3058@columbia.edu.

Date: November 6, 2023.

The authors are grateful to Jean-Philippe Bouchaud for fruitful discussions.

[1]For instance, Frazzini, Israel, and Moskowitz (2018) use proprietary data of AQR, a major quantitative fund, to estimate their trades' price impact. They find price impact costs to be an order of magnitude larger than other costs such as the effective bid-ask spread. These results align with many other practitioner studies. For instance, Nasdaq's "Intern's Guide to Trading" (Mackintosh, 2022) also finds price impact for institutional trades to be an order of magnitude larger than spreads.

[2]Cf., e.g., Loeb (1983); Hasbrouck (1991); Hasbrouck and Seppi (2001); Lillo et al. (2003); Bouchaud et al. (2004); Almgren et al. (2005); Gabaix et al. (2006); Bershova and Rakhlin (2013); Frazzini et al. (2018) or the textbooks Bouchaud et al. (2018); Webster (2023) and the references therein.

[3]Cf., e.g., Biais et al. (1995); Coppejans et al. (2004); Degryse et al. (2005); Bouchaud et al. (2009); Bacry et al. (2015); Brokmann et al. (2015) and the references therein.

A key practical challenge is how to make these robust statistical findings "actionable" by embedding them into a consistent stochastic control problem. There is a large and active literature that studies optimal trading with price impact but, for tractability, these studies typically either assume price impact to be linear (Gârleanu and Pedersen, 2013, 2016) or to decay instantaneously (Almgren, 2003). In contrast, for models with nonlinear and transient price impact, even basic qualitative properties such as the absence of price manipulation are typically poorly understood (Gatheral, 2010). A notable exception is the model of Alfonsi, Fruth, and Schied (2010) (henceforth AFS), where an optimal execution problem is solved explicitly.

The present study shows that the AFS model, with nonlinear price impact and impact decay, also admits closed-form solutions for general statistical arbitrage problems with arbitrary alpha signals and stochastic liquidity parameters.[4] For capacity- rather than risk-constrained traders,[5] this yields simple and intuitive trading rules that apply to general nonparametric price and liquidity forecasts in a straightforward manner, bypassing the need for any brute-force optimization. We also show how to fit models of this type to a proprietary dataset of metaorders. This bridges the gap to the empirical literature by allowing to capture the main stylized facts present in the data while retaining inherent tractability of optimal trading strategies.

We derive our theoretical results by a change of variables to "impact space", where the trader's control variable is the aggregate impact of their current and past trades rather than the position held. This change of perspective was pioneered by Fruth, Schöneborn, and Urusov (2013) for linear impact models. Here we show that this approach also allows one to reduce the analysis of general AFS models to simple pointwise optimizations, that in turn lead to closed-form expressions for the optimal trading rules. An auxiliary benefit of the method is the derivation of sharp conditions to rule out price-manipulation strategies.[6] Finding such practical, measurable, and implementable conditions guaranteeing the good behavior of live trading algorithms is another core concern of execution teams at major financial institutions.

The passage to impact space is crucially tied to the existence of a one-to-one map between holdings and the corresponding impact. This is guaranteed when impact decays at a constant exponential rate as in Obizhaeva and Wang (2013). However, many empirical studies find price impact to decay over multiple timescales not captured by a single exponential rate, cf., e.g., Bacry et al. (2015); Brokmann et al. (2015). To incorporate this, we show that our approach can be extended to price impact models with multiple decay timescales. The core idea is to switch to impact space and optimize pointwise separately on each timescale, but simultaneously enforce a consistency constraint that the respective impacts correspond to the same trades. The optimal impact state is then again available in closed form, up to solving an autonomous "decoupling" ODE for the constraint's Lagrange multiplier.

[4]Stochastic liquidity parameters are a tractable proxy for the "local concavity" of the price impact of individual trades (Muhle-Karbe et al., 2023), which can in turn be aggregated into the analysis of impact's "global concavity" at the metaorder level in the present study.

[5]Busse et al. (2020) show that capacity constraints are a concern even for mutual funds, forcing them to reduce their rebalancing frequencies and shift investments to more liquid instruments.

[6]Gatheral (2010) first defined and studied price manipulation conditions for a class of price impact models.

To illustrate the relevance of our modeling choices and explore the implications of our results for optimal trading strategies, we complement this theoretical analysis with a detailed empirical study on proprietary Capital Fund Management (CFM) metaorder data. Specifically, we fit impact levels, concavities, and magnitudes for AFS models across multiple decay timescales. This bridges the gap between sophisticated nonparametric empirical studies (which are difficult to translate into optimal trading strategies) and the stochastic control literature (which often studies models lacking empirical foundation). A model with two impact timescales (one fast and one slow) generally offers the best tradeoff between accuracy and parsimony. For the corresponding concavities, the best power-law specification is close to a square-root law across all timescales.

Using our theoretical results, optimal trading strategies taking into account all of these empirical features can be derived in a straightforward manner. Figure 1 illustrates this. The left panel plots the optimal peak impact I_T at the end of the trading interval $[0, T]$ against the corresponding optimal order size Q_T. This illustrates that square-root concavities in the AFS model indeed generate the square-root law for metaorder impact observed empirically. The right panel of Figure 1 displays the temporal evolution of the optimal impact state I_t during and after the completion of the trading period $[0, T]$. During the trading period, impact builds up as the alpha signal at hand is gradually exhausted. For the fast and slow impact decay timescales calibrated in our empirical analysis, a substantial proportion of the peak impact at the end of the trading period then decays very quickly, but the remaining long-term impact lingers much longer.

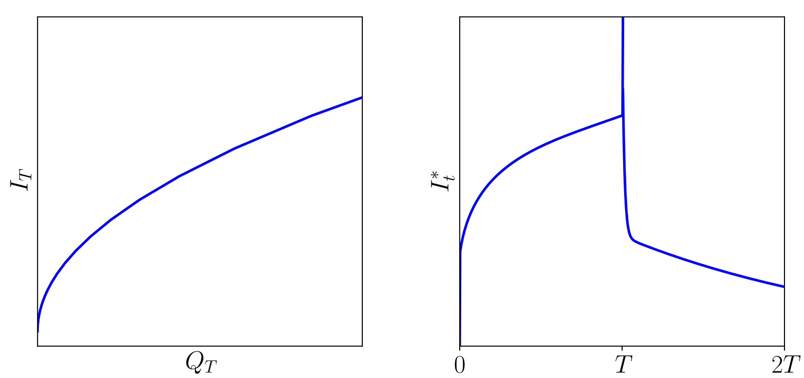


FIGURE 1. Optimal impact for constant alpha signals as a function of optimal traded volume (left panel) and time (right panel).

Metaorder data is proprietary and often difficult to access for academic researchers or small trading firms. Therefore, we also analyze to what extent our fitting results can be recovered from the public trading tape, using "proxy metaorders" constructed from the aggregate order-flow imbalance. These proxy metaorders generally underestimate the magnitude of price impact. For models with a single timescale for impact decay, the public tape allows to recover similar values for decay timescales and concavities, but proprietary metaorders are required to correctly pin down multiple decay timescales correctly.

Application to Transaction Cost Analysis (TCA). To give a non-technical introduction to our optimal trading rules and illustrate the ease with which these can be implemented in practice, we now discuss a concrete application, "Transaction Cost Analysis" (TCA).[7]

The primary purposes of TCA are to establish, ex-post, whether a set of orders traded optimally and to investigate inconsistent behavior. The European Commission (2014) defines MiFID II requirements for best execution in Article 27: "Obligation to execute orders on terms most favourable to the client":

> "[Regulators] require investment firms who execute client orders to monitor the effectiveness of their order execution arrangements and execution policy in order to identify and, where appropriate, correct any deficiencies."

TCA uncovers incorrect model assumptions, algorithm implementation errors, or poorly calibrated alpha signals. Moreover, researchers and traders must communicate TCA to stakeholders, including investors, clients, and regulators:

> "[Regulators] require investment firms to be able to demonstrate to their clients, at their request, that they have executed their orders in accordance with the investment firm's execution policy and to demonstrate to the competent authority, at its request, their compliance with this Article."

Therefore, TCA must streamline communication, conveying core algorithm trade-offs without getting lost in implementation details, or conversely, obfuscating best execution requirements.

The optimal trading formula in this paper fulfills these requirements by expressing the optimal strategy as a balance between impact, alpha, and alpha decay. To illustrate this, first suppose for simplicity that liquidity is constant over the trading horizon. If

(a) price impact decays over a timescale τ,
(b) price impact is a power function of order size with exponent $c \in (0, 1]$, and
(c) the trader's alpha signal is α_t with alpha decay $-\alpha'_t$,

then the impact I_t of the optimal trading strategy satisfies the linear relationship

$$I_t = \frac{1}{1+c}\left(\alpha_t - \tau\alpha'_t\right). \tag{1.1}$$

At any given time, an algorithm thus reacts to signals by trading in a direction that balances these three core variables. For instance, an execution algorithm accelerates when detecting larger alpha decay. Concavity of the impact function ($c < 1$) simply implies that alpha signals can be traded more aggressively than for linear impact. For example, the optimal impact state with square root impact exhausts two thirds of the alpha signal, rather than half in the linear case.

We now outline how to use the simple relationship (1.1) for TCA; throughout, we focus on the empirically most relevant case $c = 1/2$ where the impact function is consistent with the square-root law.

The Baseline Scenario. First consider a trader executing orders with a constant alpha signal $\alpha_t = \alpha$. For instance, a long-term trader determines their alpha signal at the start of the

[7] Another important application, the opportunity costs of misspecified price impact models is discussed in the companion paper of the present study (Hey et al., 2023), based on a preview of some of the results derived in the present study.

day and does not update the signal intraday. In particular, the long-term trader assumes no intraday alpha decay, $\alpha'_t = 0$. In this scenario, the optimal execution strategy satisfies

$$I_t = \frac{2}{3}\alpha. \tag{1.2}$$

Strategy	α (bps)	I (bps)
A (Macro)	60	40
B (Technicals)	45	30
C (News)	60	50

TABLE 1. Mock TCA report for three strategies under the baseline scenario.

Consider Table 1 under this baseline scenario providing the average alpha and impact of three different trading strategies A, B, and C. From an execution perspective, the trader's first-order question is:[8]

Are the strategies correctly balancing alpha and impact?

Given our baseline assumptions, strategy C did not balance correctly its alpha and impact: it paid too much impact and should have traded more slowly because $I > \frac{2}{3}\alpha$.

Of course, traders should take care when translating such inconsistencies into actions. Indeed, the model quickly and intuitively detects inconsistencies within a trading strategy. However, without further analysis, it is unclear which of the trader's assumptions is false:

(a) Did the strategy experience alpha decay during the execution?
(b) Is the price impact model incorrect? For example, did liquidity vary during the execution?
(c) Is there a code or data error in the execution algorithm's implementation?

The next sections apply our impact formula to dig deeper into Strategy C.

News Trigger	α (bps)	$-\tau\alpha'$ (bps)	I (bps)
Earnings news	90	30	80
Other news	45	0	30

TABLE 2. TCA breakdown of Strategy C across alpha triggers

Adding Alpha Decay. Given strategy C's focus on news, a trader may investigate which news event triggers their trading strategy and if some events experience alpha decay. Table 2 explains the abnormal behavior in Row C of Table 1.

[8]From an alpha research perspective, an earlier question to answer of course is: *Are the alpha signals correctly calibrated?* This leads to the distinction between the strategies' *predicted* and *realized* alpha. We assume in this study that all alphas are correctly calibrated and that predicted and realized alphas match in expectation.

Indeed, while most news events in strategy C experience no alpha decay, the earnings news signal decays during the execution. Therefore, the trading strategy behaved correctly and accelerated into the alpha decay to capture more alpha, raising impact costs:

$$I_t = \frac{2}{3}\left(\alpha_t - \tau\alpha_t'\right).$$

In this scenario, the trader cannot blame the execution strategy for the high impact costs. If they wish to increase strategy C's profitability (net alpha capture after transaction costs), the trader must reduce the strategy's alpha decay, for instance, by improving or removing the signal triggered by earnings news.

Liquidity Condition	α (bps)	$-\tau\alpha'$ (bps)	I (bps)
Constant liquidity	90	30	80
Dropping liquidity	90	30	84
Rising liquidity	90	30	75

Table 3. TCA breakdown of Strategy C on earnings news across liquidity conditions.

Adding Dynamic Liquidity Conditions. A trader that is unhappy with strategy C's low profitability during earnings events, 10bps on 90bps of alpha, may nevertheless not want to completely turn off the strategy. Therefore, they may further decompose the strategy's performance to determine when it may recover a more comfortable profitability.

Table 3 breaks down the earnings news bucket considering whether liquidity was constant, rising, or dropping. Section 4.2 extends the balancing formula (1.1) to consider such dynamic liquidity conditions. The table is consistent with the formula, quantifying the following trading intuition:

(a) The strategy's alpha level matters more when liquidity decreases.
(b) The strategy's alpha decay matters more when liquidity increases.

Therefore, if the trader wishes to improve strategy C's profitability, one option is to only respond to earnings news events when forecasting increasing liquidity, where the profitability rises from 10bps to 15bps on 90bps of alpha.

Outline. This paper is organized as follows. Section 2 introduces the price impact model, and Section 3 formulates the corresponding risk-neutral stochastic control problems. Section 4 describes the explicit optimal strategies that can be obtained by changing variables to "impact space", and Section 5 extends this method to multiple impact decay timescales. These theoretical results are complemented by the empirical analysis in Section 6, where the models are fit to proprietary trading data. For better readability, the derivations of the results are collected in the appendix.

Notation. Throughout, we fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ with finite time horizon $T > 0$.

2. Price Impact Model

Price impact models describe how prices causally depend on trades. To formalize this, let $(S_t)_{t\in[0,T]}$ be the "unaffected" (or "fundamental") mid-price process in the absence of trading. If $(Q_t)_{t\in[0,T]}$ denotes the holdings of one (or several) large trader(s), the observed market mid-price is

$$P_t(\omega, Q) = S_t(\omega) + I_t(\omega, Q).$$

Here, the notation stresses that $S_t(\omega)$ describes price changes that happen independent of the large trader's actions, e.g., due to external news. In contrast, the price impact term $I_t(\omega, Q)$ can depend both on external randomness *and* large traders' present and past actions $(Q_s)_{s\in[0,t]}$. (We now suppress the dependence on the random state $\omega \in \Omega$ as usual.)

Alfonsi, Fruth, and Schied (2010) (henceforth AFS) proposed a price impact model that captures the nonlinear and transient nature of price impact (cf. the references in the introduction) while remaining analytically tractable:[9]

Definition 2.1 (AFS price impact model)**.** *The price impact of a strategy* $(Q_t)_{t\in[0,T]}$ *is*

$$I_t = h(J_t).$$

Here, the impact function $h \in C^2$ *is increasing, odd, and concave on* $[0, \infty)$. *Its argument*

$$dJ_t = -\frac{1}{\tau_t} J_t dt + \lambda_t dQ_t, \quad J_0 = 0,$$

is an exponential moving average of current and past trades. The timescale $(\tau_t)_{t\in[0,T]}$ *over which impact decays and the* push factor $(\lambda_t)_{t\in[0,T]}$ *can be time dependent and random.*

When the price impact function is the identity ($h(x) = x$), this recovers the model of Obizhaeva and Wang (2013), where each trade causes *linear* price impact proportional to "Kyle's lambda" λ_t, and subsequently decays at a timescale governed by τ_t.

If, more generally, the price impact function h is smooth and concave on $[0, \infty)$, then small trades dQ_t still have approximately linear impact $dI_t = h'(J_t)\lambda_t dQ_t$, but the overall impact of large trades Q_t is sublinear in line with the crossover from linear to square-root impact documented empirically by Bucci et al. (2019b). Indeed, as trades and in turn the moving average J_t accumulate, the linear impact $h'(J_t)\lambda_t dQ_t$ becomes smaller by concavity of the impact function $h(x)$. This also leads to sublinear impact for metaorders that are executed gradually over time (large block trades evidently have a direct sublinear impact).

Remark 2.2. *Obizhaeva and Wang (2013) motivated their linear price impact model with a flat limit-order book. Analogously, nonlinear price impact functions can be derived from a limit order book with non-constant density (Alfonsi et al., 2010; Carmona and Webster, 2019).*

The connection between the order book and price impact is that the order book maps prices to marginal trading volumes. When one derives the price impact model from an order book shape, J measures a trade's volume impact *on the order book. The exponential moving average corresponds to the "resilience" of the order book, which gradually recovers due to new incoming limit orders. One then uses the order book shape to map this volume impact back to a price impact.*

[9] Alfonsi et al. (2010) leverage this tractability to solve an optimal execution problem. In the present study we show that statistical arbitrage problems with general alpha signals and stochastic liquidity parameters also admit closed-form solutions.

Note, however, that the impact function $h(x)$ cannot *simply be read off the cumulated state of the current order book. Instead, at lower frequencies, it is a reduced form model for the* latent order book *(Tóth et al., 2011; Donier et al., 2015), which encodes the latent portion of the demand and supply curve that has not yet materialized into the visible order book.*

Example 2.3. *As a concrete example, fix $x_0 > 0$ and suppose that*

$$h(x) = \begin{cases} x, & |x| \leq x_0, \\ \operatorname{sgn}(x)\sqrt{2|x|x_0 - x_0^2}, & |x| > x_0. \end{cases}$$

Then, volume impacts smaller than the threshold x_0 shift prices linearly, whereas the price impact of large trade imbalances scales with the square root of the volume impact. The location and scale parameters in the square root function are chosen to ensure value matching and smooth pasting between these two regimes, compare Figure 2.

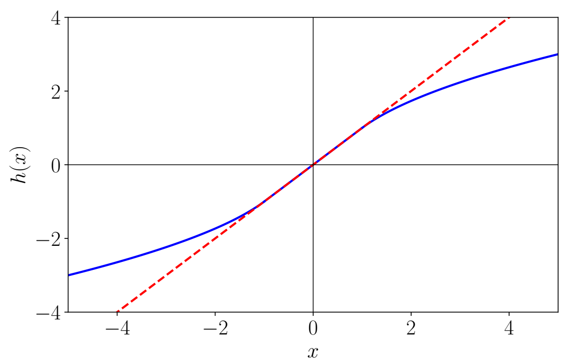


FIGURE 2. The price impact function $h(x)$ from Example 2.3 for $x_0 = 1$ (solid blue) and the identity function $h(x) = x$ (dashed red) from the linear impact model of Obizhaeva and Wang (2013).

3. A Stochastic Control Problem for Trading

We now turn to optimal trading with concave price impact. For simplicity, we focus on a single large trader who controls the holdings $(Q_t)_{t\in[0,T]}$. Additional "external flow" of other market participants could be added similarly as in Muhle-Karbe et al. (2023), but we do not spell this out here to not overload the notation.

3.1. **Self-Financing Equation.** When the large trader's holdings vary smoothly, that is $Q_t = \int_0^t \dot{Q}_s ds$ for a finite "trading rate" $\dot{Q}_t = dQ_t/dt$, then trades $dQ_t = \dot{Q}_t dt$ are settled at $S_t + h(J_t + dQ_t) = S_t + h(J_t) + O(dt)$. Accordingly, the trader's cash balance from continuous trading on $[0,t]$ is given by[10]

$$Y_t = -\int_0^t (S_t + I_t)dQ_t = -Q_T S_T + \int_0^t Q_t dS_t - \int_0^t h(J_t)dQ_t.$$

[10]Here, the second equality follows from integration by parts. Note that for smooth trades the distinction between the price impact before and after their execution vanishes. This differs for discrete block trades or holdings "wiggling like Brownian motion". Section 4 avoids the corresponding cumbersome bookkeeping equations by directly approximating the performance of such strategies "in impact space".

Here, the first integral describes the standard gains and losses from frictionless continuous-time trading. The second integral accounts for additional price impact costs.[11]

3.2. **Risk-Neutral Objective Function.** The most tractable objective function in this context is to maximize expected returns net of transaction costs. Note that superlinear price impact costs automatically impose an endogenous capacity constraint. Therefore, such risk-neutral optimization problems are typically well-posed.

Definition 3.1 (Risk-neutral intraday trading)**.** *A risk-neutral* intraday trader *solves the stochastic control problem*

$$\sup_Q \mathbb{E}\left[Y_T + Q_T S_T\right] = \sup_Q \mathbb{E}\left[\int_0^T Q_t dS_t - \int_0^T h(J_t) dQ_t\right]. \tag{3.1}$$

Here, the trader values their end-of-day position at the unaffected price rather than the market price. This avoids illusory gains caused by pushing up prices when entering a position. Such illusory mark-to-market gains due to price impact are studied by Caccioli et al. (2012); Kolm and Webster (2023). To avoid such misleading profits, traders minimize arrival slippage and ignore mark-to-market P&L, in line with (3.1)*.*

A helpful statistic to simplify (3.1) is the so-called *intraday alpha signal*

$$\alpha_t = \mathbb{E}_t\left[S_T - S_t\right] = \mathbb{E}_t[S_T] - S_t.$$

This signal predicts intraday returns the trader *doesn't* cause. For a risk-neutral trader, an alpha signal is a sufficient statistic of the unaffected price S_t. Indeed, an integration by parts and $\alpha_T = 0$ yield $\mathbb{E}[\int_0^t Q_s dS_s] = \mathbb{E}[\int_0^t \alpha_s dQ_s]$ for smooth strategies $dQ_t = \dot{Q}_t dt$. The risk-neutral intraday objective (3.1) in turn can be written as

$$\sup_Q \mathbb{E}\left[\int_0^T (\alpha_t - h(J_t)) dQ_t\right]. \tag{3.2}$$

That is, in expectation, each trade earns alpha but pays price impact.

A long-term trader may also have views on returns beyond the trading day. The risk-neutral objective (3.1) can be extended to incorporate such long-term views as follows:

Definition 3.2 (Risk-neutral long-term trading)**.** *A risk-neutral* long-term trader *solves the stochastic control problem*

$$\sup_Q \mathbb{E}\left[Y_T + S_\tau Q_T\right] \quad \textit{for some time } \tau > T.$$

A risk-neutral long-term trader still tracks an alpha signal

$$\alpha_t = \mathbb{E}\left[S_\tau - S_t \middle| \mathcal{F}_t\right],$$

but the end-of-day constraint $\alpha_T = 0$ no longer applies in this case. With this notation, the representation (3.2) of the risk-neutral goal functional still applies. The only difference is that the long-term alpha does not vanish at the end of the trading day.

[11]This budget equation assumes that trades settle at the mid-price and thus neglects the contribution of bid-ask spreads to transaction costs. As discussed in the introduction, this is justified for large, capacity- rather than risk-constrained actors, for which impact costs dominate spread costs.

For simplicity, we henceforth assume that the alpha signal α_t is an Itô process and denote its drift rate by μ_t^α. Traders typically refer to α_t as the *alpha level* and to $-\mu_t^\alpha$ as the *alpha decay*. Examples 3.3 and 3.4 below specify standard parametric alpha signals. Example 3.5 covers the increasingly common non-parametric case. In particular, this applies to modern machine-learning approaches to alpha research and showcases the advantage of our explicit trading formulas from Theorem 4.2 for general non-parametric signals.

Example 3.3 (Deterministic alpha)**.** *The trading signal α_t can be deterministic even if the fundamental price S_t is stochastic. For instance, a constant drift rate of the fundamental price $dS_t = \mu dt + \sigma dW_t$ translates to the deterministic intraday alpha signal $\alpha_t = \mu(T-t)$. In the long-term case, the alpha signal equals $\alpha_t = \mu(\tau - t)$.*

Example 3.4 (Ornstein-Uhlenbeck alpha)**.** *If the unaffected price process has Itô dynamics $dS_t = \mu_t dt + \sigma_t dW_t$, then the* alpha decay *$-\mu_t^\alpha$ equals the drift rate μ_t. The most common specification is to model this as an Ornstein-Uhlenbeck process $d\mu_t = -\theta^{-1}\mu_t dt + \eta dW_t$.*[12] *Then, the intraday alpha is*

$$\alpha_t = \int_t^T \mathbb{E}_t[\mu_s]ds = \int_t^T e^{-(s-t)/\theta}\mu_t ds = (1 - e^{-(T-t)/\theta})\theta\mu_t.$$

For a long-term trader predicting the steady-state alpha ($\tau \to \infty$), we have $\alpha_t = \theta\mu_t$.

Example 3.5 (Non-parametric alpha)**.** *Examples 3.3 and 3.4 are convenient to derive analytical results. However, practitioners rarely fit two-parameter alpha signals. Instead, they increasingly rely on non-parametric models with hundreds or thousands of features. In particular, the last decade saw the rise of machine learning models for alpha research. For example, cf. Cont et al. (2021); Kolm et al. (2023) for applications of various neural network architectures to build alpha signals from limit order book events.*

Therefore, most quantitative strategies treat alpha signals as generic functions or processes. For instance, a portfolio optimization or trading algorithm may abstract away an alpha model's parameters and rely on a stream of alpha levels α_t and decays $-\mu_t^\alpha$. Solving control problems in this non-parametric regime therefore is a crucial differentiation between toy models and practical algorithms.

4. Solution by Mapping to Impact Space

We now turn to the solution of the risk-neutral trader's optimization problem (3.2). Section 3.2 derives the objective function (3.2) for smooth strategies $Q_t = \int_0^t \dot{Q}_s ds$. This is sufficient when trading signals α_t vary smoothly over time. However, diffusive trading signals as in Example 3.4 naturally lead to diffusive trades. Moreover, bulk trades also naturally appear unless the trader's initial holdings perfectly align with the initial signal.

At first glance, it seems appealing to approximate the P&L of such more general strategies by the P&Ls (3.2) of an approximating sequence of smooth strategies. However, the trading rates $\dot{Q}_t = dQ_t/dt$ blow up in any such approximation. Consequently, one can no longer neglect the impact of the current trades on the corresponding execution price. Therefore, without a more precise micro-description of price impact, the classic self-financing equation breaks down for such approximating sequences. However, the approximation argument carries through without problems if one first recasts the problem in "impact space":

[12] For example, such specifications are used by Gârleanu and Pedersen (2013) and many other academic and practitioner papers.

Theorem 4.1 (Mapping to impact space)**.** *Assume the push factor is of the form $\lambda_t = e^{\gamma_t}$ for a smooth process γ_t. The trader's stochastic control problem* (3.2) *in holdings Q_t is equivalent to the following control problem in "volume impact" J_t:*

$$\sup_{(J_t)_{t\in[0,T]}} \mathbb{E}\Bigg[\int_0^T e^{-\gamma_t}\left(-\mu_t^\alpha J_t + (\tau_t^{-1}+\gamma_t')\alpha_t J_t - \tau_t^{-1}h(J_t)J_t - \gamma_t' H(J_t)\right)dt + e^{-\gamma_T}\left(\alpha_T J_T - H(J_T)\right)\Bigg]. \tag{4.1}$$

(Here, $H(x) = \int_0^x h(y)dy$ is the antiderivative of the price impact function $h(x)$.) For a volume impact process $(J_t)_{t\in[0,T]}$, one recovers the corresponding holdings via the one-to-one map

$$Q_t = \int_0^t \frac{1}{\lambda_s} dJ_s + \int_0^t \frac{1}{\tau_s \lambda_s} J_s ds. \tag{4.2}$$

Only the volume impact states J_t appear in the reformulation (4.1) of the objective function, but not their derivatives. Hence, this representation of the trader's expected P&L naturally extends to general strategies with jumps and/or nontrivial quadratic variation.[13] Moreover, switching the control variable from the trader's holdings Q_t to the corresponding impact state J_t massively simplifies the trading problem. Indeed, in impact space, the goal functional can simply be optimized *pointwise*, circumventing the need for dynamic programming or other advanced methods.[14]

Theorem 4.2 (Pointwise maximization in impact space)**.** *Suppose the integrand in* (4.1) *is strictly concave in volume impact J_t. Then, pointwise maximization determines the optimal J^* as $h(J_T^*) = \alpha_T$ and*

$$0 = -\mu_t^\alpha + (\tau_t^{-1}+\gamma_t')\alpha_t - (\tau_t^{-1}+\gamma_t')h(J_t^*) - \tau_t^{-1}h'(J_t^*)J_t^*, \quad \text{for } t \in [0,T). \tag{4.3}$$

The corresponding optimal holdings can be recovered via (4.2).

4.1. **No Price Manipulation.** We now discuss the wellposedness of the pointwise maximization (4.1). More specifically, we link the concavity of the integrand in (4.1) to the absence of "price manipulation" (i.e., round trip trades with a positive expected cost). For linear price impact, this link was first established by Fruth et al. (2013), who observed:

> "Time-dependent liquidity can potentially lead to price manipulation. In periods of low liquidity, a trader could buy the asset and push market prices up significantly; in a subsequent period of higher liquidity, he might be able to unwind this long position without depressing market prices to their original level, leaving the trader with a profit after such a round trip trade."

The same intuition also applies in the present context with concave price impact. Indeed, differentiating twice shows that the integrand in (4.1) is strictly concave in J_t if

$$2\tau_t^{-1} + \gamma_t' > \max_x \left\{-\frac{h''(x)x}{h'(x)}\right\}. \tag{4.4}$$

[13]For linear price impact models, this extension argument first appears in Ackermann et al. (2021). The self-financing condition for general strategies can in turn be backed out in a second step, see Corollary A.2.

[14]For linear price impact models, this approach has been pioneered by Fruth et al. (2013, 2019). A related change of variable to *integrated* impact is used by Gârleanu and Pedersen (2016); Isichenko (2021).

If this condition is satisfied and the first-order condition (4.3) admits a solution,[15] then the latter identifies the unique optimizer of (4.1).

When the impact function h is the identity as in Obizhaeva and Wang (2013), then the right-hand side of (4.4) is zero, and this wellposedness condition reduces to the no price manipulation condition from Fruth et al. (2013). For general concave impact functions h, the zero lower bound is replaced by the curvature of the impact function, measured by the "Arrow-Pratt measure of relative risk aversion" of h. This lower bound makes it harder to avoid price manipulation. For example, if h follows a power law $\propto x^c$, $c \in (0,1)$, then the lower bound in (4.4) is $1-c$ rather than 0. The condition's interpretation remains: liquidity cannot increase faster than price impact decays: $\tau_t \gamma'_t > -(1+c)$.

Remark 4.3 (Necessary condition)**.** *If the impact function is a pure power law ($h(x) = x^c$ for $x \geq 0$), then Condition* (4.4) *is both necessary and sufficient. Indeed, if* (4.4) *is not satisfied then an explicit price manipulation strategy can be constructed just as in the linear case (Muhle-Karbe et al., 2023, Section 4.3).*

For general concave impact functions $h(x)$ the global concavity condition (4.4) *is only sufficient to rule out price manipulation but not necessary. For example, if the integrand fails to be strictly concave everywhere, then it may still have a unique global maximum, unlike for power law functions.*

4.2. **Examples.** We now discuss the properties of the optimal policies implied by Theorem 4.2.

Corollary 4.4. *Consider the case where the price impact function is a pure power law $h(x) = x^c$ for $x \geq 0$.*[16] *Under the no price manipulation condition $(1+c)\tau_t^{-1} + \gamma'_t > 0$, the optimal impact state then is*

$$I_t^* = \frac{\tau_t^{-1} + \gamma'_t}{(1+c)\tau_t^{-1} + \gamma'_t}\alpha_t - \frac{1}{(1+c)\tau_t^{-1} + \gamma'_t}\mu_t^\alpha. \tag{4.5}$$

The Baseline Scenario. First assume the general alpha signal to be constant over $[0,T]$. For instance, the long-term trader determines their alpha signal at the start of the day and does not update the signal intraday. In particular, the long-term trader assumes no intraday alpha decay. The optimal impact state then equals

$$I_t^* = \frac{\alpha}{1+c}, \quad t \in (0,T); \qquad I_T^* = \alpha.$$

The corresponding smooth trades are

$$dQ_t^* = \frac{\alpha^{1/c}}{\lambda\tau(1+c)^{1/c}}dt, \quad t \in (0,T),$$

and the initial and terminal bulk trades are

$$\Delta Q_0^* = \frac{\alpha^{1/c}}{\lambda(1+c)^{1/c}}; \quad \Delta Q_T^* = \frac{((1+c)^{1/c}-1)\alpha^{1/c}}{\lambda(1+c)^{1/c}}.$$

Unsurprisingly, the optimal trades are a complex non-linear function of the model and the alpha because the price impact model is non-linear. In contrast, when expressed in impact, the optimal trading strategy is a surprisingly concise linear function of alpha.

[15] In particular, such a solution always exists in the empirically relevant case where the price impact functions behave like a power function x^c, $c \in (0,1]$ for large x.

[16] Due to lack of smoothness one cannot apply Theorems 4.1 and 4.2 in this case. However, one can first apply these results to a smoothed impact function as in Example 2.3 and then send the mollification to zero.

This pattern repeats itself in the following sections. Every time one adds model features, both the optimal trades and the optimal impact increase in complexity. However, the trades' complexity increases massively. While numerically implementable, these trade formulas are challenging to grasp and communicate concisely. In contrast, the impact formulas increase only mildly in complexity and remain intuitive as one adds model features.

Adding Alpha Decay. To illustrate this, now assume that α_t is an Itô process with drift μ_t^α. The optimal impact state then equals

$$I_t^* = \frac{1}{1+c}\left(\alpha_t - \tau\mu_t^\alpha\right), \quad t \in (0,T); \qquad I_T^* = \alpha_T.$$

Therefore, a trader considering a dynamic alpha signal only needs to correct their strategy in impact space by the alpha's decay, measured by μ_t^α, relative to the impact decay rate. Furthermore, this adjustment remains linear.

Contrast this simple expression for the optimal impact to the corresponding trading speed over $(0,T)$. For simplicity, assume α is deterministic, so that $\mu_t^\alpha = \alpha_t'$. The optimal trading speed then is

$$dQ_t^* = \frac{(\alpha_t - \tau\alpha_t')^{(1-c)/c}}{\lambda\tau(1+c)^{1/c}}(\alpha_t - \tau^2\alpha_t'')dt.$$

Not only does the formula in trade space depend on higher derivatives of the alpha signal, the complexity increase also compounds with the non-linear relationship and leads to an unwieldy trading formula.

Alpha Decay and Dynamic Liquidity. With alpha decay ($-\mu_t^\alpha > 0$), one must trade more aggressively to exploit the trading signal before it disappears. We now discuss how this tradeoff between alpha level and decay is modulated by changing liquidity conditions (recall that $\gamma_t = \log\lambda_t$, where λ_t is Kyle's lambda, a measure of illiquidity):

(a) When liquidity changes are small, $\gamma_t' \ll \tau_t^{-1}$, the optimal impact state simply adds the alpha level α_t and its decay $-\tau_t\mu_t^\alpha$ over the impact's timescale:

$$I_t^* = \frac{1}{1+c}\left(\alpha_t - \tau_t\mu_t^\alpha\right).$$

(b) When liquidity decreases by a sizable amount, then the alpha level gains in importance. For instance, if $\gamma_t' = \tau_t^{-1}$, then

$$I_t^* = \frac{2}{2+c}\alpha_t - \frac{1}{2+c}\tau_t\mu_t^\alpha.$$

(c) Conversely, when liquidity increases by a sizable amount, then alpha decay gains in importance. For instance, if $\gamma_t' = -\frac{1}{2}\tau_t^{-1}$, then

$$I_t^* = \frac{1}{2+2c}\alpha_t - \frac{2}{1+2c}\tau_t\mu_t^\alpha.$$

Remark 4.5 (Liquidity droughts and floods)**.** *The alpha level and decay tradeoff is agnostic to the* level *of liquidity: only its* changes *matter. Therefore, practitioners should focus on their trade's alpha level during liquidity droughts and their trade's alpha decay during liquidity floods.*

Transition from Linear to Square-root Impact. Let us now discuss how the results above adapt to more general price impact functions such as the crossover from linear to square-root impact in Example 2.3. To this end, the crucial observation is that the optimality condition (4.3) is *local*, in that that the optimal impact state only depends on the corresponding local behavior of the price impact function. For example, if the impact functions is a concatenation of power laws (up to smooth interpolation), then the corresponding optimal impact state switches between the corresponding power law regimes as the alpha signal, its decay, and the liquidity parameters vary over time.

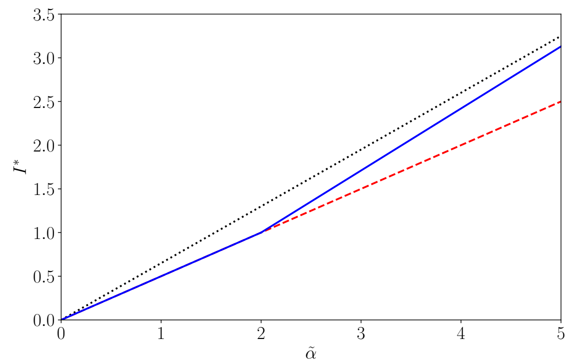


FIGURE 3. The optimal impact I_t^* as a function of the "adjusted alpha" $\tilde{\alpha}_t = \alpha_t - \tau_t \mu_t^\alpha$ for the price impact function $h(x) = \text{sgn}(x)\sqrt{2|x|x_0 - x_0^2}$ from Example 2.3 (solid blue), for the linear model of Obizhaeva and Wang (2013) (dashed red), and for a pure square-root model (dotted black).

To illustrate this, consider the price impact function from Example 2.3 and suppose for simplicity that liquidity is constant ($\gamma_t' = 0$). Then, the first-order condition (4.3) simplifies to

$$0 = \alpha - \tau_t \mu_t^\alpha - h(J_t^*) - h'(J_t^*)J_t^* \quad \text{and in turn} \quad J_t^* = g(\alpha_t - \tau_t \mu_t^\alpha),$$

where g is the inverse of $h(x)+h'(x)x$. For weak "adjusted" alpha signals $0 \leq \alpha_t - \tau_t \mu_t^\alpha \leq x_0$, the relevant part of the price impact function from Example 2.3 then is the linear one, so that the optimal impact is half of the alpha signal. For stronger adjusted alpha signals $\alpha_t - \tau_t \mu_t^\alpha > x_0$, the nonlinear part of the impact function applies. For very large adjusted alpha signals, the latter approaches a square root function so that the optimal impact state tends to two thirds of the alpha signal. For intermediate values of the alpha signal, the optimal impact state smoothly interpolates between these extreme cases, as illustrated in Figure 3.

5. Optimal Trading with Multiple Decay Timescales

Empirical studies have shown that impact decay initially follows a power-law and eventually converges to a permanent level (Brokmann et al., 2015; Bucci et al., 2019a). To approximate such multiscale dynamics, we now consider a more general version of the model from Section 2, where price impact decays at multiple different exponential timescales (and each of the corresponding concavities can potentially also be different).

5.1. **Mapping to impact space.** For simplicity, we focus on a smooth alpha signal $(\alpha_t)_{t\in[0,T]}$. The price impact of a smooth trading strategy $(Q_t)_{t\in[0,T]}$ now is the convex combination of N standard AFS models, with different time scales for impact decay and, potentially, also different concavities:

$$I_t = \sum_{n=1}^{N} w_n h_n(J_t^n). \tag{5.1}$$

Here, the normalized weights satisfy $w_n \in [0,1]$ with $\sum_{n=1}^N w_n = 1$, the impact functions h_n are increasing, odd and concave on $[0,\infty)$ (e.g., power laws), and their arguments J_t^n are exponential moving averages of current and past trades with different decay timescales $\tau_n > 0$:

$$dJ_t^n = -\tau_n^{-1} J_t^n dt + \lambda dQ_t, \quad J_0^n = 0.$$

As in Section 3.2, a risk-neutral trader maximizes alpha capture net of impact:

$$\sup_{(Q_t)_{t\in[0,T]}} \mathbb{E}\left[\int_0^T (\alpha_t - I_t)\, dQ_t\right].$$

For simplicity, we focus on a deterministic alpha signal and constant impact parameters τ_n, λ. Then there is no external randomness, so we can focus on deterministic trading strategies without loss of generality. In view of (5.1), their expected P&L equals

$$\int_0^T (\alpha_t - I_t)\, dQ_t = \sum_{n=1}^N w_n \int_0^T (\alpha_t - h_n(J_t^n))\, dQ_t.$$

For each term in this sum, we now switch to impact space using the change of variable

$$dQ_t = \lambda^{-1}\tau_n^{-1} J_t^n dt + \lambda^{-1} dJ_t^n. \tag{5.2}$$

Then, just like for a single impact decay timescale in Section 4, integration by parts applied separately for each time scale leads to a weighted sum of pointwise optimization problems:

$$\sup_{J^1,\dots,J^N} \lambda^{-1} \sum_{n=1}^N w_n \mathbb{E}\left[\int_0^T \left(\left(\tau_n^{-1}\alpha_t - \alpha_t'\right) J_t^n - \tau_n^{-1} h_n(J_t^n) J_t^n\right) dt + \alpha_T J_T^n - H_n(J_T^n)\right]. \tag{5.3}$$

(Here, $H_n(x) = \int_0^x h_n(y)dy$ is the antiderivative of the price impact function h_n.) However, we cannot simply optimize each term separately in (5.3). Instead, we also have to ensure that the volume impacts J_t^n at each different timescale correspond to the same trades:

$$\lambda^{-1}\left(dJ_t^1 + \tau_1^{-1} J_t^1 dt\right) = \lambda^{-1}\left(dJ_t^n + \tau_n^{-1} J_t^n dt\right), \quad n = 2,\dots,N. \tag{5.4}$$

5.2. **Solution.** We enforce the linear constraints (5.4) using suitable Lagrange penalties η_t^n, $n = 2,\dots,N$ (see Appendix A.1). For given η_t^n, the impact states J_t^n, $n = 1,\dots,N$ decouple, and the trading problem can once again be solved by myopic impact formulas computed via pointwise maximizations. The Lagrange multipliers guaranteeing that the consistency condition (5.4) also holds solve a nonlinear second-order ODE provided in Appendix A.1.

Theorem 5.1 (Optimal impact states)**.** *Given Lagrange multipliers η_t^n, $n = 2,\dots,N$ the optimal volume impacts are*

$$h_1(J_T^1) = \alpha_T + \frac{1}{w_1}\sum_{n=2}^N \eta_T^n, \quad h_n(J_T^n) = \alpha_T - \frac{1}{w_n}\eta_T^n, \quad n = 2,\dots,N, \tag{5.5}$$

and, for $t \in [0,T)$:

$$J_t^1 = g_1\left(\alpha_t - \tau_1\alpha_t' + \frac{1}{w_1}\sum_{n=2}^{N}(\eta_t^n - \tau_1\eta_t^{n\prime})\right), \tag{5.6}$$

$$J_t^n = g_n\left(\alpha_t - \tau_n\alpha_t' - \frac{1}{w_n}(\eta_t^n - \tau_n\eta_t^{n\prime})\right), \quad n = 2,\dots,N. \tag{5.7}$$

(Here, $g_n(x)$ is the inverse of $h_n'(x)x + h_n(x)$.) The trade consistency constraint (5.4) *is satisfied if η_t^n, $n = 2,\dots,N$ solve the system of nonlinear second-order ODEs* (A.4) *with boundary conditions* (A.3)-(A.5)*.*

5.3. **Examples.** By summing over the terminal volume impacts (5.5), we see that the optimal impact generally fully exhausts the alpha signal at the terminal time ($I_T^* = \alpha_T$). At intermediate times $t \in (0,T)$ the representation from Theorem 5.1 simplifies considerably when the price impact function follows the same power law across all decay time scales, which is supported by our empirical results in Section 6:

Example 5.2 (Empirically relevant case)**.** *The data in Section 6 suggests that the concavities at all impact timescales are similar and close to a square-root law ($c_n = 0.5$). The optimal impact state then becomes*

$$I_t^* = \frac{2}{3}\left(\alpha_t - \tau_w\alpha_t' + \sum_{n=2}^{N}(\tau_n - \tau_1)\eta_t^{n\prime}\right), \quad I_T^* = \alpha_T, \tag{5.8}$$

where $\tau_w = \sum_{n=1}^{N} w_n\tau_n$. Thus, the trading strategy behaves as if it traded under the weighted timescale τ_w, plus additional decay terms induced by the Lagrange multipliers.

For the special case of linear price impact ($h_n(x) = x$), the ODEs from Theorem 5.1 become linear, leading to explicit solutions that can be applied to arbitrary alpha signals.[17] The simplest case of two impact decay timescales and constant alpha already illustrates several key effects:

Example 5.3 (OW model with two timescales)**.** *When price impact is linear ($h_1(x) = h_2(x) = x$), then η_t satisfies a* linear *second-order ODE. Indeed, setting*

$$\bar\tau_{w_1} = w_1\tau_1 + w_2\tau_2, \quad \bar\tau_{w_2} = w_2\tau_1 + w_1\tau_2, \quad \bar\tau_{w_3} = (\tau_2 - \tau_1)w_1w_2, \quad \bar\tau = \sqrt{\tau_1\tau_2},$$

the ODE for the Lagrange multiplier then simplifies to

$$\bar\tau^2\eta_t'' - \frac{\bar\tau_{w_1}}{\bar\tau_{w_2}}\eta_t = \frac{\bar\tau_{w_3}}{\bar\tau_{w_2}}\left(\alpha_t + \bar\tau^2\alpha_t''\right), \quad \eta_0' - \frac{1}{\bar\tau_{w_2}}\eta_0 = \frac{\bar\tau_{w_3}}{\bar\tau_{w_2}}\alpha_0', \quad \eta_T' + \frac{1}{\bar\tau_{w_2}}\eta_T = \frac{\bar\tau_{w_3}}{\bar\tau_{w_2}}\alpha_T'.$$

For constant α, this equation has the explicit solution

$$\eta(t) = \alpha\left(C_+e^{Ct} + C_-e^{-Ct} - \frac{\bar\tau_{w_3}}{\bar\tau_{w_1}}\right),$$

where

$$C = \sqrt{\frac{\bar\tau_{w_1}}{\bar\tau^2\bar\tau_{w_2}}}, \quad C_+ = \frac{\bar\tau\bar\tau_{w_3}}{\bar\tau_{w_1}\left(\bar\tau(e^{CT}+1) + \sqrt{\bar\tau_{w_1}\bar\tau_{w_2}}(e^{CT}-1)\right)}, \quad C_- = C_+e^{CT}.$$

[17] Optimization problems with linear impact and general decay kernels are studied using other methods by Gatheral et al. (2012); Abi Jaber and Neuman (2022).

When the trading horizon is long ($T \to \infty$), $C_+ \to 0$ and $C_- \to \frac{\bar{\tau}\bar{\tau}_{w_3}}{\bar{\tau}_{w_1}(\bar{\tau}+\sqrt{\bar{\tau}_{w_2}\bar{\tau}_{w_3}})} > 0$. The optimal impact (5.8) in turn tends to

$$I_t^* = \frac{\alpha}{2}(1 - (\tau_2 - \tau_1)C_- C e^{-Ct}).$$

Over time, this converges to the same stationary level $\alpha/2$ obtained in models with a single decay timescale. However, instead of moving the impact to this level using a single bulk trade at time $t = 0$, the optimal policy with multiple decay timescales consists of a smaller initial jump complemented by a subsequent smooth adjustment. Put differently, for multiple decay timescales the optimal impact for a single time scale is smoothed out to a certain extent. This "transient" build-up of the optimal impact state is required to satisfy the trade consistency conditions (5.4), which depend both on the current levels and the histories of the different volume impact J_t^n and whence would not be satisfied if the overall impact state would immediately be moved to its stationary level by an initial bulk trade like for a single impact decay timescale.

With a finite trading horizon $T < \infty$, a similar smoothening is applied near the terminal time. This is illustrated in Figure 4 for different impact decay timescales and trade durations T. One takeaway from Figure 4 is that two regimes can occur depending on T:

(a) If T is closer to the long timescale τ_2, then the solution behaves like a smoothened version of the optimal strategy on timescale τ_2. The smoothening comes from the temporary buildup of the much faster decaying impact on timescale τ_1, and is reminiscent of optimal impact profiles with instantaneous impact costs.

(b) If T is closer to the short timescale τ_1, then the solution behaves like the optimal strategy on the timescale τ_1 with a linear offset to the impact. The latter stems from the linear impact build-up of the (nearly) permanent second timescale τ_2.

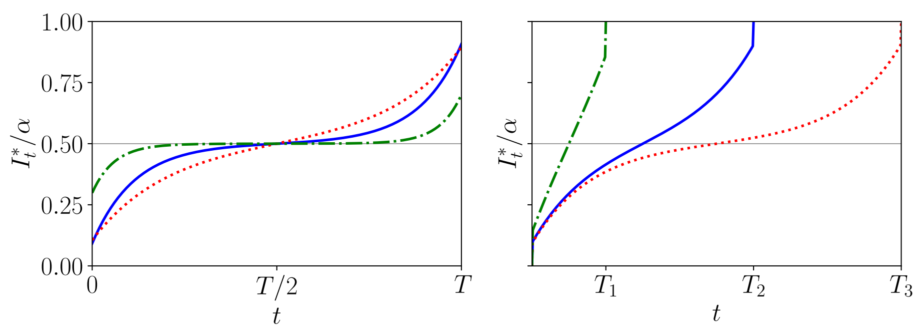


FIGURE 4. Optimal impact I_t^* over time for linear impact $h(x) = x$ and two decay timescales with weights $w_1 = 2/3$, $w_2 = 1/3$. In the left panel the order duration is $T = 70$d and the impact timescales (in days) are $(\tau_1, \tau_2) =$ (1,10) (green dashed), (1,100)(red dotted), and (0.5,65) (solid blue) in line with our empirical estimates. In the right panel T is varied from $T_3 = 50$d to $T_2 = 30$d and $T_1 = 10$d, for $(\tau_1, \tau_2) = (0.5,65)$.

If the price impact is a concave function such as $h(x) = \sqrt{x}$, then the ODE from Theorem 5.1 needs to be solved numerically, but this can be easily implemented in any standard solver. Figure 5 compares the optimal impact of the concave and linear models. We see that the initial and terminal jumps of the optimal impact state are also partially absorbed into smooth trades for square-root impact across multiple timescales. However, the optimal impact profile becomes much more asymmetric, in that the initial bulk trade is reduced to a much larger extent than its counterpart at the end of the trading interval. Moreover, whereas the average impact is the same for linear models with one or two timescales, the presence of a second timescale considerably reduces the average impact with concave price impact.

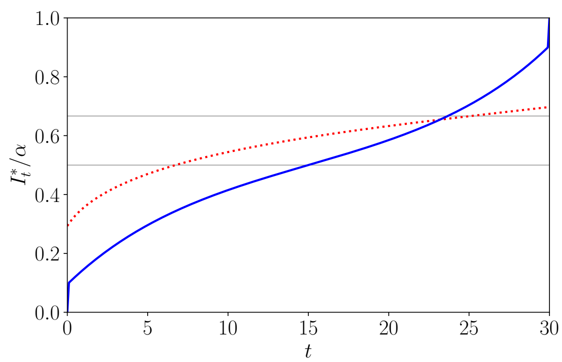


FIGURE 5. Optimal impact I_t^* over time for linear impact $h(x) = x$ (solid blue) and square-root impact $h(x) = \sqrt{x}$ (dotted red), and the optimal impacts $I_t = \alpha/(1+c)$ for a single decay timescale and $c = 1$ and $c = 0.5$ (solid grey).

6. Empirical Results

We now turn to the empirical estimation of the models studied in the previous sections. We seek to answer questions such as "What timescales does price impact decay over?" or "Do all timescales share the same concavity parameter?" Given the difficulty of accessing sufficiently large metaorder datasets for academic researchers or small trading firms, we also quantify to what extent price impact models fit on proprietary data can be recovered using the public trading tape alone.

6.1. **Dataset.** Researchers calibrate price impact models on various datasets to analyze transaction costs at the order or portfolio level. Data may include proprietary orders at large financial institutions (Almgren et al., 2005; Bershova and Rakhlin, 2013; Toth et al., 2017; Frazzini et al., 2018), and public trades on the market tape (Bouchaud et al., 2004; Cont et al., 2013; Chen et al., 2019; Muhle-Karbe et al., 2023). In this paper, we use a proprietary dataset of meatorders provided by CFM that comprises roughly 10^5 metaorders of future contracts traded over 2012-2022. The time at the start and the end of each meta-order is indicated, as well as the mid-price and the number of child-orders. All meta-orders were executed through at least three child-orders and accounted for a fraction between 0.01% and 10% of the average daily volume; the average order size was of the order of 0.1%. No meta-order was traded longer than one day and the average execution time is 3h. In line with Almgren et al. (2005), we normalize trade sizes dQ_t by the average daily traded volume of the respective contracts.

6.2. **Fitting Methodology.** Price returns are fitted against the increments of a power-law price impact of the form

$$I_t = \sigma \sum_{n=1}^{N} w_n \text{sgn}(J_t^n)|J_t^n|^{c_n}, \tag{6.1}$$

for concavities c_n, (unnormalized) weights w_n, and volume impacts J_t^n with decay timescales τ_n.[18] In line with Almgren et al. (2005), we normalize the coefficient for each timescale n using the daily price volatility σ; recall the trade sizes have already been normalize by daily trading volumes.[19] In the following, we consider a grid of impact decays τ_n and concavity parameters c_n and denote the best point estimates by $\hat{\tau}_n$ and $\hat{c}_n$ respectively.

Since i) no meta-order in the sample lasts more than one trading day, and ii) each of them is executed with a profile close to a TWAP, the volume impacts J_t^n are computed under the assumption that trades are executed uniformly during the execution time of each meta-order, and there is no need to consider overnight effects.

To implement the fitting, for each value of the concavity c_n and impact timescales τ_n on the parameter grid, we then compute the respective volume impacts state $h_n(J_t^n)$, and in turn fit their coefficients w_n by a linear regression. The optimal values of the concavity and decay parameters are in turn determined by optimizing over the grid.[20]

Together with our results from the previous sections, this allows researchers to fit nonlinear, multi-timescale models while retaining inherently tractable solutions to statistical arbitrage problems with general nonparametric alpha and liquidity signals. Furthermore, this grid-search provides a sensitivity analysis across different parameter values as a byproduct.

6.3. **Summary of Results.** Table 4 summarizes our parameter estimate across various model specifications. The following sections then delve deeper into different aspects of this analysis.

The main takeaways of our analysis of the "term structure of metaorder impact" are:

(a) Two timescales fit the data well. In order of economic importance, these are a fast timescale measured in hours and a very slow timescale measured in weeks (resembling permanent impact).
(b) Concavity is uniform across all of these timescales, in line with the square-root law.

When the aggregate order flow imbalance is used to create "proxy metaorders" from the public trading tape (see Section 6.6 for more details), then the main takeaways are:

(a) When restricted to models with one timescale, public and proprietary data retrieve the same concavities and similar decay parameters. However, the public trading tape substantially underestimates the magnitude of price impact. Taken at face value, this model misspecification can in fact turn the P&L of an otherwise profitable strategy negative, compare Hey et al. (2023).

[18] For simplicity, we focus on constant liquidity parameters. See Cont et al. (2013); Min et al. (2022); Muhle-Karbe et al. (2023) for empirical studies with dynamic liquidity parameters using the public trading tape.

[19] For power impact functions, the push factor λ_n of each volume impact J_t^N can be absorbed into the *unnormalized* weights, so we set $\lambda_n = 1$ without loss of generality.

[20] A nonparametric approach for estimating general decay kernels for linear price impact is proposed and studied by Neuman et al. (2023).

(b) When fitting multiple timescales, both the impact parameters and the corresponding decay timescales estimated from the public trading data do *not* match their counterparts derived from metaorder data. The public trading tape still underestimates the magnitude of price impact in this case.

(A) One timescale τ

Dataset	c	τ in days	w
Metaorders only	0.5	0.5	1.7
Public market tape	0.5	0.3	1.1

(B) Multiple Timescales $\vec{\tau}$, single concavity $c = 0.5$

Dataset	$\vec{\tau}$	$\vec{w}/\|\|\vec{w}\|\|$	$\|\|\vec{w}\|\|$
Metaorders only	0.5, 65, 7	0.6, 0.3, 0.1	2
Public market tape	0.3, 2, 14	0.55, 0.25, 0.2	1.3

(C) Two timescales $\vec{\tau}$ and concavities $\vec{c}$

Dataset	$\vec{c}$	$\vec{\tau}$	$\vec{w}/\|\|\vec{w}\|\|$	$\|\|\vec{w}\|\|$
Metaorders only	0.45, 0.5	0.5, 65	0.6, 0.4	1.75
Public market tape	0.65, 0.35	2, 0.3	0.75, 0.25	1.3

TABLE 4. Price impact parameter estimates across datasets and models. The elements of the vectors $\vec{\tau} = (\tau_1, ..., \tau_N)$ and $\vec{c} = (c_1, c_2, ..., c_N)$ are sorted by descending weight w_n in $\vec{w} = (w_1, ..., w_N)$.

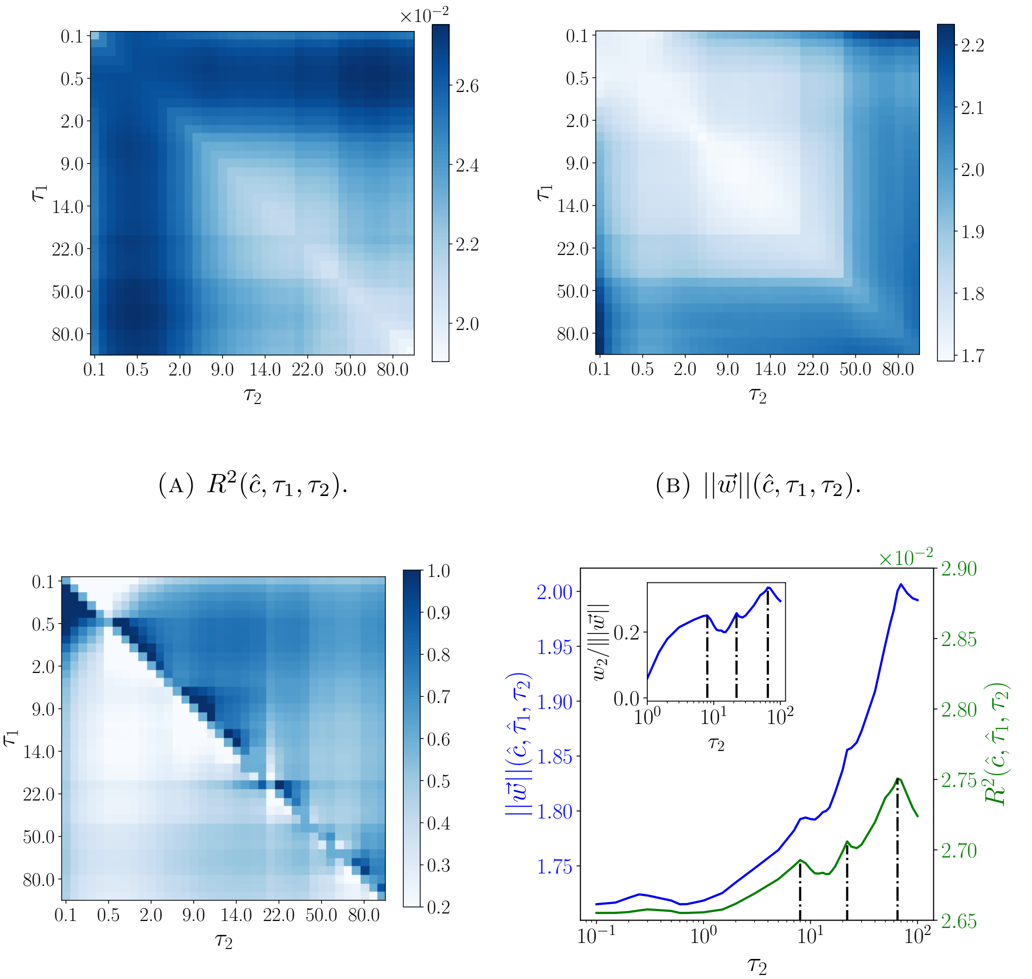
6.4. **Understanding Multiple Timescales.** We now look into the fitting of multiple impact decay timescales in more detail.

Two timescales: To show that impact decays on multiple timescales, we start by fitting two decay timescales τ_1, τ_2 while keeping the same concavity parameter $c_1 = c_2 = 0.5$ fixed. Figure 6a displays a symmetric heatmap of the statistical sensitivity to (τ_1, τ_2). The R^2 peaks at $\hat{\tau}_1 = 0.5$ days, matching the one-dimensional decay fit in Hey et al. (2023).

Figure 6d fixes $\tau_1 = \hat{\tau}_1$ and displays $R^2(\hat{c}, \hat{\tau}_1, \tau_2)$. Three peaks appear at $7, 14$ and 65 days, allowing to significantly increase the fraction of explained variance: despite the modest absolute change (from 2.65% of price variance up to 2.75%), the change is significant – for our sample size, the standard error of $R^2(\hat{c}, \hat{\tau}_1, \tau_2)$ is $5 \cdot 10^{-5}$. The largest improvement obtains for the longest timescale $\hat{\tau}_2 = 65$ days, with associated weight $w_2 = 0.3$. (As for principal component analysis, we henceforth sort the timescales by descending weights w_n.)

This slow impact decay effectively corresponds to *permanent* price impact. Using the ANcerno database, Bucci et al. (2019a) show that permanent impact is about 1/3 of peak impact. Our study matches this since $w_1(\hat{c}, \hat{\tau}_1 = 0.5, \hat{\tau}_2 = 65) = 0.7$, cf. Figure 6c and the inset of Figure 6d.

Three timescales: When the first two timescales are fixed, one fits the third timescale by scanning across candidate τ_3. Figure 7 displays this procedure. $R^2(\hat{c}, \hat{\tau}_1, \hat{\tau}_2, \tau_3)$ peaks at $\tau_3 = \hat{\tau}_3 = 7$ days, in line with the intermediate peak displayed in Figure 6d. The inset of Figure 7 plots the estimated weights w_3 for different τ_3. For the optimal value $\hat{\tau}_3 = 7$ days, $\hat{\tau}_1, \hat{\tau}_2$ contribute about 60% and 30% to the total impact. In contrast, $\hat{\tau}_3$ only contributes



(A) $R^2(\hat{c}, \tau_1, \tau_2)$.

(B) $||\vec{w}||(\hat{c}, \tau_1, \tau_2)$.

(C) $w_1(\hat{c}, \tau_1, \tau_2)/||\vec{w}||(\hat{c}, \tau_1, \tau_2)$.

(D) $R^2(\hat{c}, \hat{\tau}_1, \tau_2)$, $||\vec{w}||(\hat{c}, \hat{\tau}_1, \tau_2)$, and $w_2/||\vec{w}||(\hat{c}, \hat{\tau}_1, \tau_2)$.

FIGURE 6. Calibration results for two decay timescales τ_1 and τ_2 with fixed concavity parameter $\hat{c} = 0.5$. Panel (A) shows the statistical fit; Panel (B) depicts the model's prefactor across τ_1, τ_2. Panel (C) displays the normalized weight of the faster timescale. Panel (D) shows the model fit across τ_2 when the first timescale is fixed to $\hat{\tau}_1 = 0.5$ days (green), as well as the corresponding overall prefactor $||\vec{w}||(c, \tau_1, \hat{\tau}_2)$ (blue) and the normalized weight $w_2/||\vec{w}||$ of the second timescale (inset).

10%. The improvement of R^2 also is only $5 \cdot 10^{-5}$, i.e., similar to the noise level. Given the third timescale's low statistical improvement on our data, we henceforth focus on two timescales.

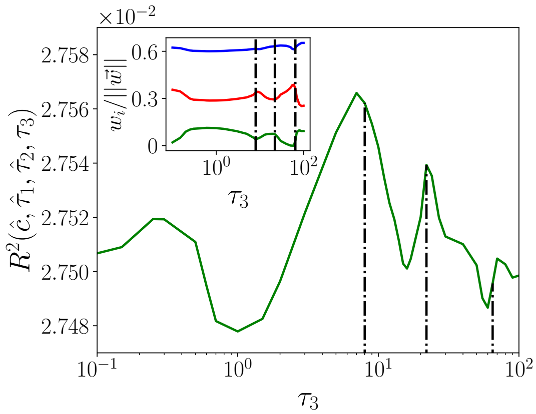


FIGURE 7. Calibration results for third decay timescale τ_3, when the concavity $\hat{c} = 0.5$ and the first two decay timescales $\hat{\tau}_1 = 0.5$ days and $\hat{\tau}_2 = 65$ days are fixed. The main plot shows the statistical fit. The inset panel displays the normalized weights of the third timescale.

6.5. **Understanding Multiple Concavities.** We now lift the assumption that $\hat{c}_1 = \hat{c}_2 = 0.5$ and explore arbitrary combinations of concavity parameters (c_1, c_2). The essential take-away is that varying concavity over timescales does not significantly improve the model for metaorder data.

More specifically, we now fix the two timescales $\hat{\tau}_1, \hat{\tau}_2$ with the largest weights obtained above, but vary the corresponding concavities c_1, c_2. Figure 8 displays the statistical fit for different concavity parameters in the left panel. The heatmap is asymmetric:

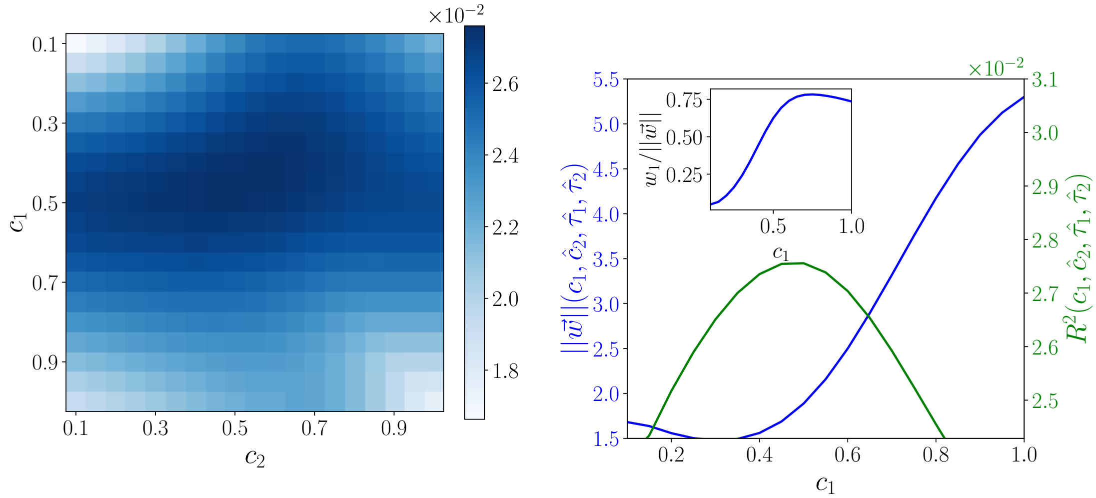
(a) For the longer timescale, $\hat{c}_2 = 0.5$ regardless of the choice for c_1.
(b) For the shorter timescale, $\hat{c}_1 < 0.5$. For example, for $\hat{c}_2 = 0.5$ we obtain $\hat{c}_1 = 0.45$, cf. Subfigure 8b.

However, compared to the previous case where one chooses the same concavity parameter for both decays, the statistical significance R^2 increases only slightly by $5 \cdot 10^{-5}$ from 2.752% to 2.759%, comparable to the standard error $5 \cdot 10^{-5}$ of the R^2. Therefore, multiple concavities only lead to relatively minor improvements in the model fit.

6.6. **Understanding the Public Trading Tape.** Metaorders are proprietary data and are not typically available to academic researchers. Trading firms (especially small ones) may also wish to compare their proprietary trades with the market. Therefore, we now assess to what extent similar results as in the previous section can also be derived from the public trading tape alone.

Many trades on the public tape belong to metaorders and, thus, are expected to have the same price impact. However, the challenge is that the public tape lacks additional metaorder information such as start and end times or even average durations.

To gain a better understanding of how the impact of the aggregate public orderflow compares to metaorders, the aggregate impact I_T can be computed over a bin of length T *conditional* on the orderflow imbalance $\sum_{t=1}^{N_T} \Delta Q_t$ where ΔQ_t is the traded quantity (normalized by average daily volume) of the t^{th} trade out of N_T total trades in the respective bin. Webster (2023) calls this approach "imbalance as an order size proxy". The intuition



(A) $R^2(c_1, c_2, \hat{\tau}_1, \hat{\tau}_2)$.

(B) $R^2(c_1, \hat{c}_2, \hat{\tau}_1, \hat{\tau}_2)$, $||w||(c_1, \hat{c}_2, \hat{\tau}_1, \hat{\tau}_2)$, and $w_1/||w||(c_1, \hat{c}_2, \hat{\tau}_1, \hat{\tau}_2)$.

FIGURE 8. Calibration results in the multi-concavity case for metaorders where $\hat{\tau}_1 = 0.5$ days and $\hat{\tau}_2 = 65$ days. Panel (A) shows the statistical sensitivity in terms of c_1 and c_2; Panel (B) shows the values for the model's prefactor and the statistical sensitivity for $\hat{c}_2 = 0.5$. $\hat{c}_1$ shifts to the left, as $\hat{c}_1 = 0.45$, and the weight $w_1 = 0.6$.

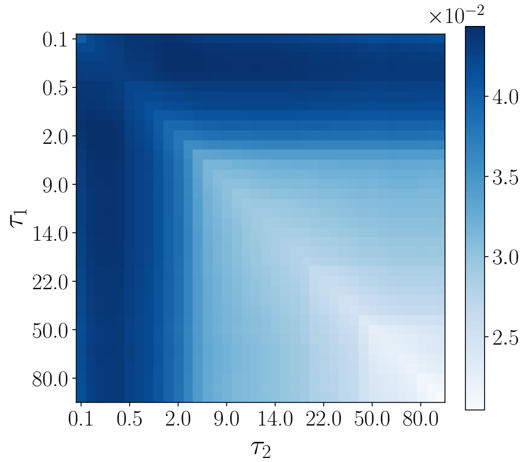
is that sizable orderflow imbalances can serve as a proxy for metaorders when these are not directly available. The proxy metaorders cover the same assets as the proprietary metaorder. We construct them to last for 3 hours each to match the average proprietary metaorder length. We also experimented with 30s metaorders, but find that this leads to quite different results.

Figure 9 collects calibration results for the public trading tape. The essential takeaway is that, when fitting a single timescale using comparable metaorder durations, the public trading tape recovers the same concavity and a similar decay timescale as the proprietary dataset. The public trading tape substantially underestimates the magnitude of price impact, which is problematic for implementing trading strategies. Nevertheless, for academic research, these results suggest that proxy metaorders allow one to obtain parameter estimates of a reasonable magnitude.

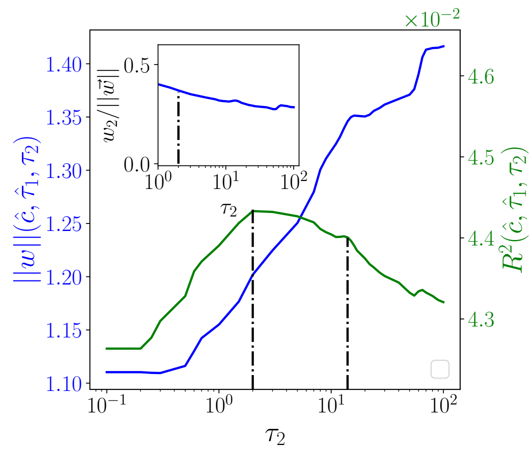
This is no longer the case when fitting multiple timescales and concavities:

(a) The public trading tape struggles to capture long-term decay. Therefore, the additional metaorder information crucially matters when discussing permanent (or slowly decaying) price impact.
(b) For the public trading tape, different timescales seem to have different concavities. In particular, the square-root law does not hold universally, but shorter timescales appear more linear and longer timescales more concave, unlike for metaorder data.

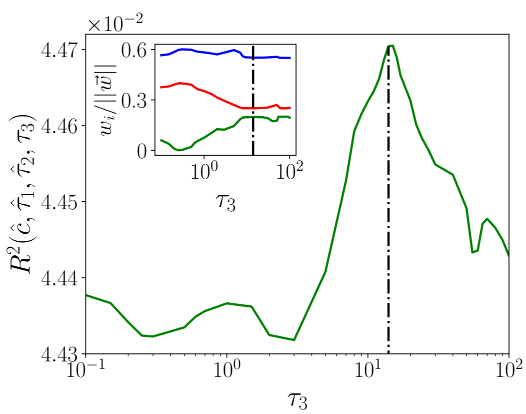
Indeed, for a single concavity but with two decay timescales, we fix $\hat{c} = 0.5$. Figure 10a displays the statistical fit. The peak occurs at $\hat{\tau}_1 = 0.3$ days and $\hat{\tau}_2 = 2$ days (with corresponding normalized weights $w_1/||w|| = 0.55$ and $w_2/||w|| = 0.45$). The next timescale



(A) $R^2(\hat{c}, \tau_1, \tau_2)$.



(B) $R^2(\hat{c}, \hat{\tau}_1, \tau_2)$, $||w||(\hat{c}, \hat{\tau}_1, \tau_2)$, and $w_2(\hat{c}, \hat{\tau}_1, \tau_2)$.

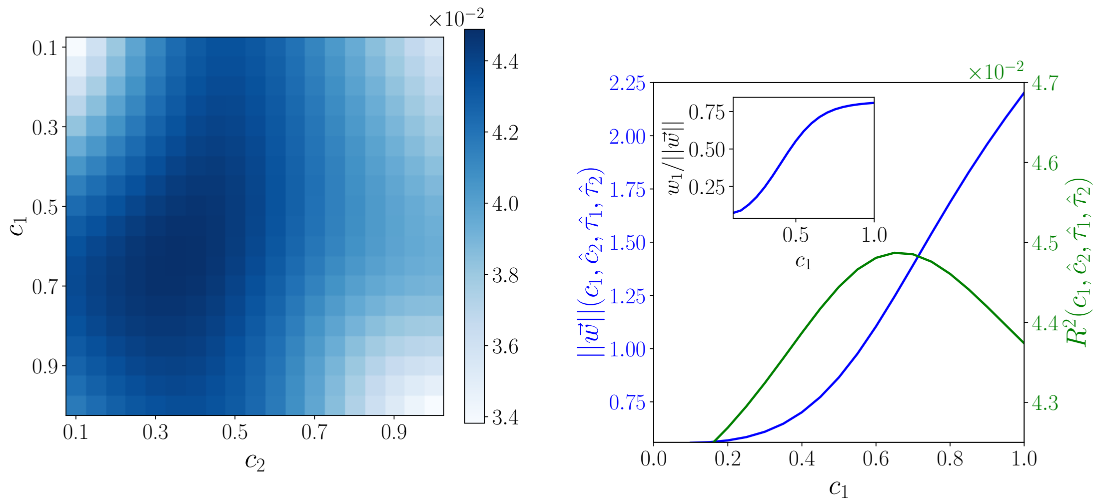


(C) $R^2(\hat{c}, \hat{\tau}_1, \hat{\tau}_2, \tau_3)$ and $w_i/||w||(\hat{c}, \hat{\tau}_1, \hat{\tau}_2, \tau_3)$.

FIGURE 9. Timescale calibration results for the public trading tape. Panel (A) shows the statistical fit across impact timescales τ_1, τ_2. Panel (B) fixes the point estimate $\hat{\tau}_1 = 0.3$ days for the first timescale and displays the statistical fit, weights, and overall level for different values of the second time scale τ_2. Panel (C) fixes the point estimates $\hat{\tau}_1$ and $\hat{\tau}_2 = 2$ days for the first two timescales and plots the statistical fit, weights and overall prefactor for different values of the third time scale τ_3. The price impact function is fixed to a square-root law throughout ($\hat{c} = 0.5$).

is $\hat{\tau}_3 = 14$ days. The three timescales then contribute 55%, 25%, and 20% to the total price impact, a more balanced account than for the metaorder data.

Figure 10 explores a model with two timescales and concavities. The two timescales are fixed to $\hat{\tau}_1 = 2$ and $\hat{\tau}_2 = 0.3$ days matching the descending order of weights $w_1/||w|| = 0.75$ and $w_2/||w|| = 0.25$; the corresponding concavities are in turn estimated as $\hat{c}_1 = 0.65$ and $\hat{c}_2 = 0.35$. The asymmetric graph indicates a strong dependence between the concavity,



(A) $R^2(c_1, c_2, \hat{\tau}_1, \hat{\tau}_2)$.

(B) $R^2(c_1, \hat{c}_2, \hat{\tau}_1, \hat{\tau}_2)$, $||w||(c_1, \hat{c}_2, \hat{\tau}_1, \hat{\tau}_2)$, and $w_1/||w||(c_1, \hat{c}_2, \hat{\tau}_1, \hat{\tau}_2)$.

FIGURE 10. Concavity calibration results for the public trading tape. Panel (A) displays the statistical fit across different values of the concavities c_1 and c_2. Panel (B) fixes the point estimate $\hat{c}_2 = 0.35$ for the concavity of the timescale $\hat{\tau}_2 = 0.3$d with the smaller weight $\hat{w}_2/||\vec{w}|| = 0.35$ and plots the fit together with the overall prefactor against the concavity c_1 of the first decay timescale. The impact decay timescales are fixed to $\hat{\tau}_1 = 2$ days and $\hat{\tau}_2 = 0.3$ days throughout.

weights, and timescale parameters. The longer timescale contributes 75% to the total price impact on the public tape.

6.7. **Conclusion.** In summary, our empirical study based on metaorders suggests that a model with square-root impact across two timescales presents a good compromise between parsimony and accuracy. In order of importance, the decay timescales are of the orders of hours and weeks.

For models with a single timescale, the public trading tape recovers the correct concavity and a reasonable estimate of the fast impact decay but underestimates the magnitude of the price impact. The longer-term impacts are difficult to extract from public data, and the public data also points towards deviations from the square-root law that cannot be found in the metaorder data. This clarifies the scope and the limitations of using publicly available trading data as a proxy for proprietary metaorders.

APPENDIX A. PROOFS FOR SECTION 4

We first establish the myopic representation of the goal functional in impact space from Theorem 4.1. Recall that we are still focusing on smooth trading strategies $Q_t = \int_0^t \dot{Q}_s ds$ at this stage.

Proof of Theorem 4.1. With the one-to-one change of variables $dQ = \tau_t^{-1} e^{-\gamma_t} J_t dt + e^{-\gamma_t} dJ_t$, the goal functional to be maximized in (3.2) can be rewritten as

$$\mathbb{E}\left[\int_0^T (\alpha_t - h(J_t))\tau_t^{-1} e^{-\gamma_t} J_t dt\right] + \mathbb{E}\left[\int_0^T (\alpha_t - h(J_t)) e^{-\gamma_t} dJ_t\right]. \tag{A.1}$$

The first term is already in a form that can be maximized pointwise in J_t. We now recast the second term in such a form, too. To this end, notice that Itô's formula gives

$$e^{-\gamma_T}\alpha_T J_T = -\int_0^T e^{-\gamma_t}\gamma_t'\alpha_t J_t dt + \int_0^T e^{-\gamma_t}\alpha_t dJ_t + \int_0^T e^{-\gamma_t} J_t d\alpha_t$$

(because $J_0 = 0$ and Q_t as well as J_t and γ_t are smooth). Moreover, Itô's formula, $J_0 = 0$ and the smoothness of J_t yield

$$e^{-\gamma_T} H(J_T) = \int_0^T e^{-\gamma_t}\gamma_t' H(J_t) dt + \int_0^T e^{-\gamma_t} h(J_t) dJ_t.$$

By substituting these two identities, the second term in (A.1) can be rewritten as

$$\begin{aligned}
&\mathbb{E}\left[-\int_0^T e^{-\gamma_t} J_t d\alpha_t + \int_0^T \gamma_t'\alpha_t J_t dt - \int_0^T e^{-\gamma_t}\gamma_t' H(J_t) dt + e^{-\gamma_T}\alpha_T J_T - e^{-\gamma_T} H(J_T)\right] \\
&= \mathbb{E}\left[\int_0^T e^{-\gamma_t}\left(-J_t\mu_t^\alpha + \gamma_t'\alpha_t J_t - \gamma_t' H(J_t)\right) dt + \alpha_T J_T - H(J_T)\right].
\end{aligned}$$

Together with the first term in (A.1), this yields the asserted myopic representation (4.1) of the goal functional in impact space. □

The next step is the observation of Ackermann et al. (2021) that, in impact space, the goal functional can be readily extended continuously to the impacts generated by general, not necessarily smooth strategies.

Proposition A.1 (Continuity of the objective function in impact space)**.** *The functional in impact space*

$$\int_0^t e^{-\gamma_s}\Big((\tau_s^{-1} + \gamma_s')(S_t - S_s)J_s - \tau_s^{-1} h(J_s) J_s - \gamma_s' H(J_s)\Big) ds + \int_0^t e^{-\gamma_s} J_s dS_s - e^{-\gamma_t} H(J_t).$$

is continuous in the volume impact $(J_t)_{t\in[0,T]}$ *generated by general semimartingale strategies with respect to the Hilbert norm*

$$||J||^2 = \mathbb{E}\left[\int_0^T e^{-\gamma_t} J_t^2 dt + e^{-\gamma_T} J_T^2\right].$$

Having constructed the goal functional for general non smooth trading strategies in impact space by continuous extension, one can back out the corresponding self-financing condition in a second step. Unlike for smooth strategies, bulk trades or holdings with nontrivial quadratic variation are no longer settled at the impact before the trade. Instead, additional Itô and jump correction terms appear.

Corollary A.2 (Continuous extension of the self-financing equation)**.** *For general semimartingale trading strategies* $(Q_t)_{t\in[0,T]}$ *with volume impact* $(J_t)_{t\in[0,T]}$*, the goal functional* (4.1) *in impact space corresponds to the self-financing equation*

$$Y_t + Q_t S_t = \int_0^t Q_{s-} dS_s - \int_0^t h(J_{s-}) dQ_s - \int_0^t \frac{\lambda_s}{2} h'(J_{s-}) d[Q^c]_s \\ - \sum_{s \le t} \left(\tfrac{1}{\lambda_s} H(J_s) - \tfrac{1}{\lambda_s} H(J_{s-}) - h(J_{s-}) \Delta Q_s \right).$$

Here, the first term corresponds to the usual gains and losses due to exogenous price changes. The second term takes into account that all (even smooth) trades incur the price impact already in place when they are executed. The third term is an Itô correction for diffusive trades.[21] The last term takes into account the extra impact of bulk trades.[22]

A.1. **Proofs for Section 5.**

Proof of Theorem 5.1. The trade consistency constraints (5.4) correspond to the Lagrange penalties

$$\lambda^{-1} \int_0^T \eta_t^n \Big(dJ_t^1 + \tau_1^{-1} J_t^1 dt - dJ_t^n - \tau_n^{-1} J_t^n dt \Big),$$

where η_t^n, $n = 2, \dots, N$ are Lagrange multipliers to be determined. Assuming the Lagrange multipliers η_t^n are smooth and, in particular, do not jump at the initial or terminal time,[23] we can rewrite the Lagrange penalties using another integration by parts as

$$\lambda^{-1} \eta_T^n (J_T^1 - J_T^n) + \lambda^{-1} \int_0^T \Big((J_t^n - J_t^1) \eta_t' + (\tau_1^{-1} J_t^1 - \tau_n^{-1} J_t^n) \eta_t \Big) dt. \tag{A.2}$$

Maximizing the goal functional in its Lagrangian form over $J_T^1, \dots, J_T^N$ directly yields (5.5). Maximizing over $J_t^1, \dots, J_t^N$ for $t \in [0, T)$ gives

$$h_1'(J_t^1) J_t^1 + h_1(J_t^1) = \alpha_t - \tau_1 \alpha_t' + \frac{1}{w_1} \sum_{n=2}^N (\eta_t^n - \tau_1 \eta_t^{n'}),$$
$$h_n'(J_t^n) J_t^n + h_n(J_t^n) = \alpha_t - \tau_n \alpha_t' - \frac{1}{w_n} (\eta_t^n - \tau_n \eta_t^{n'}), \quad n = 2, \dots, N.$$

This in turn leads to (5.6) as well as (5.7).

We now turn to the characterization of the Lagrange multipliers that guarantee that the trade consistency constraints (5.4) hold. To ease notation, set

$$g_t^1 = g_1 \left(\alpha_t - \tau_1 \alpha_t' + \frac{1}{w_1} \sum_{m=2}^N \left(\eta_t^m - \tau_1 \eta_t^{m\prime} \right) \right); \quad g_t^n = g_n \left(\alpha_t - \tau_n \alpha_t' - \frac{1}{w_n} \left(\eta_t^n - \tau_n \eta_t^{n\prime} \right) \right)$$

[21] This corresponds to execution at the average between the prices $S_t + h(J_{t-})$ before and $S_t + h(J_{t-} + \lambda_t dQ_t) \approx S_t + h(J_{t-}) + h'(J_{t-}) \lambda_t dQ_t$ after the trade. Locally, this is analogous to the accounting in the linear model of Obizhaeva and Wang (2013), but with a smaller extra price impact when price dislocations are already large.

[22] The third component of this term cancels the jumps of the stochastic integral $\int h(J_{s-}) dQ_s$. The extra impact cost of a bulk trade ΔQ_t thus is the average $\frac{1}{\lambda_t} \int_{J_{t-}}^{J_t} h(x) dx$ along the price impact function. One can show that this coincides with the costs for such jump trades in the limit-order book model of Alfonsi et al. (2010).

[23] This is a conjecture a priori as the optimal strategy jumps at these times, but turns out to be correct.

for $n = 2, \dots, N$ and similarly for the derivative and inverse functions $g_t^{n\prime}, (g_t^n)^{-1}$, e.g.,

$$g_t^{1\prime} = g_1' \left(\alpha_t - \tau_1 \alpha_t' + \frac{1}{w_1} \sum_{m=2}^{N} (\eta_t^m - \tau_1 \eta_t^{m\prime}) \right).$$

By (5.2), the jumps ΔJ_T^n of all volume impacts must match the optimal strategy's final bulk trade. To achieve this, the Lagrange multiplier have to be chosen to satisfy the terminal conditions (here, we have used that α_t is smooth and the η_t^n are also smooth by assumption):

$$h_1^{-1} \left(\alpha_T + \frac{1}{w_1} \sum_{m=2}^{N} \eta_T^m \right) - g_T^1 = h_n^{-1} \left(\alpha_T - \frac{1}{w_n} \eta_T^n \right) - g_T^n. \tag{A.3}$$

For intermediate times $t \in (0, T)$, the Lagrange multiplier needs to ensure that

$$\tau_1^{-1} J_t^1 dt + dJ_t^1 = dQ_t = \tau_n^{-1} J_t^n dt + dJ_t^n, \quad n = 2, \dots, N.$$

Equivalently:

$$\tau_1^{-1} J_t^1 + \dot{J}_t^1 = \tau_n^{-1} J_t^n + \dot{J}_t^n.$$

After plugging in (5.6) and (5.7) and their derivatives, we see that this is tantamount to

$$\frac{g_t^1}{\tau_1} + \left(\alpha_t' - \tau_1 \alpha_t'' + \frac{1}{w_1} \sum_{m=2}^{N} (\eta_t^{m\prime} - \tau_1 \eta_t^{m\prime\prime}) \right) g_t^{1\prime} = \frac{g_t^n}{\tau_n} + \left(\alpha_t' - \tau_n \alpha_t'' - \frac{1}{w_n} (\eta_t^{n\prime} - \tau_n \eta_t^{n\prime\prime}) \right) g_t^{n\prime}.$$

After rearranging, this leads to the $N-1$ coupled nonlinear second-order ODEs:

$$\begin{aligned} &\frac{\tau_n}{w_n} g_t^{n\prime} \eta_t^{n\prime\prime} + \frac{\tau_1}{w_1} g_t^{1\prime} \sum_{m=2}^{N} \eta_t^{m\prime\prime} - \frac{1}{w_n} g_t^{n\prime} \eta_t^{n\prime} - \frac{1}{w_1} g_t^{1\prime} \sum_{m=2}^{N} \eta_t^{m\prime} \\ &\quad = \tau_1^{-1} g_t^1 - \tau_n^{-1} g_t^n + \left(g_t^{1\prime} - g_t^{n\prime} \right) \alpha_t' - \left(\tau_1 g_t^{1\prime} - \tau_n g_t^{n\prime} \right) \alpha_t'', \quad \text{for } t \in (0, T). \end{aligned} \tag{A.4}$$

The initial condition for this equation is now pinned down by the consistency requirement that the initial jumps $\Delta J_0^n = J_0^n$ all have to match the initial bulk trade of the optimal strategy. In view of (5.6) and (5.7), this requires the initial conditions

$$g_0^1 = g_0^n. \tag{A.5}$$

□

References

E. Abi Jaber and E. Neuman. Optimal liquidation with signals: the general propagator case. *Preprint*, available at arXiv.org, 2022.

J. Ackermann, T. Kruse, and M. Urusov. Càdlàg semimartingale strategies for optimal trade execution in stochastic order book models. *Finance and Stochastics*, 25(4):757–810, 2021.

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

R. F. Almgren. Optimal execution with nonlinear impact functions and trading-enhanced risk. *Applied Mathematical Finance*, 10(1):1–18, 2003.

R.F. Almgren, C. Thum, E. Hauptmann, and H. Li. Direct estimation of equity market impact. *Risk*, 18(7):58–62, 2005.

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

N. Bershova and D. Rakhlin. The non-linear market impact of large trades: Evidence from buy-side order flow. *Quantitative Finance*, 13(11):1759–1778, 2013.

B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the Paris bourse. *Journal of Finance*, 50(5):1655–1689, 1995.

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

J.-P. Bouchaud, D. Farmer, and F. Lillo. How markets slowly digest changes in supply and demand. In *Handbook of Financial Markets: Dynamics and Evolution*, pages 57–160. North-Holland, Amsterdam, 2009.

J.-P. Bouchaud, J. Bonart, J. Donier, and M. Gould. *Trades, Quotes and Prices.* Cambridge University Press, Cambridge, UK, 2018.

X. Brokmann, E. Serie, J. Kockelkoren, and J. P. Bouchaud. Slow decay of impact in equity markets. *Market Microstructure and Liquidity*, 1(2):1550007, 2015.

F. Bucci, M. Benzaquen, F. Lillo, and J.-P. Bouchaud. Slow decay of impact in equity markets: insights from the ANcerno database. *Market Microstructure and Liquidity*, 4(03n04):1950006, 2019a.

F. Bucci, M. Benzaquen, F. Lillo, and J.-P. Bouchaud. Crossover from linear to square-root market impact. *Physical Review Letters*, 122(10):108302, 2019b.

J. A. Busse, T. Chordia, L. Jiang, and Y. Tang. Transaction costs, portfolio characteristics, and mutual fund performance. *Management Science*, 67(2):1227–1248, 2020.

F. Caccioli, J.P. Bouchaud, and J.D. Farmer. A proposal for impact-adjusted valuation: Critical leverage and execution risk. *Preprint*, available at `arXiv.org`, 2012.

R. Carmona and K. Webster. The self-financing equation in limit order book markets. *Finance and Stochastics*, 23(3):729–759, 2019.

Y. Chen, U. Horst, and H. Hai Tran. Portfolio liquidation under transient price impact – theoretical solution and implementation with 100 NASDAQ stocks. *Preprint*, available at `arXiv.org`, 2019.

R Cont, A. Kukanov, and S. Stoikov. The price impact of order book events. *Journal of Financial Econometrics*, 12(1):47–88, 2013.

R. Cont, M. Cucuringu, and C. Zhang. Cross-impact of order flow imbalance in equity markets. *Quantitative Finance*, 23(10):1373–1393, 2021.

M. Coppejans, I. Domowitz, and A. Madhavan. Resiliency in an automated auction. *Preprint*, available at `researchgate.net`, 2004.

H. Degryse, F. De Jong, M. Ravenswaaij, and G. Wuyts. Aggressive orders and the resiliency of a limit order market. *Review of Finance*, 9(2):201–242, 2005.

J. Donier, J. Bonart, I. Mastromatteo, and J.-P. Bouchaud. A fully consistent, minimal model for non-linear market impact. *Quantitative Finance*, 15(7):1109–1121, 2015.

European Commission. Directive 2014/65/EU of the European Parliament and of the Council. Available at `eur-lex.europa.eu/`, 2014.

A. Frazzini, R. Israel, and T. J. Moskowitz. Trading costs. *Preprint*, available at `ssrn.com`, 2018.

A. Fruth, T. Schöneborn, and M. Urusov. Optimal trade execution and price manipulation in order books with time-varying liquidity. *Mathematical Finance*, 24(4):651–695, 2013.

A. Fruth, T. Schöneborn, and M. Urusov. Optimal trade execution and price manipulation in order books with stochastic liquidity. *Mathematical Finance*, 29(2):507–541, 2019.

X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley. Institutional investors and stock market volatility. *Quarterly Journal of Economics*, 121(2):461–504, 2006.

N. Gârleanu and L. H. Pedersen. Dynamic trading with predictable returns and transaction costs. *Journal of Finance*, 68(6):2309–2340, 2013.

N. Gârleanu and L. H. Pedersen. Dynamic portfolio choice with frictions. *Journal of Economic Theory*, 165:487–516, 2016.

J. Gatheral. No-dynamic-arbitrage and market impact. *Quantitative Finance*, 10(7):749–759, 2010.

J. Gatheral, A. Schied, and A. Slynko. Transient linear price impact and Fredholm integral equations. *Mathematical Finance*, 22(3):445–474, 2012.

C. R. Harvey, A. Ledford, E. Sciulli, P. Ustinov, and S. Zohren. Quantifying long-term market impact. *Journal of Portfolio Management*, 48(3):25–46, 2022.

J. Hasbrouck. Measuring the information content of stock trades. *Journal of Finance*, 46(1):179–207, 1991.

J. Hasbrouck and D. J. Seppi. Common factors in prices, order flows, and liquidity. *Journal of Financial Economics*, 59(3):383–411, 2001.

N. Hey, J.-P. Bouchaud, I. Mastromatteo, J. Muhle-Karbe, and K. Webster. The cost of misspecifying price impact. *Risk*, to appear, 2023.

M. Isichenko. *Quantitative Portfolio Management: The Art and Science of Statistical Arbitrage.* John Wiley & Sons, Hoboken, NJ, 2021.

P. N. Kolm and K. Webster. Do you really know your P&L? The importance of impact-adjusting the P&L. *Preprint*, available at `ssrn.com`, 2023.

P. N. Kolm, J. Turiel, and N. Westray. Deep order flow imbalance: Extracting alpha at multiple horizons from the limit order book. *Mathematical Finance*, 33(4):1044–1081, 2023.

F. Lillo, J.D. Farmer, and R.N. Mantegna. Master curve for price-impact function. *Nature*, 421: 129–130, 2003.

T. F. Loeb. Trading cost: The critical link between investment information and results. *Financial Analysts Journal*, 39(3):39–44, 1983.

P. Mackintosh. The 2022 Intern's Guide to Trading. Available at `nasdaq.com`, 2022.

S. Min, C. Maglaras, and C. C. Moallemi. Cross-sectional variation of intraday liquidity, cross-impact, and their effect on portfolio execution. *Operations Research*, 70(2):830–846, 2022.

J. Muhle-Karbe, Z. Wang, and K. Webster. Stochastic liquidity as a proxy for nonlinear price impact. *Operations Research*, to appear, 2023.

E. Neuman, W. Stockinger, and Y. Zhang. An offline learning approach to propagator models. *Preprint*, available at arXiv.org, 2023.

A. Obizhaeva and J. Wang. Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32, 2013.

B. Tóth, Y. Lemperiere, C. Deremble, J. De Lataillade, J. Kockelkoren, and J.-P. Bouchaud. Anomalous price impact and the critical nature of liquidity in financial markets. *Physical Review X*, 1 (2):021006, 2011.

B. Toth, Z. Eisler, and J.P. Bouchaud. The short-term price impact of trades is universal. *Market Microstructure and Liquidity*, 3(2), 2017.

K. Webster. *Handbook of Price Impact Modeling.* CRC Press, Boca Raton, FL, 2023.