INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2025IPPAX039

Thèse de doctorat

ÉCOLE POLYTECHNIQUE

IP PARIS

Trading with Concave (Cross-)Impact

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à l'École Polytechnique

École doctorale n°626 École doctorale de l'Institut Polytechnique de Paris (EDIPP)
Spécialité de doctorat : Physique

Thèse présentée et soutenue à Palaiseau, le 20 juin 2025, par

NATASCHA HEY

Composition du Jury :

Olivier Guéant Université Paris 1 Panthéon-Sorbonne	Rapporteur
Ciamac Moallemi Columbia Business School	Rapporteur
Aurélien Alfonsi École Nationale des Ponts et Chaussées	Président du jury
Eyal Neuman Imperial College	Examinateur
Michael Benzaquen CNRS & École Polytechnique	Directeur de thèse
Jean-Philippe Bouchaud Capital Fund Management & Académie des Sciences	Co-directeur de thèse
Iacopo Mastromatteo Capital Fund Management	Co-encadrant de thèse
Johannes Muhle-Karbe Imperial College London	Invité
Kathryn Zhao Cantor Fitzgerald & Cornell University	Invitée

Acknowledgement

I would like to begin by warmly thanking my three supervisors: Michael Benzaquen, Jean-Philippe Bouchaud, and Iacopo Mastromatteo.

I first met Jean-Philippe during his course on econophysics at ENS. Impressed by his vision, I was fortunate to join the EconophysiX chair at CFM under the supervision of Michael. From the beginning, I felt welcomed, and receiving the WIQF scholarship allowed me to pursue my PhD in this stimulating environment. I deeply admire Jean-Philippe for building a space where academics and practitioners collaborate so freely.

I worked most closely with Iacopo, whose impressive intuition about markets and ability to articulate complex phenomena in simple terms deeply influenced my approach to research. I hope this clarity of thought is evident throughout my thesis.

Michael, as head of the lab, was always present with thoughtful guidance and support. His insights and commitment to his students were essential throughout this journey.

Although not officially a supervisor, Johannes Muhle-Karbe has been a mentor in the truest sense. He encouraged me to write proper mathematical proofs (not always easy as a physicist in a math-finance world!) and supported every major decision I made, including pursuing a postdoc at Columbia University. My decision to remain in academia is largely thanks to him.

I also thank Kevin Webster for his exceptional market understanding and inspiring guidance, which greatly influenced the development of two chapters in this thesis. His profound contributions to the field, particularly through his "Handbook of Price Impact," have been a consistent source of inspiration. I continue to aspire to his level of insight and dedication as both a collaborator and role model.

I am grateful to the members of my jury. Olivier Guéant and Ciamac Moallemi kindly accepted to act as rapporteurs – thank you for your careful reading and feedback. Aurélien Alfonsi and Eyal Neuman, thank you for serving as examiners and for your interest in my work. I am especially honored that Kathryn Zhao joined the jury. When I began my Ph.D., I dreamed of having her on my committee. Seeing her among the top quants worldwide was one of the reasons why I pursued econophysics. In a field with so few women, her presence as a role model means a lot to me and I look forward to more discussions in New York.

I also had the privilege of engaging in constructive exchanges with academic re-

searchers Mathieu Rosenbaum, Thierry Foucault, Jim Gatheral and Eduardo Abi Jaber, whose perspectives greatly enriched my work.

My gratitude goes to the EconophysiX Chair for their generous support. Cecilia, who supervised my first internship, inspired me with her diligence. Nirbhay provided motivation and shared support early in my journey. Fabian, Victor and Swann, I am thankful for the great conversations on science and beyond. Mehdi, I particularly appreciate our in-depth discussions on cross-impact. Guillaume's insights into market impact and calibration have been helpful. To the rest of the lab – Ana, Anirudh, Antoine, Elia, Jerome, Jutta, Karl, Max, Michele, Pierre, Ruben, Rudy, Salma, Samy, and Thomas – thank you for creating such a fun, supportive, and intellectually rich atmosphere.

At École Polytechnique, my gratitude goes to Sandrine for her flawless management of travel arrangements, Lutz for always providing swift responses as the head of LadHyX, and Maxine for her commitment to the bachelor's admission committee, which I was pleased to assist throughout my PhD.

I would also like to extend my appreciation to the people at CFM who consistently provided constructive feedback and support. The quant researchers Julius Bonart, Bence Toth, Raphaël Benichou, and James Ridgway offered invaluable insights. My gratitude extends to the IT team – Kenji Lis, Arthur Hinfray – and the HR team, Pierina Taya and Vincent Smadja, for their patience in addressing my many questions.

In addition, I am grateful to Louis, Nathan, and Shaun for making my time in New York truly memorable; I appreciated our conversations on (concave) price impact, volatility modeling and beyond. My thanks also go to Bruno Dupire for his supervision at Bloomberg. I am grateful for his inquisitive remarks and joyful humor, which enriched my learning and made each interaction memorable.

Finally, I want to express my gratitude to my family and those closest to me. Beginning this PhD under challenging personal circumstances, I would not have made it through without the persistent support of my mother. I also wish to honor my stepfather and grandfather, who both passed away during the last year of my PhD. My stepfather's constant encouragement and my grandfather's generosity in allowing me to stay at his home during my undergraduate years were pivotal in my academic journey. I carry their memory with pride.

Résumé

Dans les marchés financiers modernes, l'exécution d'ordres de grande taille influence significativement les prix des actifs, un phénomène connu sous le nom d'impact de prix. Comprendre et modéliser cet impact est essentiel pour les investisseurs institutionnels, les market makers et les traders algorithmiques, car il affecte directement les coûts d'exécution, les stratégies de trading et l'efficience des marchés. Les modèles analytiquement tractables supposent souvent une fonction d'impact (croisé) linéaire, mais les observations empiriques montrent que l'impact de prix est généralement concave et transitoire, décroissant dans le temps selon une loi de puissance plutôt qu'une décroissance exponentielle.

Cette recherche examine les fondements théoriques et empiriques de l'impact de prix sous quatre angles :

1. Stratégies d'exécution optimales avec un impact concave et transitoire.
La première partie analyse les stratégies optimales lorsque l'impact de prix est non linéaire et transitoire, c'est-à-dire que les transactions passées influencent encore les prix. Les résultats théoriques montrent que des règles de trading simples peuvent être dérivées même pour des signaux d'alpha et de liquidité non paramétriques. L'analyse empirique, fondée sur des données propriétaires, quantifie la concavité et la décroissance de l'impact et leur influence sur l'exécution optimale.

2. Coût de la mauvaise spécification des modèles d'impact en gestion de portefeuille.
Les gestionnaires de portefeuille s'appuient sur des modèles d'alpha pour prévoir les rendements et des modèles d'impact de prix pour estimer les coûts d'exécution. Une mauvaise spécification de l'impact peut conduire à une suroptimisation ou une sous-optimisation des ordres, réduisant ainsi la performance. Cette étude dérive des formules analytiques pour évaluer ces coûts asymétriques : sous-estimer la concavité ou la décroissance réduit les profits, tandis qu'une surestimation peut transformer une stratégie rentable en pertes. Ces résultats fournissent des lignes directrices clés pour calibrer les modèles de trading.

3. Extension aux marchés multi-actifs avec impact croisé.
Dans la pratique, l'exécution d'un ordre modifie non seulement le prix de l'actif concerné, mais aussi celui des actifs corrélés, un phénomène appelé impact croisé. Cette étude étend les modèles d'impact concave au cas multi-actifs, garantissant la cohérence du

marché et excluant toute possibilité d'arbitrage ou de manipulation. En imposant des contraintes minimales, cette approche aboutit à des modèles parcimonieux, calibrables sur des données réelles, offrant une représentation efficace et réaliste de l'impact multi-actifs.

4. Une estimation non paramétrique du cross-impact sur des données propriétaires enrichies par des données publiques.
Les modèles analytiques supposent souvent une décroissance exponentielle de l'impact dans le temps, alors que les données empiriques suggèrent une décroissance en loi de puissance. Cela soulève la question suivante : faut-il imposer une forme fonctionnelle ? Pour répondre à cela, un estimateur non paramétrique de noyaux d'impact concaves et décroissants est généralisé au cas multivarié, en étendant une méthode d'apprentissage hors-ligne au cadre du cross-impact. Afin d'élargir la couverture des données, un proxy synthétique de méta-ordres fondé sur un échantillonnage aléatoire est introduit, en complément des données propriétaires de CFM. Appliqué aux contrats à terme sur le maïs, l'estimateur met en évidence une décroissance en loi de puissance de l'auto-impact, des gains prédictifs liés au cross-impact, ainsi que des asymétries dues aux différences de liquidité.

Cette recherche apporte des avancées théoriques et appliquées en modélisation de l'impact de prix. Elle permet aux acteurs du marché d'affiner leurs stratégies d'exécution, d'améliorer la gestion des risques, et d'adopter des modèles plus robustes face à la complexité des dynamiques d'impact de prix.

Abstract

In modern financial markets, the execution of large orders significantly affects asset prices, a phenomenon known as price impact. Understanding and modeling this impact is crucial for institutional investors, and algorithmic traders, as it directly influences execution costs, trading strategies, and market efficiency. Analytically tractable models often assume a linear (cross-)impact function, but empirical evidence suggests that price impact is typically concave in order size and exhibits transient effects that decay over time, often following a power-law rather than an exponential form.

This research explores the theoretical foundations and empirical properties of price impact, addressing four key aspects:

1. Optimal execution strategies in the presence of concave and transient price impact.

The first part of this research examines optimal trading strategies when price impact is both nonlinear and transient, meaning that past trades continue to influence prices over time. Theoretical results demonstrate that simple and explicit trading rules can be derived even in general settings with nonparametric alpha and liquidity signals. Empirical analysis using proprietary trading data provides key insights into the levels, concavity, and decay properties of price impact, showing how they shape optimal execution strategies.

2. The costs of using a misspecified price impact model in portfolio management.

Portfolio managers rely on alpha models to predict returns and on price impact models to quantify execution costs. However, if the impact model is incorrect, traders may over- or under-trade their signals, leading to suboptimal performance. This study derives explicit formulas for the cost of impact model misspecification and shows that these costs are inherently asymmetric: underestimating impact concavity or decay erodes profits, while overestimating them can turn otherwise profitable strategies into loss-making trades. These insights provide crucial guidance for the construction and calibration of trading models.

3. The extension of concave price impact models to multi-asset settings with cross-impact.

In practice, the execution of trades in one asset affects not only its own price but also the prices of other correlated assets, a phenomenon known as cross-impact. This study

extends existing concave impact models to the multi-asset case, developing a framework that ensures market consistency and rules out arbitrage or price manipulation. By imposing minimal theoretical constraints, this approach leads to parsimonious model specifications that can be effectively calibrated to trading data, providing a realistic and computationally tractable representation of multi-asset price impact.

4. A non-paramteric estimation of cross-impact on proprietary data enhanced with public data.

While analytically tractable models often assume an exponential decay for the persistence of price impact, empirical evidence suggests a power-law decay. This raises the question: why assume any specific functional form at all? To address this, a non-parametric estimator for concave, decaying impact kernels is extended to the multivariate setting, extending an offline learning method to the cross-impact case. To enhance data coverage, a synthetic meta-order proxy based on random sampling is introduced alongside CFM's proprietary trade data. Applied to Corn futures, the estimator reveals power-law self-impact decay, predictive gains from cross-impact, and liquidity-driven asymmetries.

These findings contribute to both academic research and practical trading applications, helping market participants refine their execution strategies and risk management frameworks in the presence of complex price impact dynamics.

List of Publications

Academic papers:

- **N. Hey**, J.-P. Bouchaud, I. Mastromatteo, J. Muhle-Karbe, and K. Webster. *The Cost of Misspecifying Price Impact.* Risk, January 2024.
- **N. Hey**, I. Mastromatteo, J. Muhle-Karbe, and K. Webster. *Trading with Concave Price Impact and Impact Decay - Theory and Evidence.* Operations Research, 2025.
- **N. Hey**, I. Mastromatteo, J. Muhle-Karbe. *Concave Cross-Impact.* Submitted to Management Sciences.

Code and Data:

- *Trading with Concave Price Impact and Impact Decay - Theory and Evidence*: Synthetic meta-order dataset and empirical implementation. `https://github.com/nataschahey/concavepriceimpactanddecay`.

Contents

Chapter 1

Introduction

> *"... what the world needs is ways to execute efficiently and to measure transaction costs."*
> — Robert Almgren (2021)

1.1 Trading costs in modern capital markets

In the high-speed, competitive world of modern finance, execution has become the new battleground. As markets become more efficient and alpha signals diminish in strength and persistence, what ultimately matters is not just finding alpha, but keeping it. Trading costs have moved from being a footnote to becoming a central determinant of portfolio performance. While linear costs such as fees and spreads, are important for high-frequency traders, market impact costs have become more significant for investors concerned with portfolio performance over longer horizons. As Almgren reflects in a podcast on the role of optimal execution, it is vital for all traders in financial markets.

Indeed, the importance of execution is no longer limited to back-office considerations or short-term tactical trades. With the rise of algorithmic trading, which is now responsible for the majority of volume in many liquid markets, execution strategies are at the heart of modern capital markets. As the European Central Bank (2019) notes:

> *"[Algorithmic] trading has been growing steadily since the early 2000s and, in some markets, is already used for around 70% of total orders. This growth has been facilitated by technological developments, such as increased computing power, reduced storage costs and the implementation of artificial intelligence and machine learning techniques."*

Fast, scalable, and theoretically sound execution code is now an essential part of any serious trading operation.

Yet many execution models remain simplistic. Linear models of market impact are analytically convenient, but fail to capture the empirical realities observed in trading data: price impact is concave in order size and decays in complex, multi-timescale patterns. Furthermore, in portfolios and correlated markets, trades in one asset often

influence the prices of others — an effect known as cross-impact. Ignoring this effect can lead to serious execution shortfalls and unintended risk exposures.

This thesis addresses these challenges by developing and empirically validating execution models that incorporate concave self- and cross-impact with decay. Central to this approach is the use of both public and proprietary data to estimate model parameters robustly, and the use of realistic structural constraints, such as the absence of price manipulation, to ensure tractability and the absence of arbitrage in the models. The result is a framework that is both empirically grounded and practically useful; one that allows traders to ask not only "What is my signal?" but "What will it cost to act on it?" and "How can I do so optimally?".

1.2 From alpha to action: how to actually trade it?

These days, hedge funds extract alpha from just about anything that flickers on a screen. Satellite images of parking lots? Check. Twitter sentiment? Of course. The angle of a CEO's smile during an earnings call? Probably. And then there is traditional market data: predictive signals in order flow, liquidity imbalances, volatility clustering — all carefully engineered into alpha.

Enter portfolio management. While the signal-generation team high-fives over a new backtest that "definitely will not overfit this time", the portfolio manager faces the real challenge: turning that alpha into actual returns through execution. This means trading under constraints, in a market that pushes back via *price impact*. The key question becomes:

> *How should a portfolio manager optimally schedule their trades to balance alpha and market impact, especially when impact is concave and slow to fade?*

A key practical challenge is how to translate robust statistical findings into actionable execution strategies by embedding them into a consistent stochastic control framework. To the best of my knowledge, Bertsimas and Lo (1998) was among the first to derive dynamic optimal trading strategies aimed at minimizing the expected cost of trading a large block of equity over a defined horizon. Their work provided a foundational deterministic framework with price-impact functions shaped by market conditions. Following this, Almgren and Chriss (2001) advanced the field by framing algorithmic execution as a control problem, assuming price impact decays instantaneously. They introduced a novel mean-variance approach, allowing for a trade-off between market impact and execution risk, which contrasted with the deterministic approach of Bertsimas and Lo (1998) by addressing stochastic elements in execution strategy optimization. It was soon recognized, however, that price impact has mean-reverting dynamical properties. Bouchaud et al. (2004) introduced the propagator model after empirically observing

from Trades and Quotes data on the Paris stock market that traded prices exhibit random walk properties due to a delicate balance between long-range correlated market orders and mean-reverting limit orders. This model captures the market's response to past trades via a time-varying propagator, often assumed to follow a power-law decay without a fixed decay law. In contrast, Obizhaeva and Wang (2013) introduced the OW model, which uses an analytically tractable linear framework with exponentially decaying impact, derived from a stylized order book model. Their approach was later generalized by Fruth et al. (2013, 2019) to a broader class of impact functions.

While concavity is a key empirical feature of market impact, even basic theoretical properties, such as the absence of price manipulation, remain poorly understood in many models. A notable exception is the concave framework of Alfonsi, Fruth, and Schied (2010) (AFS), where optimal execution is solved explicitly.

In the AFS model, price impact is modeled as a concave function of an exponential moving average of past trades:

$$I_t = h(J_t), \quad dJ_t = -\frac{1}{\tau} J_t dt + \lambda_t dQ_t, \tag{1.2.1}$$

where Q_t is the cumulative trade size, λ_t is the push factor that may depend on the asset's liquidity, τ is the decay timescale, and $h(\cdot)$ is an increasing, concave function, typically modeled to follow square-root behavior. Figure 1.1 illustrates a sample impact trajectory I_t (blue) created through uniform trading over two time intervals T_1 and T_2 under a concave power-law function $h(x) = \text{sgn}(x)|x|^c$, where $c = 0.5$. When the first trade stops, impact decays rapidly, but it is still lingering in the market when the next metaorder starts. This is why it is important to model impact dynamics to take into account the past trades.

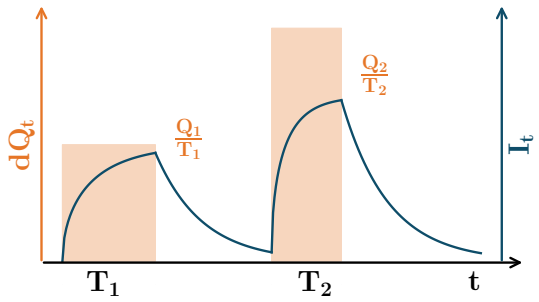
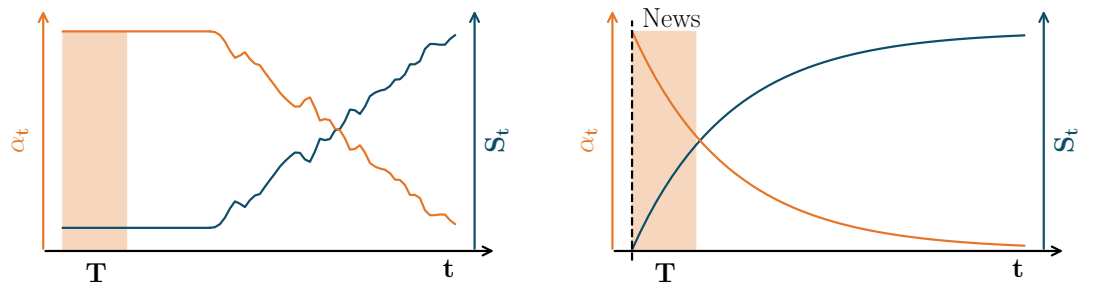


Figure 1.1: Impact trajectory I_t (blue) under a uniform execution profile where dQ_t (orange) is piecewise constant and I_t following the AFS model in Eq. 1.2.1 for $c = 0.5$.

The price forecast is commonly called an alpha signal and we define it as

$$\alpha_t = \mathbb{E}_t \left[S_{T'} - S_t \right]$$

for some time $T' > T$, where T is the execution horizon usually smaller than one day and S_t is the unaffected price. We consider two types of alpha signals to model realistic



(a) Constant: Long-term traders with macro signals do not assume intra-day decay.

(b) Decaying: News signals may decay during execution.

Figure 1.2: Sample trajectories of two alpha signals (orange) corresponding to the change in the fundamental price (blue). The execution of the metaorder takes place during the execution horizon T. Left: The alpha is constant during execution, which corresponds to a macro (long-term) signal. Right: During execution alpha is decaying which is a characteristic of signals generated from earnings news.

trading scenarios, ranging from constant signals $\alpha_t = \alpha$ to those that decay over time, e.g. $d\alpha_t = -\theta^{-1}\alpha_t dt + \sigma dW_t$ where θ characterizes the alpha decay induced by the change in the unaffected price σdW_t. Figure 1.2a shows an alpha-signal that does not decay throughout execution. These could be signals that fundamental traders commonly use over a long prediction horizon. On the other hand, an example for decaying alphas can be signals that are constructed from earnings news; they have a quick decay between open and close of the market which need to be taken into account.

Given the unperturbed price S_t, a trading strategy Q_t and an impact model $I_t(Q)$, a smooth strategy's P&L Y_t has the dynamics.

$$dY_t = Q_t dS_t - I_t dQ_t.$$

The trader's cash balance can be derived by integration as shown in Chapter 2, Section 2.3. The trader's objective is to maximize the risk neutral expected P&L $\mathbb{E}[Y_T]$ which solves the stochastic control problem

$$\sup_Q \mathbb{E}\left[\int_0^T (\alpha_t - I_t)\, dQ_t\right].$$

By switching the control variables from positions Q_t and I_t to the corresponding impact J_t and assuming a push-factor of form $\lambda_t = e^{\gamma_t}$ for a smooth process γ_t, the problem becomes a simple pointwise maximization. Note, we deliberately focus on a risk-neutral objective without an inventory penalty for two reasons. First, this set-up reflects common practice in large hedge funds and principal trading firms, where risk constraints are less restrictive. Second, mapping to impact space, which simplifies the problem into pointwise maximization, relies on the linearity of the objective in Q_t and cannot be applied directly when a quadratic inventory cost is included. Under these constraints, the optimal impact state is computed:

Table 1.1: Transaction Cost Analysis for two different alpha signals that are either constant (Macro) or decay exponentially with a timescale θ (News). Both α and optimal impact I^* are in bps with respect to the initial price. Left: Liquidity is constant ($\gamma'_t = 0$). Right: Liquidity changes.

(a) Constant Liquidity

Strategy	α (bps)	τ/θ	I^* (bps)
A (Macro)	60	0	40
B (News)	90	1/5	72

(b) Varying Liquidity

Strategy	α (bps)	τ/θ	I^* (bps)
Constant Liquidity	90	1/15	64
Dropping Liquidity	90	1/15	72
Raising Liquidity	90	1/15	60

Key Result 1: A tractable solution to a concave impact model with decay

In the AFS model, the optimal impact state is given in closed form by:

$$I_t^* = \frac{(\tau^{-1} + \gamma'_t)\,\alpha_t - \alpha'_t}{(1+c)\tau^{-1} + \gamma'_t}. \tag{1.2.2}$$

where $c < 1$ is the concavity measure, τ is the impact decay timescale and $\alpha'_t = d\alpha_t/dt$. The optimal trade rate dQ_t/dt can then be recovered from I_t^* using Eq. 1.2.1.

The closed-form expression in Equation (1.2.2) illustrates how optimal execution balances several forces: the alpha level, its decay with respect to transient, concave market impact and the changing liquidity in the market. Table 1.1 offers a mock transaction cost analysis across different trading environments to help interpret these results.

Table 1.1 (a) considers two prototypical alpha signals under constant liquidity: a constant and an exponentially decaying alpha. When alpha is constant, the optimal impact simplifies to $I_t^* = \alpha/(1+c)$. For a concavity parameter of $c = 0.5$, this results in spending approximately two-thirds of the available alpha to overcome market impact. In contrast, a trader operating under the false assumption of linear impact ($c = 1$) would estimate higher costs and trade less, thus leaving alpha unexploited. When the alpha signal decays, the optimal impact increases. The trade-off now depends on the relative speed of signal decay versus impact decay, encoded in the ratio τ/θ: more rapidly decaying signals justify more aggressive trading to capture value before it vanishes.

This behavior is illustrated in Figure 1.3. The first and third panels show the optimal impact trajectories for constant and decaying alphas during constant liquidity. Blue lines correspond to $c = 0.5$ (concave) and orange to $c = 1$ (linear). The concave model consistently yields higher optimal impact, reflecting more aggressive trading. In the presence of alpha decay, the optimal strategy accelerates early in the horizon to front-run the signal's decay.

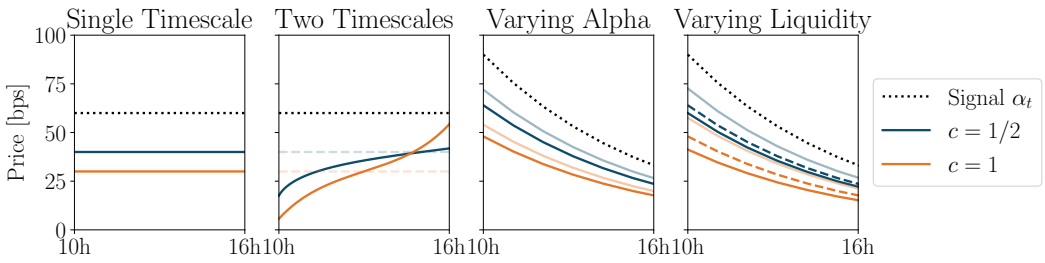


Figure 1.3: Four subplots showing optimal impact I_t^* under and a constant alpha for one (panel one) and two (panel two) impact decay timescale, decaying alpha with one impact decay timescale (panel three) and lastly decaying alpha with varying liquidity (panel four). In all cases, concave impact ($c = 0.5$, blue) leads to more aggressive trading compared to linear impact ($c = 1$, orange). The dashed lines denote the baseline for constant liquidity; deviations from this line indicate the effect of dynamic market conditions.

Table 1.1 (b) explores the effect of changing liquidity, modeled through the term γ_t' in Equation (1.2.2). Importantly, the optimal trade-off between alpha level and decay is independent of the absolute level of liquidity and instead depends on its dynamics. A rising λ_t (worsening liquidity, i.e., $\gamma_t' > 0$) increases the weight of the current alpha level, favoring slower execution. Conversely, improving liquidity conditions ($\gamma_t' < 0$) amplify the effect of alpha decay, calling for faster trading.

The final panel in Figure 1.3 captures this effect visually. The dashed lines show the baseline scenario of constant liquidity, while solid lines indicate time-varying conditions. When liquidity drops (light curves), the optimal impact rises above the baseline; when liquidity improves (dark curves), it falls below. These dynamics demonstrate that execution timing should adjust not only to the alpha signal but also to the evolving state of market liquidity.

1.3 Impact and integrity: price manipulation in execution models

When impact decays and liquidity fluctuates, a dangerous possibility emerges: manipulation. Huberman and Stanzl (2004) were among the first to identify this issue, introducing the concept of quasi-arbitrage in their work. They demonstrated that when the price impact of trades is time stationary, only linear price-impact functions can eliminate quasi-arbitrage opportunities, ensuring viable market prices. Building on these foundational ideas, Gatheral (2010) formulated the price manipulation condition by defining round-trip trades that should inherently bear positive costs. While Huberman and Stanzl (2004) focused on the absence of arbitrage in stationary impact functions, Gatheral's work provided a mathematical framework to ensure that trading strategies could not exploit impact decay for cost-free gains, thus maintaining market integrity.

Fruth et al. (2013) also noted that time-dependent liquidity can potentially lead to price manipulation. In particular, a trader could exploit periods of low liquidity to create upward price pressure and then unwind their position when liquidity improves to effectively generate profits without informational advantage. This leads to a critical requirement: models must be robust to price manipulation.

Some models inadvertently allow price manipulation. In fact, machine learning-based execution strategies occasionally report unrealistically high Sharpe ratios, only to be revealed as exploiting such loopholes in the modeling framework.

> *How can we ensure that execution models remain arbitrage-free, even when impact is transient and liquidity is time-varying?*

An arbitrage-free model only allows round-trip trades $Q_0 = Q_T = 0$ over a trade horizon $[0, T]$ that have positive expected costs, i.e. $\mathbb{E}[\int_0^T I_t dQ_t] > 0$. To prevent arbitrage through liquidity timing, the impact model must remain strictly concave in the control variable.

Key Result 2: No-Price Manipulation Condition

In the AFS-model a sufficient no-arbitrage condition is

$$\tau\gamma_t' > -(1 + c). \tag{1.3.1}$$

This condition provides a simple rule-of-thumb for practitioners and modelers alike: if liquidity improves too rapidly relative to the decay and concavity of impact, manipulation strategies may arise and must be ruled out by design.

1.4 The temporal structure of impact: balancing decay and execution timing

Markets are not only sensitive to the size of a trade, but also to its timing. Once a transaction is executed, its influence on prices does not remain static; it begins to decay, gradually fading as new information and trades enter the market. This transient nature of impact is a fundamental feature of real-world trading, and understanding its temporal structure is key to developing robust execution strategies.

Empirically, price impact decays across a broad range of time scales. Some components vanish quickly, within minutes or even seconds, while others persist for hours or more. While traditional models often assume an exponential decay structure, several studies have shown that this may be overly simplistic. Notably, Bouchaud et al. (2004, 2009a) and Bucci et al. (2015, 2019a) provide evidence of power-law decay in financial markets, suggesting a richer, more nuanced temporal behavior that cannot be captured by a single exponential mode.

This complexity introduces a central challenge for optimal trading: while impact decays over time, so too does the expected return from an alpha signal. Execution strategies must therefore navigate the tension between two competing forces – one pushing to wait and reduce costs, the other urging speed to capture value before the signal dissipates.

What are the dynamical properties of price impact, and how can they be used to balance impact decay against alpha decay in execution strategies?

To address this trade-off, we extend the classical Alfonsi–Fruth–Schied (AFS) framework to a setting with multiple decay rates. Specifically, we introduce a model with N exponential decay kernels, each characterized by its own decay timescale τ_n and concavity parameter c_n. The total impact experienced by the trader is modeled as a weighted sum of these components:

$$\begin{aligned} I_t &= \sum_{n=1}^{N} w_n h_n(J_t^n), \\ dJ_t^n &= -\tau_n^{-1} J_t^n dt + \lambda dQ_t, \quad J_0^n = 0. \end{aligned} \tag{1.4.1}$$

Here, J_t^n denotes the decaying memory of past trading activity on timescale τ_n, λ is the instantaneous impact coefficient, and h_n represents the concave impact function associated with each kernel. The weights w_n reflect the contribution of each mode.

This flexible structure captures multiple layers of decay simultaneously and reflects heterogeneous market responses from both fast and slow liquidity providers. It is particularly useful for modeling how different trading strategies and information flows interact with market depth across timescales.

To empirically calibrate this model, we use a proprietary dataset of $\sim 10^5$ metaorders on futures contracts executed by CFM between 2012 and 2022. This long time span spans different market regimes and evolving execution practices, which raises natural concerns about stationarity. While the model is estimated globally, we emphasize that some parameters, particularly decay rates and push-factor vary over time and across products due to changes in liquidity, market microstructure and CFM's trading behavior.

We also ask whether similar insights can be drawn from public data. Since metaorder datasets are typically inaccessible to researchers and smaller institutions, we assess whether properly aggregated public trade flow can approximate the multi-timescale structure seen in the proprietary data.

For estimation, we normalize all quantities to control for volatility and liquidity seasonality. We then fit a power-law impact model with multiple exponential decay components, as specified in (1.4.1). Estimation proceeds via linear regression over a fixed grid of candidate (τ_n, c_n) values, enabling us to recover the impact profile and relative importance of each timescale.

Key Result 3: Metaorder Impact Decays on Muti-Timescales

Empirical fits reveal that price impact decays over (at least) two dominant timescales:

- A short timescale $\tau_1 \approx 0.5$ day (70% of total impact),
- A long timescale $\tau_2 \approx 65$ days (30%).

Both decay components share the same concavity $c \approx 0.5$, validating the square-root law.

To understand whether public data on the same asset class can be used to calibrate impact models, we apply the same fitting procedure to "proxy metaorders" derived from order flow imbalance on the public trading tape. Proxy trades are constructed over 3-hour bins to match the average metaorder duration in our dataset. The results of the same fitting procedure are

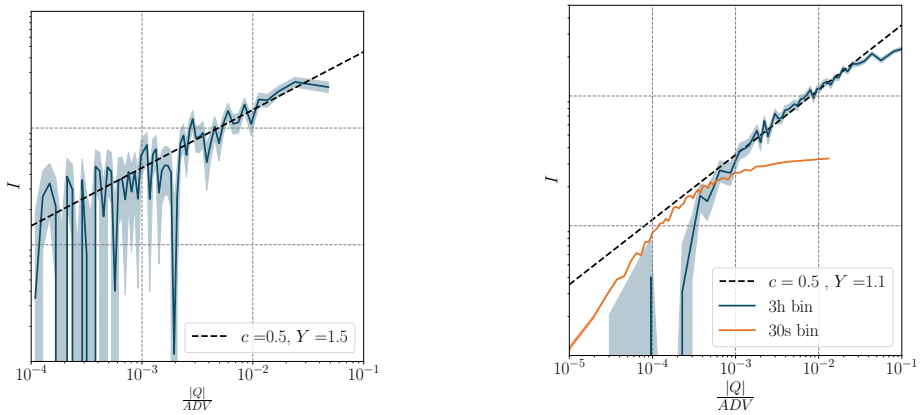
Key Result 4: Public Data Misspecifies Impact Magnitude and Curvature

Calibration on order flow imbalance retrieves similar short-term decay ($\tau \approx 0.3$ days) and concavity ($c \approx 0.5$) for a single-scale model, but:

- It underestimates the overall impact magnitude, leading to smaller weights w_n (see Table 2.4).
- For multi-scale models, it fails to capture long-horizon decay ($\tau_1 = 2$, $\tau_2 = 0.3$ day).
- The curvature is scale-dependent: more linear at short timescales ($c_1 \approx 0.65$), more concave at long timescales ($c_2 \approx 0.35$).

To shed light on these results, Figure 1.4 compares peak impact from proprietary metaorders and public order flow imbalance. The metaorder impact follows a clean power-law with constant concavity, consistent with the square-root law which has been confirmed in Sato and Kanazawa (2024). In contrast, impact from order flow imbalance for two different time intervals (30s and 3h comparable to average metaorder duration) displays a sigmoidal shape that has already been observed in Patzelt and Bouchaud (2018): it increases linearly at small sizes and then saturates, suggesting that the public tape is not scale invariant.

The proprietary dataset reveals a long-term impact decay that aligns with the concept of traded alpha signals, as discussed in Bacry et al. (2015). At a daily time scale, price movements post-metaorder divide into expected returns initiating investment decisions and an idiosyncratic impact that reduces over time. Public datasets, however, mix these idiosyncratic elements, leading to unclear impact decay parameters. This entangle-



(a) Metaorder Impact

(b) Aggregated Market Impact

Figure 1.4: Impact is plotted in log-log scale as a function of (left) the metaorder volume where the average execution horizon is $\bar{T} \approx 3$ h and (right) the order flow imbalance over two different time intervals $T_1 = 30$ s (orange) and $T_2 = 3$ h (blue).

ment results in different decay parameters when using public data, potentially skewing model accuracy. Thus, proprietary data is crucial for capturing the true multi-scale nature of market impacts and ensuring more accurate execution strategies.

Given that the pre-factor can both underestimate and potentially overestimate trading costs depending on the bin size and the asset features, this mismatch implies that using public data for model calibration can lead to overly optimistic or pessimistic assumptions about cost scaling. This is especially relevant when applying such models in automated execution systems, where even minor misestimations can systematically distort trade scheduling over repeated runs. Consequently, execution strategies based on these models can lead to trading too aggressively or too conservatively, ultimately resulting in net financial losses (negative P&L) or in lost opportunities for profit due to overly cautious trading.

By calibrating the correct parameters from the proprietary dataset, we can integrate them into the trader's objective function, which then decomposes into a sum of all impact components:

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

To ensure that all impact components are driven by the same trading trajectory (Q_t), a consistency constraint must be imposed:

$$dQ_t = \lambda^{-1}\tau_n^{-1} J_t^n, dt + \lambda^{-1} dJ_t^n \quad \text{for all } n.$$

This constraint is enforced via time-dependent Lagrange multipliers (η_t^n), leading to the following key result for the optimal impact trajectory.

Key Result 5: Optimal Trading with multi-impact decay

In the multi-timescale AFS model with uniform concavity c, the optimal impact trajectory I_t satisfies:

$$I_t^* = \frac{1}{1+c}\left[\alpha_t - \sum_{n=1}^{N}\left(w_n\tau_n\alpha_t' + (\tau_n - \tau_1)\eta_t^{n\prime}\right)\right]. \tag{1.4.2}$$

The Lagrange multipliers η_t^n enforce consistency across the N moving average states J_t^n so that they all correspond to the same trading strategy Q_t. They solve a (non-) linear ODE system that depends on the timescale configuration and alpha dynamics. While the full solution generally requires numerical methods, the case $c = 1$ (i.e., linear impact) allows for explicit analytical formulas.

Figure 1.3 (b) illustrates the impact trajectory for a two-timescale model for $c_1 = c_2 = 0.5$ and $\tau_1 = 0.5$ and $\tau_2 = 65$ days. Compared to the single-timescale case, the solution no longer consists of block trades at the beginning and end of the interval. Instead, the impact builds up and decays more smoothly, reflecting the ability to exploit the fast-decaying impact early on while gradually offsetting slower components. This effect can be interpreted from equation 1.4.2 where the Lagrange multiplier in the last term couples the two timescales, ensuring a gradual build-up of impact.

1.5 Trading with the wrong model: the cost of misspecifying impact

> *Since all models are wrong the scientist must be alert to what is importantly wrong.*
> — George Box (1976)

No matter how sophisticated the strategy, if it relies on the wrong impact model, it can misfire. Assume linear impact when the reality is concave, and you may overestimate costs. Assume impact decays too fast, and you may trade too aggressively. These mismatches are not benign; instead, they can systematically undermine performance.

> *What happens when traders optimize their strategies using a misspecified impact model, and how costly are these mistakes?*

Having established the optimal execution strategies under the AFS model with one exponential decay timescale, we now ask: *What happens when these strategies are implemented with a wrong price impact model?* In particular, we quantify how misspecifying either the concavity c or the impact decay timescale τ affects profitability.

The benchmark model used to evaluate the cost of misspecification is the AFS model with a single exponential decay timescale and a concave power-law impact function.

Based on empirical calibration results from the previous section, the best-fitting parameters for the price impact kernel for a single exponential decay are $c_* = 0.5$ and $\tau_* = 0.5$ day. These values serve as the ground truth when evaluating how suboptimal it is to implement trading policies optimized under incorrect assumptions.

This framework allows us to precisely isolate and quantify the performance loss incurred when the execution engine relies on incorrect assumptions about market frictions. By doing so, we can assess the sensitivity of optimal execution to each model component. Suppose the true impact model is given by $I_t = \lambda_* \operatorname{sign}(J_t)|J_t|^{c_*}$, where J_t satisfies $dJ_t = -\frac{1}{\tau_*} J_t dt + dQ_t$. Here, λ_* represents the push factor, which is no longer part of the concave function simplifying the notation for subsequent calculations. The trader then optimizes execution based on incorrect parameters $c \neq c_*$ or $\tau \neq \tau_*$. Such a mismatch between belief and reality leads to a suboptimal impact trajectory $J(c, \tau)$, whose associated performance we now evaluate under the true market dynamics. The P&L of the misspecified policy $J(c, \tau)$ under the actual price impact model is:

$$U\Big(J(c,\tau); c_*, \tau_*\Big) = \frac{1}{\lambda_*}\mathbb{E}\Bigg[\frac{1}{\tau_*}\int_0^T \Big[(\alpha_t - \tau_*\alpha_t')J_t(c,\tau) - \lambda|J_t(c,\tau)|^{1+c_*}\Big]\,dt + \alpha_T J_T(c,\tau) - \frac{1}{1+c_*}|J_T(c,\tau)|^{1+c_*}\Bigg],$$

To isolate the roles of misspecifying concavity and decay we denote the realized P&L of such a strategy under the true model as $U(J(c,\tau_*); c_*, \tau_*) = U(J(c); c_*)$ and $U(J(c_*,\tau); c_*, \tau_*) = U(J(\tau); \tau_*)$, respectively.

Key Result 6: Cost of Concavity Misspecification

The expected P&L for the misspecified policy due to an incorrect estimate c of the impact concavity c_* is:

$$U\Big(J(c); c_*\Big) = \frac{\sigma V}{g(c)^{1/c}}\Bigg[\left(\frac{\alpha}{\sigma}\right)^{1+1/c}\left(\frac{T}{\tau_*(1+c)^{1/c}} + 1\right) - \frac{g(c_*)}{g(c)^{c_*/c}}\left(\frac{\alpha}{\sigma}\right)^{(1+c_*)/c}\left(\frac{T}{\tau_*(1+c)^{(1+c_*)/c}} + \frac{1}{1+c_*}\right)\Bigg],$$

where $g(c) = \lambda(c,\tau_*)\cdot V^c/\sigma$, σ is the daily return volatility and V is the daily traded volume.

Implication: It is better to over-estimate c than to under-estimate it. An excessively concave model leads to overly aggressive trading and potential losses.

Key Result 7: Cost of Impact Decay Misspecification

The expected P&L for the misspecified policy due to an incorrect decay τ when the true decay is τ_* is:

$$U\Big(J(\tau);\tau_*\Big) = \frac{\sigma\, V}{\tau_*}\left(\frac{1+\tau/\theta}{g(\tau)(1+c_*)}\right)^{1/c_*}\left(1+\frac{\tau_*}{\theta}-\frac{g(\tau_*)(1+\tau/\theta)}{g(\tau)(1+c_*)}\right) \times \lim_{T\to\infty}\frac{1}{T}\left[\int_0^T\left|\frac{\alpha_t}{\sigma}\right|^{1+1/c_*}dt\right],$$

where $g(\tau) = \lambda(c_*,\tau)\cdot V^c/\sigma$.

Implication: It is better to under-estimate τ than to over-estimate it. Assuming slow decay can result in overly aggressive trading that seeks to exploit inflated prices from impact.

The analysis begins with the significance ratios as depicted in Figures 1.5 and 1.6 on the left. These figures illustrate how the R^2 values are derived from models utilizing different concavity c and decay timescale τ parameters, with the best point estimates identified as $\hat{c} = 0.5$ and $\hat{\tau} = 0.3$ days. For concavity, the significance ratio $R^2(c)/R^2(\hat{c})$ is nearly symmetric around its optimal estimate, suggesting that the fit of the statistical model is relatively stable across a range of c. In contrast, the significance ratio for the decay timescale $R^2(\tau)/R^2(\hat{\tau})$ displays more pronounced asymmetry, indicating a greater sensitivity to variations in the decay parameter.

Next, the impact of these parameter misspecifications on profitability is shown. As illustrated in Figures 1.5 and 1.6 on the right, the P&L ratios reveal a stark asymmetry in economic outcomes. In Figure 1.5 (right), the P&L ratio $U(J(c);c_*)/U(J(c_*);c_*)$ drops sharply when c is underestimated, reflecting losses due to overly aggressive trading under the false belief of weaker price impact. Overestimating c results in more cautious strategies with relatively limited downside. This supports the intuition that assuming a more linear (i.e., stronger) impact is safer from a profitability perspective.

In contrast, Figure 1.6 (right) shows that overestimating the impact decay timescale τ, i.e. believing impact persists longer than it actually does, results in losses. Underestimating τ leads to more conservative strategies which, while suboptimal, do not incur large losses. Again, the plot is asymmetrical, reinforcing the guidance to err on the conservative side in model assumptions.

Together, these results show that high-quality model calibration is essential and statistical goodness-of-fit is not enough. Using the wrong model, even if it is statistically close, can have severe economic consequences.

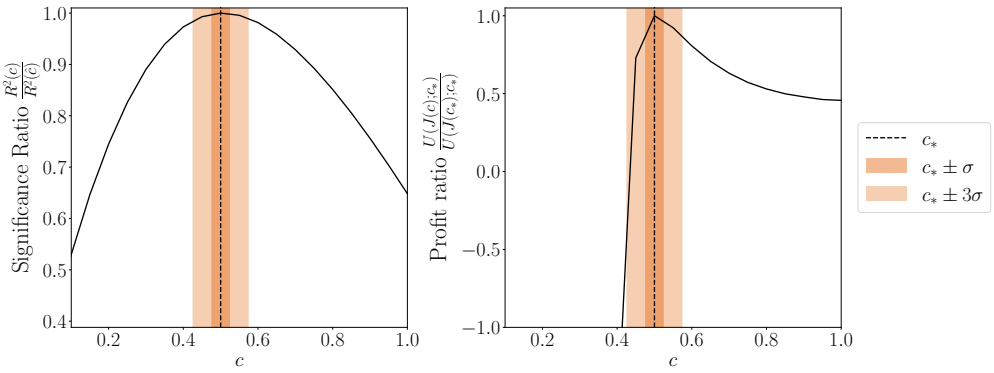


Figure 1.5: Left: Statistical fit $R^2(c)/R^2(\hat{c})$ is symmetric around the best estimate $\hat{c} = 0.5$. Right: Profit ratio $U(J(c); c_*)/U(J(c_*); c_*)$ is highly asymmetric — underestimating c is more costly than overestimating.

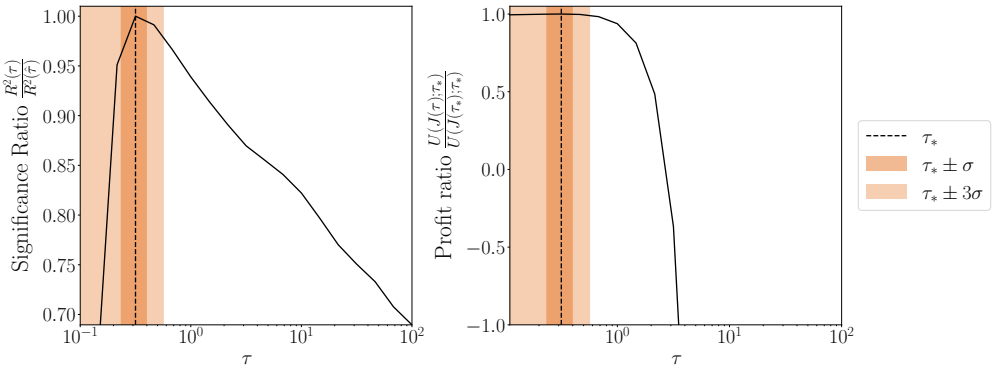


Figure 1.6: Left: Significance ratio $R^2(\tau)/R^2(\hat{\tau})$ is asymmetric around the best estimate $\hat{\tau} = 0.3$ days. Right: Profit ratio $U(J(\tau); \tau)/U(J(\tau_*); \tau_*)$ remains roughly constant for fast decay but drops for overly slow decay estimates.

1.6 When assets talk to each other: impact in portfolios

Most real-world strategies involve trading multiple assets. These assets are often correlated by economic exposure, sector, or investor behavior and trading one can move the prices of others. This effect, known as *cross-impact*, was first measured by Pasquariello and Vega (2013) using daily order imbalances and has since been further explored at the intraday level in studies by Mastromatteo et al. (2017); Wang and Guhr (2017); Tomas et al. (2022a); Le Coz et al. (2024). Additionally, Cont et al. (2021) examined it in the context of combined order imbalance with top levels of the limit order book. These findings suggest that cross-impact should be considered when executing portfolios. All previously mentioned estimates were only performed on linear models with public data, but instead our work made it evident that cross-impact of metaorders is indeed a concave function. Figure 1.7 compares self- and cross-impact estimates on a log-log scale for highly correlated asset pairs, showing that both are clearly concave functions.

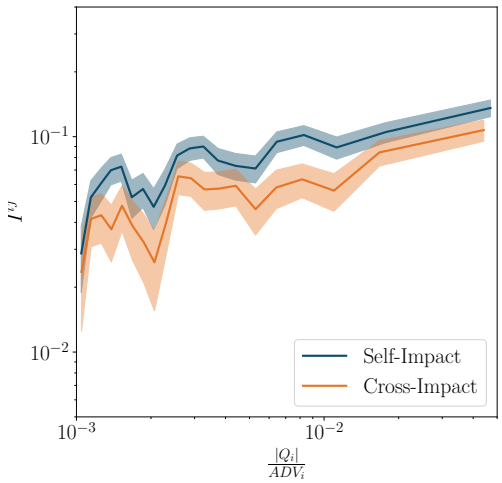


Figure 1.7: Average signed differences between the prices of asset j at the beginning and end of metaorders for asset i, plotted against the size of the metaorders volume fraction $(|Q|/ADV)_i$ in log-log scale. All asset pairs have a return correlation of more that 90%. Self-Impact (blue) corresponds to $i = j$ and cross-impact (orange) to $i \neq j$.

Gârleanu and Pedersen (2016); Schied et al. (2010); Tsoukalas et al. (2019); Horst and Xia (2019); Abi Jaber and Neuman (2022) derive optimal execution principles under linear cross-impact models. However, no derivation has yet been made on how to trade with a concave cross-impact function.

> *How should one optimally balance alpha and concave price impact across a portfolio of correlated assets?*

Motivated by the empirical fact in Figure 1.7, we propose a nonlinear, multivariate generalization of the AFS model. Trades in d risky assets with positions

$\boldsymbol{Q}_t = (Q_t^1, \dots, Q_t^d)^\top$ drive an exponentially weighted moving average

$$d\boldsymbol{J}_t = -\boldsymbol{B}\boldsymbol{J}_t dt + \boldsymbol{\Lambda} d\boldsymbol{Q}_t, \quad \boldsymbol{B}, \boldsymbol{\Lambda} \in \mathbb{R}^{d\times d}. \tag{1.6.1}$$

The price impact across d assets $\boldsymbol{I}_t = (I_t^1, \dots, I_t^d)^\top$ is then given by

$$\boldsymbol{I}_t = \boldsymbol{L} h(\boldsymbol{J}_t), \quad \mathbf{L} \in \mathbb{R}^{d\times d} \tag{1.6.2}$$

where $h(\cdot)$ is applied element wise and typically takes the form $h(x) = \operatorname{sgn}(x)|x|^c$ with $c \in (0, 1]$. The matrices $\boldsymbol{L}$ and $\boldsymbol{\Lambda}$ encode the structure of the liquidity factors and return correlations, whereas $\mathbf{B}$ captures the dynamical properties of impact decay.

To derive the optimal execution strategy, we consider again the risk-neutral objective that needs to be optimized with respect to trade volume:

$$\begin{aligned}\sup_{(\boldsymbol{Q}_t)_{t\in[0,T]}} \mathbb{E}\left[\int_0^T (\boldsymbol{\alpha}_t - \boldsymbol{I}_t)^\top d\boldsymbol{Q}_t\right] = \sup_{(\boldsymbol{J}_t)_{t\in[0,T]}} \mathbb{E}\Big[\int_0^T \left(\bar{\boldsymbol{\alpha}}_t^\top \boldsymbol{J}_t - h(\boldsymbol{J}_t)^\top \boldsymbol{\zeta} \boldsymbol{J}_t\right) dt \\ - \int_0^T h(\boldsymbol{J}_t)^\top \boldsymbol{\theta} d\boldsymbol{J}_t + \boldsymbol{\alpha}_T^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{J}_T\Big].\end{aligned} \tag{1.6.3}$$

where the second equation mapped again to impact space and the following matrices are introduced

$$\boldsymbol{\zeta} = \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{B}, \quad \boldsymbol{\theta} = \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1}, \quad \bar{\boldsymbol{\alpha}}_t = \boldsymbol{\zeta}^\top \boldsymbol{L}^{-1} \boldsymbol{\alpha}_t - \boldsymbol{\theta}^\top \boldsymbol{L}^{-1} \boldsymbol{\alpha}_t'. \tag{1.6.4}$$

The term $h(J_t^i) dJ_t^j$ has no common anti-derivative when $i \neq j$, making the general problem intractable.

Cross-impact models inflate the number of fitting parameters and can exploit arbitrage opportunities if not modeled properly, particularly when liquidity differs across assets. If trades in one asset can move others, then price manipulation can be possible; a trader could push the price of an illiquid instrument and exploit the resulting move in a correlated more liquid one. Abi Jaber et al. (2024); Alfonsi et al. (2016); Schneider and Lillo (2019); Tomas et al. (2022b) study the properties of linear impact models to be arbitrage-free. It was still left to answer

> *What conditions must a concave cross-impact model satisfy to remain manipulation-free?*

These constraints serve as necessary guidelines for building realistic, tractable and arbitrage-free models for multi-asset execution and, in fact, reduce the number of possible models by a significant amount. In the framework of the multivariate AFS-model we derive conditions on $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ that ensure the absence of price manipulation.

Once the no-manipulation conditions are imposed, the optimization problem becomes well defined. However, it still involves a nonlinear and generally non-convex objective function. In this setting, closed-form solutions can only be derived under

structural assumptions that simplify the interaction between assets.

Key Result 8: No-Manipulation Constraints

Absence of price manipulation implies:

- For symmetric round-trip strategies that the matrix $\boldsymbol{\zeta}$ satisfies

$$0 < \bar{\zeta}_{aa} + \phi^{1+c} - \phi^c \bar{\zeta}_{ab} - \phi \bar{\zeta}_{ba}, \quad \forall \phi \geq 0, \tag{1.6.5}$$

  where $\bar{\zeta}_{ab} = \zeta_{ab}/\zeta_{bb}$. If $\boldsymbol{\zeta}$ has positive eigenvalues, then this condition always holds.

- For impulsive pump-and-dump strategies that the matrix $\boldsymbol{\theta}$ satisfies

$$0 < -\theta_{ab}(\mu_a - \mu_b) - \theta_{ba}(\mu_b - \mu_a) \left(\frac{\sigma^b}{\sigma^a}\right)^{1-c} \left(\frac{(\sigma^a)^2 + c(\sigma^b)^2}{c(\sigma^a)^2 + (\sigma^b)^2}\right)^{3/2}. \tag{1.6.6}$$

  where σ_a and σ_b control the trade speed and μ_a and μ_b the execution time in asset a and b respectively. This requires $\boldsymbol{\theta}$ to be symmetric if $c = 1$ and diagonal if $c < 1$

Key Result 9: Optimal Trading Strategies in Solvable Cases

- **Decoupled Case:** If $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ are both diagonal,

$$\boldsymbol{I}_t^* = \frac{1}{1+c}\left(\boldsymbol{\alpha}_t - \boldsymbol{L}\boldsymbol{\zeta}^{-1}\boldsymbol{\theta}^\top \boldsymbol{L}^{-1}\boldsymbol{\alpha}_t'\right). \tag{1.6.7}$$

- **Bivariate Case:** With two assets, the optimal solution satisfies a cubic equation in $\phi_t = J_t^1/J_t^2$:

$$0 = \phi_t + \operatorname{sgn}(\phi_t)|\phi_t|^c k_t^1 + \operatorname{sgn}(\phi_t)|\phi_t|^{c-1} k_t^2 + k_t^3. \tag{1.6.8}$$

  where k_t^1, k_t^2, k_t^3 are functions of $\boldsymbol{\zeta}$ and $\bar{\boldsymbol{\alpha}}_t$. Uniqueness of the solution cannot be guaranteed.

- **General Case with Small Cross Terms:** For small off-diagonal elements $\zeta_{ij} = \zeta, \quad \forall i \neq j$,

$$\boldsymbol{I}_t^* \approx \sum_{a=1}^{d} \boldsymbol{L}_a \frac{\bar{\alpha}_t^a}{\zeta_{aa}(1+c)} \left[1 - \sum_{b \neq a} \frac{\operatorname{sgn}(\bar{\alpha}_t^b)\zeta}{c(1+c)} \left(c\left|\frac{\bar{\alpha}_t^b}{\bar{\alpha}_t^a}\right|^{1/c} + \left|\frac{\bar{\alpha}_t^b}{\bar{\alpha}_t^a}\right|\right)\right]^c. \tag{1.6.9}$$

If both matrices $\boldsymbol{\theta} = \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1}$ and $\boldsymbol{\zeta} = \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1}\boldsymbol{B}$ are diagonal, the multivariate optimization problem decomposes into a system of d independent univariate problems. This case, although seemingly restrictive, still captures rich cross-impact effects through the structure of $\boldsymbol{L}$ and $\boldsymbol{\Lambda}$, such as when both are derived from the asset return covariance

matrix. Although the optimization problem decouples, the trading strategy itself still reflects cross-asset interactions: price impact in each asset aggregates effects from all latent liquidity factors.

When $d = 2$, the pointwise maximization problem remains analytically tractable despite the presence of off-diagonal cross-impact. The solution is governed by a nonlinear implicit equation for the ratio of the two impact states. The resulting problem admits a finite number of candidate solutions (e.g., six for $c = 1/2$), although uniqueness may not always hold. In practice, this regime is particularly relevant as cross-impact between two assets already arises in many execution settings (e.g., calendar spreads, sector hedging).

When $\boldsymbol{\zeta}$ is close to diagonal (i.e., off-diagonal entries are small), the multivariate objective remains strictly concave on a compact set, ensuring the existence and uniqueness of the optimizer. The first-order conditions can be solved perturbatively using the implicit function theorem. The solution shows how small levels of cross-impact lead to lower optimal impact states when latent alpha signals are aligned, which in turn allows more efficient trading across correlated instruments.

1.7 From model to market: estimating cross-impact empirically

The dynamic decay of self-impact has been widely documented as mentioned before. But the properties of cross-impact concavity and decay were still not fully evaluated, which leads us to ask:

> *What are the dynamical properties of cross-impact induced by metaorders, and how does it depend on return correlation?*

The no-manipulation constraints, derived earlier, ensure that only an arbitrage-free model can be calibrated, which is crucial for execution strategies that must avoid market manipulation. Additionally, these constraints significantly reduce the number of fitting parameters, enabling us to achieve the first reliable estimation of a concave, dynamic cross-impact model. This approach allows for the joint calibration of the concavity, decay, and amplitude of cross-impact using proprietary metaorder data.

For simplicity and to ensure a tractable calibration procedure that respects the no-manipulation conditions, we restrict our focus to fitting impact pairs and assuming the matrix of decay rates is in a diagonal form, $\boldsymbol{\theta}^{-1}\boldsymbol{\zeta} = \beta \cdot \boldsymbol{Id}_2$, which introduces a single shared decay parameter β across the two liquidity factors. This assumption decouples the latent impact dynamics across factors and leads to the following simplified model:

$$\boldsymbol{I}_t = \sum_{a=1}^{2} \boldsymbol{L}_a \, h(\theta_{aa}^{-1}) \, h(J_t^a), \quad \text{where } J_t^a = \int_0^t e^{-\beta(t-s)} d(\boldsymbol{L}^T \boldsymbol{Q}_s)^a. \tag{1.7.1}$$

where only the 2 diagonal elements of θ need to be fitted. There still remains flexibility in choosing $\boldsymbol{L}$ to satisfy the no price manipulation constraints. To ensure that

the model admits desirable properties, such as treating fully correlated assets as identical and the ability to handle large heterogeneity in liquidity, we derive a method inspired by del Molino et al. (2020), which bases $\boldsymbol{L}$ on total order flow and return covariances.

Using proprietary metaorder data, we quantify concave cross-impact and its decay dynamics across asset pairs. Our main empirical findings are as follows:

Key Result 10: Measuring Concave Cross-Impact

- Cross-impact is concave.
- Cross-impact decays over time.
- Cross-impact contributions grow with correlation.

The best-fit concavity lies between $c = 0.5$ and 0.7, consistent with the square-root law. The decay is well captured by an exponential function with rate $\beta \in [0.1, 0.9]$, corresponding to half-lives of 0.7 to 7 days. For highly correlated pairs, cross-impact can represent up to 50% of total trading costs.

1.8 When data is not enough: synthetic metaorders and data enhancement

Estimating impact kernels, especially cross-impact, requires a large amount of data. But in many cases, proprietary datasets contain only sparse metaorder activity across assets. This makes a reliable estimation difficult. Inspired by the empirical findings of Maitrier et al. (2025b); Sato and Kanazawa (2024) that the square-root law can be recovered when knowing the Trader ID in the public trade data and reassembling metaorders, Maitrier et al. (2025a) introduced a novel method to generate synthetic metaorder data. These metaorders have not been used to measure cross-impact yet, so the question arises

> *Can we enhance the dataset through synthetic metaorders or public order flow?*

To answer this question, a non-parametric impact model is introduced next to estimate cross-impact on a sparse dataset.

1.9 Beyond the model: estimating impact without assumptions

Most execution strategies begin with an assumption about impact, i.e. linear, exponential, permanent, transient. But with sufficient data, we can now let the data speak for itself. In Neuman and Zhang (2023), a non-parametric estimation method for a linear impact kernel was introduced, where the estimation occurs in an online learning framework. Neuman et al. (2023) proposed an offline estimator based on historical trade and

price data and derived optimal convergence rates for this estimator. This framework allows to estimate an impact kernel non-parametrically, directly from trade data.

> *What does price impact look like when we do not assume a model – can impact decay really described by a power-law?*

In the univariate case, the observed price P_{t_i} evolves as a function of past trade sizes Q_{t_j} via a discrete-time convolution model:

$$P_{t_{i+1}} - P_{t_1} = \sum_{j=1}^{i} G_{i-j}\, h(Q_{t_j}) + \epsilon_{t_i}, \quad i = 1, \ldots, M, \tag{1.9.1}$$

where $h(x) = \mathrm{sgn}(x)|x|^c$ is a concave impact function and G_{i-j} represents a convolution kernel evaluated at lag $i-j$ and there are M timesteps in total. No assumption is made on the shape of the kernel beyond convexity constraints. Estimation is performed via constrained least squares with regularization, and a confidence bound is derived for the resulting estimator. Non-parametric kernel estimation reveals the empirical structure of price impact without assuming a model.

To overcome data scarcity in the CFM proprietary data, we design a metaorder proxy mechanism that reconstructs synthetic metaorders from public trade data. The augmented dataset is used to stabilize the estimation and reveal temporal structures in the kernel that are otherwise obscured by noise.

Key Result 11: Impact Without Structural Assumptions

- Concave impact functions ($c \approx 0.6$) consistently outperform linear ones in predictive accuracy;
- The estimated impact kernel decays smoothly over time and follows a power-law form;
- Proxy metaorders improve kernel smoothness and capture long-range temporal effects otherwise obscured by noise.

Figure 1.8 shows how kernel estimation improves when the dataset is augmented with synthetic metaorders. The estimate based solely on proprietary Corn futures data (orange) is noisy and irregular, whereas the enhanced dataset (blue) yields a smoother propagator that more clearly captures the temporal decay of impact.

1.10 From one to many: a non-parametric estimate of a cross-impact model

Although most real-world strategies involve portfolios of correlated assets, most empirical and theoretical work on price impact has focused on single-asset settings or makes

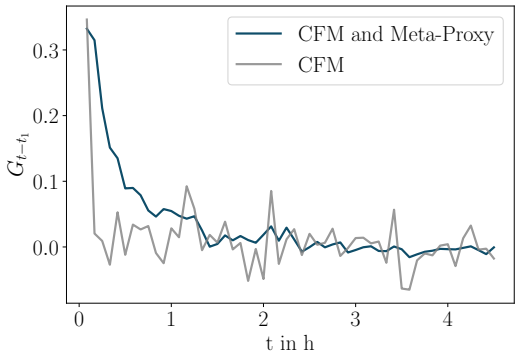


Figure 1.8: Estimated price impact kernel G_{t-t_1} using CFM proprietary data only (gray) and the enhanced dataset including synthetic metaorders (blue).

strong assumptions on the cross-impact model. Cross-impact has been observed empirically in Le Coz et al. (2024) where they reveal clear asymmetries: liquid assets exert stronger influence on illiquid ones than the reverse. Yet, existing models often impose restrictive structural assumptions, such as linearity Tomas et al. (2022a), symmetry Hey et al. (2024b), or specific decay forms Hey et al. (2025), limiting their descriptive power and flexibility.

In a sense, what we seek is a non-parametric lens through which to observe the true shape of cross-impact including its inherent asymmetries driven by liquidity imbalances. This step is critical if we hope to relax the structure of current parametric models, just as we extended the AFS model in the single-asset case to capture decay across multiple time scales.

To that end, we extend the non-parametric estimator of Neuman et al. (2023) to the multivariate setting with concave price impact. This yields the first fully data-driven estimation of cross-impact kernels across multiple assets without assuming linearity, symmetry, or functional decay. The only structural constraint imposed is the convexity of the convolution kernel, which is a necessary condition for ruling out price manipulation.

> *How does cross-impact behave in practice when we do not impose properties or assume a model and how can we estimate it reliably in a multivariate setting?*

To estimate cross-impact, we model the return of an asset $a_1 \in \{0, d\}$ via a multivariate convolution of lagged volumes:

$$P^{a_1}_{t_{i+1}} - P^{a_1}_{t_1} = \sum_{a=1}^{d} \sum_{j=1}^{i} G^{(a_1,a)}_{i-j}\, h(Q^a_{t_j}) + \epsilon^{a_1}_{t_i}, \quad i = 1, \dots, M, \tag{1.10.1}$$

where $h(x) = \operatorname{sgn}(x)|x|^c$ is the concave impact function and $G^{(a_1,a)}_{i-j}$ is the impact of past trades in asset $a \in \{0, d\}$ on asset a_1. The price impact contribution therefore becomes

a weighted sum of trades over all d assets under consideration.

The kernel is estimated from N metaorders across d assets by solving a regularized least-squares problem:

$$(\boldsymbol{G}_v)_{N,\lambda} := \arg\min_{\boldsymbol{G}_v \in \mathcal{G}_{\text{ad}}} \left(\sum_{n=1}^{N} \left\| \mathbf{y}^{(n)} - \mathbf{U}^{(n)} \boldsymbol{G}_v \right\|^2 + \lambda \left\| \boldsymbol{G}_v \right\|^2 \right), \tag{1.10.2}$$

where $\mathbf{U}^{(n)}$ encodes the lagged, concave-transformed volume vectors across all assets for metaorder n, $\mathbf{y}^{(n)}$ is the vector of observed returns, $\boldsymbol{G}_v$ is the vectorized form of the multivariate kernel belonging to a set of admissible kernels $\mathcal{G}_{\text{ad}}$ that enforces convexity and decay and λ is the regularization parameter. The findings on the confidence bound and the empirical estimation are summarized as follows:

Key Result 12: Estimating a Non-Parametric Cross-Impact Kernel

- The estimation error satisfies a confidence bound that scales in $\mathcal{O}(\sqrt{Md^2 \log(1/\lambda)})$, highlighting the increased data requirement for multivariate kernels;
- Including cross-impact terms improves predictive accuracy: out-of-sample R^2 increases from 4.8% (self-only) to 6.3% (2 assets) and 6.5% (3 assets);
- The estimated kernels reveal asymmetries: more liquid contracts exert stronger and more persistent impact on less liquid ones.

These results confirm that cross-impact is not only present but informative: including other assets' trade flow improves forecasts. Moreover, relaxing symmetry assumptions reveal structural asymmetries aligned with market microstructure: front-month contracts tend to lead, and less liquid contracts absorb more impact. The framework provides a practical tool for portfolio execution and lays the groundwork for optimal control in concave multivariate impact models.

1.11 Outline

This manuscript is organized as follows.

Chapter 2 introduces the theoretical framework of concave price impact with transient effects. Section 2.3 formulates the optimal trading problem as a stochastic control problem, and Section 2.4 derives explicit solutions by mapping the problem into impact space, discussing price manipulation constraints, risk aversion. Section 2.5 generalizes the model to multiple decay timescales. The chapter concludes with an empirical analysis in Section 2.6 that validates these theoretical results on proprietary data and also explores the impact properties when using the public tape in Section 2.6.6.

Chapter 3 focuses on the consequences of model misspecification. In practical set-

tings, execution strategies often rely on simplified impact assumptions. This chapter quantifies the cost of using an incorrect model by first introducing an analytically tractable control problem in Section 3.2 and then estimates the parameters of the impact model on proprietary data in Section 3.3. The sensitivity of performance to misestimated concavity and decay is analyzed in Section 3.4 where explicit formulas for performance loss and highlights the asymmetry in misspecification costs are derived.

Chapter 4 extends the single-asset framework to the multi-asset setting with concave cross-impact. The AFS model is extended to a multi-dimensional framework in Section 4.2 and the trader's objective function is presented in Section 4.3. Since this function is not generally pointwise solvable, Section 4.5 discusses sufficient conditions for the absence of price manipulation in this nonlinear cross-impact setting. Section 4.6 provides a decomposition of solutions to the optimal control problem. The empirical section 4.7 applies the model to metaorder data and identifies the dependence of cross-impact with respect to return correlation. It also validates that cross-impact is indeed concave and it decays.

Chapter 5 turns to a non-parametric estimation of cross-impact kernels. Rejecting predefined decay shapes, Section 5.2 extends an offline learning method to estimate impact directly from metaorder data. To address fitting problems such as data scarcity, the proprietary dataset is enhanced by synthetic metaorders whose synthetic generation is introduced in 5.3.2. Other aspects of the fitting procedure are introduced in 5.3.3. The results in 5.3.4 confirm power-law decay in self-impact, document asymmetric cross-impact, and demonstrate that correlation between assets amplifies cross-effects. This chapter bridges the gap between flexible empirical fitting and the more structured models of previous chapters.

Chapter 2

Trading with Concave Price Impact and Impact Decay

Summary

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.

Based on Hey et al. (2025): ***N. Hey****, I. Mastromatteo, J. Muhle-Karbe, and K. Webster. Trading with Concave Price Impact and Impact Decay – Theory and Evidence. Operations Research, 2025.*

2.1 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

[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.

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]

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 derive these 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.

[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., Hasbrouck (1991); Biais et al. (1995); Coppejans et al. (2004); Degryse et al. (2005); Bouchaud et al. (2009b); Bacry et al. (2015); Brokmann et al. (2015) and the references therein.

[4]Stochastic liquidity parameters are a tractable proxy for the "local concavity" of the price impact of individual trades (Muhle-Karbe, Wang, and Webster, 2024), 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 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.

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 trade-off between accuracy and parsimony. For the corresponding concavities, the best power-law specification is close to a square-root law across all timescales. These empirical results emphasize that the scope of our theoretical analysis is not merely an academic exercise, but incorporates salient features of empirical data.[7]

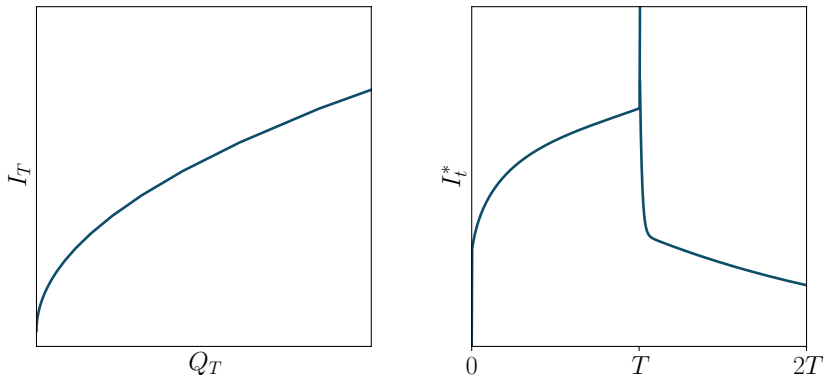


Figure 2.1: Optimal impact for constant alpha signals as a function of optimal traded volume (left panel) and time (right panel).

Using our theoretical results, optimal trading strategies taking into account these empirical features can be derived in a straightforward manner. Figure 2.1 illustrates their main qualitative properties. 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 2.1 displays the temporal evolution of the optimal impact state I_t for a constant alpha signal during

[7] Previous empirical studies of impact concavity often focus on "static" settings where the price change during the metaorder is regressed against its size without taking into account impact decay. Conversely, impact decay is typically studied in propagator models where concavities are considered at the fill- rather than the metaorder level. The empirical results we report provide consistent estimates for both effects in the AFS framework.

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: there is an initial jump caused by an initial block trade, then impact increases gradually until another jump corresponding to a second block trade at the terminal time T. After the completion of the order, this peak impact decays. 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 decays very quickly, but the remaining impact decays much more slowly, in that it is still substantial at time $2T$, for example.

Outline

This paper is organized as follows. Section 2.2 introduces the price impact model, and Section 2.3 formulates the corresponding risk-neutral stochastic control problems. Section 2.4 describes the explicit optimal strategies that can be obtained by changing variables to "impact space", and Section 2.5 extends this method to multiple impact decay timescales. These theoretical results are complemented by the empirical analysis in Section 2.6, where the models are fit to proprietary trading data. For better readability, the derivations of the theoretical results are collected in Appendix A. Finally, Appendix B outlines to what extent fitting results obtained from proprietary metaorder data can be reproduced using suitable proxies derived from the public trading tape alone.

Notation

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

2.2 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 independently 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 suppress the dependence on the random state $\omega \in \Omega$ from now on as usual.)

Alfonsi, Fruth, and Schied (2010) (henceforth AFS) proposed a price impact model that captures the nonlinear and transient nature of price impact while remaining analytically tractable:[8]

Definition 2.2.1 (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 is defined as an exponential moving average of current and past trades*

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

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 *instantaneous* 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 2.2.2. *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_t *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.*

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*

[8] 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.

latent portion of the demand and supply curve that has not yet materialized into the visible order book.

Example 2.2.3. *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 2.2.

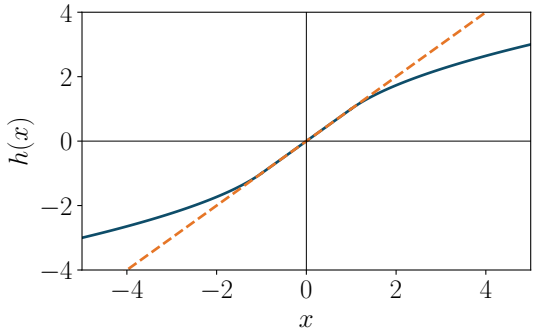


Figure 2.2: The price impact function $h(x)$ from Example 2.2.3 for $x_0 = 1$ (solid blue) and the identity function $h(x) = x$ (dashed orange) from the linear impact model of Obizhaeva and Wang (2013).

2.3 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. (2024), but we do not spell this out here to not overload the notation.

2.3.1 Self-Financing Equation

When the large trader's holdings vary smoothly, i.e., $Q_t = \int_0^t \dot{Q}_s ds$ 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 con-

tinuous trading on $[0,t]$ is given by[9]

$$Y_t = -\int_0^t (S_u + I_u) dQ_u = -Q_t S_t + \int_0^t Q_u dS_u - \int_0^t h(J_u) dQ_u.$$

Here, the first integral describes the standard gains and losses from frictionless continuous-time trading. The second integral accounts for additional price impact costs.[10]

2.3.2 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 2.3.1 (Risk-neutral intraday trading)**.** *A risk-neutral* intraday trader *solves the stochastic control problem*

$$\begin{aligned} &\sup_Q \mathbb{E}\left[Y_T + Q_T S_T\right] \\ &\quad = \sup_Q \mathbb{E}\left[\int_0^T Q_t dS_t - \int_0^T h(J_t) dQ_t\right]. \end{aligned} \tag{2.3.1}$$

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 (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 (2.3.1)*.*

Remark 2.3.2. *The portfolio optimization problem* (2.3.1) *has three main limitations: (i) there is a fixed terminal time T, (ii) positions still held at time T are valued at the unaffected price rather than penalized with an additional liquidation cost (or required to be zero), and (iii) there are no size or risk constraints on positions held over time.*

Whereas the fixed time horizon is natural for optimal execution problems, one can relax it for "stat arb" problems by passing to the ergodic version of the problem (2.3.1)*. This means that one considers the performance per unit time, and then postpones the terminal time indefinitely.*

Requiring the terminal position to be zero can be incorporated through a Lagrange multiplier. This does not complicate the analysis much if all model parameters are

[9] 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 2.4 avoids the corresponding cumbersome bookkeeping equations by directly approximating the performance of such strategies "in impact space".

[10] 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 high-frequency strategies that are even more capacity constrained due to the fast decay of the corresponding trading signals, spread rather than impact costs become the main concern.

deterministic, but rules out explicit solutions with stochastic liquidity parameters for example. Requiring liquidation over a longer time horizon, directly, imposing a liquidation cost, or adding a penalty on inventory held over time all make it impossible to reduce the maximization to a pointwise problem as in Section 2.4. We comment on the main qualitative changes induced by such risk penalties for models with linear price impact in Section 2.4.3 below.

A helpful statistic to simplify the risk-neutral portfolio optimization problem (2.3.1) is the so-called *intraday alpha signal*[11]

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

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 (2.3.1) in turn can be written as

$$\sup_Q \mathbb{E}\left[\int_0^T (\alpha_t - h(J_t))dQ_t\right]. \tag{2.3.2}$$

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 (2.3.1) can be extended to incorporate such long-term views as follows:

Definition 2.3.3 (Risk-neutral long-term trading)**.** *A risk-neutral* long-term trader *solves the stochastic control problem* $\sup_Q \mathbb{E}\left[Y_T + S_{T'}Q_T\right]$ *for some time* $T' > T$.

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

$$\alpha_t = \mathbb{E}_t\left[S_{T'} - S_t\right],$$

but the end-of-day constraint $\alpha_T = 0$ no longer applies in this case. With this notation, the representation (2.3.2) 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.

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 2.3.5 and 2.3.6 below specify standard parametric alpha signals. Example 2.3.7 covers the increasingly common non-parametric case. In particular, this applies to modern machine-learning approaches in alpha research and showcases the advantage of our explicit trading formulas from Theorem 2.4.2 for general non-parametric signals.

[11] Some traders refer to α_t as the *cumulative* alpha over $[t, T]$.

Example 2.3.4 (Constant alpha)**.** *Constant signals arise naturally when modeling trading catalysts. Indeed, suppose that on the trading interval* $[0, T]$*, the price* S_t *is a martingale and reveals no further information about a price change that happens afterward,* $\mathbb{E}_t[S_{T'} - S_T] = \mathbb{E}[S_{T'} - S_T] > 0$*. Then:*

$$\alpha_t = \mathbb{E}\left[S_{T'} - S_T\right], \quad \textit{for all } t \in [0, T].$$

Signals based on financial catalysts such as earnings announcements naturally exhibit such sharp boundaries.

Example 2.3.5 (Deterministic alpha)**.** *The trading signal* α_t *can be deterministic even if the unaffected price* S_t *is stochastic. For instance, a constant drift rate of the unaffected 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* $\alpha_t = \mu(T' - t)$*.*

Example 2.3.6 (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*

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

For a long-term trader predicting the steady-state alpha $(T' \to \infty)$*, we have* $\alpha_t = \theta\mu_t$*. In particular, both the alpha and the alpha decay are stochastic in this case.*

Example 2.3.7 (Non-parametric alpha)**.** *Examples 2.3.5 and 2.3.6 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, Cont et al. (2021); Kolm et al. (2023) use various neural network architectures to construct 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.*

2.4 Solution by Mapping to Impact Space

We now turn to the solution of the risk-neutral trader's optimization problem (2.3.2). Section 2.3.2 derives the objective function (2.3.2) for smooth strategies $Q_t = \int_0^t \dot{Q}_s ds$.

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

This is sufficient when trading signals α_t vary smoothly over time. However, diffusive trading signals as in Example 2.3.6 naturally lead to diffusive trades. Moreover, block 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 (2.3.2) 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, it is not straightforward to extend the classic self-financing equation to the limits of such approximating sequences. However, the approximation argument carries through without problems if one first recasts the problem in "impact space":

Theorem 2.4.1 (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* (2.3.2) *in holdings Q_t is equivalent to the following control problem in "volume impact" J_t:*

$$\sup_{(J_t)_{t\in[0,T]}} \mathbb{E}\Bigg[\int_0^T e^{-\gamma_t}\Big(-\mu_t^\alpha J_t + (\tau_t^{-1}+\gamma_t')\alpha_t J_t - \tau_t^{-1}h(J_t)J_t - \gamma_t' H(J_t)\Big)dt + e^{-\gamma_T}\left(\alpha_T J_T - H(J_T)\right)\Bigg]. \tag{2.4.1}$$

[13] *For a volume impact process $(J_t)_{t\in[0,T]}$, one recovers the corresponding holdings via the one-to-one map*

$$Q_t = \int_0^t \frac{1}{\lambda_s} dJ_s + \int_0^t \frac{1}{\tau_s\lambda_s} J_s ds. \tag{2.4.2}$$

The myopic representation of the goal functional in impact space is derived in 2.A.1 . Only the volume impact states J_t appear in the reformulation (2.4.1) 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. 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 2.A.2. 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 (2.4.1) can simply be optimized *pointwise*, circumventing the need for dynamic programming or other advanced methods.[14]

[13] Here, $H(x) = \int_0^x h(y)dy$ is the antiderivative of the price impact function $h(x)$.

[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).

Theorem 2.4.2 (Pointwise maximization in impact space)**.** *Suppose the integrand in* (2.4.1) *is strictly concave in volume impact J_t. Then, pointwise maximization determines the optimal J^* as $h(J_T^*) = \alpha_T$ and*

$$\begin{aligned} 0 = &- \mu_t^\alpha + (\tau_t^{-1} + \gamma_t')\alpha_t - (\tau_t^{-1} + \gamma_t')h(J_t^*) \\ &- \tau_t^{-1} h'(J_t^*) J_t^*, \quad \textit{for } t \in [0, T). \end{aligned} \tag{2.4.3}$$

The corresponding optimal holdings can be recovered via (2.4.2)*.*

2.4.1 No Price Manipulation

We now discuss the wellposedness of the pointwise maximization (2.4.1). More specifically, we link the concavity of the integrand in (2.4.1) to the absence of "price manipulation" (i.e., trades that generate expected profits not from accurate price forecasts but from turning price impact into gains). 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, if

$$2\tau_t^{-1} + \gamma_t' > \tau_t^{-1} \max_x \left\{ -\frac{h''(x)x}{h'(x)} \right\}. \tag{2.4.4}$$

then differentiating twice shows that the integrand in (2.4.1) is strictly concave in J_t. When this sufficient condition is satisfied and the first-order condition (2.4.3) admits a solution,[15] then the latter identifies the unique optimizer of (2.4.1). In particular, without an alpha signal, the optimizer is $J_t = 0$. This means that without a predictive signal about future price changes it is not possible to make positive expected profits by manipulating the price impact dynamics.

When the impact function h is the identity as in Obizhaeva and Wang (2013), then the right-hand side of (2.4.4) 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. Just as risk aversion in utility theory affects economic behaviors by guiding investment and consumption decisions based on wealth changes, this curvature informs trading strategy considerations indicating how aggressively the market impact diminishes as trade size increases.

[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.

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 (2.4.4) 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 2.4.3 (Necessary condition)**.** *If the impact function is a pure power law ($h(x) = x^c$ for $x \geq 0$), then Condition* (2.4.4) *is both necessary and sufficient. Indeed, if* (2.4.4) *is not satisfied then an explicit price manipulation strategy can be constructed just as in the linear case (Muhle-Karbe et al., 2024, Section 4.3).*

For general concave impact functions $h(x)$ the global concavity condition (2.4.4) *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.*

2.4.2 Examples

We now discuss the properties of the optimal policies implied by Theorem 2.4.2.

Corollary 2.4.4. *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*

$$I_t^* = \frac{\tau_t^{-1} + \gamma_t'}{(1+c)\tau_t^{-1} + \gamma_t'} \alpha_t - \frac{1}{(1+c)\tau_t^{-1} + \gamma_t'} \mu_t^{\alpha}. \tag{2.4.5}$$

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 block trades are

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

[16] Due to lack of smoothness one cannot apply Theorems 2.4.1 and 2.4.2 in this case. However, one can first apply these results to a smoothed impact function as in Example 2.2.3 and then send the mollification to zero.

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.

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 trade-off 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):

1. 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:

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

2. When liquidity decreases by a sizable amount, then the alpha level gains in im-

portance. For instance, if $\gamma_t' = \tau_t^{-1}$, then

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

3. 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 2.4.5 (Liquidity droughts and floods)**.** *The alpha level and decay trade-off 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 2.2.3. To this end, the crucial observation is that the optimality condition (2.4.3) is *local*, in 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.

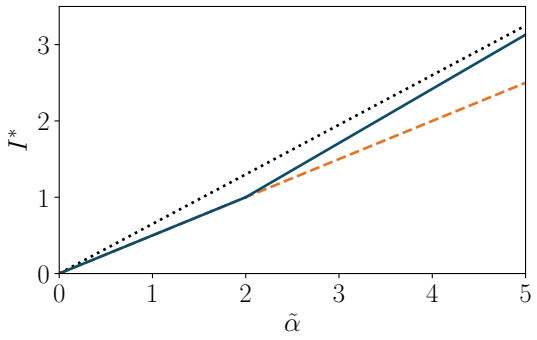


Figure 2.3: The optimal impact I_t^* as a function of the "adjusted alpha" $\tilde{\alpha}_t = \alpha_t - \tau_t\mu_t^\alpha$ for the price impact function $h(x) = \operatorname{sgn}(x)\sqrt{2|x|x_0 - x_0^2}$ from Example 2.2.3 (solid blue), for the linear model of Obizhaeva and Wang (2013) (dashed orange), and for a pure square-root model (dotted black).

To illustrate this, consider the price impact function from Example 2.2.3 and suppose for simplicity that liquidity is constant ($\gamma_t' = 0$). Then, the first-order condition (2.4.3) 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 2.2.3 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 2.3.

2.4.3 Effects of Risk Aversion?

Models with risk penalties have been studied extensively with *linear* price impact (cf., e.g., Gârleanu and Pedersen (2013, 2016) and various more recent papers). With a quadratic holding cost, optimization without price impact then is also well-posed; the solution is myopic and the holding cost just acts as a scaling parameter for the current alpha signal that allows to set a given average risk level, for example. With transient linear price impact and a linear mean-reverting signal, the optimal holdings become a linear feedback function of the current signal and the current impact (for more general signal dynamics the relationship is more involved). As the holding cost increases, the same alpha signals are exploited less aggressively and the effect of past impact diminishes.

With nonlinear price impact, we expect these qualitative properties to remain true. In particular, the optimal holdings should depend on the same state variables, but in a nonlinear manner described by a PDE with two space dimensions and unknown boundary behavior.[17]

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).[18]

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."

[17] The case of a constant alpha signal and small price impact that decays instantaneously – for which there is only one space dimension – has been studied by Guasoni and Weber (2020)

[18] Another important application, the opportunity costs of misspecified price impact models is discussed in the companion paper of the present study (Hey, Bouchaud, Mastromatteo, Muhle-Karbe, and Webster, 2024a), based on a preview of some of the results derived in the present study.

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

$$I_t = \frac{1}{1+c} \left(\alpha_t - \tau \alpha_t' \right). \tag{2.4.6}$$

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 (2.4.6) 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 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

$$I_t = \frac{2}{3}\alpha. \tag{2.4.7}$$

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

Table 2.1: Mock TCA report for three strategies under the baseline scenario.

Consider Table 2.1 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:[19]

> 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.

Adding Alpha Decay

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

Table 2.2: TCA breakdown of Strategy C across alpha triggers

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 2.2 explains the abnormal behavior in Row C of Table 2.1.

[19] 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.

Adding Dynamic Liquidity Conditions

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 2.3: TCA breakdown of Strategy C on earnings news across 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 2.3 breaks down the earnings news bucket considering whether liquidity was constant, rising, or dropping. Section 2.4.2 extends the balancing formula (2.4.6) 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.

2.5 Optimal Trading with Multiple Decay Timescales

Empirical studies suggest that impact decay initially follows a power-law and eventually converges to a permanent level (Brokmann et al., 2015; Bacry et al., 2015; Bucci et al., 2019a). To approximate such multiscale dynamics, we now consider a more general version of the model from Section 2.2, where price impact decays at multiple different

exponential timescales (and each of the corresponding concavities can potentially also be different).

2.5.1 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:

$$I_t = \sum_{n=1}^{N} w_n h_n(J_t^n). \tag{2.5.1}$$

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 2.3.2, 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 (2.5.1), 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

$$dQ_t = \lambda^{-1}\tau_n^{-1} J_t^n dt + \lambda^{-1} dJ_t^n. \tag{2.5.2}$$

Then, just like for a single impact decay timescale in Section 2.4, integration by parts applied separately for each time scale leads to a weighted sum of pointwise optimization problems:

$$\sup_{J^1,\ldots,J^N} \lambda^{-1} \sum_{n=1}^{N} w_n \mathbb{E}\Bigg[\int_0^T \Big(\left(\tau_n^{-1}\alpha_t - \alpha_t'\right) J_t^n - \tau_n^{-1} h_n(J_t^n) J_t^n\Big) dt + \alpha_T J_T^n - H_n(J_T^n)\Bigg]. \tag{2.5.3}$$

[20] However, we cannot simply optimize each term separately in (2.5.3). Instead, we also have to ensure that the volume impacts J_t^n at each different timescale correspond to the same trades:

$$\lambda^{-1}\left(dJ_t^1 + \tau_1^{-1} J_t^1 dt\right) = \lambda^{-1}\left(dJ_t^n + \tau_n^{-1} J_t^n dt\right) \tag{2.5.4}$$

for $n = 2, \ldots, N$.

2.5.2 Solution

We enforce the linear constraints (2.5.4) using suitable Lagrange multipliers η_t^n, $n = 2, \ldots, N$ (see Appendix 2.B). For given η_t^n, the impact states J_t^n, $n = 1, \ldots, N$ 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 (2.5.4) also holds solve a nonlinear second-order ODE provided in Appendix 2.B.

Theorem 2.5.1 (Optimal impact states)**.** *Given Lagrange multipliers η_t^n, $n = 2, \ldots, N$ the optimal volume impacts are*

$$h_1(J_T^1) = \alpha_T + \frac{1}{w_1}\sum_{n=2}^{N} \eta_T^n, \quad h_n(J_T^n) = \alpha_T - \frac{1}{w_n}\eta_T^n, \tag{2.5.5}$$

for $n = 2, \ldots, N$ and, for $t \in [0, T)$:

$$J_t^1 = g_1\left(\alpha_t - \tau_1 \alpha_t' + \frac{1}{w_1}\sum_{n=2}^{N}(\eta_t^n - \tau_1 \eta_t^{n'})\right), \tag{2.5.6}$$

$$J_t^n = g_n\left(\alpha_t - \tau_n \alpha_t' - \frac{1}{w_n}(\eta_t^n - \tau_n \eta_t^{n'})\right), \tag{2.5.7}$$

for $n = 2, \ldots, N$.[21] *The trade consistency constraint* (2.5.4) *is satisfied if η_t^n, $n = 2, \ldots, N$ solve the system of nonlinear second-order ODEs* (2.B.3) *with boundary conditions* (2.B.2)-(2.B.4)*.*

2.5.3 Examples

By summing over the terminal volume impacts (2.5.5), 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 2.5.1 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 2.6:

Example 2.5.2 (Empirically relevant case)**.** *The data in Section 2.6 suggests that the concavities at all impact timescales are similar and close to a square-root law ($c_n = 0.5$).*

[20] Here, $H_n(x) = \int_0^x h_n(y)dy$ is the antiderivative of the price impact function h_n.
[21] Here, $g_n(x)$ is the inverse of $h_n'(x)x + h_n(x)$.

The optimal impact state then becomes

$$I_t^* = \frac{2}{3}\left(\alpha_t - \tau_w \alpha_t' + \sum_{n=2}^{N} (\tau_n - \tau_1)\eta_t^{n\prime}\right), \quad I_T^* = \alpha_T, \tag{2.5.8}$$

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 2.5.1 become linear, leading to explicit solutions that can be applied to arbitrary alpha signals.[22] The simplest case of two impact decay timescales and constant alpha already illustrates several key effects:

Example 2.5.3 (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*

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

the ODE for the Lagrange multiplier then simplifies to

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

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

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

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 (2.5.8) 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 block 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

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

to a certain extent. This "transient" build-up of the optimal impact state is required to satisfy the trade consistency conditions (2.5.4), which depend both on the current levels and the histories of the different volume impact J_t^n and hence would not be satisfied if the overall impact state would immediately be moved to its stationary level by an initial block trade like for a single impact decay timescale.

With a finite trading horizon $T < \infty$, a similar smoothing is applied near the terminal time. This is illustrated in Figure 2.4 for different impact decay timescales and trade durations T. One takeaway from Figure 2.4 is that two regimes can occur depending on T:

1. 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 smoothing 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.

2. 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.

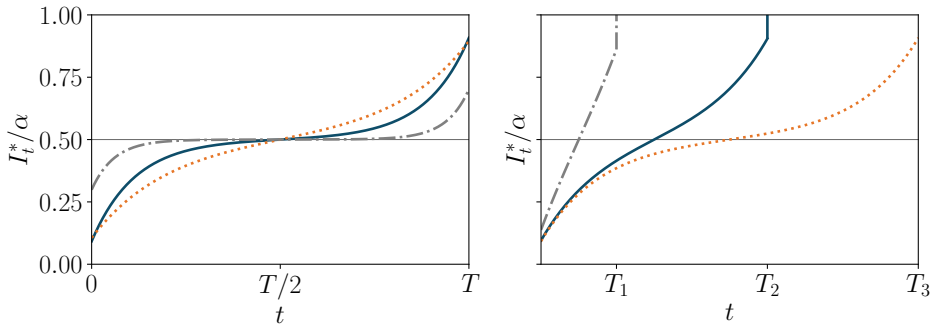


Figure 2.4: Optimal impact I_t^* over time for linear impact $h(x) = x$ and two decay timescales with weights $w_1 = 2/3$, $w_2 = 1/3$. In the left panel the order duration is $T = 70$d and the impact timescales (in days) are $(\tau_1, \tau_2) = (1,10)$ (gray dashed), $(1,100)$ (orange dotted), and $(0.5,65)$ (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 2.5.1 needs to be solved numerically, but this can be easily implemented in any standard solver. Figure 2.5 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 block 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.

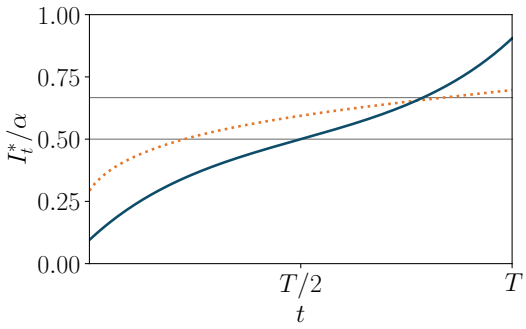


Figure 2.5: Optimal impact I_t^* over time for linear impact $h(x) = x$ (solid blue) and square-root impact $h(x) = \sqrt{x}$ (dotted orange), and the optimal impacts $I_t = \alpha/(1+c)$ for a single decay timescale when $c = 1$ and $c = 0.5$ (solid gray).

2.6 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.

2.6.1 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; Tóth 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., 2024). 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 future contracts include commodities, resources and indices. The time at the start and the end of each metaorder is indicated, as well as the mid-price and the number of child orders. All metaorders 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 metaorder was traded longer than one day and the average execution time is 3h. In line with Almgren et al. (2005), we normalize trade sizes Q_t by the average daily traded volume of the respective contracts. This normalization facilitates comparison across assets with very different liquidity profiles and ensures that the estimated impact kernels are dimensionless and interpretable.

2.6.2 Fitting Methodology

Price returns are fitted against the increments of a power-law price impact of the form

$$I_t = \sigma \sum_{n=1}^{N} w_n \text{sgn}(J_t^n)|J_t^n|^{c_n}, \tag{2.6.1}$$

for concavities c_n, (unnormalized) weights w_n, and volume impacts J_t^n with decay timescales τ_n.[23] 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 normalized by daily trading volumes.[24] 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 metaorder 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 metaorder, 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.[25]

Together with our results from the previous sections, this allows researchers to fit non-linear, 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.

2.6.3 Summary of Results

Table 2.4 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.

[23] For simplicity, we focus on constant liquidity parameters. See Cont et al. (2013); Min et al. (2022); Muhle-Karbe et al. (2024) for empirical studies with dynamic liquidity parameters using the public trading tape.

[24] 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.

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

When the aggregate order flow imbalance is used to create "proxy metaorders" from the public trading tape (see Section 2.6.6 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, see Hey et al. (2024a).

(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 2.4: 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 $\vec{w} = (w_1, ..., w_N)$.

2.6.4 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 2.6a 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. (2024a).

Figure 2.6d 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 and the average volume fraction attributed to CFM trades, the $R^2(\hat{c}, \hat{\tau}_1, \tau_2)$ can be reliably estimated up to $5 \cdot 10^{-5}$ – variations exceeding this level are considered statistically significant, indicating a meaningful improvement in model performance. The largest improvement obtains for the longest timescale $\hat{\tau}_2 = 65$ days, with associated weight $w_2/||\vec{w}|| = 0.3$. (As for principal component analysis, we henceforth sort the timescales by descending weights $w_n/||\vec{w}||$.)

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 $w_1(\hat{c}, \hat{\tau}_1 = 0.5, \hat{\tau}_2 = 65)/||\vec{w}|| = 0.7$, cf. Figure 2.6c and the inset of Figure 2.6d. Following the framework discussed by Bacry et al. (2015), this permanent component of impact can be interpreted as the alpha traded, assuming the impact itself is purely mechanical and reverting over time.

Three timescales:

When the first two timescales are fixed, one fits the third timescale by scanning across candidate τ_3. Figure 2.7 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 2.6d. The inset of Figure 2.7 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 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.

2.6.5 Understanding Multiple Concavities

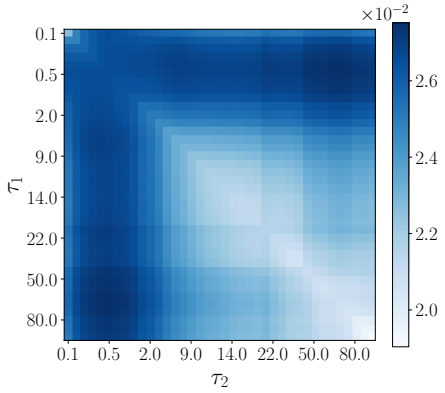
We now lift the assumption that $\hat{c}_1 = \hat{c}_2 = 0.5$ and explore arbitrary combinations of concavity parameters (c_1, c_2). The essential takeaway is that varying concavity over timescales does not significantly improve the model for metaorder data.

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 2.8 displays the statistical fit for different concavity parameters in the left panel. The heatmap is asymmetric:

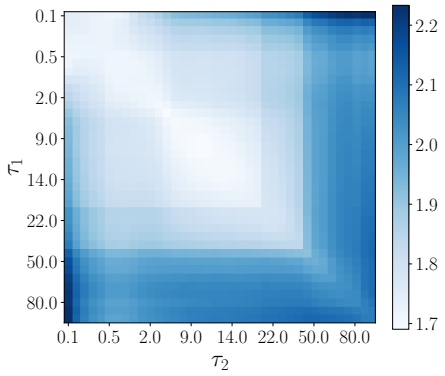
(a) For the longer timescale, $\hat{c}_2 = 0.5$ regardless of the choice for c_1.

(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 2.8b.

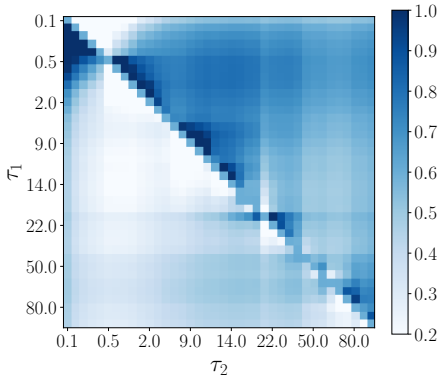
However, compared to the previous case where one chooses the same concavity parameter for both decays, the statistical significance R^2 increases only slightly by $5 \cdot 10^{-5}$ from 2.752% to 2.759%, comparable to the standard error $5 \cdot 10^{-5}$ of the R^2. Therefore, multiple concavities only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



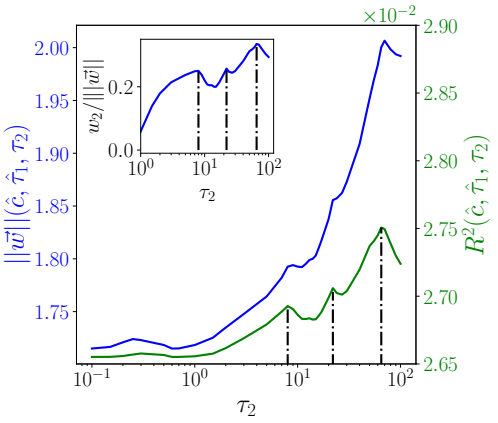
(a) $R^2(\hat{c}, \tau_1, \tau_2)$.



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



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



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

Figure 2.6: 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).

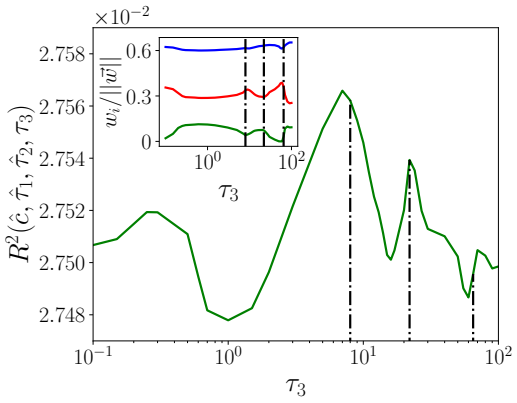


Figure 2.7: Calibration results for third decay timescale τ_3, when the concavity $\hat{c} = 0.5$ and the first two decay timescales $\hat{\tau}_1 = 0.5$ days and $\hat{\tau}_2 = 65$ days are fixed. The main plot shows the statistical fit. The inset panel displays the normalized weights of the third timescale.

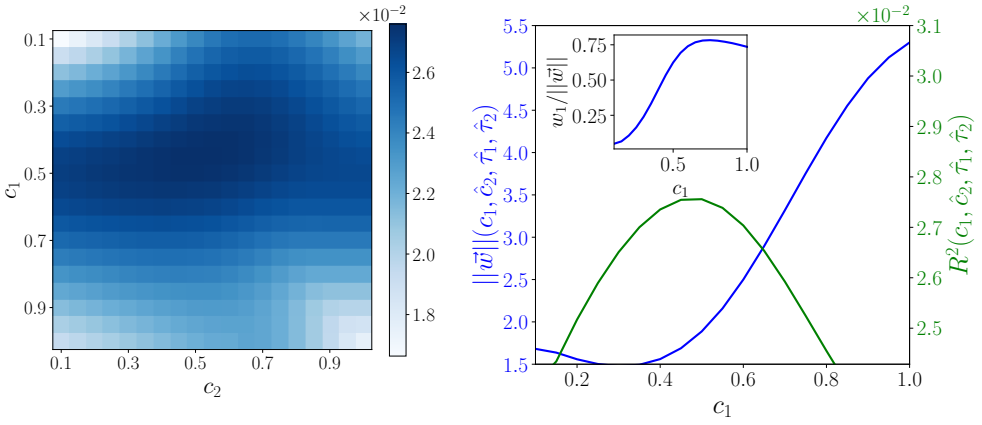
2.6.6 Understanding the Public Trading Tape

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 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 3h each to match the average proprietary metaorder length. We also experimented with 30s metaorders, but find that this leads to quite different results.

Figure 2.9 collects calibration results for the public trading tape that contains the same universe of future contracts as the proprietary dataset. 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.



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

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

Figure 2.8: 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 $w_1 = 0.6$.

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.

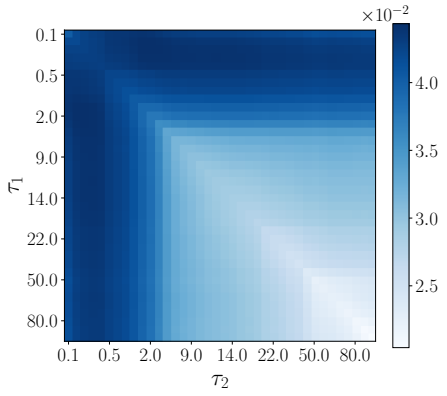
(b) For the public trading tape, different timescales seem to have different concavities. In particular, the square-root law does not hold universally, but shorter timescales appear more linear and longer timescales more concave, unlike for metaorder data.

Indeed, for a single concavity but with two decay timescales, we fix $\hat{c} = 0.5$. Figure 2.9a 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 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.

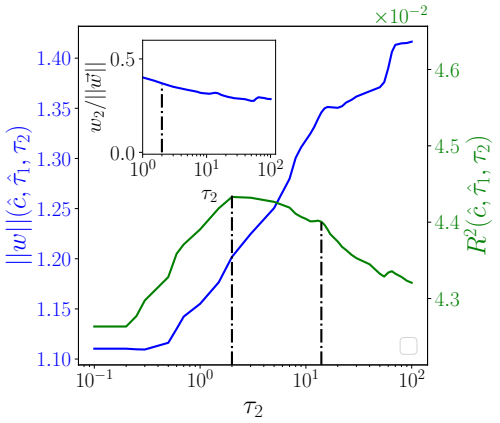
Figure 2.10 explores a model with two timescales and concavities. The two timescales are fixed to $\hat{\tau}_1 = 2$ and $\hat{\tau}_2 = 0.3$ days matching the descending order of weights $w_1/||w|| = 0.75$ and $w_2/||w|| = 0.25$; the corresponding concavities are in turn estimated as $\hat{c}_1 = 0.65$ and $\hat{c}_2 = 0.35$. The asymmetric graph indicates a strong dependence between the concavity, weights, and timescale parameters. The longer timescale contributes 75% to the total price impact on the public tape.

2.6.7 Conclusion

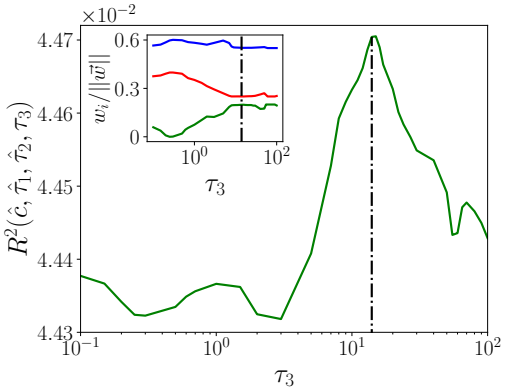
In summary, our empirical study based on metaorders suggests that a model with square-root impact across two timescales presents a good compromise between parsimony and



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

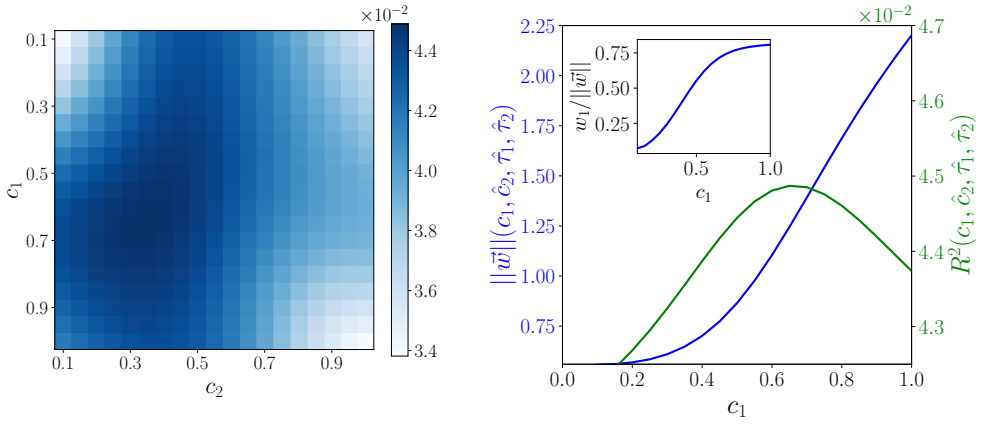


(b) $\Big(R^2, ||w||, w_2/||w||\Big)(\hat{c}, \hat{\tau}_1, \tau_2)$.



(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 2.9: 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 ($\hat{c} = 0.5$).



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

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

Figure 2.10: 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.

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.

2.7 Code and Data Disclosure

The code and data to support the numerical experiments in this paper can be found at `https://github.com/nataschahey/concavepriceimpactanddecay`.

Appendix

2.A Proofs to Theorem 2.4.1

We first establish the myopic representation of the goal functional in impact space from Theorem 2.4.1. Recall that we are still focusing on smooth trading strategies $Q_t = \int_0^t \dot{Q}_s ds$ at this stage.

2.A.1 Myopic Representation of the Goal Functional

Proof of Theorem 2.4.1. 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 (2.3.2) can be rewritten as

$$\mathbb{E}\left[\int_0^T (\alpha_t - h(J_t))\tau_t^{-1} e^{-\gamma_t} J_t dt\right] + \mathbb{E}\left[\int_0^T (\alpha_t - h(J_t)) e^{-\gamma_t} dJ_t\right]. \tag{2.A.1}$$

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 (2.A.1) can be rewritten as

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

Together with the first term in (2.A.1), this yields the asserted myopic representation (2.4.1) 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.

2.A.2 Continuity of the Goal Functional

Proposition 2.A.1 (Continuity of the objective function in impact space)**.** *The functional in impact space*

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

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

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

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 2.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* (2.4.1) *in impact space corresponds to the self-financing equation*

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

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.[26] The last term takes into account the extra impact of bulk trades.[27]

[26] 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.

[27] 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).

2.B Proof of Theorem 2.5.1

Proof of Theorem 2.5.1. The trade consistency constraints (2.5.4) correspond to the Lagrange penalties

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

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,[28] we can rewrite the Lagrange penalties using another integration by parts as

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

Maximizing the goal functional in its Lagrangian form over $J_T^1,\dots,J_T^N$ directly yields (2.5.5). Maximizing over $J_t^1,\dots,J_t^N$ for $t\in[0,T)$ gives

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

This in turn leads to (2.5.6) as well as (2.5.7).

We now turn to the characterization of the Lagrange multipliers that guarantee that the trade consistency constraints (2.5.4) hold. To ease notation, set

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

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

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

By (2.5.2), 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):

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

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

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 (2.5.6) and (2.5.7) and their derivatives, we see that this is tantamount to

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

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

$$\begin{aligned} &\frac{\tau_n}{w_n} g_t^{n\prime} \eta_t^{n\prime\prime} + \frac{\tau_1}{w_1} g_t^{1\prime} \sum_{m=2}^{N} \eta_t^{m\prime\prime} - \frac{1}{w_n} g_t^{n\prime} \eta_t^{n\prime} - \frac{1}{w_1} g_t^{1\prime} \sum_{m=2}^{N} \eta_t^{m\prime} \\ &\quad = \tau_1^{-1} g_t^1 - \tau_n^{-1} g_t^n + \left(g_t^{1\prime} - g_t^{n\prime} \right) \alpha_t' - \left(\tau_1 g_t^{1\prime} - \tau_n g_t^{n\prime} \right) \alpha_t'', \quad \text{for } t \in (0,T). \end{aligned} \tag{2.B.3}$$

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 (2.5.6) and (2.5.7), this requires the initial conditions

$$g_0^1 = g_0^n. \tag{2.B.4}$$

□

Chapter 3

Misspecification of Price Impact

Summary

Portfolio managers' orders trade off return and trading cost predictions. Return predictions rely on alpha models, whereas price impact models quantify trading costs. This paper studies what happens when trades are based on an incorrect price impact model, so that the strategy either over- or under-trades its alpha signal. We derive tractable formulas for these misspecification costs and illustrate them on proprietary trading data. The misspecification costs are naturally asymmetric: underestimating impact concavity or impact decay shrinks profits, but overestimating concavity or impact decay can even turn profits into losses.

Based on Hey et al. (2024a): ***N. Hey****, J.-P. Bouchaud, I. Mastromatteo, J. Muhle-Karbe, and K. Webster. The Cost of Misspecifying Price Impact. Risk, January 2024.*

3.1 Introduction

Price impact refers to price movements induced by trading flows, independently of their information content. For large investors, such adverse price moves caused by their own trades are the main source of transaction costs.[1] Price impact models are thus essential tools in algorithmic trading, allowing investment teams to estimate the effect of their trades on asset prices and thereby design, size and deploy systematic strategies. For instance, (capacity constrained) statistical arbitrage strategies seek to achieve the best trade-off between price predictions and trading costs.[2]

The price forecast is commonly called an alpha signal. Turning the latter into optimal

[1] In this regime, other sources of costs such as proportional bid-ask spreads are of secondary concern and hence will be disregarded in the following; the interested reader is referred to Martin and Schöneborn (2011) for a detailed discussion of the role of such linear costs for the design of trading strategies that operate on a smaller scale.

[2] If impact costs do no constrain the size of the portfolio tightly enough, it may also be risk constrained, cf., e.g., Gârleanu and Pedersen (2013) and the references therein.

trades in turn requires an appropriate price impact model. Many empirical studies find a concave *nonlinear* relationship between order size and the price impact of sizeable trades, see, e.g., Bouchaud, Bonart, Donier, and Gould (2018); Webster (2023) and the references therein for an overview. More specifically, two essential parameters emerge when determining the evolution of price impact: *concavity* and *impact decay*.

Concavity describes how the marginal build-up of price impact slows beyond a certain order size: larger orders are cheaper than a linear model suggests. *Impact decay* describes how long trade-induced price moves linger in the market and affects how quickly portfolios can turn over.

This paper quantifies how incorrect calibration of these price impact parameters leads to significant misspecification costs for traders and portfolio managers. This allows to go beyond purely statistical parameter estimates by establishing P&L-driven error bounds for price impact parameters, which are crucial to any team estimating the actual trading capacity of their alpha signals.

We carry out this analysis using closed-form expressions for misspecification costs in a nonlinear but nevertheless tractable price impact model introduced by Alfonsi, Fruth, and Schied (2010). This model was originally proposed and studied for optimal execution problems, but in fact admits explicit formulas for optimal trades with general alpha signals, derived in the companion paper of the present study (Hey et al., 2025).

These formulas allow one to quantify how wrong model parameters can be before turning a profitable trading strategy for a given alpha signal into an unprofitable one. For example, linear price impact models as in Obizhaeva and Wang (2013) inflate the transaction costs of large trades and thereby imply overly conservative trading strategies. However, they rarely turn a winning strategy into a losing one. Conversely, even relatively small overestimates of impact concavity can lead to overly aggressive trading that leads to losses despite an accurate alpha signal. Similarly, underestimating impact decay leads to conservative strategies that trade slower than optimal, which sacrifices some trading opportunities but does not lead to dramatic losses. In contrast, overly aggressive trades based on overestimates of impact decay can quickly turn a strategy's P&L negative.

These fundamental asymmetries are illustrated in Figure 3.1, which contrasts the symmetric shape of the statistical model fit with the highly asymmetric nature of the corresponding P&Ls. Such asymmetric behavior underscores that statistical accuracy alone is not a sufficient benchmark for execution model performance. What ultimately matters to the trader is not just how well the model fits the data, but how sensitive trading outcomes are to any deviations from it.

The remainder of this chapter is organized in three parts:

- *Optimal Trading Strategy:* Section 3.2 derives closed-form formulas for trading general alpha signals with non-linear price impact.

- *Empirical Estimation:* Section 3.3 estimates the impact decay and concavity pa-

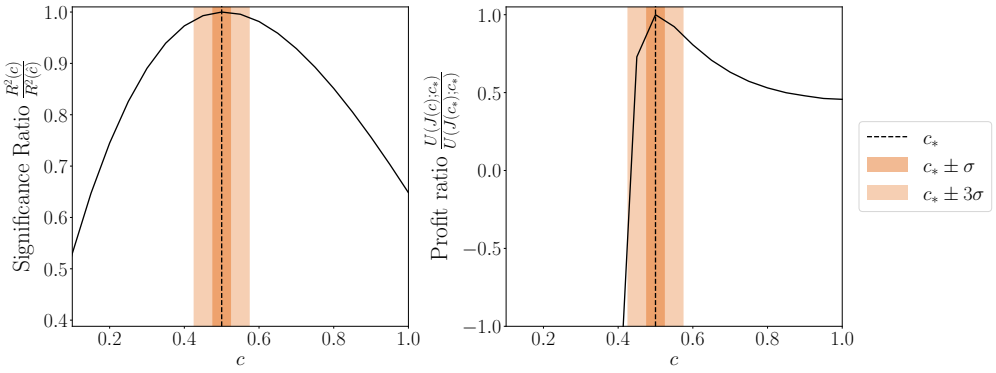


Figure 3.1: Left panel: Symmetric statistical significance ratio $R^2(c)/R^2(\hat{c})$ for point estimate $\hat{c} = 0.5$ plotted against misspecified impact concavity c. Right panel: Asymmetric profit ratio $U(c)/U(c_*)$ for ground truth $c_* = 0.5$ plotted against misspecified impact concavity c. The dark and light shaded areas are bootstrap confidence intervals.

rameters of the non-linear price impact model using proprietary trading data.

- *Sensitivity Analysis:* Section 3.4 combines the results from the previous two sections. Using the explicit trading strategies from Section 3.2, we evaluate the P&L implications of the parameter estimates from Section 3.3.

Both optimal trading and parameter estimation in models with nonlinear price impact were studied in greater detail in the last Chapter 2 to which we also refer for the derivations of the results presented here.

3.2 Optimal Trading Strategy

This section takes a price impact model as given and derives the trading strategy that maximizes an alpha signal's value net of price impact costs. We write Q_t for a strategy's position at time t. Therefore, dQ_t represents the trade at time t. An Itô process $(S_t)_{t\in[0,T]}$ describes the "unperturbed" midprice in the absence of trading. The observed midprice at time t is

$$P_t = S_t + I_t(Q),$$

where $I_t(Q)$ is the price impact caused by the trades $(Q_s)_{s\leq t}$ up to time t.

3.2.1 The AFS model

We focus on the price impact model introduced by Alfonsi et al. (2010) (henceforth AFS) and studied further by Gatheral et al. (2012), which captures the nonlinear and transient nature of price impact, but nevertheless remains remarkably tractable.

In the AFS model, price impact is a nonlinear function of a moving average of current

and past order flow:[3]

$$I_t = \lambda \operatorname{sign}(J_t)|J_t|^c, \tag{3.2.1}$$

where J_t is the exponentially weighted moving average

$$dJ_t = -\frac{1}{\tau} J_t dt + dQ_t, \quad J_0 = 0. \tag{3.2.2}$$

Here, $\lambda > 0$[4] describes the magnitude of price impact, $c \in (0, 1]$ measures its concavity, and $\tau > 0$ describes the timescale over which impact decays.[5]

For $c = 1$, price impact is linear and one recovers the model of Obizhaeva and Wang (2013). $c = 0.5$ corresponds to the "square-root law" practitioners commonly use in Transaction Cost Analysis.

3.2.2 The optimization problem

Given a model for the unperturbed price S_t, a trading strategy Q_t and an impact model $I_t(Q)$, a (smooth) strategy's P&L Y_t has the dynamics[6]

$$dY_t = Q_t dS_t - I_t dQ_t.$$

For many statistical arbitrage strategies, price impact rather than risk is the main capacity constraint. In this regime, one maximizes the expected P&L, given future return predictions. Such predictions take the form of an alpha signal:

$$\alpha_t = \mathbb{E}_t \left[S_{T'} - S_t \right]$$

for some time $T' > T$, where T is the execution horizon usually smaller than one day and S_t is the unaffected price. In addition to the *level* α_t of the current alpha prediction, another crucial statistic in this context is its *decay*, captured by its drift rate μ_t^α. (For smooth alphas, this is simply the derivative.)

[3] This paper presents results for price impact functions of power form; these are special cases of the results for AFS models with general impact functions derived in the companion paper (Hey et al., 2025). The analysis there also incorporates time-dependent and even stochastic liquidity, but still leads to explicit formulas. Another important generalization are additional linear trading costs corresponding to bid-ask spreads, for example. These only have a second-order effect for large investors (for which the superlinear price impact costs are the main concern). For more risk-constrained investors or large-spread instruments, the joint analysis of price impact and spread costs requires finer tools.

[4] Note that in the companion paper (Hey et al., 2025), the push-factor is integrated within the differential equation, forming part of the concave function. In the current framework, to align with the previous definition, the concave function would need to be applied to λ.

[5] This specification implies that impact I_t relaxes all the way to zero over long time horizons, i.e., trades have no *permanent* impact. But see Gabaix and Koijen (2021); Bouchaud (2022) for a detailed discussion of this point.

[6] Diffusive or bulk trades lead to extra correction terms (Hey et al., 2025), but their detailed treatment can be avoided by recasting the optimization in impact space (3.2.4).

For a given alpha signal, a risk-neutral statistical arbitrage strategy maximizes

$$\mathbb{E}\left[Y_T\right] = \mathbb{E}\left[\int_0^T (\alpha_t - I_t) dQ_t\right]. \tag{3.2.3}$$

That is, each trade captures the expected alpha but pays price impact, since the latter eventually disappears.

3.2.3 Solution in impact space

The statistical arbitrage problem (3.2.3) has a straightforward, closed-form solution for arbitrary alpha signals, even for non-linear price impact described by the AFS model. This result hinges on a simple but powerful observation, proven in Hey et al. (2025) using a technique introduced by Fruth et al. (2013). The key insight is that, for any position Q or, *equivalently*, the corresponding impact J (resp., I),

$$\mathbb{E}\left[Y_T\right] = \mathbb{E}\left[\frac{1}{\tau}\int_0^T (\alpha_t J_t - \lambda |J_t|^{1+c}) dt - \int_0^T J_t d\alpha_t + \alpha_T J_T - \frac{1}{1+c}|J_T|^{1+c}\right]. \tag{3.2.4}$$

Whence, by switching the control variables from positions Q_t to the corresponding impact J_t ("passing to *impact space*"), the complex control problem (3.2.3) becomes a simple pointwise maximization.[7] The optimal impact in turn is

$$I_t^* = \lambda \operatorname{sign}\left(J_t^*\right) |J_t^*|^c = \begin{cases} \frac{1}{1+c}\left(\alpha_t - \tau \mu_t^\alpha\right), & t \in (0, T), \\ \alpha_T, & t = T. \end{cases} \tag{3.2.5}$$

One recovers the corresponding optimal positions via

$$Q_t^* = J_t^* + \frac{1}{\tau}\int_0^t J_s^* ds. \tag{3.2.6}$$

Even though the optimal trading strategy Q^* is complex, switching to impact space allows one to retain a straightforward linear relationship between the optimal policy as well as the current alpha signal and its decay. Up to a bulk trade at the terminal time exhausting all remaining alpha, it is optimal to keep impact equal to a constant fraction of the alpha signal, adjusted for alpha decay relative to impact decay. The intuition for this adjustment is that one has to trade more aggressively and accept higher impact costs for signals that decay quickly compared to impact, as Gârleanu and Pedersen (2013) highlight for linear impact models:

> "The alpha decay is important because it determines how long time the investor can enjoy high expected returns and, therefore, affects the trade-off between returns and transactions costs."

[7] A similar change of variable to *integrated* impact is used in Gârleanu and Pedersen (2016); Isichenko (2021).

The fraction of the adjusted alpha to pay in impact depends on the concavity parameter c of the price impact function. For linear price impact ($c = 1$), the optimal impact equals one-half the adjusted alpha (Isichenko, 2021). For strictly concave price impact functions, this ratio increases and reaches two thirds for $c = 1/2$ compatible with square-root impact. As a result, linear models prescribe to trade less aggressively than is optimal in their strictly concave counterparts. The impact scale λ does not appear in the formula for the optimal impact – it only determines the corresponding trades needed to attain the optimal impact state.

3.2.4 Implications for Trading

In trading applications, it is common to normalize the impact scale λ as

$$\lambda = \frac{\sigma}{V^c}\, g(c, \tau), \tag{3.2.7}$$

where V is the average daily volume (Almgren et al., 2005). This normalization expresses the quantities of interest in trader-friendly units: return predictions compare to the volatility σ of price changes and trading volumes are expressed as fractions of the average daily volume V. With this normalization, the price impact coefficient also becomes comparable across different assets; our notation $g(c, \tau)$ highlights that this prefactor depends on the corresponding impact concavity c and decay timescale τ.

Long-term alpha signal

We first illustrate the application of the general trading rule (3.2.5) for the simplest case where $\alpha_t = \alpha$ is constant. This means that the signal predicts a return that happens in the distant future, e.g., an event at the end of the month when focusing on today's trading. For square-root impact ($c = 1/2$), Equation (3.2.5) then simplifies to

$$I^* = \frac{2}{3}\alpha. \tag{3.2.8}$$

That is, the optimal trading strategy (for sizeable orders) pays two thirds of the constant alpha in impact. The corresponding optimal order size is

$$\frac{Q_T}{V} = \Lambda^{-2} \cdot \operatorname{sign}(\alpha) \left(\frac{\alpha}{\sigma}\right)^2, \quad \text{where} \quad \Lambda = \frac{g(1/2, \tau)}{\sqrt{1 + \frac{4}{9}\frac{T}{\tau}}}. \tag{3.2.9}$$

Conversely, one implies a constant alpha from a long-term order of size Q via the formula

$$\frac{\alpha}{\sigma} = \Lambda \cdot \operatorname{sign}(Q) \sqrt{\frac{|Q|}{V}}. \tag{3.2.10}$$

The order size formula 3.2.9 helps portfolio managers size trades considering price impact's square-root law. Then, traders use the implied alpha formula 3.2.10 to agree on a

baseline alpha with the portfolio team. After establishing a baseline alpha, the trading team can linearly add their own short-term signals into the trading strategy.

Mean-reverting alpha signal:

Another standard specification assumes that the alpha signal α_t is an Ornstein-Uhlenbeck process with relaxation time θ:

$$d\alpha_t = -\frac{1}{\theta}\alpha_t dt + \sigma dW_t.$$

Then, $\mu_t^\alpha = -\theta^{-1}\alpha_t$, i.e., the alpha decays exponentially in the absence of new information. In this model, a statistical arbitrage strategy continuously updates its signals based on a steady information stream. The optimal impact in this context is

$$I^* = \frac{2}{3}\left(1 + \frac{\tau}{\theta}\right)\alpha. \tag{3.2.11}$$

Whence, mean-reverting alphas should always be traded more aggressively than constant ones, and the size of this adjustment depends on the relative magnitudes of impact decay τ and alpha decay θ.

If $\tau > c\theta$, then $I^* > \alpha$, implying that impact should decay at least twice as fast as alpha. This observation suggests that impact decay, driven by changing liquidity, should indeed occur on the fastest timescale. When $\theta \to \tau$, the risk of losses due to impact costs increases. At these timescales, microstructure effects may become significant. Therefore, we focus on intraday alpha decay, which typically exceeds impact decay.

3.3 Empirical Estimation

3.3.1 Description of the data

Our empirical analysis uses proprietary Capital Fund Management (CFM) trading data comprised of $1.9 \cdot 10^5$ metaorders of future contracts traded over 2012-2022. The time at the start and the end of each metaorder is indicated, as well as the mid-price and the number of child orders. All metaorders were executed through at least three child orders and accounted for a fraction between 0.01% and 10% of ADV; the average order size was of the order of 0.1%. No metaorder was traded longer than one day and the average execution time was 3h.

3.3.2 Fitting methodology

Price returns are fitted against the increments of a power-law price impact function of the form (3.2.1) with parameters c, τ and λ (normalized as in (3.2.7)). Since i) no metaorder 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 impact J is computed under the assumption

that trades are executed uniformly during the execution time Δt of each metaorder. Note that the alpha signals driving the trade decisions are typically realized over a time window much larger than one day, implying that it is appropriate to attribute to impact, rather than alpha, the average price variation occurring during the execution of metaorders in the sample.

3.3.3 Estimating impact concavity

To estimate the price impact parameters with impact decay, we consider a grid of impact decays τ and concavity parameters c. For each pair, we then first compute the volume impact J_t corresponding to the impact timescale τ according to (3.2.2); then the price impact I_t is computed for the corresponding concavity parameter c. With the impact variable at hand, we then run a linear regression of price changes against impact to determine the prefactor (3.2.7).

Figure 3.2 shows the model's statistical fit $R^2(c, \tau)$ and prefactor $g(c, \tau)$ across a broad range of concavity and decay parameters. The heatmap of R^2 in Subfigure 3.2a shows an ellipsoid around the point estimates $(\hat{c}, \hat{\tau})$. Figure 3.2c plots the statistical sensitivity with respect to the concavity parameter c, when the point estimate of the decay parameter $\hat{\tau} = 0.3$ days remains fixed.

As the concavity parameter c varies, the model's R^2 peaks at $\hat{c} = 0.50$ and decreases markedly around this point estimate.[8] The prefactor $g(c, \hat{\tau})$ increases with increasing concavity parameter. This behavior arises because, when normalized by V^c, the linear model tends to underestimate the impact costs compared to a concave model. Consequently, the prefactor adjusts more than quadratically to account for this difference. Notably, when $c = 0.5$, the prefactor $g(c, \hat{\tau})$, aligning well with established literature, as this configuration accurately captures impact costs.

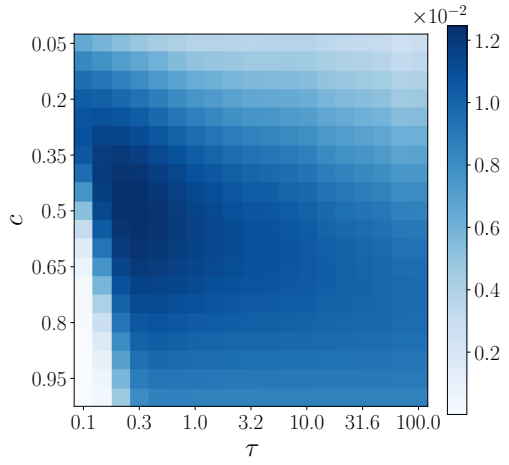
3.3.4 Estimating impact decay

To investigate the statistical sensitivity with respect to the impact decay, the concavity parameter is fixed at the point estimate $\hat{c} = 0.50$ and the impact decay τ varies. Figure 3.2d displays $R^2(\hat{c}, \tau)$ and $g(\hat{c}, \tau)$. There is a peak at $\hat{\tau} = 0.3$ days, indicating that one measures the short-time scale of impact decay. The model's prefactor remains relatively stable for timescales larger than 0.3 days.

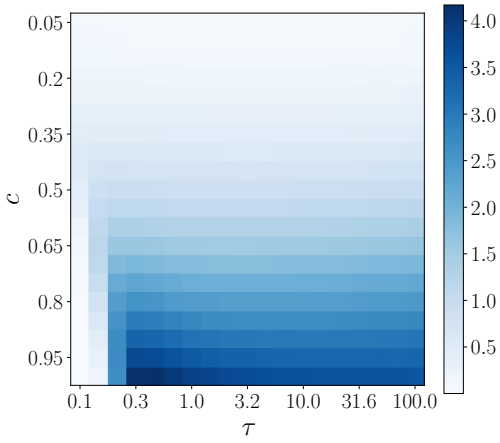
The primary takeaway is that the *statistical* loss is significantly more sensitive to concavity than to decay. For example, misspecifying $c = 1$ instead of the point estimate $\hat{c} = 0.50$ reduces R^2 by more than 30%. Conversely, getting τ wrong by a factor of ten reduces R^2 by less than 10%.[9]

[8] The magnitude of the model's R^2 is consistent with results published in papers using the public trading tape. Indeed, when one runs the price impact model on the full trading tape, the R^2 varies between 10% and 20%. However, in this study, only CFM's metaorders are used. Therefore, the R^2 drops in line with the fraction of the public tape CFM's orders represent.

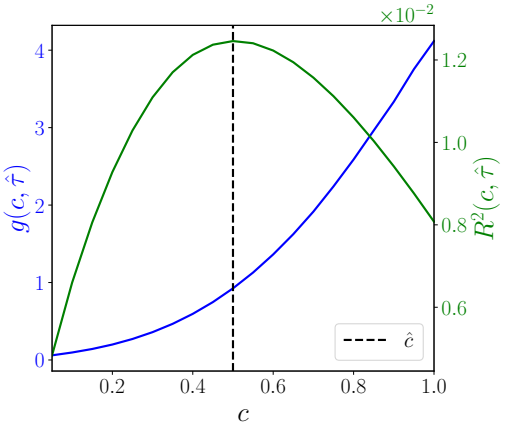
[9] This is not a new observation: studies across many datasets and asset classes agree on price impact's concavity and general order of magnitude but disagree on decay. The reason is that different orders



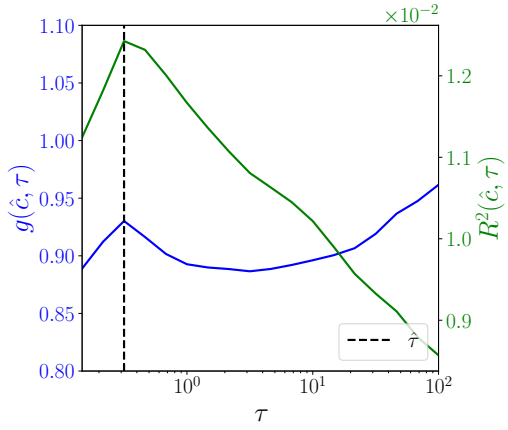
(a) $R^2(c, \tau)$.



(b) $g(c, \tau)$.



(c) $R^2(c, \hat{\tau})$ and $g(c, \hat{\tau})$.



(d) $R^2(\hat{c}, \tau)$ and $g(\hat{c}, \tau)$.

Figure 3.2: Calibration results. Panel (a) shows the statistical sensitivity; Panel (b) depicts the model's prefactor across $\hat{c}, \hat{\tau}$. The sensitivity analysis yields an ellipsoid around the point estimates. The model's prefactor remains roughly constant along the $\hat{\tau}$-axis and increases with concavity c. Panel (c) shows the sensitivity across c for $\hat{\tau} = 0.3$ days in green. The corresponding fitted prefactor $g(c, \hat{\tau})$ is displayed in blue. Panel (d) displays the sensitivity and the model's pre-factor across τ for the point estimate $\hat{c}$.

These empirical results do *not* imply that calibrating impact decay correctly is unimportant. Indeed, a central result in Section 3.4 is that a strategy's P&L can be *highly* sensitive to τ, even if *statistically*, R^2 is less sensitive to τ.

3.4 Sensitivity Analysis

We now combine the previous theoretical and empirical results to quantify the performance losses caused by trading based on a misspecified model.

3.4.1 Quantifying misspecification costs

Recall that, for given estimates c, τ of impact concavity and decay, the corresponding fitted prefactor is

$$\lambda(c,\tau) = \frac{\sigma}{V^c} g(c,\tau).$$

Therefore, assuming c, τ are the correct model parameters, a strategy trading an alpha signal α_t with drift rate μ_t^α will implement

$$I_t(c,\tau) = \frac{\alpha_t - \tau\mu_t^\alpha}{1+c} \quad \text{or, equivalently,} \quad J_t(c,\tau) = \left(\frac{\alpha_t - \tau\mu_t^\alpha}{(1+c)\lambda(c,\tau)}\right)^{1/c}.$$

We call $I(c,\tau)$ or, equivalently, $J(c,\tau)$, the *misspecified* policy. This is the optimal policy when the impact parameters c, τ correctly describe price impact, but not if these parameters differ from the price impact parameters c_*, τ_* of the *actual* data generating model. To ease notation, we omit the corresponding variables when one of the parameters is held fixed at its actual value, e.g., we use the shorthand $I(\tau) = I(c_*, \tau)$.

The P&L of the *misspecified* policy $J(c,\tau)$ under the *actual* price impact model with parameters c_*, τ_* is

$$\begin{aligned} U\Big(J(c,\tau); c_*, \tau_*\Big) = &\frac{1}{\lambda_*}\mathbb{E}\Bigg[\frac{1}{\tau_*}\int_0^T ((\alpha_t - \tau_*\mu_t^\alpha)J_t(c,\tau) - \lambda|J_t(c,\tau)|^{1+c_*})dt \\ &+ \alpha_T J_T(c,\tau) - \frac{1}{1+c_*}|J_T(c,\tau)|^{1+c_*}\Bigg]. \end{aligned}$$

This general formula is straightforward to implement numerically in a backtest. For instance, statistical arbitrage teams can plug in historical alpha signals to measure the expected P&L of trading based on c, τ when the actual parameters are c_*, τ_*.

To illustrate the misspecification formula's implications, Sections 3.4.2 and 3.4.3 cover two common cases. By specifying concrete parametric alpha signals, we quantify the cost of misspecifying concavity or decay in closed-form. This reveals the dependence

decay at different timescales, which are better captured by a multi-exponential or power law decay kernel. With linear price impact, there are some optimization results for such decay kernels (Gatheral et al., 2012; Abi Jaber and Neuman, 2022). The companion paper Hey et al. (2025) reports first results for *nonlinear* impact and multiscale impact decay.

on alpha signal characteristics without relying on a backtest.

3.4.2 Misspecification cost of impact concavity

We first consider the simplest case of a constant alpha signal: $\alpha_t = \alpha$, so that $\mu_t^\alpha = 0$. For example, this assumption is reasonable when trading intraday based on a signal that predicts a return past today's close, such as an event taking place the next day or week.

As impact decay only plays a minor role in this context,[10] we assume for simplicity that it is correctly specified ($\tau = \tau_*$) and focus on the misspecification cost of the concavity parameter $c \neq c_*$. In words:

> How much P&L does the strategy lose if it trades with the wrong price impact concavity?

For example, one canonical case is when $c_* = 0.5$ and $c = 1$: one estimates the cost of implementing a linear price impact model when actual impact follows a square-root law.

The expected P&L for the *misspecified* policy is

$$U\Big(J(c);c_*\Big) = \frac{\sigma V}{g(c)^{1/c}}\Bigg[\left(\frac{\alpha}{\sigma}\right)^{1+1/c}\left(\frac{T}{\tau_*(1+c)^{1/c}}+1\right) - \frac{g(c_*)}{g(c)^{c_*/c}}\left(\frac{\alpha}{\sigma}\right)^{(1+c_*)/c}\left(\frac{T}{\tau_*(1+c)^{(1+c_*)/c}}+\frac{1}{1+c_*}\right)\Bigg].$$

The natural comparison for this is the expected P&L $U(J(c_*);c_*)$ of the optimal policy for the actual concavity parameter c_*:

$$U\Big(J(c_*);c_*\Big) = \frac{\sigma V}{g(c_*)^{1/c_*}}\left(\frac{\alpha}{\sigma}\right)^{1+1/c_*}\frac{c_*}{1+c_*}\left(\frac{T}{\tau(1+c_*)^{1/c_*}}+1\right).$$

In this setting, the primary alpha characteristic is the signal's Sharpe ratio, α/σ. A core result from Figure 3.3 is that misspecification costs are more sensitive to c as a signal's Sharpe ratio increases. Therefore, the stronger a team's alpha signal, the more important it becomes to correctly estimate the price impact model's concavity accurately.

In addition to quantifying the empirical sensitivity of P&L to misspecifications of the concavity parameter c, the closed-form formulas also allow one to compute *how wrong* a parameter can become before turning profitable strategies unprofitable. Indeed, for each true parameter c_*, there exists a critical value $c_{\min}$ such that, for any choice $c < c_{\min}$, the expected P&L becomes a *decreasing* function of the alpha signal. Naturally, statistical arbitrage strategies crucially need to avoid this critical regime where adding more predictive power leads to worse outcomes. Figure 3.3 illustrates the sharpness of the phase transition between profitable and non-profitable misspecified trading strategies.

[10] Indeed, for constant alpha, the P&L of the believed strategy only depends on the impact decay parameter τ through the estimate of the magnitude of price impact $\lambda(c,\tau)$. As this dependence is rather weak, impact decay only plays a minor role in the absence of alpha decay.

Because misspecification costs become more sensitive with higher alpha Sharpe ratios, the critical value becomes tighter as the Sharpe increases. Again, this highlights the importance of correctly estimating price impact's concavity when trading high Sharpe ratio signals.

A crucial insight is that this misspecification risk is asymmetric: even for large overestimates $c > c_*$, the expected P&L typically remains positive (except for very small Sharpe ratios). The intuition is that overestimating the concavity parameter c leads the strategy to submit smaller trades. While this is suboptimal and sacrifices some profit opportunities, it will only lead to shrinking profits but not losses. Conversely, underestimates below the critical value $c_{\min} < c$ can lead to significant losses, as the corresponding trading strategy submits outsized trades leading to excessive trading costs.

While these results are derived in a concrete model with a convenient closed-form solution, we expect them to remain true qualitatively across a much wider range of models. This leads to a practical takeaway when fitting price impact models: when considering confidence intervals for a price impact model's concavity, the robust solution is to deploy a more conservative estimate from the upper part of the confidence interval, i.e., use a impact model that is somewhat less concave than the point estimate.

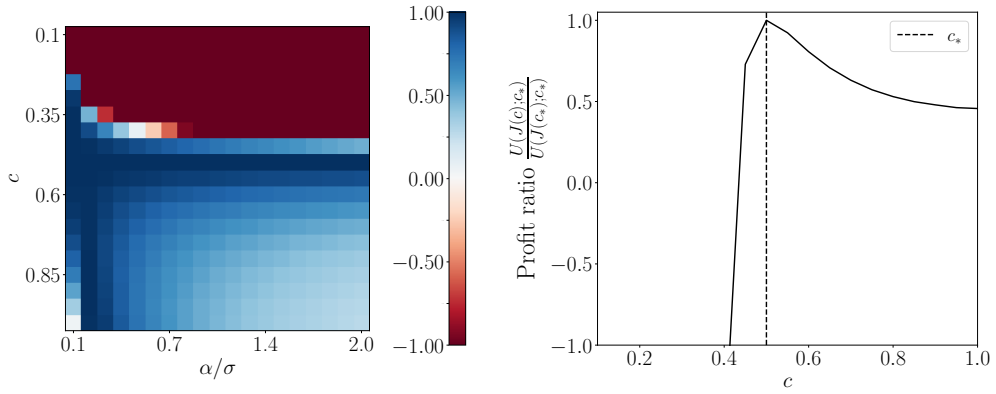


Figure 3.3: Profit ratios between correctly and incorrectly specified strategies when the ground truth is $c_* = 0.5$. Left panel: Heat map of the profit ratio across c and α/σ. Right panel: Profit ratio across c for an alpha signal with Sharpe $\alpha/\sigma = 1$. The profits are asymmetric: over-estimating concavity (low c) quickly leads to losses, while underestimating concavity only shrinks profits.

3.4.3 Misspecification cost of impact decay

We now turn to misspecification of the impact decay parameter τ, while keeping the concavity parameter c fixed to its correct value c_*. As impact decay predominantly interacts with the alpha decay, we study this for a mean-reverting alpha with $\mu_t^\alpha = -\alpha_t/\theta$. To obtain crisp results, we again focus on the steady-state limit of the P&L. Unlike the model's statistical R^2, it turns out that the expected P&L is surprisingly sensitive to misspecifications of τ relative to the ground truth τ_*.

For the *misspecified* policy, we have

$$U\Big(J(\tau);\tau_*\Big) = \frac{\sigma\, V}{\tau_*}\left(\frac{1+\tau/\theta}{g(\tau)(1+c_*)}\right)^{1/c_*}\left(1+\frac{\tau_*}{\theta}-\frac{g(\tau_*)(1+\tau/\theta)}{g(\tau)(1+c_*)}\right) \times \lim_{T\to\infty}\frac{1}{T}\left[\int_0^T\left|\frac{\alpha_t}{\sigma}\right|^{1+1/c_*}dt\right].$$

The optimal value derived from the policy for the actual impact decay τ_* in turn is

$$U\Big(J(\tau_*);\tau_*\Big) = \frac{\sigma\, V}{\tau_*}\frac{c(1+\tau_*/\theta)^{1+1/c}}{g(\tau_*)^{1/c}(1+c)^{1+1/c}} \times \lim_{T\to\infty}\frac{1}{T}\left[\int_0^T\left|\frac{\alpha_t}{\sigma}\right|^{1+1/c}dt\right].$$

In this setting, the key alpha characteristic is its decay θ. More specifically, the ratio of performances for the misspecified and optimal policies mostly depends on the estimated and true impact decays through their ratios relative to alpha decay, τ/θ and τ_*/θ. Indeed, as depicted in Figure 3.2d, $g(\tau_*) \approx g(\tau)$ is a good approximation for a wide range of decay parameters; with this, the ratio of performance simplifies to

$$\frac{U\Big(J(\tau);\tau_*\Big)}{U\Big(J(\tau_*);\tau_*\Big)} = \frac{1}{c}\left(\frac{1+\tau/\theta}{1+\tau_*/\theta}\right)^{1/c}\left(1+c-\frac{1+\tau/\theta}{1+\tau_*/\theta}\right).$$

The profit ratio between misspecified and optimal policy is displayed in Figure 3.4. There is a sharp boundary between the profitable and unprofitable regions that depends linearly on the ratios τ/θ and τ_*/θ. This boundary appears then curved in log-log scale on the left panel in Figure 3.4 separating the positive and negative profit regions. A wrong estimation of impact decay for a long-lasting alpha signal (small θ) is less costly than for a fast decaying one.

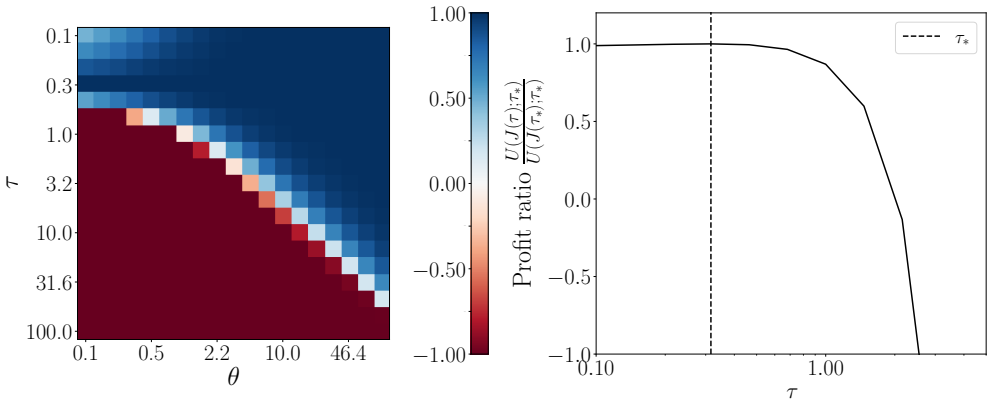


Figure 3.4: Profit ratio between correctly and incorrectly specified strategies when $\tau_* = 0.3$ days. Left panel: Heat map of the profit ratio across τ and θ in log-log scale. Right panel: Profit ratio across τ in log-scale for an alpha signal with a decay timescale of 1 day. The profits are asymmetric: overestimating decay (high τ) quickly leads to losses, while underestimating decay only shrinks profits.

Like for impact concavity, the costs of misspecifying impact decay are highly asymmetric. Indeed, overestimating impact decay quickly turns profitable alpha signals unprofitable, as illustrated in the right panel of Figure 3.4 for a moderate alpha decay time scale of 1 day. The intuition is that overestimates $\tau > \tau_*$ lead to larger than optimal impacts, i.e., overly aggressive trading just like underestimates of the concavity parameter.

3.5 Conclusion

We investigated the effect that the misspecification of the price impact model has on the performance of a trader. The P&L cost of getting an impact parameter wrong can be evaluated using the AFS model's tractable *misspecification cost* formula. This P&L-driven approach complements a statistical approach and shows that the opportunity costs of getting model parameters wrong are asymmetric:

- It is better to over- than under-estimate c. An excessively concave price impact model can lead to losses.
- It is better to under- than over-estimate τ. A price impact model with excessively fast impact decay can lead to losses.

In both cases, overly aggressive trading has a substantially bigger effect than an overly cautious approach. While we illustrate this with the closed-form solutions for the AFS model, these insights are expected to remain qualitatively true for many models. Therefore, a general approach when considering confidence intervals for price impact model parameters is to deploy parameters within the error band that lead to more conservative trades.

Our results thereby support the relevance of a robust approach to optimal trading problems in the presence of trading costs, and supports the idea of jointly tackling parameter estimation and alpha exploitation.

Chapter 4

Concave Cross Impact

Summary

The price impact of large metaorders is well known to be a concave function of their size. We discuss how to extend models consistent with this "square-root law" to multivariate settings with cross impact, where trading each asset also impacts the prices of the others. In this context, we derive consistency conditions that rule out price manipulation. These basic requirements make risk-neutral trading problems tractable and also naturally lead to parsimonious model specifications that can be calibrated to historical data. We illustrate this with a case study using proprietary CFM metaorder data.

Based on Hey et al. (2024b): ***N. Hey****, I. Mastromatteo, J. Muhle-Karbe. Concave Cross-Impact. 2024. (submitted to Management Sciences)*

4.1 Introduction

Price impact is the main source of trading costs for large institutional investors. The impact of large metaorders builds up over time and then gradually decays (Hasbrouck, 1991; Biais et al., 1995). Crucially, the magnitude of this effect is not linear, but better described by a "square-root law" (Loeb, 1983; Hasbrouck, 1991).

When trading several securities simultaneously, it is natural to expect that trades in one asset do not only affect its own price ("self impact") but also shift the prices of other related securities ("cross impact"). In particular, for assets that are closely linked – such as futures on the same underlying with different maturities – cross impact is bound to play a major role. For example, rolling over futures positions appears costly if one considers each trade's impact separately. However, accounting for cross impact reduces the trading costs for such strategies significantly, as selling one contract partially offsets the impact of buying the other. Conversely, if several futures contracts are traded based on similar alpha signals, then cross impact compounds the trading costs across assets.

Over the last decade, a number of studies have investigated *linear* cross impact models.[1] In contrast, nonlinear cross impact models consistent with square-root self impact are virtually uncharted territory, both in terms of theory and empirical analysis. A key reason for this impasse is that with several traded assets, guaranteeing the absence of "price manipulation" becomes highly nontrivial. Ruling out the existence of such strategies that turn price impact into profits is a basic consistency requirement any price impact model needs to satisfy. Otherwise, numerical optimizers naturally converge towards extreme strategies that seek to exploit these inconsistencies in the model, but are unlikely to be effective in practice.

Similarly as for no-arbitrage conditions in option pricing models (Schönbucher, 1999), the absence of price manipulation is relatively easy to characterize for a single traded asset (Fruth et al., 2013, 2019; Hey et al., 2025). In particular, no conditions are required when the price impact parameters do not change over time. In contrast, nontrivial conditions are required for multivariate models with linear cross impact already with constant parameters (Alfonsi et al., 2016; del Molino et al., 2020; Tomas et al., 2022b; Rosenbaum and Tomas, 2022; Abi Jaber et al., 2024). This raises the natural question if and how these consistency conditions can be extended to nonlinear models compatible with the univariate square-root law.

On the empirical side, a first basic question is whether cross impact can be reliably measured from price and trading data, and whether it displays the same nonlinear form as self impact. The next key challenge in turn is to build parsimonious models for cross impact that guarantee the absence of price manipulation and can be fitted efficiently to data.

The present study breaks new ground in all of these directions using a multi-asset version of the nonlinear price impact model of Alfonsi, Fruth, and Schied (2010). By considering suitable parametric families of trading strategies as in Gatheral (2010), we derive necessary conditions for the absence of price manipulation that substantially narrow down the parameter space. Once these conditions are imposed, risk-neutral optimal trading problems can in fact be reduced to simple pointwise maximizations by "passing to impact space" as in (Fruth et al., 2013; Hey et al., 2025). More specifically, switching control variables from positions held to impact caused does not directly lead to a pointwise problem here due to some intractable cross terms. However, as in Bilarev (2018), absence of price manipulations dictates that these intractable terms have to vanish. Whence, for all well behaved models, risk-neutral optimization problems can be solved by pointwise maximization. This in turn allows one to determine whether a given model indeed guarantees that price manipulation is not possible.

In particular, we find that there is a natural subclass of models for which the multivariate optimal trading problem decouples into simple one-dimensional subproblems, for which wellposedness and optimal trading strategies are well understood (Hey et al.,

[1] See, e.g., (Schied et al., 2010; Gârleanu and Pedersen, 2013, 2016; Alfonsi et al., 2016; Benzaquen et al., 2017; Tsoukalas et al., 2019; Horst and Xia, 2019; Tomas et al., 2022a; Abi Jaber et al., 2024).

2025). Even though the optimal impact states for each asset do not depend on the magnitude of cross impact in this case, the corresponding optimal trades evidently do. Indeed, with positive cross impact, much less trading in the same direction is needed to create the same amount of impact, but much bigger trades of opposite signs can be implemented.

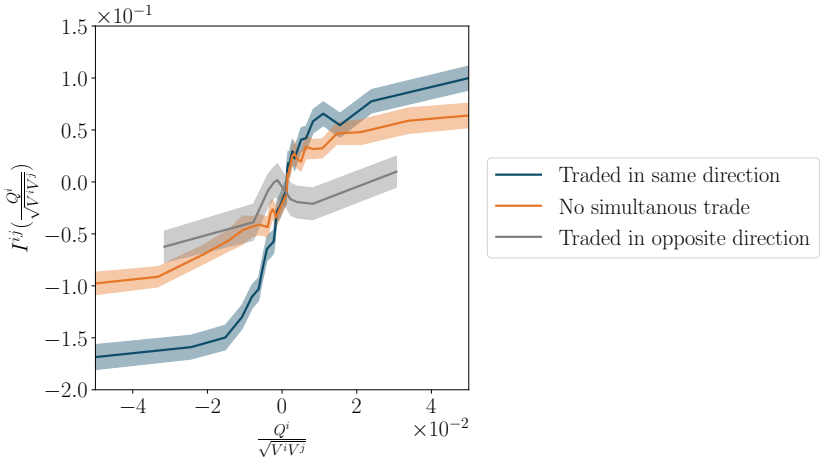


Figure 4.1: Average signed differences I^{ij} between the prices of asset j at the beginning and end of metaorders for asset i, plotted against the size of the metaorders Q^i (normalized by the geometric mean $\sqrt{V^iV^j}$ of the average daily trading volumes of assets j and i). The shaded regions are bootstrap confidence intervals.

The relevance and applicability of this model class is in turn illustrated by an empirical case study based on proprietary metaorder data from Capital Fund Management (CFM). We first perform a simple comparison of the arrival prices at the beginning of each metaorder to the peak impact incurred at their completion. As illustrated in Figure 4.1, this demonstrates that for highly correlated assets[2] cross impact measurements are highly significant and depend on metaorder sizes in the same concave manner as for self impact. Moreover, the figure clearly illustrates the impact of different trading scenarios. Indeed, when both assets are traded in the same direction, then self and cross impact compound. In contrast, they largely offset each other when the assets are traded in opposite directions.

Building on these findings, we then show that it is also possible to fit our dynamic cross impact model to the data. To this end, the consistency conditions derived in the theoretical part of the paper play a key role. On the one hand, these hard code the absence of price manipulation strategies. On the other hand, they substantially narrow down the parameter space and thereby lead to much more parsimonious models that can be calibrated efficiently.

We validate the feasibility and flexibility of this approach by fitting bivariate impact models to pairs of assets. Figure 4.2 displays how the fitted model parameters depend

[2]Here, we focus on futures contracts with the same underlying but different maturities so that the average return correlations are above 90%.

on the return correlation of the assets. We see in the top panel that the concavity of the impact function is largely insensitive to the correlation parameter and in line with other studies corroborating the square root law. The middle panel shows that impact decay tends to become slower for highly correlated assets. The intuition for this is that many of the highly correlated assets are commodity futures contracts, which are not as liquid as the index futures that make up many of the less correlated asset pairs. The bottom panel of Figure 4.2 plots the proportion of total impact accounted for by self impact. We see that for assets with low correlation, cross impact play only a minor role but, for highly correlated assets, self and cross impact become almost interchangeable.

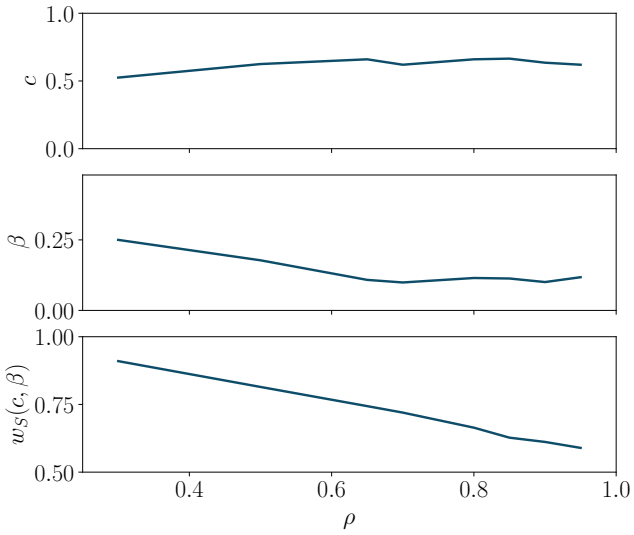


Figure 4.2: Parameter estimates for a bivariate cross impact model. The common impact concavity c, impact decay rate β and proportion $w_S(c, \beta)$ of total impact caused by self impact is plotted against the correlation between the assets.

In summary, this paper proposes a general consistent framework for modeling the concave cross impact of trading multiple assets simultaneously. In this setting, the absence of price manipulation can be guaranteed, risk-neutral trading problems can be solved in closed form, and the resulting models can be estimated efficiently from data.

The remainder of this article is organized as follows. Section 4.2 introduces our multivariate nonlinear price impact model. Subsequently, in Section 4.3 we formulate risk-neutral optimal trading problems in this context and then reformulate these "in impact space" in Section 4.4. In Section 4.5, we in turn derive necessary conditions for the absence of price manipulation and then show in Section 4.6 that these conditions allow to reduce the risk-neutral trading problems to simple pointwise optimizations. Finally, our empirical case study is described in Section 4.7. For better readability, the derivations of the no-price-manipulation conditions are delegated to the appendix.

4.2 Modeling Concave Cross Impact

We consider a financial market with $1 + d$ assets. The first one is safe, with price normalized to one. The other d assets are risky: their *unaffected prices* are modeled by an $\mathbb{R}^d$-valued Itô process S_t. This process describes price changes due to exogenous events such as news or the trades of other market participants.

The focus of the present study is how the transactions of a large trader shift these baseline prices, both directly through the "self impact" on the securities purchased or sold, but also through the "cross impact" trades in one asset have on the prices of the others. Self impact is well known to be a nonlinear function of trade sizes, and gradually decays from its peak value (Hasbrouck, 1991; Hasbrouck and Seppi, 2001). These stylized facts are captured in a parsimonious manner by the model of Alfonsi, Fruth, and Schied (2010) (henceforth AFS), where the price impact of the trades $(dQ_s)_{s\leq t}$ until time t is a nonlinear function $h(J_t)$ of an exponentially weighted moving average $dJ_t = -\frac{1}{\tau}J_t dt + \lambda dQ_t$ of current and past trades.[3] Here, the exponential smoothing captures impact decay, whereas a nonlinear impact function $h(\cdot)$ allows to account for a concave relationship between impact and executed volume.

We extend this model to a multi-asset setting with cross impact as follows.[4] Trading strategies are described by the trader's holdings $\boldsymbol{Q}_t = (Q_t^1, \dots, Q_t^d)^\top$ in the d risky assets. These in turn drive a multivariate exponential moving average $\boldsymbol{J}_t = (J_t^1, \dots, J_t^d)^\top$:

$$d\boldsymbol{J}_t = -\boldsymbol{B}\boldsymbol{J}_t dt + \boldsymbol{\Lambda} d\boldsymbol{Q}_t, \quad \boldsymbol{B}, \boldsymbol{\Lambda} \in \mathbb{R}^{d\times d}. \tag{4.2.1}$$

For a scalar function $h : \mathbb{R} \to \mathbb{R}$ that is increasing, odd, as well as nonnegative and concave on $\mathbb{R}_+$, the *price impact* $\boldsymbol{I}_t = (I_t^1, \dots, I_t^d)^\top$ of the large trader is in turn given by

$$\boldsymbol{I}_t = \sum_{a=1}^{d} \boldsymbol{L}^a h(J_t^a), \quad \text{where } \boldsymbol{L}^a \in \mathbb{R}^d \text{ for } a = 1, \dots, d. \tag{4.2.2}$$

This means that the price impact in each asset $i = 1, \dots, d$ is a linear combination $I_t^i = \sum_{a=1}^d L^{ia} h(J_t^a)$ of concave functions of the *liquidity factors* J_t^a, $a = 1, \dots d$. The liquidity factors can be the moving averages of the individual assets, for example, or also moving averages of portfolio trades, e.g., in the overall market. With the matrix of factors $\boldsymbol{L} = (\boldsymbol{L}^1, \dots, \boldsymbol{L}^d) \in \mathbb{R}^{d\times d}$ and writing, with a slight abuse of notation, $h(\boldsymbol{J}_t) = (h(J_t^1), \dots, h(J_t^d))^\top \in \mathbb{R}^d$, we can then represent the price impact concisely in

[3] An apparently similar but fundamentally different phenomenon is the nonlinear price impact of individual child orders documented empirically in Bouchaud et al. (2004), for example. Muhle-Karbe et al. (2024) show that such "local concavities" can be proxied by a linear price impact model on a mesoscopic scale in line with empirical results of Patzelt and Bouchaud (2018). In contrast, there is no such effective linear model for the "global concavities" observed at the metaorder level and described by the AFS model.

[4] Our model is a special case of the general framework proposed by (Bilarev, 2018, Chapter 5), where impact can also depend on the level of the unaffected price but no concrete nonlinear models are specified for multiple assets.

matrix-vector notation as

$$\boldsymbol{I}_t = \boldsymbol{L}h(\boldsymbol{J}_t).$$

Remark 4.2.1. *Suppose the price impact function is of the standard power form* $h(x) = \operatorname{sgn}(x)|x|^c$, $c \in (0, 1]$. *Then, for a single asset, its homotheticity implies that changing the outside multiplier* $\boldsymbol{L}$ *has the same effect as rescaling the push factor* $\boldsymbol{\Lambda}$ *by an appropriate power of the same factor. In contrast, in the multi-asset case, sums of powers and powers of sums generally lead to different models.*

A key question for the cross impact model (4.2.2) is whether it can guarantee the absence of "price manipulation". These are trading strategies that produce positive expected profits not because of accurate forecasts about the unaffected price, but by combinations of purchases and sales that turn price impact into profits. Such strategies are highly model dependent and unlikely to be effective in practice. Ruling them out therefore is a basic requirement any price impact model should satisfy, similar to the absence of arbitrage for option pricing models.

For a single asset, the AFS model with constant impact parameters does not allow price manipulation (Hey et al., 2025). However, with several assets, avoiding price manipulation becomes much more delicate already when impact is linear (Alfonsi et al., 2016; del Molino et al., 2020; Tomas et al., 2022b; Rosenbaum and Tomas, 2022; Abi Jaber et al., 2024; Muhle-Karbe and Tracy, 2024). In addition to understanding how to turn price forecasts into trades, characterizing the absence of price manipulation strategies in turn is another major motivation for studying the risk-neutral optimization problems that we turn to next.

4.3 Risk-Neutral Goal Functional

We now derive the trader's profits and losses (P&L) when trading with nonlinear price impact of the form (4.2.2). To this end, we first focus on smooth trading strategies $d\boldsymbol{Q}_t = \dot{\boldsymbol{Q}}_t dt$, for which the trade at time t is executed at $\boldsymbol{S}_t + \boldsymbol{I}_t$, the unaffected price shifted by the price impact accumulated so far.[5] If the trader's position $\boldsymbol{Q}_T$ at the terminal time T is evaluated with the unaffected price to avoid illusionary profits, then the P&L accumulated over the trading interval $[0, T]$ is

$$Y_T = \boldsymbol{Q}_T^\top \boldsymbol{S}_T - \int_0^T (\boldsymbol{S}_t + \boldsymbol{I}_t) d\boldsymbol{Q}_t.$$

Writing

$$\boldsymbol{\alpha}_t = \mathbb{E}_t \left[\boldsymbol{S}_T - \boldsymbol{S}_t\right]$$

[5] In contrast, discrete block trades require a delicate specification of where they need to be settled between the pre- and post-trade prices to be consistent with approximations of the block trade by smooth strategies. We sidestep this technical issue by first focusing on smooth strategies only and then reformulating the corresponding expected P&L's "in impact space", where the extension to general strategies is straightforward.

for the trader's price forecast at time t ("alpha") and integrating by parts, the *expected P&L* then is

$$\mathbb{E}\left[\int_0^T (\boldsymbol{\alpha}_t - \boldsymbol{I}_t)^\top d\boldsymbol{Q}_t\right]. \tag{4.3.1}$$

That is, each trade earns alpha and pays impact. The P&Ls are simply added across assets, but interact through the cross impact that trades in one asset may have on the execution prices of the others.

Remark 4.3.1. *Suppose the alpha signal*

$$\alpha_t = \mathbb{E}_t[S_\tau - S_t]$$

forecasts price changes until a time τ larger than the endpoint T of the trading interval. A typical example is a long-term alpha signal that does not change at all over a trading day. Then, the risk-neutral goal functional (4.3.1) *remains unchanged if the terminal position is valued with the forecast $\mathbb{E}_T[S_\tau]$ at time T.*

4.4 Passage to Impact Space?

For single-asset models, the risk-neutral goal functional (4.3.1) can be optimized by a straightforward *pointwize* maximization after "passing to impact space", i.e., switching the control variable from the risky positions Q_t to the corresponding moving averages J_t (Fruth et al., 2013; Bilarev, 2018; Ackermann et al., 2021; Hey et al., 2025).

In our multi-asset setting, as long as the push factor $\boldsymbol{\Lambda}$ is invertible (which we assume from now on), positions $\boldsymbol{Q}_t$ and the corresponding moving averages $\boldsymbol{J}_t$ are still in a one-to-one correspondence:

$$d\boldsymbol{Q}_t = \boldsymbol{\Lambda}^{-1}\boldsymbol{B}\boldsymbol{J}_t dt + \boldsymbol{\Lambda}^{-1} d\boldsymbol{J}_t. \tag{4.4.1}$$

Using this identity to replace the trades $d\boldsymbol{Q}_t$ in (4.3.1), the expected P&L becomes

$$\mathbb{E}\left[\int_0^T (\boldsymbol{\alpha}_t - \boldsymbol{L}h(\boldsymbol{J}_t))^\top \left(\boldsymbol{\Lambda}^{-1}\boldsymbol{B}\boldsymbol{J}_t dt + \boldsymbol{\Lambda}^{-1} d\boldsymbol{J}_t\right)\right].$$

Via integration by parts, this can be rewritten as

$$\mathbb{E}\left[\int_0^T \left(\bar{\boldsymbol{\alpha}}_t^\top \boldsymbol{J}_t - h(\boldsymbol{J}_t)^\top \boldsymbol{\zeta}\boldsymbol{J}_t\right) dt - \int_0^T h(\boldsymbol{J}_t)^\top \boldsymbol{\theta} d\boldsymbol{J}_t + \boldsymbol{\alpha}_T^\top \boldsymbol{\Lambda}^{-1}\boldsymbol{J}_T\right]. \tag{4.4.2}$$

Here (assuming invertibility of $\boldsymbol{L}$), we have defined

$$\boldsymbol{\zeta} = \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1}\boldsymbol{B}, \qquad \boldsymbol{\theta} = \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1}, \quad \text{and} \quad \bar{\boldsymbol{\alpha}}_t = \boldsymbol{\zeta}^\top \boldsymbol{L}^{-1}\boldsymbol{\alpha}_t - \boldsymbol{\theta}^\top \boldsymbol{L}^{-1}\boldsymbol{\mu}_t^\alpha, \tag{4.4.3}$$

for the drift rate $\boldsymbol{\mu}_t^\alpha$ ("alpha decay") of the alpha signal $\boldsymbol{\alpha}_t$.

Remark 4.4.1. *The transformations* (4.4.3) *describe a change of variable from physical space to liquidity factor space. More specifically,* $\boldsymbol{L}^{-1}$ *maps prices into factor space;* $\boldsymbol{\theta}$ *in turn is the push factor in these new coordinates and* $\boldsymbol{\zeta}$ *accounts for the contribution of impact decay.*

For a single risky asset, one can replace the term $h(J_t)dJ_t$ in the P&L (4.4.2) by applying Itô's formula to the antiderivative $H(J_T)$ of the impact function. The integrand of the dt terms and the terms associated with the terminal time T can in turn be maximized pointwise in a straightforward manner (Hey et al., 2025). In the multi-asset version of the model we consider here, this trick no longer works, as one cannot replace the cross terms $h(J_t^a)dJ_t^b$ for $a \neq b$ in this way.

We therefore first approach the problem from a somewhat less ambitious angle. To wit, in the spirit of Gatheral (2010); Bilarev (2018); Schneider and Lillo (2019), we consider some concrete parametric families of trading strategies and analyze what restrictions need to be imposed on the matrices $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ from (4.4.3) to rule out price manipulation, i.e., trades for which a positive expected P&L is generated by price impact rather than the presence of an alpha signal.

4.5 Necessary Conditions for the Absence of Price Manipulation

To derive necessary conditions for the absence of price manipulation, we suppose there is no alpha signal and focus on smooth deterministic trading strategies, for which the associated moving averages also are smooth. Then, the expected impact cost simplifies to

$$C_T = \int_0^T \left(h(\boldsymbol{J}_t)^\top \boldsymbol{\zeta} \boldsymbol{J}_t + h(\boldsymbol{J}_t)^\top \boldsymbol{\theta} \frac{d\boldsymbol{J}_t}{dt} \right) dt. \tag{4.5.1}$$

The principle of "no-dynamic-arbitrage" (Gatheral, 2010) states that price manipulation is not possible, in that this cost of trading indeed is positive for any nontrivial round-trip strategy. Unlike for a single risky asset, many different combinations of buying and selling actions need to be considered in the present context. To derive separate conditions on the elements of the matrices $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$, we design strategies in the space of liquidity factors J_t^a for $a = 1, \dots, d$, for which either the first or the second term (4.5.1) becomes negligible.

More specifically, to isolate the role of the matrix $\boldsymbol{\zeta}$, we consider trading strategies that are *symmetric* around a time point $T_\star/2 > 0$. These allow to cancel the $\boldsymbol{\theta}$-term in (4.5.1) and in turn yield conditions on the matrix $\boldsymbol{\zeta}$. Specifically, we construct strategies such that for each asset a, $J^a_{T_\star/2-\epsilon} = J^a_{T_\star/2+\epsilon}$, for all $\epsilon \in [0, T_*/2]$. Note that this does *not* mean that the corresponding trades Q_t^a are symmetric. This provides another illustration how the passage to impact space simplifies calculations – not just for pointwise maximization but also to construct convenient test strategies.

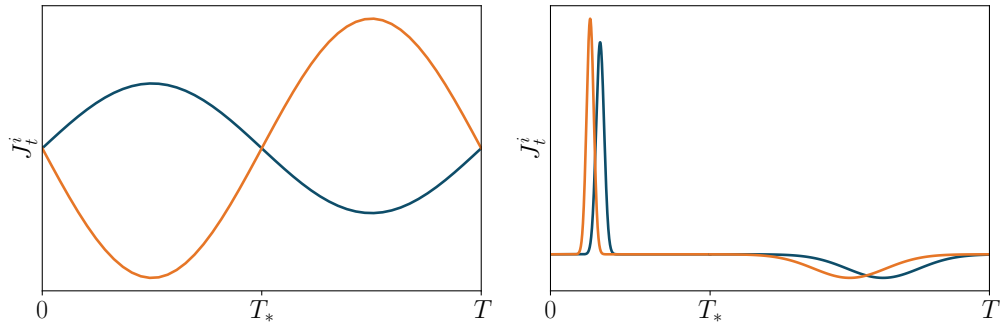


Figure 4.3: Symmetric strategy (4.5.4) (left panel) and impulsive strategy (4.5.8) (right panel).

Under sufficient regularity, such symmetry in J_t^a implies that its time derivative is antisymmetric:

$$\frac{dJ^a_{T_\star/2-\epsilon}}{dt} = -\frac{dJ^a_{T_\star/2+\epsilon}}{dt}. \tag{4.5.2}$$

As a consequence, the $\boldsymbol{\theta}$-terms in the cost functional (4.5.1) in turn vanish on each interval $[T_\star/2-\epsilon, T_\star/2+\epsilon]$ around the point $T_\star/2$.

Conversely, to focus on the matrix $\boldsymbol{\theta}$, we can consider "impulsive" strategies that quickly build up and liquidate positions in pairs of the assets. Indeed, for such fast trading strategies the derivatives dJ_t^a/dt of the impact factors states become larger and larger and therefore dominate the impact costs (4.5.1).

In the next two sections, we consider specific examples for such symmetric and impulsive strategies (illustrated in Figure 4.3), for which the impact costs can be computed in closed form. This in turn allows us to derive explicit conditions that are necessary to rule out price manipulation.

4.5.1 Symmetric strategies

For a single risky asset, gradually building up a target position and then reverting the trade always leads to positive trading costs (Gatheral, 2010). With multiple risky assets, however, price manipulation can be possible even with such a simple strategy. To rule this out, nontrivial conditions have to be imposed on the matrix $\boldsymbol{\zeta}$.

To derive such conditions, we think of the two liquidity factors a and b as virtual assets that are traded using a strategy (J_t^a, J_t^b) with the symmetries illustrated in the left panel of Figure 4.3:

- On $[0, T_\star]$, impact in one of the assets is first built up and then reversed in a symmetric manner. The other asset is traded in exactly the opposite direction.

- On $[T_\star, T]$, the pattern is the same but the direction of trade is reversed.

Observe that by (4.4.1) and (4.4.3),

$$\boldsymbol{L}^\top \boldsymbol{Q}_T = \int_0^T \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{B} \boldsymbol{J}_t dt + \int_0^T \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1} d\boldsymbol{J}_t = \boldsymbol{\zeta} \int_0^T \boldsymbol{J}_t dt + \boldsymbol{\theta} \int_0^T \frac{d\boldsymbol{J}_t}{dt} dt. \tag{4.5.3}$$

Due to the symmetry (4.5.2), the $\boldsymbol{\theta}$-term vanishes. We can therefore always choose a suitable magnitude of the trade reversal for which the round-trip condition $0 = \boldsymbol{Q}_T = (\boldsymbol{L}^\top)^{-1} \boldsymbol{L}^\top \boldsymbol{Q}_T$ holds. A particularly convenient parametrization to compute the corresponding impact costs in closed form is

$$(J_t^a, J_t^b) = (j_a \sin(t), -j_b \sin(t)), \quad 0 \leq t < 2\pi. \tag{4.5.4}$$

In the impact costs of the round-trip trade, the time-dependent terms factor out, and the sign in turn only depends on the volume ratio $\phi = j_b/j_a$. Varying this parameter in turn leads to the following necessary conditions for the absence of price manipulation, derived in Appendix 4.A:

Lemma 4.5.1. *Suppose the price impact function is of power form, $h(x) = \operatorname{sgn}(x)|x|^c$ for $0 < c \leq 1$. Then, to avoid price manipulation, the entries of the matrix $\boldsymbol{\zeta}$ need to satisfy:*

$$0 < \zeta_{aa} + \phi^{1+c} \zeta_{bb} - \phi^c \zeta_{ab} - \phi \zeta_{ba}, \quad \text{for all } \phi \geq 0 \text{ and } a, b = 1, \ldots, d. \tag{4.5.5}$$

Specialized to small values of ϕ, Condition (4.5.5) implies that the diagonal elements ζ_{aa} of $\boldsymbol{\zeta}$ all need to be nonnegative. For linear price impact ($c = 1$), the right-hand side of (4.5.5) in turn is a quadratic function of ϕ whose unique minimum allows to simplify this constraint to $\bar{\zeta}_{ab}, \bar{\zeta}_{ba} > 0$ as well as

$$\bar{\zeta}_{aa} > \frac{1}{4} (\bar{\zeta}_{ab} + \bar{\zeta}_{ba})^2, \tag{4.5.6}$$

where $\bar{\zeta}_{aa} = \zeta_{aa}/\zeta_{bb}$, $\bar{\zeta}_{ab} = \zeta_{ab}/\zeta_{bb}$, and $\bar{\zeta}_{ba} = \zeta_{ba}/\zeta_{bb}$. This condition is satisfied for positive eigenvalues of matrix $\boldsymbol{\zeta}$. To wit, so that the off-diagonal elements corresponding to cross impact have to be small enough relative to the diagonal elements describing self impact (both in factor space).

For strictly concave price impact ($c < 1$), the constraint (4.5.5) is more involved but qualitatively and quantitatively rather similar. Let us illustrate this for the case $c = 1/2$ corresponding to the "square-root law" that is well established for self impact. Then (in addition to again requiring all elements to be positive), we need

$$\bar{\zeta}_{aa} > \frac{2}{27} \left(\left(3\bar{\zeta}_{ab} + \bar{\zeta}_{ba}^2 \right)^{3/2} + \bar{\zeta}_{ba}^3 \right) + \frac{1}{3} \bar{\zeta}_{ab} \bar{\zeta}_{ba}. \tag{4.5.7}$$

[6] The left panel of Figure 4.4 visually compares the conditions for $c = 1$ and $c = 1/2$

[6] This inequality emerges from the positivity condition of the real root of a cubic equation that characterizes the optimal cost under square-root impact. Its structure is closely related to the discriminant

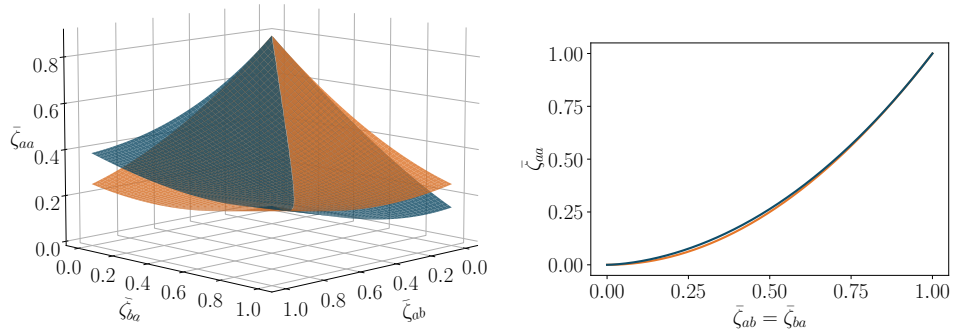


Figure 4.4: The constraints for linear price impact ($c = 1$, orange), and square-root impact ($c = 1/2$, blue): admissible values of $\bar{\zeta}_{aa}$ need to lie above the surface (left panel) and curve (right panel, for symmetric $\boldsymbol{\zeta}$), respectively.

by plotting the surfaces that correspond to their right-hand sides. We see that the constraints are qualitatively and quantitatively rather similar. In particular, in the symmetric case $\bar{\zeta}_{ab} = \bar{\zeta}_{ba}$, the conditions are virtually the same as illustrated in the right panel of Figure 4.4.

4.5.2 Impulsive Strategies

Next, we derive conditions on the matrix $\boldsymbol{\theta}$. To this end, we consider the following family of "impulsive" trading strategies:[7]

$$(J_t^a, J_t^b) = \begin{cases} \left(\frac{1}{\sqrt{2\pi(\sigma_1^a)^2}} e^{-\frac{(t-\mu_1^a)^2}{2(\sigma_1^a)^2}}, \quad \frac{1}{\sqrt{2\pi(\sigma_1^b)^2}} e^{-\frac{(t-\mu_1^b)^2}{2(\sigma_1^b)^2}} \right) & 0 \le t < T_*, \\ \left(\frac{-1}{\sqrt{2\pi(\sigma_2^a)^2}} e^{-\frac{(t-\mu_2^a)^2}{2(\sigma_2^a)^2}}, \quad \frac{-1}{\sqrt{2\pi(\sigma_2^b)^2}} e^{-\frac{(t-\mu_2^b)^2}{2(\sigma_2^b)^2}} \right) & T_* \le t \le T. \end{cases} \tag{4.5.8}$$

As illustrated in Figure 4.3, we choose $\sigma_1^{a,b} \ll \sigma_2^{a,b}$ so that the "impulsive" trades corresponding to the first humps dominate the overall trading costs (4.5.1). The smaller second humps then correspond to a slower unwinding of the position built up in the first ones.

The necessary conditions to avoid price manipulation are derived in Appendix 4.B and they demand a strikingly strict shape of matrix $\boldsymbol{\theta}$ when impact is strictly concave rather than linear:

Lemma 4.5.2. *To avoid price manipulation:*

(i) For a linear impact function $h(x) = x$, the matrix $\boldsymbol{\theta}$ must be symmetric*;*

in *Cardano's formula* for solving cubic equations.

[7] Here, the trade direction is reverted at $\mu_2 \gg T_* \gg \mu_1$ such that $J_{T_*}^{a,b} \to 0$ and the two parts of the strategy are smoothly pasted together. Alternatively, one could directly apply a mollification operator around T_*.

(ii) For a concave impact function $h(x) = \operatorname{sgn}(c)|x|^c$ *where* $0 < c < 1$*, the matrix* $\boldsymbol{\theta}$ *must be* diagonal*.*

For linear price impact models, condition (i) reproduces the results of Schneider and Lillo (2019). However, when price impact is strictly concave, then the corresponding condition in (ii) turns out to be much stronger, in that off-diagonal elements do not only have to be symmetric but instead have to vanish. Whence, to avoid price manipulation the cross terms that prevented us from rewriting the goal functional (4.4.2) in impact space in fact have to vanish. Put differently, the models for which the optimization problem (4.4.2) is intractable are ruled out already by imposing no price manipulation.

Remark 4.5.3. *The intuitive meaning of restricting* $\boldsymbol{\theta}$ *to be diagonal becomes apparent when rewriting the dynamics of* $\boldsymbol{J}_t$ *in factor space:*

$$d\boldsymbol{J}_t = \boldsymbol{\theta}^{-1}\left(-\boldsymbol{\zeta}\boldsymbol{J}_t dt + d\boldsymbol{L}^T\boldsymbol{Q}_t\right). \tag{4.5.9}$$

This shows that – in factor space – instantaneous cross-factor impact should be zero in order to avoid price manipulation. It is important to note, however, that this does not mean that there is no instantaneous cross impact (e.g., through the matrix $\boldsymbol{L}$*) in physical space.*

Remark 4.5.4. *In terms of modelling, ensuring that* $\boldsymbol{\theta}$ *is diagonal imposes constraints on the choice of the model parameters* $\boldsymbol{L}$ *and* $\boldsymbol{\Lambda}$*. Indeed, as*

$$\boldsymbol{\Lambda} = \boldsymbol{\theta}^{-1}\boldsymbol{L}^\top,$$

we see that once the matrix $\boldsymbol{L}$ *has been fixed, there are only* d *(rather than* $d \times d$*) degrees of freedom to be fixed when choosing* $\boldsymbol{\Lambda}$*, corresponding to the diagonal elements of* $\boldsymbol{\theta}$*.*

4.6 Solution of the Risk-Neutral Optimization

In view of Lemma 4.5.2, we henceforth assume that the matrix $\boldsymbol{\theta} = \boldsymbol{L}^\top\boldsymbol{\Lambda}^{-1}$ is diagonal to rule out price manipulation. Then, the cross terms $\theta^{ab}h(J_t^a)dJ_t^b$, $a \neq b$ disappear in the risk-neutral goal functional (4.4.2). Using Itô's formula to replace the terms $h(J_t^a)dJ_t^a$ with $H(J_T^a)$, where $H(\cdot)$ is the antiderivative of the price impact function $h(\cdot)$, (4.4.2) can therefore can be reduced to a simple pointwise maximization just like in the single-asset case (Hey et al., 2025):

$$\mathbb{E}\left[\int_0^T \left(\bar{\boldsymbol{\alpha}}_t^\top\boldsymbol{J}_t - h(\boldsymbol{J}_t)^\top\boldsymbol{\zeta}\boldsymbol{J}_t\right)dt + \boldsymbol{\alpha}_T^\top\boldsymbol{\Lambda}^{-1}\boldsymbol{J}_T - \mathbf{1}^\top\boldsymbol{\theta}H(\boldsymbol{J}_T)\right]. \tag{4.6.1}$$

Remark 4.6.1. *Unlike its counterpart* (4.4.2) *in trade space, the goal functional* (4.6.1) *in impact space only depends on the liquidity factors* J_t^a*, but not their derivatives.*

Whence, as in Becherer et al. (2019); Ackermann et al. (2021), it can easily be extended to general strategies in a consistent manner, by defining their P&L as the limit of the P&Ls of a sequence of approximating smooth strategies. In trade space, such an approach does not work because the derivatives of the approximating strategies typically blow up.

At the terminal time T – when neither impact on future trades nor alpha decay needs to be considered anymore – one can check that the optimal impact state always exhausts the entire available alpha signal ($I_T = \alpha_T$), just like in the single-asset version of the model. The optimization at intermediate times $t \in (0, T)$ does not generally admit a closed-form solution, but can be solved explicitly in an important special case that we consider first.

4.6.1 Decomposition into Univariate Subproblems

Suppose that the matrix $\boldsymbol{\zeta} = \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{B}$ is also diagonal, e.g., because not just $\boldsymbol{\theta} = \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1}$ is diagonal (as required for the absence of price manipulation) but the impact decay matrix $\boldsymbol{B}$ is diagonal as well. Then, the multivariate optimization problem (4.6.1) decomposes into d separate univariate subproblems. Each of these can in turn be solved as in the single-asset case (Hey et al., 2025, Theorem 4.2). In particular, the necessary conditions from Lemmas 4.5.1 and 4.5.2 indeed suffice to rule out price manipulation in this case.

Crucially, assuming the matrices $\boldsymbol{\theta} = \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1}$ and $\boldsymbol{\zeta} = \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{B}$ to be diagonal does *not* mean that the model has no cross impact. Indeed, if $\boldsymbol{\Lambda}$ and $\boldsymbol{L}$ are multiples of the same symmetric matrix (e.g., the covariance matrix of asset returns as in (Gârleanu and Pedersen, 2013, Assumption 1) or its square root) and the decay matrix is diagonal (as in Gârleanu and Pedersen (2016)), then both $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ are clearly diagonal. However, the price impact $I_t^i = \sum_{a=1}^d L^{ia} h(J_t^a)$ in asset i then still depends on all the liquidity factors, because the matrix $\boldsymbol{L}$ does not have to be diagonal. In the case where $\boldsymbol{L}$ and $\boldsymbol{\Lambda}$ are both multiples of the covariance matrix of positively correlated assets (or its square root), this leads to positive cross impact through two channels: on the one hand, trades in one asset not only affect the corresponding liquidity factor but also shift the other ones in the same direction (though the matrix $\boldsymbol{\Lambda}$). The impact on each asset then is obtained as a positive combination of the positively correlated impact factors (through the matrix $\boldsymbol{L}$).

As a concrete example, suppose both $\boldsymbol{\theta} = \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1}$ and $\boldsymbol{\zeta} = \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{B}$ are diagonal and the impact function $h(x) = \operatorname{sgn}(x)|x|^c$, $c \in (0, 1]$ is of power form. Then, pointwise maximization of (4.6.1) yields an explicit formula for the optimal impact state

$$\boldsymbol{I}_t^* = \begin{cases} \frac{1}{1+c} \left(\boldsymbol{\alpha}_t - \boldsymbol{L} \boldsymbol{\zeta}^{-1} \boldsymbol{\theta}^\top \boldsymbol{L}^{-1} \boldsymbol{\mu}_t^\alpha \right), & t \in (0, T), \\ \boldsymbol{\alpha}_T, & t = T. \end{cases}$$

In particular, without alpha decay ($\boldsymbol{\mu}_t^\alpha = 0$), we recover the same optimal impact

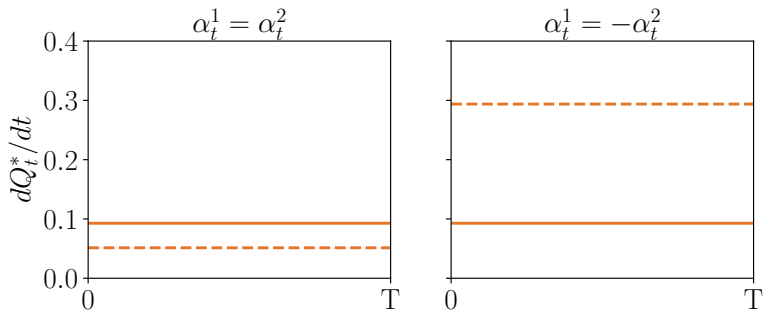


Figure 4.5: Comparison of optimal trading speeds at time $t \in (0, T)$ without cross impact (solid lines) and with cross impact (dashed lines), with parameters estimated for assets with correlation 0.6 in Section 4.7. In the left panel, the (constant) alpha signals for both assets are the same and given in bps ($\boldsymbol{\alpha}_t = (1, 1)$), so less trading is possible with the same optimal impact state. In the right panel, the signal of the alpha signals are opposite ($\boldsymbol{\alpha}_t = (1, -1)$), so that cross impact increases the optimal trading volumes.

states as in a collection of single asset versions of the model: a first bulk trade pushes the optimal impact state to a fraction $1/(1+c)$ of the corresponding alpha signal at the initial time $t = 0$. Subsequently, one trades to maintain this impact state (by continuing to trade in the same direction to offset impact decay) until the terminal time T, where the remaining signal is exhausted with another bulk trade. These optimal impact states do not change here due to the presence of cross impact, but the same is *not* true for the corresponding trades. Indeed, at time $t \in (0, T)$, the optimal trading rate depends on the matrix $\boldsymbol{L}$ and the alpha signals in all d assets:

$$\frac{d\boldsymbol{Q}_t^*}{dt} = (\boldsymbol{L}^\top)^{-1} \boldsymbol{\zeta} h^{-1} \left(\frac{\boldsymbol{L}^{-1} \boldsymbol{\alpha}_t}{1+c} \right), \tag{4.6.2}$$

where the inverse $h^{-1}(x) = \operatorname{sgn}(x)|x|^{1/c}$ of the impact function is applied componentwise. Figure 4.5 illustrates the implications of this formula for $d = 2$ assets with parameters estimated from asset pairs with return correlation 0.6 in Section 4.7. More specifically, we compare the optimal trading rates in the calibrated model with cross impact (i.e., with a nondiagonal matrix $\boldsymbol{L}$) to the optimal trading rate in an otherwise identical model where the off-diagonal elements of $\boldsymbol{L}$ are set to zero. We see that for aligned alpha signals cross impact reduces the trading speed substantially. Conversely, for anti-aligned signals, the optimal trading rate is substantially larger with cross impact.

4.6.2 The Bivariate Case

When the risk-neutral problem (4.6.1) does not decompose into separate univariate subproblems, it remains easy to solve numerically via pointwise maximization of the integrand. However, its analytical analysis becomes considerably more involved. Indeed, simple numerical examples show that already for two risky assets ($d = 2$), the goal functional (4.6.1) generally is *not* a concave function of the controls (J_t^1, J_t^2). Whence,

there is little hope to establish uniqueness in general.

Example 4.6.2. *As the alpha signal does not affect the concavity of the goal functional, we focus on the impact terms in* (4.6.1)*. For square-root impact* $h(x) = \operatorname{sgn}(x)|x|^{1/2}$ *and* $\zeta_{11} = \zeta_{22} = 1$, $\zeta_{12} = 0.1$, $\zeta_{21} = 0.9$, *the constraint* (4.5.7) *is satisfied so that price manipulation is not possible with the symmetric strategies from Lemma 4.5.1.*

The left panel of Figure 4.6 plots the integrand $-h(\boldsymbol{J}_t)^\top \boldsymbol{\zeta} \boldsymbol{J}_t$ *of the goal functional* (4.6.1) *as a function of the liquidity factors* J_t^1, J_t^2*. This function clearly has a unique maximum at* $J_t^1 = J_t^2 = 0$*, consistent with the absence of price manipulation. However, it is* not *globally concave. This is illustrated in the right panel of Figure 4.6, which plots* $(J_t^1, J_t^2) H (J_t^1, J_t^2)^\top$ *for the Hessian matrix* H *of the integrand at the point* $(J_t^1, J_t^2) = (15, 3)$*. This function takes some positive values, so the integrand is not a globally concave function of the impact states.*

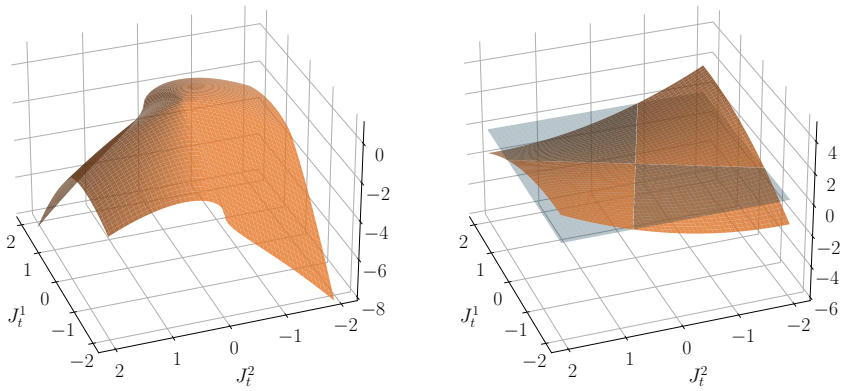


Figure 4.6: Panel on the left: the integrand $-h(\boldsymbol{J}_t)^\top \boldsymbol{\zeta} \boldsymbol{J}_t$ of the goal functional (4.6.1) plotted against the liquidity factors J_t^1, J_t^2. Panel on the right: $(J_t^1, J_t^2) \boldsymbol{H} (J_t^1, J_t^2)^\top$ for the Hessian matrix $\boldsymbol{H}$ of the integrand at the point $(J_t^1, J_t^2) = (15, 3)$, plotted against the liquidity factors. The blue plane separates the positive and negative domain. Model parameters are chosen as in Example 4.6.2.

However, for two risky assets, it is easy to check that when the price impact function is of power form ($h(x) = \operatorname{sgn}(x)|x|^c$, $c \in (0, 1]$) then the constraint (4.5.5) from Lemma 4.5.1 is exactly what is needed to guarantee that the goal function (4.6.1) is bounded from above and becomes negative for sufficiently large absolute values of J_t^1 or J_t^2.[8] As a consequence, a global optimum always exists in this case, but may not be unique.[9]

With some algebraic manipulations (cf. Appendix 4.C), the first order conditions that any maximum (J_t^1, J_t^2) must satisfy can be reduced to a single autonomous equation for

[8] Indeed, this is clear when J_t^1 and J_t^2 have the same sign. When they have opposite signs, this follows from (4.5.5) by changing variables from J_t^2 to $\kappa_t J_t^1$ and using the homotheticity of the power function.

[9] In the symmetric case ($\zeta_{11} = \zeta_{22}$, $\zeta_{12} = \zeta_{21}$, $\theta_{11} = \theta_{22}$, $\bar{\alpha}^1 = \bar{\alpha}^2$, and $\bar{\mu}^1 = \bar{\mu}^2$), any maximizer then needs to be symmetric by a classical result of Bouniakovsky (1854) when the first order condition is a cubic polynomial for square-root impact ($c = 1/2$). This again reduces the problem at hand to a one-dimensional optimization, for which uniqueness follows from concavity.

the ratio $\phi = J_t^2/J_t^1$: [10]

$$0 = \phi_t + \text{sgn}(\phi_t)|\phi_t|^c k_t^1 + \text{sgn}(\phi_t)|\phi_t|^{c-1} k_t^2 + k_t^3, \tag{4.6.3}$$

where

$$k_t^1 = \frac{1}{c} - \frac{1+c}{c}\frac{\bar{\alpha}_t^1}{\bar{\alpha}_t^2}\frac{\zeta_{bb}}{\zeta_{ab}}, \quad k_t^2 = \frac{\bar{\alpha}_t^1}{\bar{\alpha}_t^2}\frac{\zeta_{ba}}{\zeta_{ab}}, \quad k_t^3 = \frac{\bar{\alpha}_t^1}{c\bar{\alpha}_t^2}\frac{\zeta_{ba}}{\zeta_{ab}} + \frac{\zeta_{aa}}{\zeta_{ab}}\frac{(1+c)}{c}. \tag{4.6.4}$$

The product $J_t^1 J_t^2$ and in turn the individual impact states are pinned down by Equation (4.C.1) in Appendix 4.C. In the empirically most relevant case of square-root impact ($c = 1/2$), changing variables to a power $1/c$ of ϕ leads to a cubic equation for positive ϕ_t, and another for negative values of ϕ_t. The three roots of each of these equations then need to be compared directly to the points where one or both of the variables vanish.

4.6.3 The General Case

For more than two risky assets, both existence and uniqueness for the maximization of (4.6.1) are challenging open problems for further research. On the one hand, it is not clear whether the necessary conditions derived by considering pairs of liquidity factors in Lemma 4.5.1 are sufficient to guarantee that the goal functional remains bounded from above in general. On the other hand, establishing uniqueness in the absence of concavity also is a wide-open problem.

One regime that can be treated directly is the case of *small* off diagonal terms for which the model is close to the decoupled case discussed in Section 4.6.1. Indeed, if the off-diaongal elements of $\boldsymbol{\zeta}$ are sufficiently small, then it is easy to check that any maximum must lie on a compact set, and that the integrand of the goal functional is strictly concave on the latter. Whence, there is a unique maximum characterized by the first-order conditions

$$\bar{\alpha}_t^a = (1+c)\zeta_{aa}\text{sgn}(J_t^a)|J_t^a|^c + \sum_{b\neq a}^{d} \zeta_{ab}\left(\text{sign}(J_t^b)|J_t^b|^c + cJ_t^b|J_t^a|^{c-1}\right).$$

These optimality equations are nonlinear and coupled, but can be solved in closed-form using the implicit function theorem when the off-diagonal elements of $\boldsymbol{\zeta}$ are small. Indeed, if $\bar{\alpha}_t^a \neq 0$, $a = 1\ldots, d$, then there exists a solution of the first-order conditions. In the case where all of diagonal elements are the same to ease notation ($\zeta_{ab} = \zeta$), the

[10] When $J_t^1 \to 0$, the ratio $\phi_t = J_t^2/J_t^1$ becomes ill-defined. In such cases, it is more appropriate to consider the inverse ratio $\tilde{\phi}_t = J_t^1/J_t^2 \to 0$, and reformulate the first-order conditions accordingly. This allows for a local analysis around $\phi_t \to \infty$ (or $\tilde{\phi}_t \to 0$), ensuring that solutions can still be characterized through a modified version of Equation (4.6.3).

corresponding optimal impact states have the leading-order asymptotics

$$\boldsymbol{I}_t = \sum_{a=1}^{d} \boldsymbol{L}_a \frac{\bar{\alpha}_t^a}{\zeta_{aa}(1+c)} \left[1 - \sum_{b \neq a}^{d} \frac{\operatorname{sgn}(\bar{\alpha}_t^b)\zeta}{c(c+1)} \left(c \left| \frac{\bar{\alpha}_t^b}{\bar{\alpha}_t^a} \right|^{1/c} + \left| \frac{\bar{\alpha}_t^b}{\bar{\alpha}_t^a} \right| \right) \right]^c . \tag{4.6.5}$$

If all alpha signals in the latent factor space have the same sign, then this implies that cross-impact leads to smaller optimal impact states, as traders reduce their aggressiveness to internalize mutual costs. However, if alpha signals have opposing signs or sufficiently unbalanced magnitudes, cross-impact may instead amplify trading, resulting in larger optimal impact states.

4.7 Empirical Analysis

With a general consistent modeling framework at hand, we now turn to its empirical validation. To implement this, the no-price-manipulation conditions derived in Section 4.5 play a key role. Indeed, by narrowing down the parameter space for sensible models, these conditions increase the robustness of the empirical calibration. Using proprietary metaorder data, this allows us to reliably identify the concave structure of cross impact as well as its decay patterns.

4.7.1 Data

In this paper, we use CFM's proprietary metaorder dataset, cf. Hey et al. (2025) for more details. Additionally, we use public data to determine the mid prices at the start and end of each metaorder and to estimate the volatilities, correlations, and average daily traded volumes of all asset pairs.

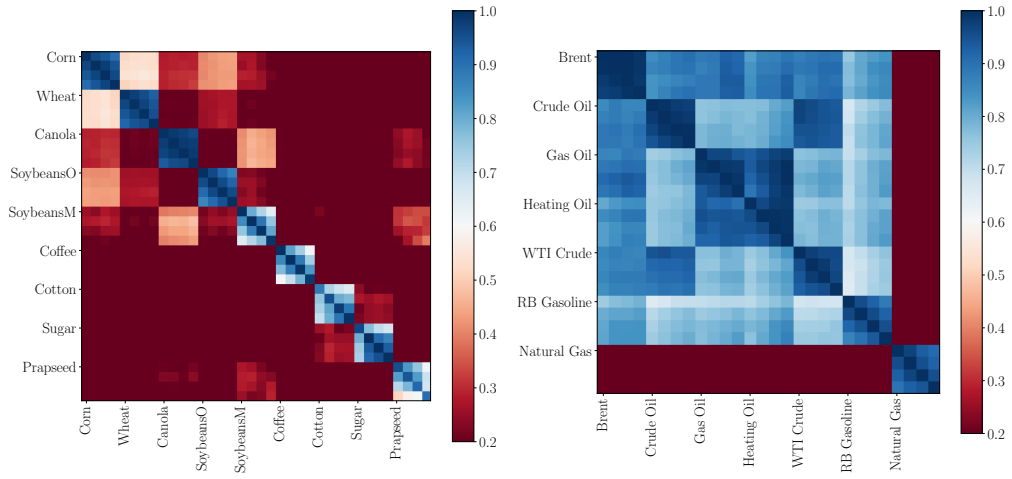
Figure 4.7 displays the return correlations of a subset of various futures contracts included in the proprietary dataset. The left panel focuses on agricultural futures, which are available with four different maturities, separated by a quarter of a year each. The corresponding returns have a high correlation, which typically decreases slightly as the distance between maturities increases. In contrast, there is not much inter product correlation.

As a complement, the right panel of Figure 4.7 plots the corresponding correlations for energy contracts. These display much larger intra-product correlations, since they mostly depend on the same underlying resources.

Other future contracts in the data set include metals and indices that offer a wide range of pairwise return correlations. This will allow us to study below how cross impact estimates depend on the corresponding asset correlations.

4.7.2 Fitting Methodology

Price returns of pairs of assets are fitted against the cross impact model (4.5.9). To obtain a system of decoupled equations as in 4.6.1, we assume that $\boldsymbol{\theta}^{-1}\boldsymbol{\zeta} = \beta \cdot \boldsymbol{Id}_2$ is



(a) Agricultural Products (b) Energy Products

Figure 4.7: Return correlations between different futures contracts.

a diagonal matrix with a single impact decay parameter β across all assets. Then the following impact formula emerges:

$$\boldsymbol{I}_t = \sum_{a=1}^{2} \boldsymbol{L}_a h(\theta_{aa}^{-1}) h(J_t^a), \quad \text{where } J_t^a = \int_0^t e^{-\beta(t-s)} d(\boldsymbol{L}^T \boldsymbol{Q}_s)^a. \tag{4.7.1}$$

Even under these assumptions that hard code the absence of price manipulation, there remains some freedom in how to choose the matrix $\boldsymbol{L}$. In a static linear model, this problem is studied by del Molino et al. (2020); a systematic extension of their results that link to our dynamic nonlinear model is an important direction for future research. In the present study, we focus on the simplest consistent extension of the typical normalizations for single-asset models. To wit, we choose $\boldsymbol{L} = \boldsymbol{\Sigma}^{1/2}$, so that impact scales with volatility for uncorrelated assets. Moreover, the trades dQ_t^i of each asset are normalized by the geometric mean of the average trading volumes of the asset pair.[11]

This bivariate impact model is in turn calibrated for eight equal-sized batches of about 100 product pairs each, sorted by correlations. The corresponding exponentially weighted moving averages J_t^a are precomputed on a grid of values for the impact decay rates β. With these moving averages at hand, we then calculate the terms $\boldsymbol{L}_a h(J_t^a)$ in (4.7.1) for a grid of different concavity coefficients c. Finally, for each pair (c, β), we regress the predicted returns from the cross impact model against the true observed returns. This allows us to fit the remaining two parameters $(h(\theta_{11}^{-1})(c,\beta), h(\theta_{22}^{-1})(c,\beta))$ by maximizing the model fit $R^2(c,\beta)$.[12]

[11] In the single-asset case, volumes are naturally expressed relative to the asset's own average volume. However, in the multivariate case, a normalization by individual volumes would typically not commute with the matrix $\boldsymbol{L}$ and is therefore not guaranteed to be consistent with the absence of price manipulation. In contrast, normalizing volumes by a single constant across both assets allows to ensure consistency with the no-manipulation condition.

[12] Due to the normalization of trading volumes, the regression coefficients are of order one. This

To assess the relative contributions of self and cross impact, we consider the "self impact weights" $w_S^i(c, \beta, \rho)$:

$$w_S^i(c, \beta, \rho) = \vec{\nabla} I^i(c, \beta, \rho)_i^2 ||\vec{\nabla} I^i(c, \beta, \rho)||^{-2}, \tag{4.7.2}$$

where $\vec{\nabla} = (\partial_{Q_1}, \partial_{Q_2})^\top$ represents the gradient operator which acts on the price impact function I^i of asset i and computes the partial derivatives of the price impact with respect to the traded volumes $\boldsymbol{Q}$ of both assets. It thereby captures the sensitivity of the price impact to local changes in trading activity.[13] These individual sensitivities are in turn normalized by the aggregate sensitivity captured by the norm of the whole gradient.

4.7.3 Results

The key findings of the empirical analysis summarized above are:

i) Cross impact is highly concave: the concavity parameter varies between 0.5 and 0.7.

ii) Cross impact decays on a daily timescale: the decay rate β varies between 0.1 and 0.9 per day, corresponding to a half-life of 0.7 to 7 days.

iii) The importance of cross impact depends on correlation: as correlation increases, the self impact weight decreases. In particular, for highly correlated asset pairs, bivariate cross impact accounts for nearly 50% of the total measured impact.

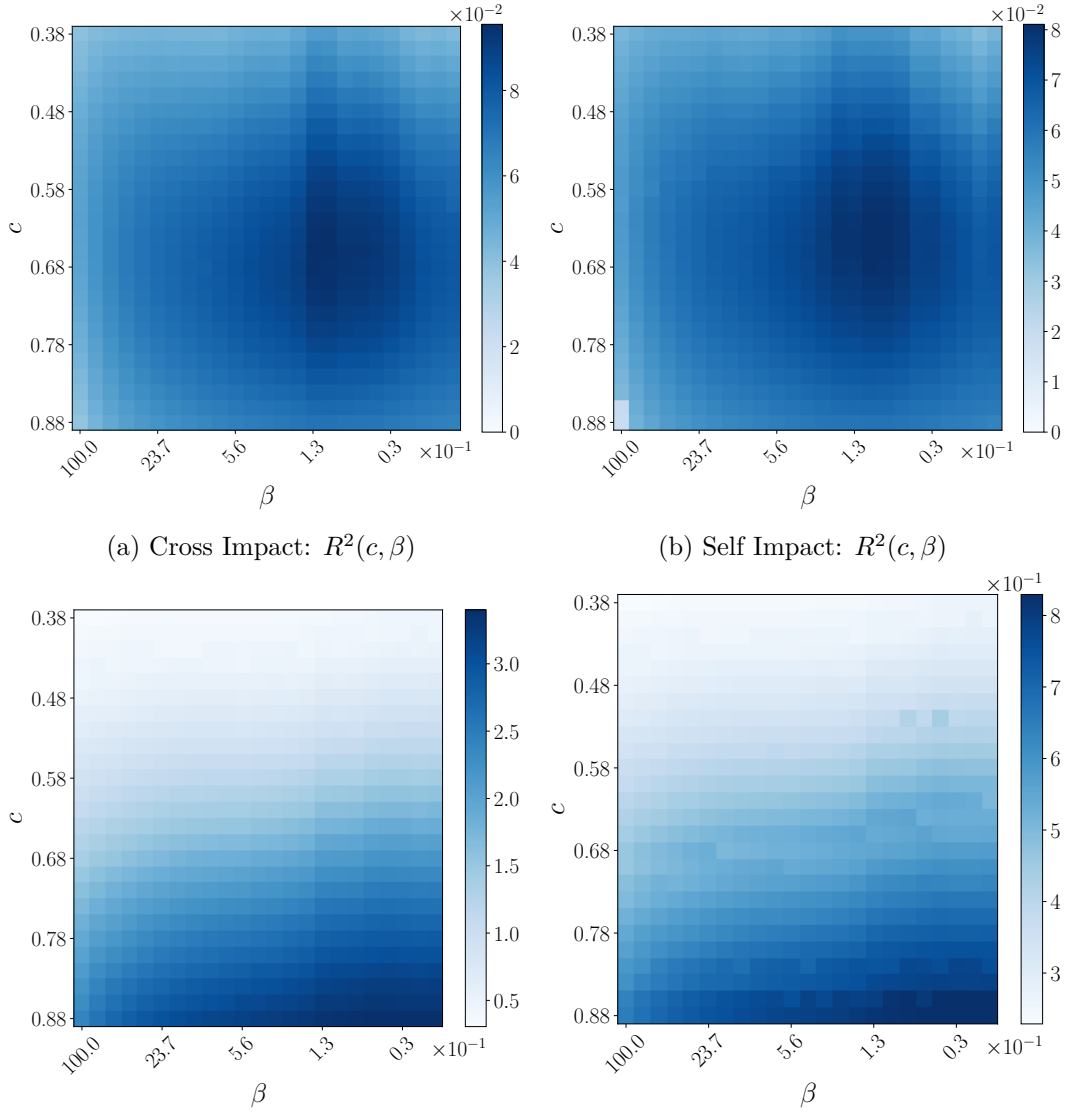
To illustrate this, Figure 4.8a shows the R^2 of the fitted cross impact model as a function of impact concavity c and impact decay β for product pairs with an average return correlation of $\rho = 0.95$. Figure 4.8b presents the R^2 values considering only self impact, where the matrix $\boldsymbol{L}$ is diagonal and each volume is normalized by its own daily volume. The maximum R^2 achieved with cross impact is approximately $9.5 \cdot 10^{-2}$, which is 18% higher than the maximum R^2 for self impact only. The prefactors $h(\theta_{11}^{-1})(c, \beta)$ and $h(\theta_{22}^{-1})(c, \beta)$ in (4.7.1) are plotted in Figures 4.8c and 4.8d, respectively.[14]

Figure 4.9 extends the analysis described in Figure 4.2 from the introduction, which examines how the fitted model parameters depend on the return correlation ρ between the assets. In addition to the results for cross impact fitting, Figure 4.9 includes the point estimates for self impact-only fits. We see that these estimates for concavity c and impact decay β are encouragingly consistent with their counterparts for the cross-impact version of the model.

simplifies the fitting procedure and reduces the sensitivity to scaling issues.

[13] Unlike for impact models, the sensitivities (4.7.2) generally depend on the trade sizes at which they are evaluated. However, this dependency turns out to be rather weak, in that the bottom panel of Figure 4.9 only changes slightly if the evaluation point is changed from anti-aligned to aligned trades.

[14] These prefactors are not directly comparable to the single asset version of the model studied in Hey et al. (2025), as each asset's trading volume is normalized by the geometric mean of both assets' average trading volumes here, rather than just its own average volume.



(a) Cross Impact: $R^2(c, \beta)$

(b) Self Impact: $R^2(c, \beta)$

(c) Coss-Impact: $h(\theta_{11}^{-1})(c, \beta)$

(d) Coss-Impact: $h(\theta_{22}^{-1})(c, \beta)$

Figure 4.8: Calibration results of the cross impact model in 4.7.1 for a single decay timescale β: Panel (A) shows the statistical sensitivity for an arbitrage-free cross impact model with pairs that have a return correlation $\rho = 0.95$. At this correlation level, R^2 peaks at $c = 0.66$ and $\beta = 0.13$ per day. For comparison, Panel (B) shows R^2 for the self impact model. The cross impact model fits the data better since the highest R^2 in Panel (A) is by 18% larger than the one in Panel (B). Panel (C) and (D) represent the calibrated parameters $h(\theta_{11}^{-1})(c, \beta) = 1.9$ and $h(\theta_{22}^{-1})(c, \beta) = 0.5$, respectively that maximize the R^2.

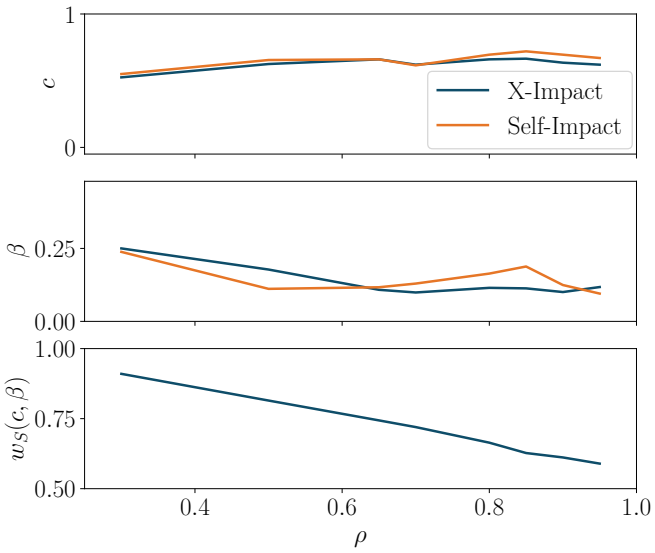


Figure 4.9: The point estimates (c, β) and the corresponding self impact weight $w_S(c, \beta)$ across ρ of the cross impact model (blue) and the self impact model (orange). The point estimates remain roughly constant and the self impact weight is inverse proportional to return correlations.

4.8 Conclusion

This paper introduces and studies a model for the concave cross impact induced by simultaneous trades of multiple assets. This framework allows to consistently address several crucially important but often conflicting requirements:

(i) The model can be used for optimization problems, whose stability hinges on ensuring that the model does not allow price manipulation;

(ii) The model exhibits full analytical tractability in some empirically relevant cases, where optimzation in impact space factorizes into univariate subproblems but cross impact nevertheless plays a key role in the sizing of the corresponding trades;

(iii) The model makes it possible to calibrate the cross impact of metaorders to empirical data, for which non-linearity and impact decay are prominent features.

More broadly, a conceptual contribution of the paper is to illustrate the interplay between model complexity and price manipulation conditions. To wit, when passing from self impact to cross impact the parameter space of the model increases quadratically in the number of assets. However, the space of manipulation strategies also grows, so that the absence of price manipulation limits the number of genuinely free parameters. Further extending this result is a key problem for future research. Indeed, any high-dimensional, possibly machine-learned model of cross impact is bound to be unsuitable for practical applications if the problem of dynamic arbitrage is not properly addressed.

Our first promising empirical results show that it is indeed possible to reliably measure the concave cross impact of metaorders and its gradual decay. An important direc-

tion for future research is to extend this proof of principle to higher-dimensional settings and to account not only for (symmetric) return correlations but also for heterogeneous asset characteristics such as the different decay rates and also the markedly different trading volumes of many highly correlated and otherwise similar assets (e.g., futures with shorter and longer maturities or on-the-run vs. off-the-run treasury bonds).

Appendix

4.A Proof of Lemma 4.5.1

To construct a roundtrip trade, we need

$$\int_0^T \frac{dQ_t^i}{dt} dt = \int_0^T \Big(\frac{dJ_t^i}{dt} + \beta J_t^i \Big) dt = 0. \tag{4.A.1}$$

Plugging in symmetric strategy (4.5.4) for asset a, we obtain

$$j_a \int_0^T (\sin(t) + \beta \cos(t))\, dt = j_a \left(-\cos(t) + \beta \sin(t)\right) \Big|_0^T.$$

Whence, the integral vanishes if we choose $T = 2n\pi$ with for an integer n. The argument for the strategy for asset b is analogous.

The impact costs (4.5.1) of the strategy (4.5.4) can be computed directly as

$$\begin{aligned} C_T = \int_0^T & [j_a \sin(t) \zeta_{aa} h\left(j_a \sin(t)\right) - j_b \sin(t) \zeta_{bb} h\left(-j_b \sin(t)\right) \\ & + j_a \sin(t) \zeta_{bb} h\left(-j_b \sin(t)\right) - j_b \sin(t) \zeta_{ba} h\left(j_a \sin(t)\right)]\, dt \\ = & \left(j_a h(j_a) \zeta_{aa} - j_b h(-j_b) \zeta_{bb} + j_a h(-j_b) \zeta_{ab} - j_b h(j_a) \zeta_{ba}\right) \int_0^T \sin(t) h(\sin(t)) dt. \end{aligned} \tag{4.A.2}$$

As $xh(x) \geq 0$ for all x, the integral term is always nonnegative, so the sign of the costs depends only on the prefactor.

When the impact function is of power form $h(x) = \operatorname{sign}(x)|x|^c$, then to guarantee nonnegative trading costs we need

$$0 \leq j_a^{1+c} \zeta_{aa} + j_b^{1+c} \zeta_{bb} - j_a j_b^c \zeta_{ab} - j_b j_a^c \zeta_{ba}. \tag{4.A.3}$$

For $j_a, j_b > 0$, we can divide this inequality by $(j_a j_b)^c$ and rewrite it in terms of the fraction $\phi = j_b / j_a$. This finally leads to the necessary condition (4.5.5) from Lemma 4.5.1.

4.B Proof of Lemma 4.5.2

The roundtrip condition for asset a requires

$$\begin{aligned} 0 = & \beta \left[\int_0^{T_*} (2\pi(\sigma_1^a)^2)^{-1/2} e^{-\frac{(t-\mu_1^a)^2}{2(\sigma_1^a)^2}} dt - \int_{T_*}^T (2\pi(\sigma_2^a)^2)^{-1/2} e^{-\frac{(t-\mu_2^a)^2}{2(\sigma_2^a)^2}} dt \right] \\ & - \int_{T_*}^T (t-\mu_1^a)(2\pi(\sigma_1^a)^2)^{-1/2} e^{-\frac{(t-\mu_1^a)^2}{2(\sigma_1^a)^2}} dt + \int_{T_*}^T (t-\mu_2^a)(2\pi(\sigma_2^a)^2)^{-1/2} e^{-\frac{(t-\mu_2^a)^2}{2(\sigma_2^a)^2}} dt. \end{aligned}$$

If we choose $0 \ll T_* \ll T$ and $\mu_1^a \in (0, T_*)$, $\mu_a^2 \in (T_*, T)$ sufficiently far away from the endpoints of these intervals, then all the integrals tend to one, so that the roundtrip

condition is satisfied in the limit (which is sufficient for the necessary condition we derive below). The argument for the strategy for asset b is analogous.

When the impact function is of the power form $h(x) = \operatorname{sgn}(x)|x|^c$ with $0 < c \leq 1$, then the impact costs (4.5.1) of the strategy (4.5.8) are given by

$$\begin{aligned} C_T = \sum_{i=1}^{2} \int_{\mathcal{I}_i} \Bigg[& \left(\zeta_{aa} - \theta_{aa}\frac{(t-\mu_i^a)}{(\sigma_i^a)^2}\right) \frac{e^{-\frac{(1+c)(t-\mu_i^a)^2}{2\sigma_i^2}}}{\sqrt{2\pi(\sigma_i^a)^2}^{1+c}} + \left(\zeta_{bb} - \theta_{bb}\frac{(t-\mu_i^b)}{(\sigma_i^b)^2}\right) \frac{e^{-\frac{(1+c)(t-\mu_i^b)^2}{2(\sigma_i^b)^2}}}{\sqrt{2\pi(\sigma_i^b)^2}^{1+c}} \\ & + \left(\zeta_{ab} - \theta_{ab}\frac{(t-\mu_i^b)}{(\sigma_i^b)^2}\right) e^{A_i^{ab}} \frac{e^{-\frac{(t-\bar{\mu}_i^{ab})^2}{2(\bar{\sigma}_i^{ab})^2}}}{\sqrt{2\pi(\sigma_i^b)^2}\sqrt{2\pi(\sigma_i^a)^2}^c} \\ & + \left(\zeta_{ba} - \theta_{ba}\frac{(t-\mu_i^a)}{(\sigma_i^a)^2}\right) e^{A_i^{ba}} \frac{e^{-\frac{(t-\bar{\mu}_i^{ba})^2}{2(\bar{\sigma}_i^{ba})^2}}}{\sqrt{2\pi(\sigma_i^a)^2}\sqrt{2\pi(\sigma_i^b)^2}^c} \Bigg] dt, \end{aligned}$$

where the integrals are computed over the intervals $\mathcal{I}_1 = [0, T_*]$ and $\mathcal{I}_2 = [T_*, T]$, respectively, and

$$\begin{aligned} (\bar{\sigma}_i^{ab})^2 &= \frac{(\sigma_i^a)^2(\sigma_i^b)^2}{c(\sigma_i^b)^2 + (\sigma_i^a)^2}, \quad \bar{\mu}_i^{ab} = \frac{\mu_i^a c(\sigma_i^b)^2 + \mu_i^b(\sigma_i^a)^2}{(\sigma_i^a)^2 + c(\sigma_i^b)^2}, \\ A_i^{ab} &= \frac{1}{2(\bar{\sigma}_i^{ab})^2}\Big[-(\bar{\mu}_i^{ab})^2\Big(c(\sigma_i^b)^2 + (\sigma_i^a)^2\Big) + (\mu_i^a)^2 c(\sigma_i^b)^2 + (\mu_i^b)^2(\sigma_i^a)^2\Big]. \end{aligned}$$

Again using that almost all mass of the Gaussians is contained on respective intervals, the impact costs become

$$\begin{aligned} C_T = \sum_{i=1}^{2} \Bigg[& \frac{1}{\sqrt{1+c}\sqrt{2\pi(\sigma_i^a)^2}^c}\zeta_{aa} + \frac{1}{\sqrt{1+c}\sqrt{2\pi(\sigma_i^b)^2}^c}\zeta_{bb} \\ & + \left(\zeta_{ab} - \theta_{ab}\frac{(\bar{\mu}_i^{ab} - \mu_i^b)}{(\sigma_i^b)^2}\right)\sqrt{\frac{(\bar{\sigma}_i^{ab})^2}{(\sigma_i^b)^2}}\frac{1}{\sqrt{2\pi(\sigma_i^a)^2}^c} e^{A_i^{ab}} \\ & + \left(\zeta_{ba} - \theta_{ba}\frac{(\bar{\mu}_i^{ba} - \mu_i^a)}{(\sigma_i^a)^2}\right)\sqrt{\frac{(\bar{\sigma}_i^{ba})^2}{(\sigma_i^a)^2}}\frac{1}{\sqrt{2\pi(\sigma_i^b)^2}^c} e^{A_i^{ba}} \Bigg]. \end{aligned}$$

Now suppose that the variances $(\sigma_1^a)^2$ and $(\sigma_1^b)^2$ in the first interval are significantly smaller than the ones in the second interval. Under this assumption, the terms associated with $\boldsymbol{\zeta}$ and the $\boldsymbol{\theta}$ terms in the second part of the strategy become negligible. This assumption allows us to focus solely on the contribution from the first term of the strategy $0 \leq t \leq T_*$. We therefore henceforth drop the subscript i to ease notation. If we assume that both $(\sigma_1^a)^2$ and $(\sigma_1^b)^2$ are small enough, then the $\boldsymbol{\zeta}$ terms become negligible and only the $\boldsymbol{\theta}$ terms remain. The no-price-manipulation condition in turn

reduces to

$$0 < -\theta_{ab}\frac{(\bar{\mu}^{ab}-\mu^b)}{(\sigma^a)^2}\sqrt{\frac{(\bar{\sigma}^{ab})^2}{(\sigma^b)^2}}\frac{1}{\sqrt{2\pi(\sigma^a)^2}^c}e^{A^{ab}} - \theta_{ba}\frac{(\bar{\mu}^{ba}-\mu^a)}{(\sigma^b)^2}\sqrt{\frac{(\bar{\sigma}^{ba})^2}{(\sigma^a)^2}}\frac{1}{\sqrt{2\pi(\sigma_i^b)^2}^c}e^{A^{ba}}. \tag{4.B.1}$$

After plugging in the definitions of $\bar{\mu}_{ab}, \bar{\mu}_{ba}, \bar{\sigma}^{ab}, \bar{\sigma}^{ba}$, we observe that $A^{ba} - A^{ab} = 0$ and, after rearranging, (4.B.1) in turn simplifies to

$$0 < -\theta_{ab}(\mu_a - \mu_b) - \theta_{ba}(\mu_b - \mu_a)\left(\frac{\sigma^b}{\sigma^a}\right)^{1-c}\left(\frac{(\sigma^a)^2 + c(\sigma^b)^2}{c(\sigma^a)^2 + (\sigma^b)^2}\right)^{3/2}. \tag{4.B.2}$$

Linear Impact: When the price impact function is linear($c = 1$), the inequality (4.B.2) further simplifies to:

$$0 \leq -\theta_{ab}(\mu_a - \mu_b) - \theta_{ba}(\mu_b - \mu_a) = (\theta_{ba} - \theta_{ab})(\mu_a - \mu_b). \tag{4.B.3}$$

This needs to hold both for $\mu_a < \mu_b$ and for $\mu_a > \mu_b$. Consequently, to prevent price-manipulation, the matrix $\boldsymbol{\theta}$ must be symmetric.

Concave Impact: We now turn to strictly concave impact functions with $c < 1$. In the limit as $\sigma_b \to 0$. the second term in the inequality (4.B.1) vanishes. Whence, to avoid price manipulation, we need

$$0 \leq -\theta_{ab}(\mu_a - \mu_b). \tag{4.B.4}$$

As this has to hold for any choice of μ_a, μ_b, it follows that $\theta_{ab} = 0$. As the indices a, b were arbitrary, the matrix $\boldsymbol{\theta}$ therefore must be diagonal to avoid price manipulation.

4.C Proof to Section 4.6.2

When the impact function is of power form with exponent c, then the partial derivatives of the goal function with respect to J_t^1 and J_t^2 lead to the first-order conditions

$$\bar{\alpha}_t^1 \text{sign}\left(\frac{\gamma_t}{\phi_t}\right)\left|\frac{\gamma_t}{\phi_t}\right|^{-\frac{c}{2}} = (1+c)\zeta_{aa} + \zeta_{ab}\left(\text{sign}\left(\phi_t\right)|\phi_t|^c + c\phi_t\right), \tag{4.C.1}$$

$$\bar{\alpha}_t^2 \text{sign}\left(\frac{\gamma_t}{\phi_t}\right)\left|\frac{\gamma_t}{\phi_t}\right|^{-\frac{c}{2}} = (1+c)\zeta_{bb}\text{sign}(\phi_t)|\phi_t|^c + \zeta_{ba}\left(1 + c\text{sign}(\phi_t)|\phi_t|^{c-1}\right). \tag{4.C.2}$$

Here, we have introduced the new variables $\phi_t = J_t^2/J_t^1$ and $\gamma_t = J_t^1 J_t^2$ (tacitly assuming $J_t^1 \neq 0$). After multiplying the first equation with $\bar{\alpha}_t^2/\bar{\alpha}_t^1$, subtracting it from the second equation and rearranging terms, we obtain an autonomous equation for ϕ_t:

$$0 = \phi_t + \text{sign}(\phi_t)|\phi_t|^c k_t^1 + \text{sign}(\phi_t)|\phi_t|^{c-1}k_t^2 + k_t^3,$$

with the coefficients k_t^i from (4.6.4). For square-root impact ($c = 1/2$), this leads to a cubic equation after another change of variable, where the signs of the coefficients depend on the sign of the variable. This in turn leads to six candidate solutions for the maximum of the goal function. These in turn need to be directly compared to the points where one or both variables are zero (so that the goal function is not differentiable).

4.D Constraints on Matrices L and θ

To motivate the structure of the concave cross-impact model, we begin by expressing the instantaneous trading cost in terms of the price variation $\Delta \boldsymbol{p}$ and the traded quantity $d\boldsymbol{Q}$:

$$\Delta \boldsymbol{p}^T d\boldsymbol{Q} = \Delta \boldsymbol{p}^T \boldsymbol{M}^{-1,T} \boldsymbol{M} d\boldsymbol{Q}.$$

Here, the matrix $\boldsymbol{M}$ plays a central role in transforming observable market quantities into a latent impact factor space. Specifically, it encodes the mapping between price changes and trading activity in a way that facilitates decorrelation and normalization.

To define a model based on observable market data, we consider returns $\Delta \boldsymbol{p}$, order flow $\boldsymbol{Q}$, and volatility $\boldsymbol{\sigma}$. From these, we construct the order flow covariance matrix $\boldsymbol{\Omega} := \mathbb{E}[\boldsymbol{Q}\boldsymbol{Q}^T]$, which we take to be diagonal in an idealized setting, and the return covariance matrix $\boldsymbol{\Sigma} := \mathbb{E}[\Delta \boldsymbol{p} \Delta \boldsymbol{p}^T]$, which captures asset return correlations.

The expected price impact is assumed to take the general form

$$\mathbb{E}[\Delta \boldsymbol{p} \mid \boldsymbol{Q}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}] = f(\boldsymbol{Q}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}).$$

Applying a linear transformation via $\boldsymbol{M}$, we express this in the transformed (latent) space:

$$\boldsymbol{M}^{-1} \mathbb{E}[\Delta \boldsymbol{p} \mid \boldsymbol{Q}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}] = f(\boldsymbol{M}\boldsymbol{Q}, \boldsymbol{M}^{-1,T} \boldsymbol{\Sigma} \boldsymbol{M}^{-1}, \boldsymbol{M} \boldsymbol{\Omega} \boldsymbol{M}^T).$$

To diagonalize the model and decouple its variables, we introduce the symmetrized matrix

$$\boldsymbol{\Gamma} := \boldsymbol{\Omega}^{1/2,T} \boldsymbol{\Sigma} \boldsymbol{\Omega}^{1/2},$$

which admits a spectral decomposition $\boldsymbol{\Gamma} = \boldsymbol{U} \boldsymbol{\mu} \boldsymbol{U}^T$ in terms of orthonormal eigenvectors $\boldsymbol{U}$ and eigenvalues $\boldsymbol{\mu}$. This motivates the choice

$$\boldsymbol{M} = \boldsymbol{U}^T \boldsymbol{\Omega}^{-1/2},$$

which leads to the simplified expression

$$\mathbb{E}[\Delta \boldsymbol{p} \mid \boldsymbol{Q}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}] = \boldsymbol{\Omega}^{-1/2,T} \boldsymbol{U} \, f(\boldsymbol{U}^T \boldsymbol{\Omega}^{-1/2} \boldsymbol{Q}, \boldsymbol{\mu}, \mathbb{I}).$$

Exploiting the additive structure of f over eigencomponents, this becomes

$$\mathbb{E}[\Delta \boldsymbol{p} \mid \boldsymbol{Q}] = \sum_a \boldsymbol{\Omega}^{-1/2,T} \boldsymbol{U}_a \, f(\boldsymbol{U}_a^T \boldsymbol{\Omega}^{-1/2} \boldsymbol{Q}, \mu_a, \mathbb{I}),$$

where $\boldsymbol{U}_a$ denotes the a-th column of $\boldsymbol{U}$. After appropriate scaling and normalization, we obtain a factorized form:

$$\mathbb{E}[\Delta \boldsymbol{p} \mid \boldsymbol{Q}] = \sum_a \boldsymbol{\Omega}^{-1/2,T} f_1(\mu_a) \boldsymbol{U}_a \, f(\boldsymbol{U}_a^T \boldsymbol{\Omega}^{-1/2} \boldsymbol{Q}, 1, \mathbb{I}).$$

This leads to the identification of the matrix $\boldsymbol{L}$, which governs the directional sensitivity of price to order flow, as

$$\boldsymbol{L} = \boldsymbol{\Omega}^{-1/2,T} \boldsymbol{U},$$

while the nonlinear scaling function f defines the matrix $\boldsymbol{\theta}$ through its dependence on the eigenvalues $\boldsymbol{\mu}$ and a model-dependent prefactor. In particular, the diagonal of $\boldsymbol{\theta}$ can be interpreted as encoding the strength of impact per factor, modulated by the eigenstructure of return correlations.

This construction has important structural implications. When volumes and volatilities are homogeneous, i.e., when $\boldsymbol{\Omega} = V\mathbb{I}$ and $\boldsymbol{\Sigma} = \sigma^2\mathbb{I}$, then $\boldsymbol{L} = \sqrt{\boldsymbol{\Sigma}} = \sigma\mathbb{I}$, and the model reduces to a fully symmetric, tractable setting where impact commutes with correlation. In this case, all traded volumes are scaled uniformly by V. However, for general (non-uniform) $\boldsymbol{\Omega}$, the factor structure must be carefully preserved. Arbitrary deviations from this structure may break symmetry and lead to models that admit price manipulation. The matrix $\boldsymbol{L}$ must therefore absorb the anisotropy in $\boldsymbol{\Omega}$ to ensure that the resulting impact model remains arbitrage-free.

Chapter 5

Estimation of a Nonparametric Cross-Impact Kernel

Summary

This chapter develops a non-parametric estimation framework for concave and decaying price impact kernels in a multivariate setting, using both proprietary and synthetically-generated metaorder data. The estimator generalizes an offline learning method to the empirically supported concave case and provides a confidence bound that scales with the squared number of assets. A synthetic metaorder generation mechanism is revisited to overcome data scarcity and improve kernel calibration. Applied to corn futures, the estimated kernels reveal power-law decay in self-impact, enhanced predictive accuracy through cross-impact, and pronounced asymmetries linked to liquidity differences across assets.

5.1 Introduction

Price impact refers to the empirical fact that the execution of a large order causes adverse and slowly decaying changes in the asset price, resulting in less favorable execution prices for the trader. It is well documented that price impact is concave in trade size: larger trades tend to move prices less per unit volume than smaller trades, a property that is not captured by linear models. Instead, the empirical literature supports a square-root law of market impact (Almgren et al., 2005; Tóth et al., 2011; Bershova and Rakhlin, 2013; Mastromatteo et al., 2014; Sato and Kanazawa, 2024), where the average signed return (or peak impact) induced by a metaorder of volume Q is given by

$$I^{\text{peak}} = Y\sigma_D \text{sign}(Q) \left|\frac{Q}{V_D}\right|^c, \tag{5.1.1}$$

where Y is a constant of order 1, V_D is the daily traded volume, σ_D the daily volatility, and $c \approx 0.5$ is a concavity parameter that is remarkably stable across asset classes, exchanges, and market structures (Tóth et al., 2016; Sato and Kanazawa, 2024;

Hey et al., 2025).

Beyond its instantaneous effect, price impact also decays after trading. This temporal relaxation has motivated a broad literature on optimal execution, where impact dynamics is modeled either exponentially Obizhaeva and Wang (2013); Gârleanu and Pedersen (2013); Hey et al. (2025) or with a power-law decay Bouchaud et al. (2009b); Gatheral et al. (2012), leading to continuous-time Volterra-type models that express price as the convolution of historical trade flow with an impact kernel. For a metaorder split into child orders across M time intervals $\{Q_{t_i}\}_{i=1}^{M}$, the price dynamics follow

$$P_{t_{i+1}} - P_{t_1} = \sum_{j=1}^{i} G_{i,j} h(Q_{t_j}) + \epsilon_{t_i}, \quad i = 1, \dots, M, \tag{5.1.2}$$

where $h(x) = \mathrm{sgn}(x)|x|^c$ is a concave impact function and ϵ_{tj} represents some asset specific noise. In stationary settings, where seasonal effects are accounted for, the kernel is often assumed to be time-homogeneous: $G_{i,j} = G_{i-j}$, simplifying estimation via convolution.

A non-parametric estimation method for the impact kernel in the linear case $c = 1$ was introduced in Neuman and Zhang (2023), where the estimation occurs in an online learning framework. While theoretically appealing, this approach assumes independence between price trajectories and heterogeneous trading strategies, which is unrealistic in practice. Furthermore, online learning is computationally expensive and difficult to implement in real time. As a result, most impact calibrations are performed offline. Neuman et al. (2023) proposed an offline estimator based on historical trade and price data and derived optimal convergence rates for this estimator. While effective on synthetic datasets, its performance on real-world metaorder data remains untested.

A central challenge in applying these methods to proprietary metaorder data is that impact is concave, with empirical $c \approx 0.5$[1], unlike the linear assumption $c = 1$. The present work generalizes the estimation framework of Neuman et al. (2023) to the concave case, incorporating the empirically supported square-root impact law. This extension significantly improves goodness-of-fit in kernel estimation.

A second difficulty lies in data scarcity. Institutional investors often execute a single metaorder per asset per day, meaning that even long historical datasets rarely contain more than 10^3 metaorders per product. This limited sample size is insufficient to reliably estimate high-dimensional impact kernels, especially when accounting for liquidity or tick-size-specific effects.[2]

To overcome this, CFM's proprietary dataset is enhanced using a synthetic metaorder generation procedure inspired by empirical findings in Maitrier et al. (2025b); Sato and Kanazawa (2024) and introduced in Maitrier et al. (2025a). These synthetic metaorders

[1] Linear models with $c = 1$ allow analytical tractability but fail to capture empirically observed saturation effects in impact. The concave case is more realistic but introduces significant estimation and optimization complexity.

[2] This issue is further exacerbated when attempting to estimate cross-sectional impact effects or calibrate volatility- or liquidity-conditioned kernels.

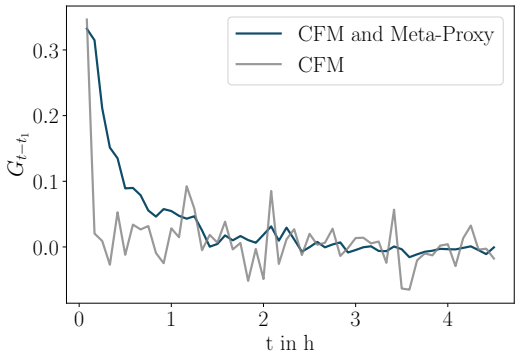


Figure 5.1: Estimated price impact kernel G_{t-t_1} using CFM proprietary data only (gray) and the enhanced dataset including synthetic metaorders (blue).

are constructed by randomly sampling from tick-by-tick data, assigning trades to proxy metaorders via a systematic prescription described in Section 5.3.2.

Figure 5.1 illustrates the improvement in kernel estimation where the propagator kernel is estimated with the non-parametric procedure described in Neuman et al. (2023) from two datasets; CFM's proprietary corn future trades over a ten-year horizon (gray) and a regularized version where the dataset is augmented with synthetic metaorders (blue). The enhanced dataset yields a smoother kernel that better captures the decay of impact.

Cross-impact, the influence of trades in one asset on the price of another, plays an essential role in multivariate price formation. Mastromatteo et al. (2017) show that ignoring cross-impact leads to misestimation of liquidity and suboptimal execution. Le Coz et al. (2024) empirically demonstrate that highly liquid assets lead the price discovery process, influencing less liquid correlated instruments with a lag constrained by trading frequency. In Hey et al. (2024b), the first consistent nonlinear cross-impact model is proposed and price manipulation conditions are derived for exponentially decaying concave kernels. They show empirically using CFM metaorder data that concavity and impact decay are significant in cross-impact settings.

Based on this literature, the framework of Neuman et al. (2023) is extended to a multivariate setting with concave price impact. The multivariate model expresses the return of asset $a_1 \in \{1, \dots, d\}$ as:

$$P^{a_1}_{t_{i+1}} - P^{a_1}_{t_1} = \sum_{a=1}^{d} \sum_{j=1}^{i} G^{(a_1,a)}_{i-j} h(Q^a_{t_j}) + \epsilon^{a_1}_{t_i}, \quad i = 1, \dots, M. \tag{5.1.3}$$

where $P^{a_1}_{t_i}$ is the mid-price of asset a_1 at time t_i, $Q^a_{t_j}$ are the traded volumes in assets $a \in 1, d$ and $G^{(a_1,a)}_{i-j}$ is the propagator that quantifies the influence of all assets a on the asset a_1. This framework is applied to three highly correlated corn futures with different expiries. The resulting kernel estimates confirm findings from Hey et al. (2024b), showing strong cross-impact contributions among highly correlated assets. They also provide

new insights: not only does incorporating metaorders from correlated assets improve predictive performance, but the estimated kernels exhibit strong asymmetries. These asymmetries reflect differences in liquidity across contracts, with more liquid assets exerting a stronger and more persistent impact on their less liquid counterparts.

In summary, this work contributes in the following ways:

1. It extends the nonparametric kernel estimation to accommodate concave impact functions in line with the square-root law;

2. It proposes a metaorder proxy to augment sparse proprietary datasets, enabling better non-parametric estimation;

3. It estimates the Kernel non-parametrically on metaorder data and confirms the universal power-law decay;

4. It generalizes the framework to a multivariate setting and derives the convergence rate for the convolution kernel;

5. It is shown on proprietary CFM data that the cross-impact kernel has more predictive power than the univariate model;[3]

6. It illustrates the importance of liquidity differences in cross-impact.

The remainder of this paper is structured as follows. Section 5.2 introduces the multivariate propagator model with concave impact functions and presents the offline estimator used for kernel calibration. The section also derives a confidence bound on the estimator under convex constraints. Section 5.3 describes the empirical implementation, including the construction of synthetic metaorders in Section 5.3.2 and useful methods needed for reliable kernel estimation in Section 5.3.3. Section 5.3.4 presents the main empirical findings: self-impact kernels follow power-law decay, proxy data improve estimation, and cross-impact also enhances predictive power. The results on cross-impact in Section 5.3.4 illustrate the asymmetric effects on the kernels with respect to liquidity differences in the calibrated assets. Finally, Section 5.4 summarizes the results and highlights the drawbacks and possible extension of the multivariate estimation.

5.2 Price Impact Estimation with offline data

This section introduces the offline dataset and methodology used to estimate the cross-impact propagator $\boldsymbol{G}$. The dataset consists of metaorders[4] and corresponding price

[3]This is consistent with microstructural theories that posit that liquidity and information flow across correlated assets is inherently multivariate, especially in futures markets where contracts on the same underlying asset co-move. See, e.g., Hasbrouck (1995).

[4]Each metaorder represents a large, strategically executed order, often spanning minutes to hours that is sliced into smaller child orders. In this dataset, metaorders are treated as exogenous realizations and are not derived from a known optimization criterion.

trajectories observed across multiple time periods for a portfolio of d assets. A least-squares framework is used to estimate $\boldsymbol{G}$, both in the unconstrained and constrained settings. The constrained estimator enforces admissibility conditions derived from no-arbitrage arguments, such as convexity and non-negativity of the kernel.

5.2.1 Dataset

The offline dataset $\mathcal{D}$ consists of $N \in \mathbb{N}$ metaorders executed across d assets. Each metaorder is observed on an equidistant time grid of M intervals over a fixed time horizon $[0, T]$. Formally,

$$\mathcal{D} = \left\{ (\boldsymbol{P}_{t_i}^{(n)})_{i=1}^{M+1}, (\boldsymbol{Q}_{t_i}^{(n)})_{i=1}^{M+1} \middle| n = 1, ..., N \right\} \tag{5.2.1}$$

where $\boldsymbol{P}_{t_i}^{(n)} = \left((P_{t_i}^1)^{(n)}, \ldots, (P_{t_i}^d)^{(n)} \right)^T$ represents the observed price for all d assets during metaorder n, and $\boldsymbol{Q}_{t_i}^{(n)} = \left((Q_{t_i}^1)^{(n)}, \ldots, (Q_{t_i}^d)^{(n)} \right)^T$ the associated trading volumes. Setting $\boldsymbol{P}^{(n)} := (\boldsymbol{P}_{t_i}^{(n)})_{i=1}^M$ and $\boldsymbol{Q}^{(n)} := (\boldsymbol{Q}_{t_i}^{(n)})_{i=1}^M$, $(\boldsymbol{P}^{(n)}, \boldsymbol{Q}^{(n)})_{n=1}^N$ are realizations of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ [5] satisfying the following properties: for each n, $\mathbf{Q}^{(n)}$ is measurable with respect to the σ-algebra $\mathcal{F}'_{n-1}$ where

$$\mathcal{F}'_{n-1} = \sigma\Big\{ \left(\boldsymbol{P}^{(k)} \right)_{k=1}^{n-1}, \left(\boldsymbol{Q}^{(k)} \right)_{k=1}^{n-1} \Big\} \tag{5.2.2}$$

and there exist $\mathcal{F}'_n$-measurable random-variables $\boldsymbol{\epsilon}_{t_i}^{(n)} = ((\epsilon_{t_i}^1)^{(n)}, ..., (\epsilon_{t_i}^d)^{(n)})^T$ such that the price evolution can be assumed to follow a propagator model with additive noise

$$\boldsymbol{P}_{t_{i+1}}^{(n)} - \boldsymbol{P}_{t_1}^{(n)} = \sum_{j=1}^{i} \boldsymbol{G}_{i-j}^* h(\boldsymbol{Q}_{t_j}^{(n)}) + \boldsymbol{\epsilon}_{t_i}^{(n)}, \quad i = 1, ..., M,$$

where $\boldsymbol{G}^* \in \mathbb{R}^{M \times d \times d}$ is the true (unknown) propagator and $h(\cdot)$ is a continuous, increasing and concave function that is applied to each element in the volume vector. The random variables satisfy $\mathbb{E}^{\mathbb{P}}[\boldsymbol{\epsilon}_{t_i}^{(n)} | \mathcal{F}'_{n-1}] = 0$ and $\mathbb{E}^{\mathbb{P}}[(\boldsymbol{\epsilon}_{t_i}^{(n)})^T \boldsymbol{\epsilon}_{t_i}^{(n)}] < \infty$ for all $i = 1, ..., M$. For each asset $a_1 \in \{1, ..., d\}$, the noise $(\boldsymbol{\epsilon}^{a_1})^{(n)} \in \mathbb{R}^M$ is assumed to be conditionally sub-Gaussian analogously to Assumption 2.8 of Neuman et al. (2023): [6]

Assumption 5.2.1. *For an offline dataset $\mathcal{D}$ of size $N \in \mathbb{N}$ there exists a known constant $R > 0$ such that for all $n = 1, ..., N$,*

$$\mathbb{E}^{\mathbb{P}}[\exp(\langle \mathbf{v}, (\boldsymbol{\epsilon}^{a_1})^{(n)} \rangle) | \mathcal{F}'_{n-1}] \leq \exp\left(\frac{R^2 ||\mathbf{v}||^2}{2} \right), \forall \mathbf{v} \in \mathbb{R}^M \text{ and } a_1 = 1, ..., d \tag{5.2.3}$$

[5] Here, $\mathcal{F}$ refers to the natural filtration generated by the dataset $\mathcal{D}$, i.e., the sigma-algebra associated with observed price and volume sequences. The filtration $\mathcal{F}'$ governs the noise process $\boldsymbol{\epsilon}^{(n)}$, which models idiosyncratic price fluctuations not explained by executed volume. These two filtrations allow us to distinguish between observed trading data and residual stochasticity.

[6] This assumption allows for concentration inequalities and high-probability bounds in the estimation procedures. The conditionality reflects that volatility may depend on previously observed metaorders.

which means that each $(\boldsymbol{\epsilon}^{a_1})^{(n)}$ *is* $R-$*conditionally sub-Gaussian with respect to* $\mathcal{F}'_{n-1}$.

Example 5.2.2. *Typical examples of impact kernels for* $d = 1$ *include strictly positive, decaying, and convex functions of time. For instance, Bouchaud et al. (2018) proposed* $G(t) = \frac{l_0}{(l_0+t)^\beta}$ *for constants* $\beta, l_0 > 0$. *Other examples include power-law kernels* $G(t) = \frac{1}{t^\beta}\mathbb{I}_{t>0}$, *for* $0 < \beta < 0.5$ *Gatheral (2010), and exponential decay* $G(t) = e^{-\rho t}$ *with* $\rho > 0$ *as in Obizhaeva and Wang (2013).*

For $d > 1$, the constraint set $\mathcal{G}_{\text{ad}}$ of admissible impact kernels ensures structural properties. For the case of linear cross-impact, conditions have been derived to prevent price manipulation in Abi Jaber et al. (2024) (Theorem 2.14), where admissible kernels must be non-negative, convex, symmetric, and non-increasing in time. However, the current work focuses on concave impact functions, for which the theoretical understanding of admissibility is limited.

In Hey et al. (2024b), price manipulation conditions are established for a class of concave impact models with exponentially decaying kernels. There, the admissibility condition corresponds to a matrix property on one of the propagator components, as specified in Lemma 5.1. When the concavity parameter $c \in [0.5, 1]$, the conditions still closely resemble those in the linear case.

In the present work, we adapt and slightly relax the structural assumptions from Abi Jaber et al. (2024). In particular, we no longer enforce symmetry of the kernel matrix at each time step. [7] Instead, we require that the kernel matrices are positive semi-definite so that all eigenvalues are non-negative. The resulting class of admissible convolution kernels includes all matrix-valued functions $\mathbf{G}_i : [1, M] \to \mathbb{R}^{d\times d}$[8] that are decreasing, convex, non-negative and semi-positive definite. Formally, the admissible set is defined as:

$$\mathcal{G}_{ad} := \left\{ \mathbf{G}_i = \left(G_i^{(a_1,a_2)}\right)_{a_1,a_2=1}^{d} \;\middle|\; \begin{array}{l} x^T\mathbf{G}_i x \geq x^T\mathbf{G}_{i+1}x, \text{and} \\ x^T\mathbf{G}_i x - x^T\mathbf{G}_{i-1}x \leq x^T\mathbf{G}_{i+1}x - x^T\mathbf{G}_i x \quad \forall x \in \mathbb{R}^d \\ G_i^{(a_1,a_2)} \geq 0 \quad \forall a_1, a_2 = 1, \ldots, d \\ \exists \mu \in \mathbb{R}_+ \text{ s.t. } \det(\mathbf{G}_i - \mu \mathbb{I}_d) = 0 \end{array} \right\}. \tag{5.2.4}$$

5.2.2 Non-parametric Estimation

The unknown propagator $\boldsymbol{G}^*$ is estimated via least squares method. For each metaorder n, define the vector of observed returns $\mathbf{y}^{(n)} \in \mathbb{R}^{M\cdot d}$ as

[7] Asymmetric cross-impact across assets (e.g., asset a_1 impacting a_2 more than vice versa) is often observed empirically in equity and futures markets.

[8] While the kernel is formulated on a discrete time grid, the underlying class of models admit continuous-time analogs via Volterra-type representations.

$$\mathbf{y}^{(n)} = \left(y^1_{t_1}, \dots, y^1_{t_M}, \dots, y^d_{t_1}, \dots, y^d_{t_M}\right)^T, \quad \left(y^{a_1}_{t_i}\right)^{(n)} = (P^{a_1}_{t_{i+1}})^{(n)} - (P^{a_1}_{t_1})^{(n)}$$

These returns must satisfy

$$\mathbf{y}^{(n)} = \mathbf{U}^{(n)} \boldsymbol{G}^*_v + \boldsymbol{\epsilon}^{(n)}, \tag{5.2.5}$$

where the propagator $\boldsymbol{G}^* \in \mathbb{R}^{M\times d\times d}$ has been reshaped to a vector $\mathbf{G}^*_v \in \mathbb{R}^{M\cdot d^2}$[9] and the matrix $\mathbf{U}^{(n)} \in \mathbb{R}^{M\cdot d\times M\cdot d^2}$ consists of block matrices $\mathbf{D}^{(n)} \in \mathbb{R}^{M\times M\cdot d}$ which are lower triangular convolution matrices formed from the concave function applied to the trade flow $h(\boldsymbol{Q}^{(n)}_{t_i})$. The vector and matrices look as follows:

$$\mathbf{G}^*_v = \left(G^{(1,1)}_1, G^{(1,2)}_1, \dots, G^{(1,d)}_1, G^{(1,1)}_2, \dots, G^{(1,d)}_M, G^{(2,1)}_1, \dots G^{(d,1)}_1, ..., G^{(d,d)}_M\right)^T,$$

$$\mathbf{D}^{(n)} = \begin{pmatrix} h(Q^{(1,n)}_{t_1}) & ... & h(Q^{(d,n)}_{t_1}) & 0 & ... & ... & ... & ... & 0 \\ h(Q^{(1,n)}_{t_2}) & ... & h(Q^{(d,n)}_{t_2}) & h(Q^{(1,n)}_{t_1}) & ... & h(Q^{(d,n)}_{t_1}) & 0 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ h(Q^{(1,n)}_{t_M}) & ... & h(Q^{(d,n)}_{t_M}) & h(Q^{(1,n)}_{t_{M-1}}) & ... & h(Q^{(d,n)}_{t_{M-1}}) & ... & ... & h(Q^{(d,n)}_{t_1}) \end{pmatrix},$$

$$\mathbf{U}^{(n)} = \begin{pmatrix} \mathbf{D}^{(n)} & 0 & \dots & 0 \\ 0 & \mathbf{D}^{(n)} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & \dots & \mathbf{D}^{(n)} \end{pmatrix}.$$

The non-parametric propagator estimator is obtained by solving a regularized least-squares problem[10] by minimizing the following quadratic loss over all admissible price impact coefficients

$$(\boldsymbol{G}_v)_{N,\lambda} := \arg\min_{\boldsymbol{G}\in\mathcal{G}_{\text{ad}}} \left(\sum_{n=1}^{N} \left\| \mathbf{y}^{(n)} - \mathbf{U}^{(n)} \boldsymbol{G}_v \right\|^2 + \lambda \left\| \boldsymbol{G}_v \right\|^2 \right), \tag{5.2.6}$$

where $\lambda > 0$ is a regularization parameter and $\|\cdot\|$ denotes the Frobenius norm.

Remark 5.2.3. *Asset-specific regularization parameters could be introduced in practice. The problem would then be*

[9] The reshaping of the 3-dimensional tensor $\boldsymbol{G}^*$ into a flat vector $\mathbf{G}^*_v$ facilitates standard least-squares formulations. Each subvector of $\mathbf{G}^*_v$ corresponds to a temporal kernel for a given asset pair (a_1, a_2).

[10] This estimation procedure extends classical ridge regression to the multivariate setting with structural constraints on the kernel shape. The regularization penalizes roughness and helps stabilize the estimator in low-sample regimes.

$$(\boldsymbol{G}_v)_{N,\lambda_{a_1,a_2,\ldots,\lambda_{d,d}}} := \arg\min_{\boldsymbol{G}\in\mathcal{G}_{\text{ad}}} \left(\sum_{n=1}^{N} \left\| \mathbf{y}^{(n)} - \mathbf{U}^{(n)} \boldsymbol{G}_v \right\|^2 + \sum_{a_1=1}^{d} \sum_{a_2=1}^{d} \lambda_{a_1,a_2} \left\| (\boldsymbol{G}_i^{(a_1,a_2)})_{i=1}^{M} \right\|^2 \right),$$

where in the last term, the vector $(\boldsymbol{G}_i^{(a_1,a_2)})_{i=1}^{M} \in \mathbb{R}^M$ *is the vector whose elements are decreasing.*

The estimator $(\boldsymbol{G}_v)_{N,\lambda}$ can thus be computed by projecting the unconstrained least-squares solution onto the set $\mathcal{G}_{\text{ad}}$[11]. Specifically,

$$(\boldsymbol{G}_v)_{N,\lambda} = \arg\min_{\boldsymbol{G}_v\in\mathbb{R}^{M\cdot d^2}} \left\| W_{N,\lambda}^{1/2} \left(\boldsymbol{G}_v - (\tilde{\boldsymbol{G}}_v)_{N,\lambda} \right) \right\|^2, \tag{5.2.7}$$

with

$$W_{N,\lambda} = \sum_{n=1}^{N} (\mathbf{U}^{(n)})^T \mathbf{U}^{(n)} + \lambda \mathbb{I}_{M\cdot d^2}, \quad (\tilde{\boldsymbol{G}}_v)_{N,\lambda} = W_{N,\lambda}^{-1} \sum_{n=1}^{N} (\mathbf{U}^{(n)})^T \mathbf{y}^{(n)}. \tag{5.2.8}$$

The confidence region of the estimated coefficient $(\boldsymbol{G}_v)_{N,\lambda}$ in terms of the observed data is derived analogously to Theorem 2.14 in (Neuman et al., 2023, Proof 5).

Corollary 5.2.4. *Suppose the noise* $\boldsymbol{\epsilon}^{(n)}$ *is conditionally sub-Gaussian and independent across* n*, and that the true propagator* $\boldsymbol{G}^* \in \mathcal{G}_{\text{ad}}$ *as defined in* (5.2.4)*. Then, for all* $\lambda > 0$ *and* $\delta \in (0,1)$*, with probability at least* $1-\delta$ *under* $\mathbb{P}$*, the constrained least-squares estimator* $(\boldsymbol{G}_v)_{N,\lambda}$ *satisfies the following bound:*

$$\left\| W_{N,\lambda} \left((\boldsymbol{G}_v)_{N,\lambda} - \boldsymbol{G}_v^* \right) \right\| \leq R \left(2 \log \left(\frac{\det(W_{N,\lambda})}{\delta^2 \lambda^{Md^2}} \right) \right)^{1/2} + \lambda \left\| W_{N,\lambda}^{-1/2} \boldsymbol{G}_v^* \right\|, \tag{5.2.9}$$

where $W_{N,\lambda} = \sum_{n=1}^{N} (\mathbf{U}^{(n)})^T \mathbf{U}^{(n)} + \lambda \mathbb{I}_{M\cdot d^2}$ *and the constant* $R > 0$ *is the sub-Gaussian parameter that controls the tail behavior of the noise* $\boldsymbol{\epsilon}^{(n)}$*.*

Remark 5.2.5. *The convergence rate remains similar to the one in (Neuman et al., 2023, Theorem 2.14), the only difference being the dimensional scaling. Since the estimator now involves* d^2 *kernels over* M *time steps, the determinant term contributes a factor of* λ^{-Md^2} *instead of* λ^{-M}*.*

The proof follows the same steps as in Neuman et al. (2023), with the additional constraint $\boldsymbol{G}_v \in \mathcal{G}_{\text{ad}}$ *taken into account. Since the constraint enforces that all functions* $x^\top \mathbf{G}_i x$ *for all* $i \in [0, M]$ *are convex and non-increasing in time, the admissible set remains nonempty, closed and convex. As the regularized objective is strongly convex and the constraint set is convex, the estimator* $(\boldsymbol{G}_v)_{N,\lambda}$ *remains uniquely defined.*

[11] Since $\mathcal{G}_{\text{ad}}$ is convex and closed, this projection problem admits a unique solution. In practice, the projection step is implemented using quadratic programming or isotonic regression with semi-definite constraints.

5.3 Empirical Analysis: Metaorder Impact

This section presents the non-parametric estimation of the impact convolution kernel associated with CFM's metaorders on corn futures contracts. Due to the limited number of observed metaorders, a proxy mechanism is employed to synthetically augment the dataset. This enables a more robust estimation of the convolution kernel in the univariate setting $d = 1$, where a grid search is performed over a set of concavity parameters $c \leq 1$. The results indicate that concave impact functions outperform their linear counterparts in the single-asset case. Consequently, the identified concavity parameter is retained for subsequent model calibration in a multivariate framework involving two and three corn futures contracts with different expiries.

5.3.1 Data

The empirical analysis is based on CFM's proprietary dataset; see Hey et al. (2025) for further details. In addition, publicly available market data is used to reconstruct mid-price trajectories for all three instruments on each trading day during which at least one contract was executed. This enables both the estimation of the intra-order price dynamics and the measurement of cross-impact effects arising from the execution of one contract on the others.

Three corn futures with different maturities are selected for this study due to the availability of multiple contracts with a shared underlying and differing maturities, each displaying return correlations that exceed 90%[12]. The selection is not driven by asset-specific properties and the methodology generalizes to any instrument in the dataset that satisfies similar structural conditions. This selection reduces the proprietary dataset to approximately $1.5 \cdot 10^3$ metaorders across the three contracts. The front-month contract exhibits the highest trading frequency, consistent with its greater liquidity relative to longer-dated instruments.[13]

Since the dataset does not include detailed information on the volumes of individual child-orders, a uniform execution profile[14] is assumed over the order duration. Specifically, each metaorder is discretized into 5-minute intervals such that the execution horizon is rounded to the nearest 5-minute mark and the total traded volume is equally distributed across the 54 bins[15]. While this assumption is well-suited to the chosen

[12]Such high return correlations are typical among futures contracts with shared underlyings but differing maturities. This structural redundancy makes them well-suited for testing cross-impact effects, which rely on information propagation between correlated assets.

[13]The Samuelson effect states that futures price volatility tends to increase as contracts approach maturity. Consequently, the front-month contract typically exhibits higher intraday volatility and trading frequency, making it better suited for impact estimation. See Samuelson (1973).

[14]When high-frequency child-order timestamps are unavailable this assumption is standard. It assumes execution follows a Volume-Weighted Average Price (VWAP) trajectory at constant rate, which aligns well with typical execution algorithms in liquid futures.

[15]This corresponds to approximately 4.5 hours of trading, excluding the first and last 30 minutes of the session. The start of each metaorder is aligned with the first bin. This avoids microstructural noise at the open and close, and prevents instability in the estimation procedure that can arise when the number of time steps M becomes too large.

temporal resolution, its accuracy may degrade for finer-grained time grids.

5.3.2 Metaorder Proxy

The number of available metaorders in the proprietary dataset is insufficient to support a reliable estimation of the impact convolution kernel on a per-product basis. As shown in Neuman et al. (2023), the convergence rate of the estimator depends both on the number of samples and on the number of trading periods M. While the dataset does not permit an increase in M, the number of effective samples can be enhanced by introducing a synthetic metaorder generation procedure, inspired by the findings in Maitrier et al. (2025b).

The approach originates from proprietary data on the Tokyo Stock Exchange, where the authors had access to full trade-level information, including trader identifiers. They observed that individual traders tend to execute consecutive trades in the same direction, forming buy or sell sequences. Metaorders were identified by aggregating consecutive trades of the same sign per trader, resetting upon sign reversal. Maitrier et al. (2025a) noticed that the presence of trader IDs was not essential to reproduce this behavior. Instead, one can simulate such identifiers by randomly assigning trades to synthetic IDs, thereby generating proxy metaorders via the following algorithm:

1. Assign to each trade a random integer $n_T \in \{0, 1, ..., N_T - 1\}$ where N_T is a chosen positive integer. The integers are sampled uniformly from this finite set, representing synthetic trader IDs.

2. Group all trades with the same assigned n_T and sort them in chronological order.

3. Partition each sequence into metaorders by aggregating trades with identical signs (buy or sell), terminating the sequence upon sign change.

Table 5.3.2 illustrates this procedure on a toy example. For instance, trades with $n_T = 0$ form a consistent sell-side sequence $\{-1, -1, -1\}$, corresponding to a single metaorder with three child-orders. In contrast, trades with $n_T = 1$ exhibit a sign change, resulting in two separate metaorders with one and two child-orders, respectively. The average length of a metaorder is influenced by both the chosen value for N_T and the empirical rate of sign changes in the underlying trade flow.

Figure 5.2 displays the peak impact, as defined in Equation 5.1.1, for synthetic metaorders on TBOND, 10USNOTE, EUROSTOXX, DAX, and CORN0 futures, filtered to include only those with at least four child-orders. All observed impact curves exhibit concave scaling behavior, consistent with a power-law relation. Empirically, the impact function fits a power-law with an exponent in the range $c \in [0.5, 0.7]$. The metaorder volume range is bounded below by microstructural properties (notably for large-tick assets such as DAX) and above by the decreasing likelihood of long, uninterrupted sign sequences in the trade flow. In practice, the upper bound is observed around 1% of the daily traded volume.

Time	Trade Sign	Trader ID (n_T)	Metaorder Assignment
10:05:011	-1	0	1
10:06:123	-1	1	2
10:06:509	-1	2	3
10:07:205	-1	0	1
10:07:388	1	2	4
10:07:434	1	3	5
10:07:786	-1	1	2
10:08:657	-1	3	6
10:09:476	-1	0	1
10:09:567	1	1	7

Table 5.1: Example of synthetic metaorder construction via randomized trader ID assignment and sign-based grouping.

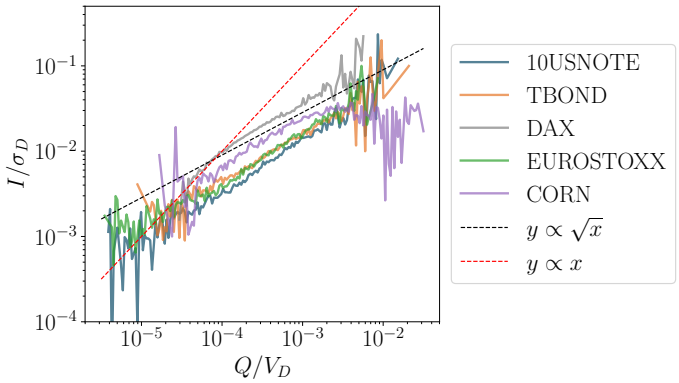


Figure 5.2: Peak impact as a function of normalized volume for synthetic metaorders across TBOND, 10USNOTE, EUROSTOXX, DAX, and CORN futures. Impact is scaled by daily volatility; volume is normalized by daily traded volume. The dashed black curve represents a square-root scaling, while the red dashed line corresponds to a linear benchmark.

Choosing N_T requires careful calibration to the microstructural properties of the asset under study. For small N_T (e.g. $N_T \approx 1$), artificially long metaorders are generated, which are unrealistic in practice. Conversely, large N_T (e.g. $N_T \approx$ Number of trades per day) results in metaorders with trades sparsely distributed over the day, diluting the temporal structure. Thus, intermediate values of N_T must be selected to capture liquidity and volatility features of the asset.[16]

The minimum and maximum feasible metaorder sizes depend on both the asset's tick size and its intraday trade frequency. Figure 5.3 demonstrates that the smallest possible metaorder volume is influenced by the average trade size in the underlying asset. For large-tick assets, where the average trade size is relatively small, fewer shares are needed

[16]Further research is needed to systematically tune N_T, potentially via minimizing the out-of-sample forecast error of peak impact. For corn futures, we use $N_T = 20$, which empirically balances the trade-off between temporal coherence and metaorder length.

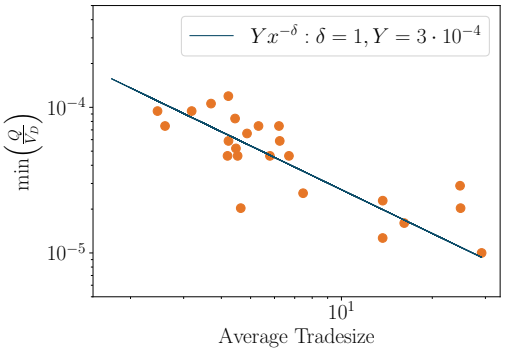


Figure 5.3: Minimum metaorder size as a function of the average trade size in the underlying asset. The metaorders with the smallest volume depend on the tick-size of the underlying asset.

to achieve a significant dollar position, resulting in smaller metaorders. This suggests an inverse relationship between the minimum volume of metaorders and the average trade size of the asset[17]. Consequently, the computational thresholds, which vary between 10^{-6} and 10^{-4}, as shown in Figure 5.2, can be explained by the average tick size. Assets such as the DAX and CORN contracts exhibit larger thresholds compared to their smaller tick counterparts.

5.3.3 Methodology

Metaorders in a Constant Timegrid.

The analysis is conducted on an equally spaced intraday time grid with a resolution of 5 minutes, consistent with the metaorder execution constraint. When constructing impact models from public order flow data, each time bin is typically populated, yielding non-zero entries throughout. In contrast, metaorders originate at arbitrary intraday timestamps that rarely coincide with the market open, introducing variability in both the start time and the effective duration of each execution.

To address this, the initial timestamp of the metaorder is chosen to be t_1, and the subsequent $M = 54$ return observations are aggregated into a column vector $\mathbf{y}^{(n)}$ as in 5.2.2. The corresponding volume process captured in matrix $\mathbf{U}^{(n)}$ is defined as zero post-execution. During execution, volume is either assumed to be uniformly distributed across bins (for CFM's metaorders) or follows observed child-order execution (in the case of metaorder proxies).

In the context of cross-impact estimation, the time origin t_1 is aligned with the start of the metaorder of the most liquid asset. This ensures consistency across the multi-asset regression framework when assessing the predictive power of the model.[18]

[17]Indeed, in large-tick assets significant positions can be built using fewer trades due to the larger notional per trade. Hence, smaller metaorders are sufficient to observe measurable impact.

[18]The alignment ensures that the multi-asset convolution model has a common time reference, which is crucial when comparing lead-lag structures or estimating asymmetric cross-impact.

Metaorder Cross-Impact.

Metaorders (both synthetic and CFM) of the most liquid contract are matched with metaorders from the remaining contracts. Since the number of available metaorders decreases with asset liquidity, it is not always possible to construct fully matched samples across all assets. The matching procedure randomly selects metaorders from each asset. If metaorders are unavailable for a given asset in a particular sample, the corresponding block in the design matrix $\mathbf{U}^{(n)}$ is set to zero. This procedure is repeated until all samples are generated. As a result, the total number of samples corresponds to the number of metaorders available in the most liquid asset.

Normalization.

The blue curve in Figure 5.4 displays the average intra-day volatility profile of the corn contracts, computed over 5-minute intervals from open to just 10 minutes before close. The clear asymmetric pattern reflects heightened volatility at market open, and lower volatility in midday periods. This profile underscores the importance of appropriately normalizing returns prior to kernel estimation, as metaorders may begin and end at different times within the trading session.

To correct for both time-of-day effects and bin duration, returns are scaled using a modified Garman–Klass estimator defined on the vector of mid-prices $(\boldsymbol{P}_{t_i}^{(n)})_{i=1}^{M+1}$. Specifically, for each metaorder n, the volatility estimate is given by:

$$\sigma_{[t_1,t_{i+1}]}^{(n)} = \frac{1}{3}\left(\max_{1\le j\le i+1}(P_{t_j}^{(n)}) - \min_{1\le j\le i+1}(P_{t_j}^{(n)})\right) + \frac{2}{3}\left|P_{t_{i+1}}^{(n)} - P_{t_1}^{(n)}\right|, \tag{5.3.1}$$

where $P_{t_1}^{(n)}$ and $P_{t_{i+1}}^{(n)}$ are the mid-prices at the start and end of the interval, respectively, and the maximum and minimum are taken over all mid-prices observed during that window. As bin duration $\Delta t = t_{i+1} - t_1$ increases, volatility scales with the square root of time, as is visible from the orange curve in Figure 5.4, where the sample mean $\bar{\sigma}_{[t_1,t_1+\Delta t]} \propto \bar{\sigma}_{[t,t+5\text{min}]}\sqrt{\Delta t}$.

Beyond return normalization, trading volume exhibits a comparable intraday pattern. During highly liquid periods, the same trade volume induces a smaller price impact than during low-liquidity intervals. To capture this effect, traded volumes are normalized by the average volume profile at the corresponding time of day. Let V_D denote the total volume traded daily and V_5 the volume in a 5-minute bin. Then, the normalized volume is computed as $\frac{Q/V_D}{V_5/V_D}$, where Q is the metaorder volume in the bin and $\overline{V_5/V_D}$ the average volume fraction. The propagator model under this normalization procedure becomes

$$\frac{P_{t_{i+1}}^{(a_1)} - P_{t_1}^{(a_1)}}{\sigma_{[t_1,t_{i+1}]}} = \sum_{a=1}^{d}\sum_{j=1}^{i} G_{i-j}^{(a_1,a)} h\Big(\frac{Q_{t_j}}{\bar{V}_{t_j}}\Big) + \boldsymbol{\epsilon}_{t_i}, \quad i = 1, ..., M,$$

where $\bar{V}_{t_j}$ denotes the stationary bin volume at time t_j, as defined above and $\sigma_{[t_1,t_{i+1}]}$

is the volatility in the time interval $[t_1, t_{i+1}]$. Note that, contrary to calibrations in Taranto et al. (2016); Patzelt and Bouchaud (2017), the propagator $\mathbf{G}$ in our setup does not absorb the volatility term. This normalization ensures robustness with respect to noisy estimates for small bin sizes, and allows the impact kernel to be estimated from inputs that are free from seasonal intraday patterns in volatility and liquidity.

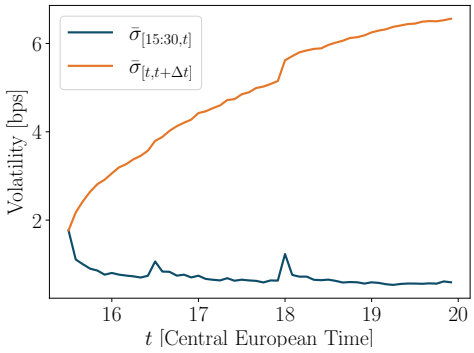


Figure 5.4: Average intra-day volatility $\bar{\sigma}_{[t,t+\Delta t]}$ defined by 5.3.1 over $\Delta t = 5$ minute bins (blue) and over increasing bin sizes $\bar{\sigma}_{[15:30,t]}$ (orange).

Predictive Power.

To assess the predictive capability of the impact kernel, the combined dataset of real and synthetic metaorders is split into training and test sets using an 80:20 ratio. The model is calibrated on the training set using the estimator defined in Equation (5.2.8), which aggregates the return responses over all observed metaorders.

Out-of-sample performance is computed by applying the kernel to predict the return trajectories in the test set. The predictive power is quantified using the coefficient of determination R^2, which allows for a direct comparison across different model specifications – for example, between linear and concave impact functions, or between models with and without cross-impact terms

5.3.4 Results

Table 5.3.4 summarizes the out-of sample performance across various model specifications. The following sections then explain thoroughly the kernel specifications. The main results of the analysis are:

1. A concave model fits better than a linear model;
2. The metaorder proxy enhances the estimation of impact decay, which is harder to calibrate than concavity alone due to lower R^2 signal contrast in the original dataset;
3. Without having any previous assumption on the shape of the kernel, it shows that impact decays as a power-law;

4. Cross-impact enhances the predictive power.

In the following two subsections, we provide additional details about our results on self- and cross-impact estimations.

Table 5.2: Out of sample R^2 for Self- and Cross-Impact Models in units of 10^{-2}.

(a) Self-Impact

Linear	Square-root	Enhanced dataset
3.2	4.6	4.8

(b) Cross-Impact

Two Assets	Three Assets
6.3	6.5

Self-Impact Estimation

The concavity of the self-impact function is first assessed using only the CFM metaorder dataset. The R^2 peaks around a concavity parameter $c = 0.6$ [19](Figure 5.5), and the square-root case ($c = 0.5$) results in a R^2 of 4.6%. This supports the use of concave impact functions over linear ones.

To improve kernel estimation, the dataset is extended using metaorder proxies. Figure 5.1 shows the kernel estimate based on the original (gray) and extended dataset (blue). The estimate based on CFM data alone appears irregular, whereas the proxy-enhanced estimate demonstrates a smooth decay consistent with a power law. This behavior becomes clearer in Figure 5.6 with log-log scale where the propagator follows a straight line.

Initially, due to our normalization, the propagator decays as t^{-1}, since returns are divided by lag-wise volatility. To align with the empirical findings in Bouchaud et al. (2004); Bucci et al. (2015), where the propagator accounts also for volatility, we rescale our estimated kernel by multiplying by $\sqrt{t}$. This adjustment recovers the conventional decay rate of $t^{-0.5}$, which is consistent with the literature.

While concavity in volume is the dominant contribution to the predictive power of the impact-model, the decay is a second-order effect, consistent with findings in Hey et al. (2025). Table 5.3.4 shows that R^2 improves slightly when including proxy data.

Cross-Impact Estimation

Pairwise and multi-asset cross-impact models are considered next. The two-asset model shows that including a second highly correlated contract increases R^2 to 6.3%, a substantial gain over the self-impact-only specification (Table 5.3.4b). The three-asset model increases R^2 further to 6.5%.

Figure 5.7 illustrates the estimated propagators for two corn futures with different expiries labeled 0 and 1. The kernels are asymmetric: the impact of Corn0 on Corn1 is stronger than the reverse. This asymmetry reflects liquidity differences across assets,

[19]This is consistent with previous empirical studies, e.g., Tóth et al. (2016), which find that $c \in [0.4, 0.6]$ across asset classes.

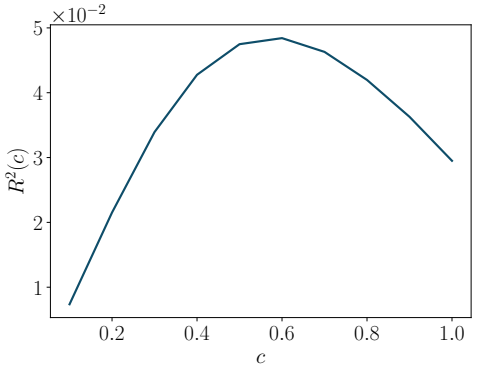


Figure 5.5: Out-of-sample R^2 as a function of concavity c for self-impact models using CFM metaorders. The peak occurs at $c = 0.6$.

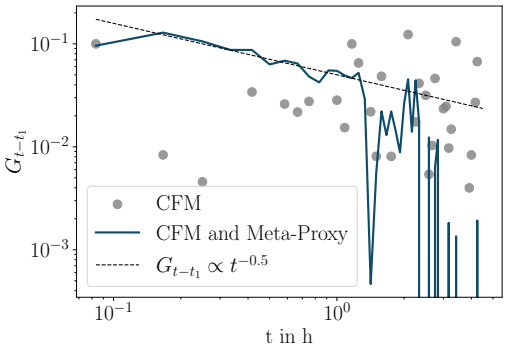


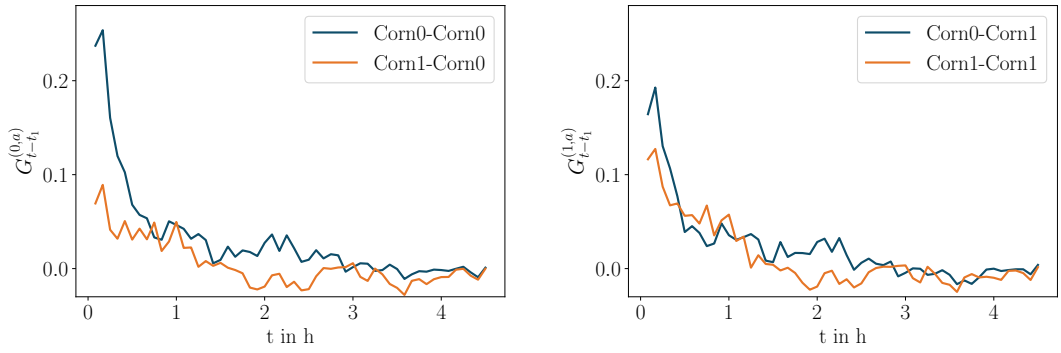
Figure 5.6: Impact kernel estimates as a function of time in hours for self-impact with (blue) and without proxy-enhanced date (gray) in log-log scale. The proxy-enhanced kernel, after rescaling, decays smoothly and can be approximated by a power-law with an exponent of -0.5.

with Corn1 being less liquid. These findings align with previous observations on aggregate order flow impact in Le Coz et al. (2024), but here they are obtained via direct calibration of the propagator for metaorders.

Figure 5.8 extends the estimation to three assets. While the qualitative patterns persist, the estimates become less stable. The width of the confidence intervals derived in 5.2.4 increases substantially from the single-asset to the multi-asset model. This confirms that multi-asset cross-impact estimation is subject to high variance, but the structure of the kernels remains interpretable. In particular, cross-impact can exceed self-impact in magnitude, depending on relative liquidity as shown in Figures 5.7b, 5.8b, and 5.8c.

5.4 Conclusion

This chapter developed a multivariate estimation framework for concave price impact using metaorder data, extending a known offline learning method to an empirically



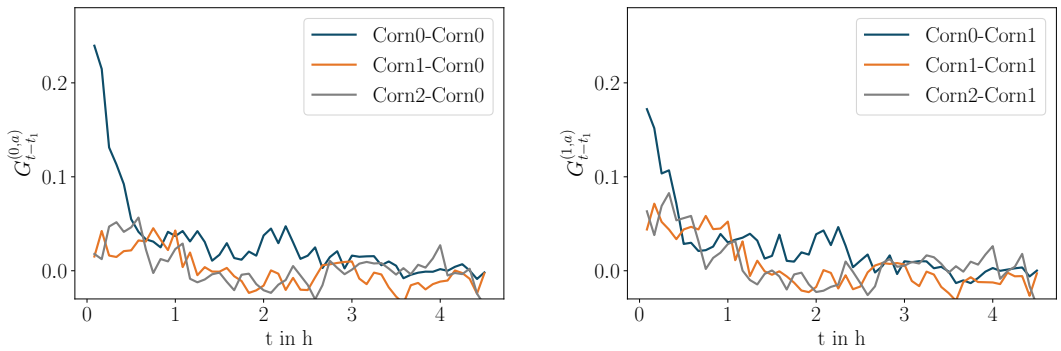
(a) $G_t^{(0,a)}$ (b) $G_t^{(1,a)}$

Figure 5.7: Cross- and self-impact kernels as a function of time in hours estimated for two assets with different expiries. Left: The estimated cross-impact from Corn1 to Corn0 (orange) and the self-impact on Corn0 (blue). Right: The estimated cross-impact from Corn0 to Corn1 (blue) and the self-impact on Corn1 (orange). The difference in ratios between cross- and self-impact in the two plots reflects the liquidity differences.

relevant setting. A confidence bound was derived, and its dependence on the number of assets d, scaling as d^2, illustrates the data requirements for high-dimensional models.

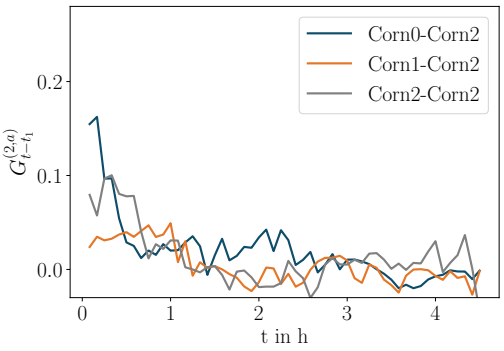
Using a novel metaorder proxy, synthetic metaorders were constructed to enrich the dataset and stabilize the estimation, particularly for the impact decay. To the author's knowledge, this is the first calibration of a metaorder-based cross-impact model involving more than two assets.

Several limitations and extensions remain. While the convexity constraints used are practically motivated, admissibility conditions for concave multivariate kernels remain an open question. Although the methodology is illustrated using corn futures, it generalizes directly to other asset classes. Finally, the dimensionality of the kernel poses practical challenges. To address this, one could introduce market-wide liquidity factors or reduce the model to pairwise interactions through a market mode decomposition, as proposed in Muhle-Karbe and Tracy (2024).



(a) $G_t^{(0,a)}$

(b) $G_t^{(1,a)}$



(c) $G_t^{(2,a)}$

Figure 5.8: Estimated cross- and self-impact kernels as a function of time in hours for the three-asset model on corn futures with different expiries. The relative importance of self- and cross-impact kernels varies according to liquidity difference. The confidence intervals from 5.2.4 increase with dimensionality causing less stable calibrations.

Chapter 6

Conclusion

Over the past decades, market microstructure research has grown increasingly sophisticated. Yet, most execution models used in practice remain crude simplifications of a more nuanced empirical reality. This thesis demonstrates that price impact is concave, decaying, and cross-asset dependent, and that these features can be meaningfully incorporated into actionable execution strategies without sacrificing tractability. By bridging the gap between theory and real data, the work offers explicit and interpretable strategies for hedge funds and algorithmic traders who face nonlinear costs in financial markets. But its relevance goes beyond optimal execution in institutional trading. For academics, it provides a blueprint for modeling impact realistically under minimal assumptions, with transparent constraints that ensure arbitrage-freeness. It also demonstrates how public trade data, when properly aggregated and in abundance, can serve as a viable alternative in the absence of meta-order datasets, especially when enhanced by synthetic proxies. More broadly, this work contributes to a principled and empirically validated framework for thinking about alpha trading as a control problem under frictions, with implications for both financial engineering and the design of regulatory impact models that better reflect actual market behavior.

At its core, this thesis provides comprehensive guidelines for trading under concave and transient price impact for both single-asset and multi-asset settings. By mapping the trading problem into impact space, we derived explicit, closed-form solutions for optimal trading strategies under realistic alpha signals, dynamic liquidity, and multiple decay timescales. These solutions reveal how impact concavity, alpha decay, and liquidity dynamics jointly shape execution decisions. Importantly, we established conditions to guarantee that these strategies remain manipulation-free to ensure consistency even in complex multi-asset environments. Our modeling choices are not just theoretically motivated but grounded in proprietary meta-order data, producing strategies that match market behavior more closely than those built on public order flow alone.

But designing a good strategy is not just about theory — it is about robustness. We quantified the costs of model misspecification, showing that underestimating concavity or decay can lead to significantly degraded P&L, even when statistical fit appears ac-

ceptable. On the empirical side, we extended kernel estimation tools to non-parametric, multivariate settings, showing that meta-order cross-impact is not only measurable but asymmetric, driven by liquidity differences across assets. Enhancing proprietary data with synthetic meta-orders further improved estimation accuracy. Altogether, this work yields a comprehensive checklist for execution strategy design:

An optimal trading strategy should ...

- balance alpha decay and impact decay across the full execution horizon;
- adapt to changing market liquidity;
- reflect the concave nature of market impact, especially in large-volume regimes;
- account for transient impact components;
- remain arbitrage-free by construction;
- account for cross-asset effects and optimize execution at the portfolio level;
- be calibrated on reliable trade data – ideally from meta-orders or enhanced public data proxies.

These principles offer both practical guidance for trading and a rigorous foundation for future academic models. Rather than assuming a fixed model, we show that good execution starts with asking the right empirical questions and then building models constrained only by what the data actually tell us.

6.1 Limitations

Despite the breadth of this study, several limitations remain. Most notably, while we successfully developed arbitrage-free models for concave self-impact and a specific concave power-law form $h(x) = \text{sgn}(x)|x|^c$, we were unable to derive necessary no-manipulation conditions for arbitrary concave functions in the cross-impact setting.

Empirically, our results are based on futures contracts, which is a liquid and homogeneous asset class. Although we expect similar behaviors in equities, applying the framework to options would require a significantly richer structure to account for the strike-maturity surface.

Furthermore, all cross-impact estimations, e.g. parametric and non-parametric, were restricted to a maximum of three assets due to data limitations. Beyond that, estimation becomes increasingly noisy, suggesting that alternative approaches, such as fitting a bivariate model between each asset and an aggregate market factor (see Mastromatteo et al. (2017); Muhle-Karbe and Tracy (2024)), may be more viable in high-dimensional settings. These challenges are compounded by the fact that reliable estimation relies

on high-quality meta-order data, or the effectiveness of proxy constructions, which may not be uniformly available across all markets.

Finally, the theoretical derivations in this thesis are based on a risk-neutral P&L objective without accounting for spread costs and other fees. This formulation makes sense for large hedge funds and principal trading firms, which typically operate under minimal inventory constraints and seek to maximize raw returns. However, many market participants, such as market makers, pension funds, or retail brokers, face risk limits, inventory penalties, and capital constraints that cannot be ignored. For market makers, in particular, trading costs such as spreads and fees are crucial, especially in high-frequency trading environments where they can significantly influence profitability. Extending the control problem to risk-sensitive objectives with linear spread costs and fees would significantly broaden the applicability of the framework. Yet, tractable formulas that optimally balance predictive trading signals, risk, and trading costs are typically limited to quadratic costs with linear price impacts.

6.2 Future Research

Concave Cross-Impact and ETFs. Building on our current results and the theory of the multivariate AFS model, a natural next step is to apply this framework to Exchange-Traded Funds (ETFs), which represent liquid baskets of underlying assets. This extension would allow us to empirically examine the cross-impact arising from ETF flows and to design trading rules that account for the underlying correlation structure and execution frictions. While del Molino et al. (2020) address linear execution, our framework introduces concavity and decay, potentially leading to more realistic cost models and improved execution strategies for ETF market makers and index traders.

Leveraging Alpha Signals. Efficiently utilizing multiple alpha signals is increasingly critical for modern funds, particularly as they face diminishing returns from price impact. As CFM notes, "*[price impact] costs always catch up and overtake expected gains. This effect of diminishing returns with ever larger orders is the principal reason why successful funds are often closed to further investment.*" Hedge funds must balance transactional efficiency while managing diverse signals with varying time horizons and stochastic properties. The framework of "A Separation Principle for Dynamic Portfolio Optimization" by Moallemi and Van Roy provides a robust foundation for addressing this challenge. By reducing complex multi-period portfolio problems to single-period deterministic optimizations, this framework allows for dynamic adaptation to evolving market conditions, making it an ideal starting point for exploring alpha signal mixing.

An extended framework would allow one to accommodate convex constraints or concave single-asset transaction costs to better reflect real-world trading conditions. It can also be extended to incorporate cross-asset transaction costs and cross-alpha dynamics. To incorporate cross-impact, the forecast return term structure could be expanded to include (linear or concave) interactions with other assets.

Dynamic Market Impact in AMMs. In automated market makers (AMMs), market impact is directly tied to the liquidity in the pool, determined by the capital allocated to the exchange. Unlike limit order books (LOBs), where liquidity varies dynamically with supply and demand, AMMs provide deterministic pricing based on invariant functions or demand curves. While this allows the immediate price impact of a trade to be predictable, the dynamics following a sequence of trades remain unclear. Key questions include: How do prices evolve after a large trade? How long does it take for prices to revert to pre-trade levels? And what factors drive this reversion or persistence? Investigating these dynamics requires analyzing on-chain data from platforms like Uniswap and Curve that provide trade sizes, pool reserves, and price trajectories to construct models for price reversion and market stability. Understanding these single-asset dynamics is important for optimizing trading rules and mitigating adverse selection within AMMs.

Cross-Impact and Causality. An interesting research direction lies in linking concave cross-impact models based on meta-order data with causal approaches to market dynamics. Recent studies on cross-impact derived from order flow imbalance (Albers et al., 2021; Cont et al., 2021) and causality-inspired financial models (Oliveira et al., 2024) have emphasized the role of impact in shaping trading costs and price formation, albeit from different perspectives. In the present work, we develop a cross-impact model that acknowledges our trades can affect the prices of correlated assets, even without directly trading them. This is distinct from the approach taken by the aforementioned authors, who incorporate high-frequency order flow data to account for these observed price changes. In their framework, trading one asset and moving its price can trigger other market participants to trade correlated assets, reflecting an indirect form of impact not labeled as cross-impact in their context.

Integrating causal inference tools with structural models could lead to a better understanding of how alpha signals, informational asymmetries, and short-term market dynamics interact. In particular, the causal regularization framework proposed in (Webster and Westray, 2022) offers a way to address latent confounding, reduce estimation bias, and improve the interpretability and robustness of cross-impact models. Such a synthesis would not only enhance predictive accuracy, but would also contribute to more reliable and theoretically grounded execution strategies.

Extending the (Linear) Latent Order Book Model. A final direction for future research concerns the extension of the nonlinear latent order book framework proposed by (Donier et al., 2015), which models price trajectories using a reaction-diffusion system inspired by agent-based dynamics. Their approach reproduces the square-root impact law, decomposes price moves into mechanical and informational components, and remains free of price manipulation. Extending this model to a multivariate setting could be achieved by introducing inter-asset interactions mediated by distinct types of agents, such as statistical arbitrageurs, who link price dynamics across order books. This would provide a natural microstructural mechanism for modeling cross-impact from first prin-

ciples and could lead to new insights into how liquidity and information propagate across correlated assets.

6.3 Final Remark

The directions outlined above are not exhaustive, but offer a glimpse into the many ways in which this work can be extended, adapted, and applied. As trading continues to evolve, so too must our models. Beyond the models and results, the work highlights a broader point: price impact is not just a nuisance to be minimized; it is a window into how information and liquidity interact. By understanding it more deeply, we also understand more about how markets work.

A better understanding of price impact is not just of academic interest. It enables more efficient trade execution, improves transaction cost estimation, and ultimately contributes to more efficient markets. As such, refining these models and extending them to new asset classes, trading environments, and data regimes remains a worthwhile direction for both research and industry practice.

Bibliography

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

E. Abi Jaber, E. Neuman, and S. Tuschmann. Optimal portfolio choice with cross-impact propagators. *Preprint*, 2024.

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.

J. Albers, M. Cucuringu, S. Howison, and A. Y. Shestopaloff. Fragmentation, price formation and cross-impact in bitcoin markets. *Applied Mathematical Finance*, 28(5): 395–448, 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.

A. Alfonsi, F. Klöck, and A. Schied. Multivariate transient price impact and matrix-valued positive definite functions. *Mathematics of Operations Research*, 41(3):914–934, 2016.

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

R. F. Almgren. A deep dive into optimal trade execution in fixed income and futures using quantitative methods. *The Wall Street Lab Podcast, Episode #77*, 2021.

R. F. Almgren and N. Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40, 2001.

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.

D. Becherer, T. Bilarev, and P. Frentrup. Stability for gains from large investors' strategies in M_1/J_1 topologies. *Bernoulli*, 25(2):1105–1140, 2019.

M. Benzaquen, I. Mastromatteo, Z. Eisler, and J.-P. Bouchaud. Dissecting cross-impact on stock markets: An empirical analysis. *Journal of Statistical Mechanics*, February: 023406, 2017.

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.

D. Bertsimas and A. W. Lo. Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50, 1998.

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.

T. Bilarev. *Feedback effects in stochastic control problems with liquidity frictions*. PhD thesis, Humboldt-Universität zu Berlin, 2018.

J.-P. Bouchaud. The inelastic market hypothesis: a microstructural interpretation. *Quantitative Finance*, 22(10):1785–1795, 2022.

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, J. D. Farmer, and F. Lillo. How markets slowly digest changes in supply and demand. In T. Hens and K. R. Schenk-Hoppé, editors, *Handbook of Financial Markets: Dynamics and Evolution*, pages 57–160. North Holland, 2009a.

J.-P. Bouchaud, J. 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, 2009b.

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

V. Bouniakovsky. Note sur les maxima et les minima d'une fonction symetrique entiere de plusieurs variables. *Bull. Classe Phys.-Math*, (12):353–361, 1854.

G. Box. Science and statistics. *Journal of the American Statistical Association*, 71(356): 791–799, 1976.

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. *Market Microstructure and Liquidity*, 4(3):1950006, 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*, 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*, 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*, 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.

L.C.G. del Molino, I. Mastromatteo, M. Benzaquen, and J.-P. Bouchaud. The multivariate Kyle model: More is different. *SIAM Journal on Financial Mathematics*, 11 (2):327–357, 2020.

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 Central Bank. Algorithmic trading – rise of the machines, 2019. *SSM Supervisory Newsletter*, February 2019.

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*, 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 and R. Koijen. In search of the origins of financial fluctuations: The inelastic markets hypothesis. *Preprint*, available at `ssrn.com`, 2021.

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.

P. Guasoni and M. H. Weber. Nonlinear price impact and portfolio choice. *Mathematical Finance*, 30(2):341–376, 2020.

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. One security, many markets: Determining the contributions to price discovery. *The Journal of Finance*, 50(4):1175–1199, 1995.

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*, January, 2024a.

N. Hey, I. Mastromatteo, and J. Muhle-Karbe. Concave cross-impact. *submitted to Management Sciences*, Decemer, 2024b.

N. Hey, I. Mastromatteo, J. Muhle-Karbe, and K. Webster. Trading with concave price impact and impact decay - theory and evidence. *Operations Research*, 2025.

U. Horst and X. Xia. Multi-dimensional optimal trade execution under stochastic resilience. *Finance and Stochastics*, 23(4):889–923, 2019.

G. Huberman and W. Stanzl. Price manipulation and quasi-arbitrage. *Econometrica*, 72(4):1247–1275, 2004.

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*, 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.

V. Le Coz, I. Mastromatteo, D. Challet, and M. Benzaquen. When is cross impact relevant? *Quantitative Finance*, 24(2):265–279, 2024.

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.

G. Maitrier, G. Loeper, and J.-P. Bouchaud. Generating realistic metaorders from public data. *Preprint*, 2025a.

G. Maitrier, G. Loeper, K. Kanazawa, and J.-P. Bouchaud. The "double" square-root law: Evidence for the mechanical origin of market impact using tokyo stock exchange data, 2025b. *Preprint*.

R. Martin and T. Schöneborn. Mean reversion pays, but costs. *Risk*, 24(2):84–89, 2011.

I. Mastromatteo, B. Tóth, and J.-P. Bouchaud. Agent-based models for latent liquidity and concave price impact. *Physical Review E*, 89(4):042805, 2014.

I. Mastromatteo, M. Benzaquen, Z. Eisler, and J.-P.. Bouchaud. Trading lightly: Cross-impact and optimal portfolio execution. *Risk*, July, 2017.

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 and C. Tracy. Stochastic liquidity as a proxy for concave multi-asset propagator models. *Preprint*, 2024.

J. Muhle-Karbe, Z. Wang, and K. Webster. Stochastic liquidity as a proxy for nonlinear price impact. *Operations Research*, 72(2):444–458, 2024.

E. Neuman and Y. Zhang. Statistical learning with sublinear regret of propagator models, 2023. *accepted in The Annals of Applied Probability*.

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

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

D. C. Oliveira, Y. Lu, X. Lin, M. Cucuringu, and A. Fujita. Causality-inspired models for financial time series forecasting. *Preprint*, 2024.

P. Pasquariello and C. Vega. Strategic cross-trading in the u.s. stock market. *Review of Finance*, 19(1):229–282, 2013.

F. Patzelt and J.-P. Bouchaud. Nonlinear price impact from linear models. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(12):123404, dec 2017.

F. Patzelt and J.-P. Bouchaud. Universal scaling and nonlinearity of aggregate price impact in financial markets. *Physical Review E*, 97(1):012304, 2018.

M. Rosenbaum and M. Tomas. A characterisation of cross-impact kernels. *Frontiers of Mathematical Finance*, 1(4):491–523, 2022.

P. Samuelson. Proof that properly discounted present values of assets vibrate randomly. *The Bell Journal of Economics and Management Science*, 4(2):369–374, 1973.

Y. Sato and K. Kanazawa. Does the square-root price impact law belong to the strict universal scalings?: quantitative support by a complete survey of the tokyo stock exchange market. *Preprint*, 2024.

A. Schied, T. Schöneborn, and M. Tehranchi. Optimal basket liquidation for CARA investors is deterministic. *Applied Mathematical Finance*, 17(6):471–489, 2010.

M. Schneider and F. Lillo. Cross-impact and no-dynamic-arbitrage. *Quantitative Finance*, 19(1), 2019.

P. J. Schönbucher. A market model for stochastic implied volatility. *Philosophical Transactions of the Royal Society of London (A)*, 357:2071–2092, 1999.

D. E. Taranto, G. Bormetti J.-P., Bouchaud, F. Lillo, and B. Tóth. Linear models for the impact of order flow on prices i. propagators: Transient vs. history dependent impact. *Preprint*, 2016.

M. Tomas, I. Mastromatteo, and M. Benzaquen. Cross impact in derivative markets. *Preprint*, 2022a.

M. Tomas, I. Mastromatteo, and M. Benzaquen. How to build a cross-impact model from first principles: Theoretical requirements and empirical results. *Quantitative Finance*, 22(6):1017–1036, 2022b.

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. Tóth, Z. Eisler, and J.-P. Bouchaud. The square-root impact law also holds for option markets, 2016. *Preprint*.

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

G. Tsoukalas, J. Wang, and K. Giesecke. Dynamic portfolio execution. *Management Science*, 65(5):2015–2040, 2019.

S. Wang and T. Guhr. Microscopic understanding of cross-responses between stocks: A two-component price impact model. *Market Microstructure and Liquidity*, 03(03n04): 1850009, 2017.

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

K. Webster and N. Westray. Getting more for less - better a/b testing via causal regularization. *Risk*, 2022.

ECOLE
DOCTORALE

Titre : Exécution d'ordres avec impact (croisé) concave

Mots clés : Exécution optimale, Impact de marché, Impact croisé, Calibrage de modèle, Trading algorithmique, Analyse des coûts de transaction

Résumé : Dans les marchés financiers modernes, l'exécution de gros ordres affecte significativement les prix des actifs, un phénomène connu sous le nom d'impact de marché. Comprendre et modéliser cet impact est essentiel pour les investisseurs institutionnels et les traders algorithmiques, car il influence directement les coûts d'exécution, les stratégies de trading et l'efficacité des marchés. Bien que de nombreux modèles analytiques supposent un impact linéaire, les données empiriques révèlent généralement un impact concave qui décroit dans le temps, suivant souvent une loi de puissance plutôt qu'une exponentielle.
Ce travail explore à la fois les fondements théoriques et les propriétés empiriques de l'impact de marché. Nous proposons d'abord une stratégie d'exécution optimale dans un cadre non linéaire avec impact transitoire, montrant comment les signaux d'alpha et la liquidité doivent être équilibrés pour minimiser les coûts. Ensuite, nous quantifions les pertes liées à une mauvaise spécification du modèle d'impact, en montrant que sous-estimer la concavité ou la vitesse de décroissance peut gravement nuire à la performance. Nous étendons ensuite ces modèles au cas multi-actifs avec impact croisé, tout en assurant des conditions de non-manipulation des prix. Enfin, nous développons une méthode non paramétrique pour estimer l'impact croisé sur des données de méta-ordres enrichies par des proxies construits à partir de données publiques. Ces résultats permettent de calibrer plus efficacement les modèles d'exécution, en capturant les effets de corrélation entre actifs et les asymétries de liquidité observées sur les marchés.

Title : Trading with Concave (Cross-) Impact

Keywords : Optimal Execution, Market Impact, Cross-Impact, Model Calibration, Algorithmic Trading, Transaction Cost Analysis

Abstract : Executing large orders in financial markets moves prices — an effect known as price impact. Understanding this phenomenon is essential for institutional investors, algorithmic traders, and researchers, as it directly influences transaction costs and the efficiency of trading strategies. While many theoretical models assume linear impact, empirical studies show that price impact is often concave and transient, decaying over time in a manner more consistent with power laws than exponential functions. This thesis develops a modeling framework that incorporates these empirical realities and translates them into practical execution strategies.
First, we derive optimal execution strategies under concave and transient impact, demonstrating that closed-form trading rules can be obtained even with time-varying alpha and liquidity. Empirical calibration on proprietary futures data shows that impact decays across multiple timescales and confirms that public trading data systematically underestimates execution costs. We also quantify the performance loss due to model misspecification and show that underestimating decay or concavity is more damaging than overestimation — providing new guidance for model calibration.
Second, we extend concave impact models to multi-asset settings. This includes a theoretical framework for arbitrage-free cross-impact, as well as non-parametric kernel estimation techniques based on meta-order data. To overcome data sparsity, we work with synthetic meta-order proxies that enhance the empirical estimation. This allows us to uncover cross-impact asymmetries driven by liquidity differences and show predictive gains in multi-asset price dynamics. The models and estimators presented here offer robust tools for both practitioners and researchers aiming to design efficient execution strategies or to analyze market impact using trade-level data.

Institut Polytechnique de Paris
91120 Palaiseau, France