CONCAVE CROSS IMPACT

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

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

Abstract. The price impact of large meta orders 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 meta order data.

Keywords: cross impact; dynamic-no-arbitrage; concave price impact; impact decay; optimal trading; model calibration

1. Introduction

Price impact is the main source of trading costs for large institutional investors. The impact of large meta orders builds up over time and then gradually decays (Biais et al., 1995; Hasbrouck, 1991). 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.

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 is that with several traded assets, guaranteeing the absence of "price manipulation"

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

Date: December 5, 2024.

The auhtors gratefully acknowledge fruitful discussions with Dirk Becherer, Jean-Philippe Bouchaud, Jim Gatheral, and Kevin Webster.

[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).

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., 2023). 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). 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 detect 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., 2023). 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.

The relevance and applicability of this model class is in turn illustrated by an empirical case study based on proprietary meta order data from Capital Fund Management (CFM). We first perform a simple comparison of the arrival prices at the beginning of each meta order to the peak impact incurred at their completion. As illustrated in Figure 1 this demonstrates that for highly correlated assets,[2] cross impact measurements are highly significant and depend on meta order 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, whereas they largely offset each other when traded in opposite directions.

[2]Here, we focus on futures contracts with the same underlying but different maturities so that the average return correlation is more than 90%.

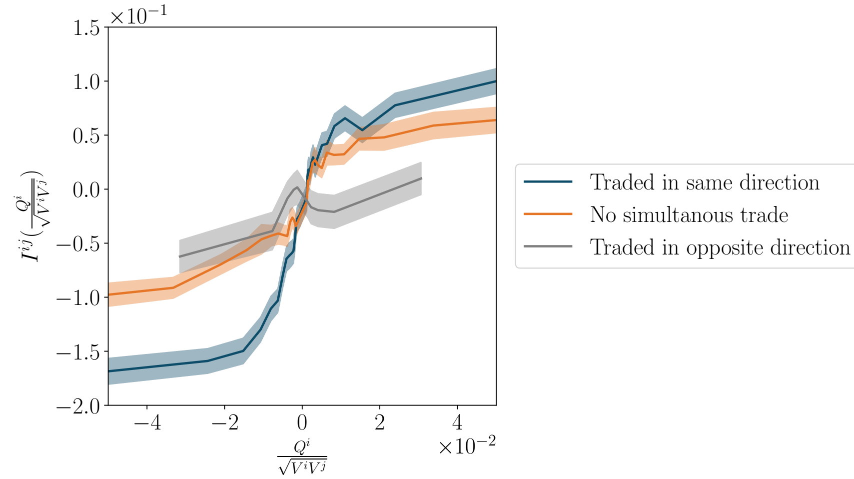


FIGURE 1. Average differences between the prices of asset j at the beginning and end of meta orders for asset i, plotted against the size of the meta orders (normalized by the geometric mean of the average daily trading volumes of assets j and i). The shaded regions are bootstrap confidence intervals.

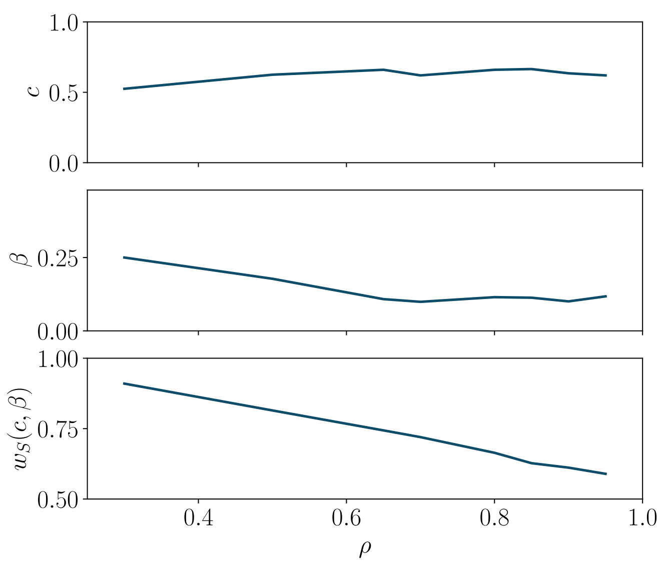


FIGURE 2. Parameter estimates for a bivariate cross impact model. The common impact concavity c, impact decay rate β and proportion $w_S(c,\beta)$ of total impact caused by self impact is plotted against the correlation between the assets.

Building on this proof of principle, 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 2 displays how the fitted model parameters depend 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 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.

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 2 introduces our multivariate extension of the nonlinear price impact model introduced by Alfonsi et al. (2010). Subsequently, in Section 3 we formulate risk-neutral optimal trading problems in this context and then reformulate these "in impact space" in Section 4. In Section 5, we in turn derive necessary conditions for the absence of price manipulation and then show in Section 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 7. For better readability, the derivations of the no-price-manipulation conditions are delegated to the appendix.

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 $dJ_t = -\frac{1}{\tau}J_t + \lambda dQ_t$ 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.

[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 meta order level and described by the AFS model.

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

$$d\boldsymbol{J}_t = -\boldsymbol{B}\boldsymbol{J}_t dt + \boldsymbol{\Lambda} d\boldsymbol{Q}_t, \quad \boldsymbol{B}, \boldsymbol{\Lambda} \in \mathbb{R}^{d\times d}. \tag{2.1}$$

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

$$\boldsymbol{I}_t = \sum_{a=1}^{d} \boldsymbol{L}^a h(J_t^a), \quad \text{where } \boldsymbol{L}^a \in \mathbb{R}^d \text{ for } a = 1, \ldots, d. \tag{2.2}$$

This means that the price impact in each asset $i = 1, \ldots, d$ 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, $a = 1, \ldots d$. 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 $\boldsymbol{L} = (\boldsymbol{L}^1, \ldots, \boldsymbol{L}^d) \in \mathbb{R}^{d\times d}$ and writing, with a slight abuse of notation, $h(\boldsymbol{J}_t) = (h(J_t^1), \ldots, h(J_t^d))^\top \in \mathbb{R}^d$, we can then represent the price impact concisely in matrix-vector notation as

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

Remark 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 (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., 2023). 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.

3. Risk-Neutral Goal Functional

We now derive the trader's profits and losses (PnL) when trading with nonlinear price impact of the form (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

[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.

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 PnL 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

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

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

$$\mathbb{E}\left[\int_0^T (\boldsymbol{\alpha}_t - \boldsymbol{I}_t)^\top d\boldsymbol{Q}_t\right]. \tag{3.1}$$

That is, each trade earns alpha and pays impact. The PnLs 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 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 (3.1) *remains unchanged if the terminal position is valued with the forecast $\mathbb{E}_T[S_\tau]$ at time T.*

4. Passage to Impact Space?

For single-asset models, the risk-neutral goal functional (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., 2023).

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:

$$d\boldsymbol{Q}_t = \boldsymbol{\Lambda}^{-1}\boldsymbol{B}\boldsymbol{J}_t dt + \boldsymbol{\Lambda}^{-1} d\boldsymbol{J}_t. \tag{4.1}$$

Using this identity to replace the trades $d\boldsymbol{Q}_t$ in (3.1), the expected PnL 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

$$\mathbb{E}\left[\int_0^T \left(\bar{\boldsymbol{\alpha}}_t^\top \boldsymbol{J}_t - h(\boldsymbol{J}_t)^\top \boldsymbol{\zeta}\boldsymbol{J}_t\right) dt - \int_0^T h(\boldsymbol{J}_t)^\top \boldsymbol{\theta} d\boldsymbol{J}_t + \boldsymbol{\alpha}_T^\top \boldsymbol{\Lambda}^{-1}\boldsymbol{J}_T\right]. \tag{4.2}$$

[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 PnL's "in impact space", where the extension to general strategies is straightforward.

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

$$\boldsymbol{\zeta} = \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{B}, \qquad \boldsymbol{\theta} = \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1}, \quad \text{and} \quad \bar{\boldsymbol{\alpha}}_t = \boldsymbol{\zeta}^\top \boldsymbol{L}^{-1} \boldsymbol{\alpha}_t - \boldsymbol{\theta}^\top \boldsymbol{L}^{-1} \boldsymbol{\mu}_t^\alpha, \tag{4.3}$$

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

Remark 4.1. *The transformations* (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 PnL (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 each be maximized pointwise in a straightforward manner (Hey et al., 2023). 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 (4.3) to rule out price manipulation, i.e., trades for which a positive expected PnL is generated by price impact rather than the presence of an alpha signal.

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

$$C_T = \int_0^T \left(h(\boldsymbol{J}_t)^\top \boldsymbol{\zeta} \boldsymbol{J}_t + h(\boldsymbol{J}_t)^\top \boldsymbol{\theta} \frac{d\boldsymbol{J}_t}{dt} \right) dt. \tag{5.1}$$

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 $a = 1, \ldots, d$, for which either the first or the second term (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 (5.1) and in turn yield conditions on the matrix $\boldsymbol{\zeta}$. Indeed, consider a trading strategy where all impact states are symmetric around some time $T_\star/2$, in that $J^a_{T_\star/2-\epsilon} = J^a_{T_\star/2+\epsilon}$ for $0 \leq \epsilon \leq T_\star/2$.[6] Then, by construction:

$$\frac{dJ^a_{T_\star/2-\epsilon}}{dt} = -\frac{dJ^a_{T_\star/2+\epsilon}}{dt}. \tag{5.2}$$

[6] Note that this does *not* mean that the corresponding trades 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.

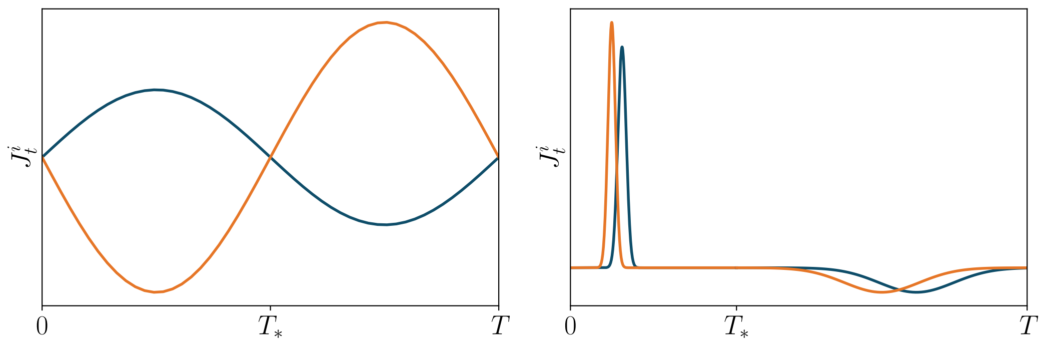


FIGURE 3. Symmetric strategy (5.4) (left panel) and impulsive strategy (5.8) (right panel).

The $\boldsymbol{\theta}$-terms in the cost functional (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 (5.1).

In the next two sections, we consider specific examples for such symmetric and impulsive strategies (illustrated in Figure 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.

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 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 (4.1) and (4.3),

$$\boldsymbol{L}^\top \boldsymbol{Q}_T = \int_0^T \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{B} \boldsymbol{J}_t dt + \int_0^T \boldsymbol{L}^\top \boldsymbol{\Lambda}^{-1} d\boldsymbol{J}_t = \boldsymbol{\zeta} \int_0^T \boldsymbol{J}_t dt + \boldsymbol{\theta} \int_0^T \frac{d\boldsymbol{J}_t}{dt} dt. \tag{5.3}$$

Due to the symmetry (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

$$(J_t^a, J_t^b) = (j_a \sin(t), -j_b \sin(t)), \quad 0 \le t < 2\pi. \tag{5.4}$$

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

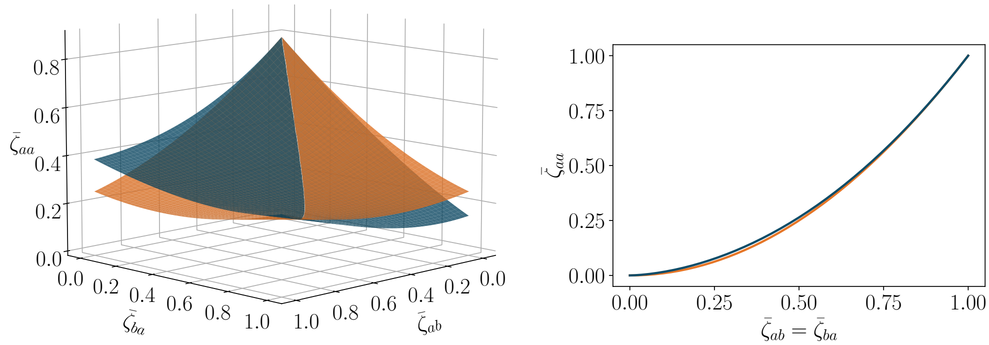


FIGURE 4. The constraints for linear price impact ($c = 1$, orange), and square-root impact ($c = 1/2$, blue): admissible values of $\bar{\zeta}_{aa}$ need to lie above the surface (left panel) and curve (right panel, for symmetric $\boldsymbol{\zeta}$), respectively.

leads to the following necessary conditions for the absence of price manipulation, derived in Appendix A:

Lemma 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:*

$$0 < \zeta_{aa} + \phi^{1+c}\zeta_{bb} - \phi^c\zeta_{ab} - \phi\zeta_{ba}, \quad \textit{for all } \phi \geq 0 \textit{ and } a, b = 1, \ldots, d. \tag{5.5}$$

Specialized to small values of ϕ, Condition (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 (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

$$\bar{\zeta}_{aa} > \frac{1}{4}(\bar{\zeta}_{ab} + \bar{\zeta}_{ba})^2, \tag{5.6}$$

where $\bar{\zeta}_{aa} = \zeta_{aa}/\zeta_{bb}$, $\bar{\zeta}_{ab} = \zeta_{ab}/\zeta_{bb}$, and $\bar{\zeta}_{ba} = \zeta_{ba}/\zeta_{bb}$. To wit, 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 (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

$$\bar{\zeta}_{aa} > \frac{2}{27}\left(\left(3\bar{\zeta}_{ab} + \bar{\zeta}_{ba}^2\right)^{3/2} + \bar{\zeta}_{ba}^3\right) + \frac{1}{3}\bar{\zeta}_{ab}\bar{\zeta}_{ba}. \tag{5.7}$$

The left panel of Figure 4 visually compares the conditions for $c = 1$ and $c = 1/2$ 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.

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]

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

$$(J_t^a, J_t^b) = \begin{cases} \left(\frac{1}{\sqrt{2\pi(\sigma_1^a)^2}} e^{-\frac{(t-\mu_1^a)^2}{2(\sigma_1^a)^2}}, \quad \frac{1}{\sqrt{2\pi(\sigma_1^b)^2}} e^{-\frac{(t-\mu_1^b)^2}{2(\sigma_1^b)^2}} \right) & 0 \le t < T_*, \\ \left(\frac{-1}{\sqrt{2\pi(\sigma_2^a)^2}} e^{-\frac{(t-\mu_2^a)^2}{2(\sigma_2^a)^2}}, \quad \frac{-1}{\sqrt{2\pi(\sigma_2^b)^2}} e^{-\frac{(t-\mu_2^b)^2}{2(\sigma_1^b)^2}} \right) & T_* \le t \le T. \end{cases} \tag{5.8}$$

As illustrated in Figure 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 (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 B and they demand a strikingly strict shape of matrix $\boldsymbol{\theta}$ when impact is strictly concave rather than linear:

Lemma 5.2. *To avoid price manipulation:*

1. *For a linear impact function $h(x) = x$, the matrix $\boldsymbol{\theta}$ must be* symmetric*;*
2. *For a concave impact function $h(x) = \text{sgn}(c)|x|^c$ 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 (4.2) in impact space in fact have to vanish. Put differently, the models for which the optimization problem (4.2) is intractable are ruled out already by imposing no price manipulation.

Remark 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:*

$$d\boldsymbol{J}_t = \boldsymbol{\theta}^{-1} \left(-\boldsymbol{\zeta} \boldsymbol{J}_t dt + d\boldsymbol{L}^T \boldsymbol{Q}_t \right). \tag{5.9}$$

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 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 $\boldsymbol{\theta}$.

6. Solution of the Risk-Neutral Optimization

In view of Lemma 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 (4.2). 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 $h(\cdot)$, (4.2) can therefore can be reduced to a simple pointwise maximization just like in the single-asset case (Hey et al., 2023):

$$\mathbb{E}\left[\int_0^T \left(\bar{\boldsymbol{\alpha}}_t^\top \boldsymbol{J}_t - h(\boldsymbol{J}_t)^\top \boldsymbol{\zeta} \boldsymbol{J}_t\right) dt + \boldsymbol{\alpha}_T^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{J}_T - \mathbf{1}^\top \boldsymbol{\theta} H(\boldsymbol{J}_T)\right]. \tag{6.1}$$

Remark 6.1. *Unlike its counterpart* (4.2) *in trade space, the goal functional* (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 PnL as the limit of the PnLs 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.

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 (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., 2023, Theorem 4.2). In particular, the necessary conditions from Lemmas 5.1 and 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 (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 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

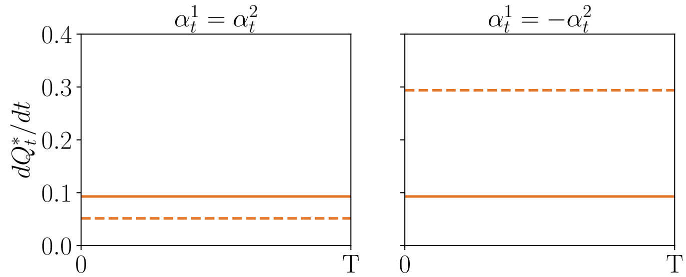


FIGURE 5. Comparison of optimal trading speeds at time $t \in (0,T)$ without cross impact (solid lines) and with cross impact (dashed lines), with parameters estimated for assets with correlation 0.6 in Section 7. In the left panel, the (constant) alpha signals for both assets are the same ($\boldsymbol{\alpha}_t = (1,1)$bps), so less trading is possible with the same optimal impact state. In the right panel, the signal of the alpha signals are opposite ($\boldsymbol{\alpha}_t = (1,-1)$bps), so that cross impact increases the optimal trading volumes.

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:

$$\frac{d\boldsymbol{Q}_t^*}{dt} = \frac{1}{(1+c)^{1/c}} (\boldsymbol{L}^T)^{-1} \boldsymbol{\zeta} h^{-1}(\boldsymbol{L}^{-1}\boldsymbol{\alpha}_t), \tag{6.2}$$

where the inverse $h^{-1}(x) = \mathrm{sgn}(x)|x|^{1/c}$ of the impact function is applied componentwise. Figure 5 illustrates the implications of this formula for $d = 2$ assets with parameters estimated from asset pairs with return correlation 0.6 in Section 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.

6.2. **The Bivariate Case.** When the risk-neutral problem (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 (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 6.2. *As the alpha signal does not affect the concavity of the goal functional, we focus on the impact terms in* (6.1)*. For square-root impact* $h(x) = \mathrm{sgn}(x)|x|^{1/2}$ *and* $\zeta_{11} = \zeta_{22} = 1$, $\zeta_{12} = 0.1$, $\zeta_{21} = 0.9$*, the constraint* (5.7) *is satisfied so that price manipulation is not possible with the symmetric strategies from Lemma 5.1.*

The left panel of Figure 6 plots the integrand $-h(\boldsymbol{J}_t)^\top \boldsymbol{\zeta} \boldsymbol{J}_t$ *of the goal functional* (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 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*

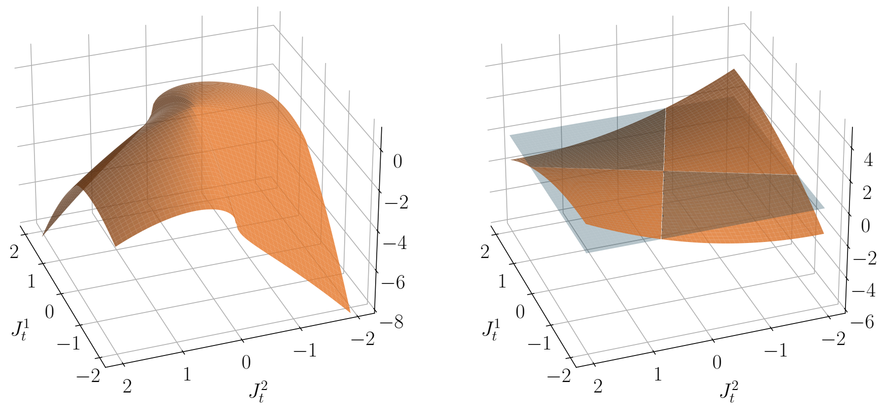


FIGURE 6. Panel (A): the integrand $-h(\boldsymbol{J}_t)^\top \boldsymbol{\zeta} \boldsymbol{J}_t$ of the goal functional (6.1) plotted against the liquidity factors J_t^1, J_t^2. Panel (B): $(J_t^1, J_t^2)\boldsymbol{H}(J_t^1, J_t^2)^\top$ for the Hessian matrix $\boldsymbol{H}$ of the integrand at the point $(J_t^1, J_t^2) = (15, 3)$, plotted against the liquidity factors. Model parameters are chosen as in Example 6.2.

takes some positive values, so the integrand is not a globally concave function of the impact states.

However, for two risky assets, it is easy to check that when the price impact function is of power form ($h(x) = \mathrm{sgn}(x)|x|^c$, $c \in (0, 1]$) then the constraint (5.5) from Lemma 5.1 is exactly what is needed to guarantee that the goal function (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 C.1), the first order conditions that any maximum (J_t^1, J_t^2) must satisfy can be reduced to a single autonomous equation for the ratio J^2/J_t^1:

$$0 = \phi_t + \mathrm{sgn}(\phi_t)|\phi_t|^c k_t^1 + \mathrm{sgn}(\phi_t)|\phi_t|^{c-1} k_t^2 + k_t^3, \tag{6.3}$$

where

$$k_t^1 = \frac{1}{c} - \frac{1+c}{c}\frac{\bar{\alpha}_t^1}{\bar{\alpha}_t^2}\frac{\zeta_{bb}}{\zeta_{ab}}, \quad k_t^2 = \frac{\bar{\alpha}_t^1}{\bar{\alpha}_t^2}\frac{\zeta_{ba}}{\zeta_{ab}}, \quad k_t^3 = \frac{\bar{\alpha}_t^1}{c\bar{\alpha}_t^2}\frac{\zeta_{ba}}{\zeta_{ab}} + \frac{\zeta_{aa}}{\zeta_{ab}}\frac{(1+c)}{c}. \tag{6.4}$$

(The product $J_t^1 J_t^2$ and in turn the individual impact states are pinned down by Equation (C.1) in Appendix C.1.) 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.

[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 (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.

6.3. **The General Case.** For more than two risky assets, both existence and uniqueness for the maximization of (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 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 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

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

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 $\bar{\alpha}_t^a \neq 0$, $a = 1\dots,d$, 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 corresponding optimal impact states have the leading-order asymptotics

$$\boldsymbol{I}_t = \sum_{a=1}^{d} \boldsymbol{L}_a \frac{\bar{\alpha}_t^a}{\zeta_{aa}(1+c)}\left[1 - \sum_{b\neq a}^{d} \frac{\mathrm{sgn}(\bar{\alpha}_t^b)\zeta}{c(c+1)}\left(c\left|\frac{\bar{\alpha}_t^b}{\bar{\alpha}_t^a}\right|^{1/c} + \left|\frac{\bar{\alpha}_t^b}{\bar{\alpha}_t^a}\right|\right)\right]^c. \tag{6.5}$$

If all alpha signals in the latent factor space have the same sign, then this implies that smaller price impacts are optimal with cross impact. However, there are also other parameter configurations the optimal impact states can be increased with cross impact.

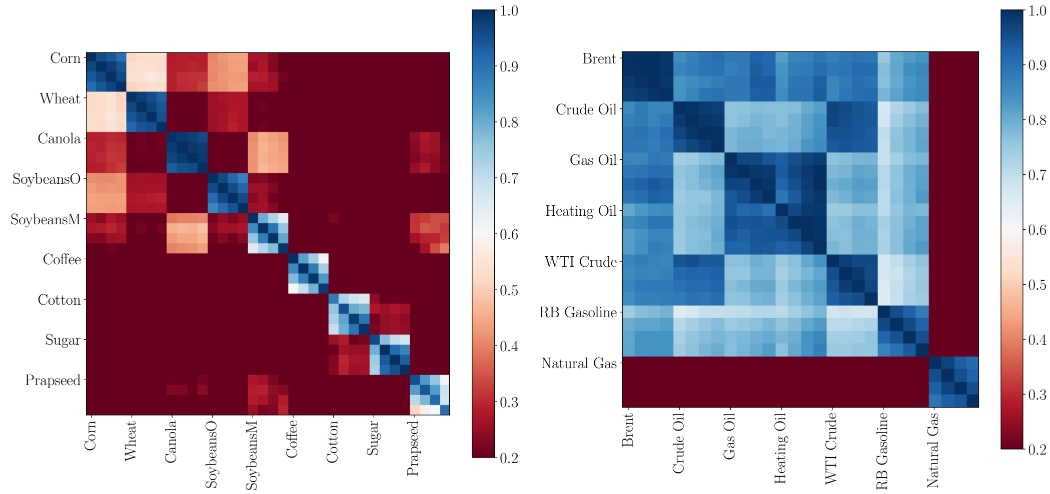
7. Empirical Analysis

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 5 play a key role. Indeed, by narrowing down the parameter space for sensible models, these increase the robustness of the empirical calibration. Using proprietary meta-order data, this allows us to reliably identify the concave structure of cross impact as well as its decay patterns.

7.1. **Data.** In this paper, we use CFM's proprietary meta-order dataset, cf. Hey et al. (2023) for more details. Additionally, we use public data to determine the mid prices at the start and end of each meta order and to estimate the volatilities, correlations, and average daily traded volumes of all asset pairs.

Figure 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 intra-product correlation.

As a complement, the right panel of Figure 7 plots the corresponding correlations for energy contracts. These display much larger intra-product correlations, since they mostly depend on the same underlying resources.



(A) Agricultural Products

(B) Energy Products

FIGURE 7. Return correlations between different futures contracts.

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.

7.2. **Fitting Methodology.** Price returns of pairs of assets are fitted against the cross impact model (5.9). To obtain a system of decoupled equations as in 6.1, we assume that $\boldsymbol{\theta}^{-1}\boldsymbol{\zeta} = \beta \cdot \boldsymbol{Id}_2$ is a diagonal matrix with a single impact decay parameter β across all assets. Then the following impact formula emerges:

$$\boldsymbol{I}_t = \sum_{a=1}^{2} \boldsymbol{L}_a h(\theta_{aa}^{-1}) h(J_t^a), \quad \text{where } J_t^a = \int_0^t e^{-\beta(t-s)} d(\boldsymbol{L}^T \boldsymbol{Q}_s)^a. \tag{7.1}$$

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.[10]

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 (7.1) for a grid

[10]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.

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 $R^2(c, \beta)$.[11]

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

$$w_S^i(c, \beta, \rho) = \vec{\nabla} I^i(c, \beta, \rho)_i^2 ||\vec{\nabla} I^i(c, \beta, \rho)||^{-2}, \tag{7.2}$$

where $\vec{\nabla} = (\partial_{dQ_1}, \partial_{dQ_2})^\top$ 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 $d\boldsymbol{Q}$ of both assets. It thereby captures the sensitivity of the price impact to local changes in trading activity.[12] These individual sensitivities are in turn normalized by the aggregate sensitivity captured by the norm of the whole gradient.

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.

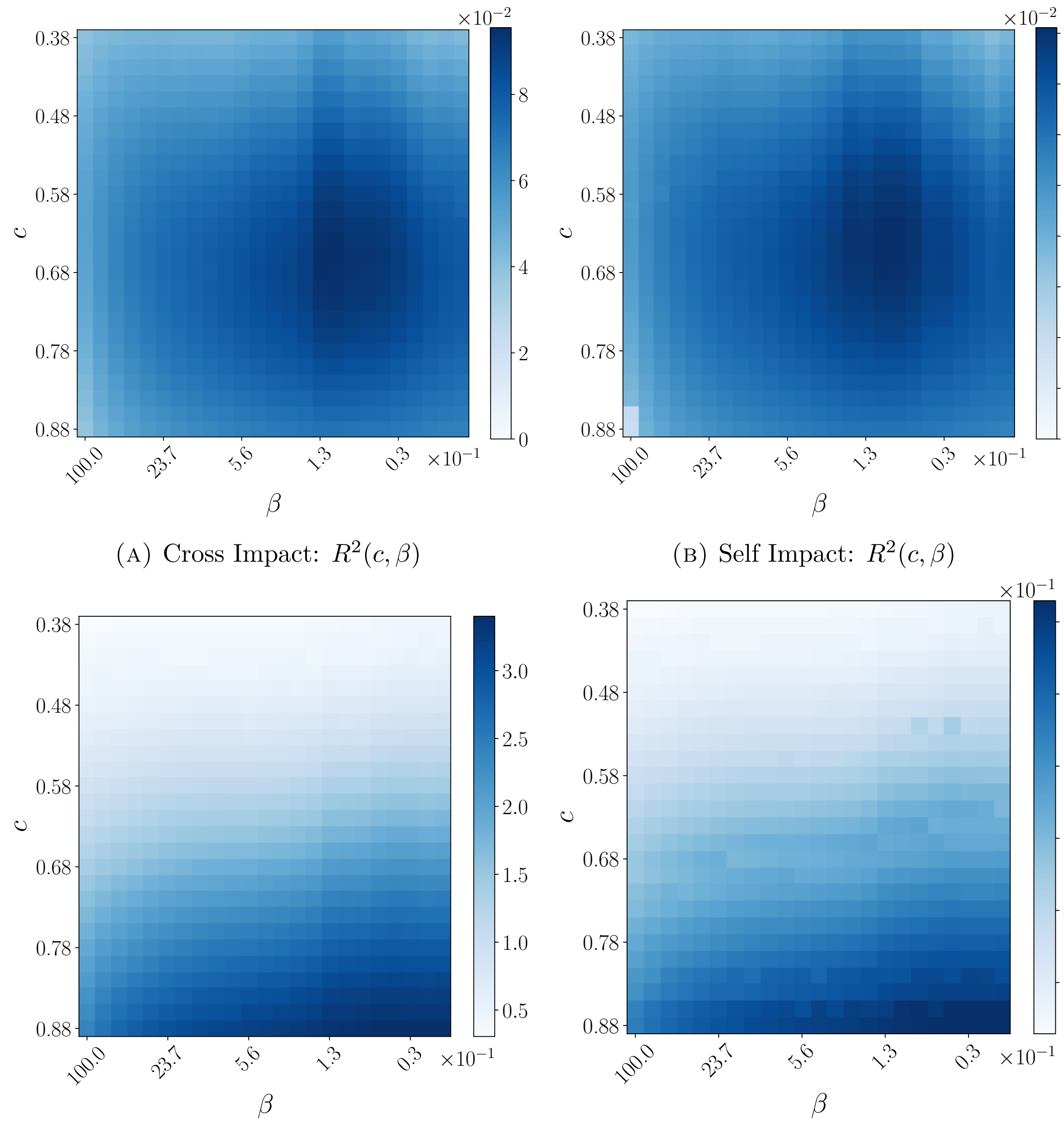
To illustrate this, Figure 8a shows the R^2 of the fitted cross impact model as a function of impact concavity c and impact decay β for product pairs with an average return correlation of $\rho = 0.95$. Figure 8b presents the R^2 values considering only self impact, where the matrix $\boldsymbol{L}$ 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 (7.1) are plotted in Figures 8c and 8d, respectively.[13]

Figure 9 extends the analysis described in Figure 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 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.

[11] Due to the normalization of trading volumes, the regression coefficients are of order one. This simplifies the fitting procedure and reduces the sensitivity to scaling issues.

[12] Unlike for impact models, the sensitivities (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 9 only changes slightly if the evaluation point is changed from anti-aligned to aligned trades.

[13] These prefactors are not directly comparable to the single asset version of the model studied in Hey et al. (2023), 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 8. Calibration results of the cross impact model in 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.

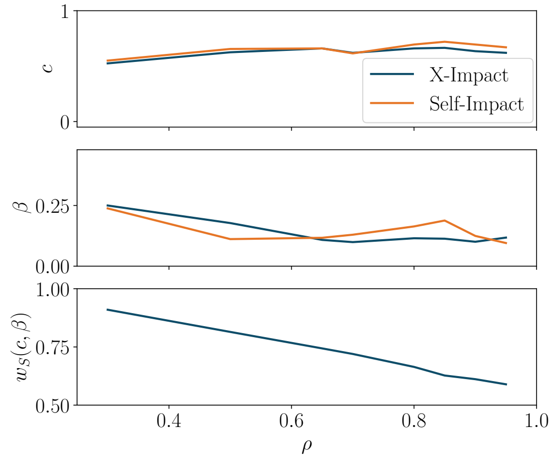


FIGURE 9. The point estimates (c, β) and the corresponding self impact weight $w_S(c, \beta)$ across ρ of the cross impact model (blue) and the self impact model (orange). The point estimates remain roughly constant and the self impact weight is inverse proportional to return correlations.

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 meta-orders 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 meta orders and its gradual decay. An important direction 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 A. Proof of Lemma 5.1

To construct a roundtrip trade, we need

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

Plugging in symmetric strategy (5.4) for asset a, we obtain

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

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 (5.1) of the strategy (5.4) can be computed directly as

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

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 $h(x) = \mathrm{sign}(x)|x|^c$, then to guarantee nonnegative trading costs we need

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

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 (5.5) from Lemma 5.1.

Appendix B. Proof of Lemma 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 (5.1) of the strategy (5.8) are given by

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

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

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

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

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

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

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

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, (B.1) in turn simplifies to

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

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

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

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 (B.1) vanishes. Whence, to avoid price manipulation, we need

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

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.

Appendix C. Proof Section 6

C.1. **Bivariate Solution.** 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

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

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

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 (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).

References

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.

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.

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.

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, 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.

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.

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.

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.

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

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

N. Hey, I. Mastromatteo, J. Muhle-Karbe, and K. Webster. Trading with concave price impact and impact decay - theory and evidence. *Preprint*, 2023.

U. Horst and X. Xia. Multi-dimensional optimal trade execution under stochastic resilience. *Finance and Stochastics*, 23(4):889–923, 2019.

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

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.

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.

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.

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.

G. Tsoukalas, J. Wang, and K. Giesecke. Dynamic portfolio execution. *Management Science*, 65(5):2015–2040, 2019.