Dynamic Portfolio Execution

Gerry Tsoukalas, Jiang Wang, Kay Giesecke*

June 25, 2014

Abstract

We analyze the optimal execution problem of a portfolio manager trading multiple assets. In addition to the liquidity and risk of each individual asset, we consider cross-asset interactions in these two dimensions, which substantially enriches the nature of the problem. Focusing on the market microstructure, we develop a tractable order book model to capture liquidity supply/demand dynamics in a multi-asset setting, which allows us to formulate and solve the optimal portfolio execution problem. We find that cross-asset risk and liquidity considerations are of critical importance in constructing the optimal execution policy. We show that even when the goal is to trade a single asset, its optimal execution may involve transitory trades in other assets. In general, optimally managing the risk of the portfolio during the execution process affects the time synchronization of trading in different assets. Moreover, links in the liquidity across assets lead to complex patterns in the optimal execution policy. In particular, we highlight cases where aggregate costs can be reduced by temporarily overshooting one’s target portfolio.

1 Introduction

This paper formulates and solves the optimal execution problem of a portfolio manager trading multiple assets with correlated risks and cross price impact. The execution process, even for a single asset, exhibits several main challenges: Generally, the available at-the-money liquidity is finite and scarce and the act of trading can influence current and future prices. For instance, a large buy order can push prices higher, making subsequent purchases more expensive. Similarly, a sell order can push prices lower, implying that subsequent sales generate less revenue. The connection between trading and price is known as price impact and its consequence on investment returns can be substantial.1 The desire to minimize the overall price impact prompts the manager to split larger orders into smaller ones and execute them over time, in order to

---

*Tsoukalas (gtsouk@wharton.upenn.edu) is from the Wharton School, University of Pennsylvania. Wang (wangj@mit.edu) is from MIT Sloan School of Management, CAFR and NBER. Giesecke (giesecke@stanford.edu) is from Stanford University, MS&E. We are grateful to seminar participants from the 2011 Annual INFORMS Meeting in Charlotte, the Management Science seminar at Rutgers University, the Operations Management seminar at the Wharton school, University of Pennsylvania, the 2012 Annual MSOM meeting at Columbia University, the IROM seminar at the University of Texas at Austin McCombs School of Business, Morgan Stanley, and AQR Capital Management, for useful comments. Tsoukalas and Giesecke are grateful to Jeff Blokker and the Mericos Foundation for a grant that supported this work.

1Perold (1988), for example, shows that execution costs can often erase true returns, leading to a significant “implementation shortfall.”
source more liquidity. However, trading over longer periods leads to more price uncertainty, increasing risk from the gap between remaining and targeted position. These considerations jointly influence the optimal execution strategy.

The execution of a portfolio generates two additional challenges: First, how to balance liquidity considerations with risks from multiple assets, which are correlated? In particular, reducing costs may require trading assets with different liquidity characteristics at different paces, while reducing risk may require more synchronized trading across assets. Second, how to manage cross-asset liquidity? To the extent that liquidity can be connected across assets, properly coordinating trades can help improve execution.

Controlling price impact is a challenging problem because it requires modeling how markets will react to one’s discrete actions. In practice, this requires a significant investment in information technology and human capital, which can be prohibitive. Therefore, many firms choose to outsource their execution needs or use black-box algorithms from specialized third parties, such as banks with sophisticated electronic trading desks. Moreover, this execution services industry has been growing rapidly over the past decade. Not surprisingly, there is a vast literature studying optimal execution. Most of the existing work focuses on a specific type of execution objective, namely, the problem of optimal liquidation for a single risky asset.

One strand of literature seeks to develop functional forms of price impact, grounded in empirical observations, such as Bertsimas & Lo (1998) and Almgren & Chriss (2000). The other focuses on the market microstructure foundations of price-impact. Recently, pro-technology regulations have continued to fuel the wide-spread adoption of electronic communication networks driven by limit order books. The order books aggregate and publish the inventory of available orders submitted by all market participants. In other words, they display the instantaneous supply/demand of liquidity available in the market. Consequently, many recent papers focus on this feature. In particular, Obizhaeva & Wang (2013) propose a market microstructure framework in which price impact can be understood as a consequence of fluctuations in the supply and demand of liquidity. One advantage of this approach is that the optimal strategies obtained are robust to different order book profiles. This literature highlights the fact that supply/demand dynamics are crucial.

The key question we seek to address in this paper is how managers can maximize their expected wealth from execution, or more generally their expected utility, when trading portfolios composed of dynamically interacting assets. As much interest as the single-asset case has generated, the multi-asset problem has been less studied, perhaps because “the portfolio setting clearly is considerably more complex than the single-stock case” (Bertsimas, Hummel & Lo (1999), page 2). Our motivation to pursue the multi-asset problem is based on the following observation: Even when the execution object is about a single asset, in the general

---

2Bertsimas & Lo (1998), Almgren & Chriss (2000), and Obizhaeva & Wang (2013) demonstrate the benefit from splitting orders over time under different assumptions regarding the impact of trading on current and future prices and liquidity.


4See also Alfonsi et al. (2008), Alfonsi et al. (2010), Bayraktar & Ludkovski (2011), Chen, Kou & Wang (2013), Cont, Stoikov & Talreja (2010), Obizhaeva & Wang (2013), Maglaras & Moallemi (2011), and Predoiu, Shaikhet & Shreve (2011). In other related work, Rosu (2009) develops a full equilibrium game theoretic framework and characterizes several important empirically verifiable results based on a model of a limit order market for one asset. Moallemi, Park & Van Roy (2012) develop an insightful equilibrium model of a trader facing an uninformed arbitrageur and show that optimal execution strategies can differ significantly when strategic agents are present in the market.
multi-asset setting, it is optimal to consider transitory trades in other assets. There are at least two reasons. First, other assets provide natural opportunities for risk reduction through diversification/hedging. Second, price-impact across assets may provide additional benefits in reducing execution costs by trading in other assets. Thus, to limit trading to the target asset is in general suboptimal. Of course, when the execution involves a portfolio, we would need to consider both effects from correlation in risk and supply/demand evolution, respectively.

To tackle the problem, we develop a multi-asset order book model with correlated risks and coupled supply/demand dynamics. Here, an order executed in one direction (buy or sell) will affect both the currently available inventory of limit orders and also future incoming orders on either side. This is in line with the empirical results in Biais, Hillion & Spatt (1995) who find that “downward (upward) shifts in both bid and ask quotes occur after large sales (purchases).” Therefore, there is a priori no reason to rule out the possibility that double-sided (buy and sell) strategies may be optimal even if the original objective is unidirectional (e.g. in the standard liquidation problem). However, allowing for arbitrary dynamics leads to some serious modeling difficulties. In particular, there is no reason to assume that the supply and demand sides of the order books are identical, implying that the manager’s buy and sell orders need to be treated separately. To this end, we need to introduce inequality constraints on the control variables, which renders the optimization computationally challenging.

To solve the problem, we first show that in our setting the optimal policy is path-independent. This allows us to find an equivalent static formulation of the original dynamic program (DP) under some mild restrictions on the price processes (namely, that they are random walks). We then provide conditions allowing us to restate the problem as a quadratic program (QP). The QP approach efficiently handles inequality constraints and time-dependent parameters, and guarantees global optimums in polynomial time.

Our model implies that managers can utilize cross-asset interactions to significantly reduce risk-adjusted execution costs. The resulting optimal policies involve advanced strategies, such as conducting a series of buy and sell trades in multiple assets. In other words, we find that managers can benefit by over-trading during the execution phase. This result may a priori seem counter-intuitive. Indeed, we demonstrate that one can lower risk-adjusted trading costs by trading more. We show that this is the case because a unique trade-off arises in the multi-asset setting. While consuming greater liquidity generally leads to higher charges, one can also take advantage of asset correlation and cross-impact to reduce risk via offsetting trades.

We show that multi-asset strategies turn out to be optimal for simple unidirectional execution objectives. Even in the trivial case where the objective is to either buy or sell units in a single asset, we find that the manager can benefit by simultaneously trading back and forth in other correlated securities. Previous work has focused on modeling the available buy-side or sell-side liquidity independently of each other. Our

---

5The existence of cross asset price-impact effects has been empirically documented and theoretically justified. It can simply result from dealers’ attempts to manage their inventory fluctuations, see for example Chordia & Subrahmanyam (2004) and Andrade, Chang & Seasholes (2008). Kyle & Xiong (2001) show that correlated liquidity shocks due to financial constraints can lead to cross-liquidity effects. King & Wadhwani (1990) argue that in the presence of information asymmetry among investors, correlated information shocks can lead to cross-asset liquidity effects among fundamentally related assets. Fleming, Kirby & Ostdiek (1998) show that portfolio rebalancing trades from privately informed investors can lead to cross-impact in the presence of risk aversion, even between assets that are fundamentally uncorrelated. Pasquariello & Vega (2012) develop a stylized model and provide empirical evidence suggesting that cross-impact may stem from the strategic trading activity of sophisticated speculators who are trying to mask their informational advantage. Lastly, evidence of comovement stemming from sentiment-based views has been studied in Barberis, Shleifer & Wurgler (2005).
results suggest that these two cannot generally be decoupled when accounting for cross-asset interactions. Furthermore, the associated strategies are often non-trivial. For instance, when liquidating (constructing) a portfolio, one can reduce execution risk by simultaneously selling (purchasing) shares in positively correlated assets. Our model explains why this type of trade provides an effective hedge against subsequent price volatility.

Extending the analysis to portfolios with heterogeneous liquidity across assets (e.g., portfolios composed of small-cap and large-cap stocks, ETFs and underlying basket securities, stocks and OTM options, etc.), we find that the presence of illiquid assets in the portfolio drastically affects the optimal policies of the liquid assets. In particular, it can be optimal to temporarily overshoot targeted positions in some of the most liquid assets in order to improve execution efficiency at the portfolio level. However, the different trading strategies associated with each asset type could leave managers over-exposed to illiquidity at certain times during the execution phase. This synchronization risk can be addressed by introducing constraints on the asset weights that synchronize the portfolio trades, while maintaining efficient execution. The constrained optimal policies obtained combine aspects of the optimal standalone policies of both liquid and illiquid assets.

Our analysis has implications for other important problems in portfolio management. The QP formulation can easily be integrated into existing portfolio optimization problems that treat transaction costs as a central theme. For example, the portfolio selection problem with transaction costs is one of the most central problems in portfolio management. Our model provides an understanding of the origin of these costs and of their propagation dynamics in the portfolio setting. The insights we develop can thus allow portfolio managers to better assess the applicability of some common cost assumptions in this strand of literature (such as assuming cost convexity and diagonal impact matrices, and prohibiting counter-directional trading).

There is prior work on the multi-asset liquidation problem. Bertsimas et al. (1999) develop an approximation algorithm for a risk-neutral agent, which solves the multi-asset portfolio problem while efficiently handling inequality constraints. Almgren & Chriss (2000) briefly discuss the portfolio problem with a mean-variance objective in their appendix and obtain a solution for the simplified case without cross-impact. Engle & Ferstenberg (2007) solve a joint composition and execution mean-variance problem with no cross-impact using the model from Almgren & Chriss (2000). They find that cross-asset trading can become optimal even without cross-impact effects. Brown, Carlin & Lobo (2010) treat a multi-asset 2-period liquidation problem with distress risk, focusing on the trade-offs between liquid and illiquid assets. In contrast to these papers, we analyze the more general multi-objective execution problem focusing on the market microstructure origins of price impact. This allows us to characterize the optimal policies as a function of intuitive order book parameters, such as inventory levels, replenishment rates and bid-ask transaction costs. These parameters could be calibrated to tick by tick high-frequency data.

The remainder of the paper is structured as follows: Section 2 details the multi-asset liquidity model.

---

6See Brown & Smith (2011) and references therein for recent advances.
7A more concrete example of how price-impact models can be integrated in a broader portfolio selection problem can be found in Iancu & Trichakis (2012), which focuses on the multi-account portfolio optimization problem. A discussion regarding the applicability of advanced cross-asset strategies and how they relate to agency trading and best execution constraints can also be found in the same paper.
8Disentangling cross-impact from correlation for individual securities is a challenging statistical problem which is beyond the scope of this paper. Empirical estimation of cross-impact is an active area of research for high-frequency trading firms and could also be an interesting direction for future academic research.
Section 3 formulates and solves the manager’s dynamic optimization problem over his terminal wealth resulting from execution. Section 4 focuses on numerical applications and economic insights. Section 5 treats mixed liquidity portfolios. Section 6 concludes. The appendix contains proofs and some additional results.

2 The Liquidity Model

In this section, we develop a model specifying how the manager’s trades affect the supply/demand and price processes of all assets. We start with the investment space and admissible trading strategies in Section 2.1. Each buy or sell order submitted to the exchange will be executed against the available inventory in the limit order books. Section 2.2 explains the instantaneous distribution of orders in the order book. Section 2.3 describes the replenishment process: Following each executed trade, new limit orders arrive, reverting prices and collapsing the bid-ask spread towards a steady state, which we define. This liquidity mean-reversion property provides an incentive for the manager to split his original order over time. Doing so, he can take advantage of more favorable limit orders arriving at future periods. However, delaying trading also introduces more price uncertainty. We formulate and eventually solve this essential trade-off between risk and liquidity.

2.1 Investment Space and Admissible Strategies

We adopt the following notation convention: vectors are in lower case bold, matrices in upper-case bold, and scalars in standard font.

Time \( t \) is discrete, with \( N \) equally spaced intervals per unit of time. The manager has a finite execution window, \([0, T]\), where the horizon \( t = T \), is normalized to 1 without loss of generality. Thus, there are \( N + 1 \), equally spaced, discrete trading times, indexed by \( n \in \{0, \ldots, N\} = I_N \), with period length \( \tau = 1/N \). Uncertainty is modeled by a probability space \((\Omega, \mathcal{F}, P)\). A filtration \((\mathcal{F}_n)_{n \in I_N}\) models the flow of information. The stochastic process generating the information flow is specified in Assumption 2.

We consider a portfolio of \( M \) assets indexed by \( i \in \{1, \ldots, M\} = I_M \). Let \( K = (N + 1) \times M \) be the dimension of the problem. Irrespective of the manager’s objective, we assume that he has the option of purchasing or selling/shorting units in any of the assets during any of the discrete times, as long as he satisfies his boundary conditions at the horizon \( N \). Let \( x_{i,n}^+ \geq 0 \) and \( x_{i,n}^- \geq 0 \) be his order sizes for buy and sell orders respectively, in asset \( i \) at time \( n \). These will constitute the control variables over the trading horizon. We also define the following corresponding buy and sell vectors stacking orders by asset and/or trading times:

\[
\text{asset } i: \mathbf{x}_i^\pm = \begin{bmatrix} x_{i,0}^\pm & \cdots & x_{i,N}^\pm \end{bmatrix}, \text{time } n: \mathbf{x}_{i,n}^\pm = \begin{bmatrix} x_{i,n}^\pm \end{bmatrix}, \text{assets and time: } \mathbf{x}^\pm = \begin{bmatrix} x_1^\pm & \cdots & x_M^\pm \end{bmatrix}, \text{aggregate: } \mathbf{x} = \begin{bmatrix} x^+ \\ x^- \end{bmatrix}.
\]

9 These will be useful in formulating the optimization problem later on. We will be adopting this vector notation convention for each variable subsequently introduced, unless otherwise specified.
Next, we define part of the execution objective by formulating the boundary conditions. Let $z_{i:n}$ represent the net amount of shares left to be purchased (or sold, if negative) in asset $i$ at time $n$, before the incoming order at time $n$. Following the vector conventions defined above, we have by definition:

$$z_{i:n} = z_{i:0} - \sum_{k=1}^{n} (x_{i:k-1}^+ - x_{i:k-1}^-) = z_{i:0} - \sum_{k=1}^{n} \Delta_M' x_{i:k-1},$$

where $z_{i:n}$ has dimension $M \times 1$, $x_{i:n} = [x_{i:n}^+; x_{i:n}^-]$ has dimension $2M \times 1$ and $\Delta_M$ is a simple matrix operator with parameter $M$ and dimension $2M \times M$, such that $\Delta_M' x_{i:n} = x_{i:n}^+ - x_{i:n}^-$. Denoting $I_M$ the identity matrix of size $M$, we have $\Delta_M = [I_M; -I_M]$. The boundary conditions which must hold $\forall i$ are $1'_{N+1}(x_{i}^+ - x_{i}^-) = z_{i:0}$, where $1'_{N+1} = [1 \ldots 1]$ is a $1 \times (N + 1)$ vector of ones. Using these notations, the boundary condition of the portfolio can be expressed as

$$1' \Delta_K' x = z_{:0},$$

(1)

where $\Delta_K = [I_K; -I_K]$ has dimension $2K \times K$ and $1' = \text{diag}(1'_{N+1}, \ldots, 1'_{N+1})$ has dimension $M \times K$.

We illustrate with an example: Consider a fund turning over a portfolio with initial positions $\{100, 0\}$ units in assets 1 and 2, and desired exposures $\{50, 50\}$ units by time $N$. It would be required to purchase $z_{1:0} = -50$ units in asset 1 (i.e. sell 50 units) and $z_{2:0} = 50$ units in asset 2 by time $N$. Therefore, $z_{:0} = [-50; 50]$ in this case.

The manager’s trades must be adapted to the filtration $(\mathcal{F}_n)_{n \in I_N}$. The set $S$ of admissible trading strategies is specified in the definition below.

**Definition 1 (Admissible Execution Strategies)** The set $S$ of all admissible trading strategies takes the form

$$S = \{ x \in \mathbb{R}^{2K} \mid (\mathcal{F}_n)_{n \in I_N} \text{ adapted}; \ 1' \Delta_K' x = z_{:0} \}.$$  

The set of strategies in Definition 1 is broad in the sense that no restrictions (such as shorting or budget constraints) are imposed during the trading window, as long as the boundary constraints are satisfied by time $N$.

Having established the preliminary notations, the next step is to model the manager’s market impact. In other words, we need to describe how his actions affect asset prices over time. The section below is dedicated to developing an adequate liquidity model, which will allow us to formulate the manager’s dynamic optimization problem.

### 2.2 Order Book

In a limit order book market, the supply/demand of each asset is described by the order book. The basic building blocks of limit order markets consist of the three order types: **Limit orders** are placed by market participants who commit their intent to buy (bids) or sell (asks) a certain volume at a specified worst-case (or limit) price. They represent the current visible and available inventory of orders in the market. **Market orders** are immediate orders placed by market participants who want to buy or sell a specific size at the
current best prices available in the market. They are executed against existing supply or demand in the limit order book. Cancelation orders remove unfilled orders from the book.

To preserve tractability, we follow the existing literature in assuming that the manager is a liquidity taker, i.e., he submits market orders that are executed against available inventory in the book on a single exchange.\(^\text{10}\) Although prices and quantities are discrete, we adopt a continuous model of the order book which is entirely described by its density functions: \(q^a_{i,n}(p)\) for the ask side and \(q^b_{i,n}(p)\) for the bid side. The density functions map available units \((q)\) to limit order prices \((p)\) and thus describe the distribution of available inventory in the order book over all price levels, at any given point in time.

To illustrate, Figure 1 displays a partial snap-shot of (a) the oil futures limit order book as of November 8, 2011 at 11:10am and as a comparison, an equivalent continuous-model (b) and a simplified continuous model (c). The continuous model along with a simplifying assumption on the order book density functions (see Assumption 1) will allow us to keep the problem tractable and focused on the multi-asset aspect of the model.

Following Huberman & Stanzl (2004) and Obizhaeva & Wang (2013), we assume that all assets in the portfolio have block-shaped order books with infinite depth and time-invariant steady-state densities:

**Assumption 1 (Order Book Shapes)** Letting \(q^a_i\), \(q^b_i\) be constants, and denoting by \(a_{i,n}\), \(b_{i,n}\) the best available ask and bid prices in each order book at \(n\), right before the trade arrives at \(n\),\(^\text{11}\) we have

\[
q^a_{i,n}(p) = q^a_i \mathbf{1}_{\{p \geq a_{i,n}\}} \quad \text{and} \quad q^b_{i,n}(p) = q^b_i \mathbf{1}_{\{p \leq b_{i,n}\}}, \quad i \in I_M. \tag{3}
\]

\(^{10}\)See Moallemi & Saglam (2013a) for a study regarding the optimal placement of limits orders. See Maglaras, Moallemi & Zheng (2012) for a study on order placement in fragmented markets.

\(^{11}\)The best available ask price, \(a_{i,n}\), is the lowest price at which a market buy order could (partially or fully) be instantaneously executed at time \(n\). Similarly, the best available limit bid price \(b_{i,n}\) is the highest price at which a market sell order could be executed.
Figure 1(c) provides an illustration of this assumption.\(^{12}\)

In addition to the shape of order book density functions, we also need to specify the location of \(a_{i,n}\) and \(b_{i,n}\) and their evolution over time. Two components are driving each asset’s best bid and ask prices: its fundamental value and the price impact of trading. We will focus on the first component and return the second later.

In absence of trading, the best bid and ask prices should be determined by the assets’ fundamental values. We will assume these are given by a vector of random walks \(u_{i,n}\):

**Assumption 2 (Random-Walk Fundamental Values)** Let \(\epsilon_{i,n} \sim N(0, \tau \Sigma_\epsilon)\) be a vector of normal random variables with covariance \(\tau \Sigma_\epsilon\), such that \(\forall n \in \mathbb{N}\), \(\mathbb{E}[\epsilon_{i,n}\epsilon_{j,n-1}] = 0\) and \(\mathbb{E}[\epsilon_{i,n}\epsilon_{j,n}] = \tau \sigma_{ij}\). We have

\[
u_{i,n} = u_{i,n-1} + \epsilon_{i,n}, \quad u_{0} > 0,\]

with \(\mathbb{E}[u_{i,n}|\mathcal{F}_{n-1}] = u_{i,n-1}.\)\(^{13}\)

The possibility of relaxing Assumption 2 is discussed in Section 3.3. Thus, we can express the best bid and ask prices as follows:

\[
a_{i,n} = u_{i,n} + \frac{1}{2} s_i, \quad b_{i,n} = u_{i,n} - \frac{1}{2} s_i, \quad \forall i,n.\]

Here, \(s_i\) gives the bid-ask spread of asset \(i\) in steady-state.

### 2.3 Order Book Dynamics

Next, we need to describe the evolution of \(a_{i,n}\) and \(b_{i,n}\) when the manager trades in the market, which impacts the supply/demand dynamics of the order books. For this purpose we extend the single-asset, one-sided, order book model in Obizhaeva & Wang (2013) in two directions. First, we develop a single-asset, two-sided, order book model with coupled bid and ask sides (i.e., a trade in one direction will affect both sides of the order book) and bid-ask transaction costs. Second, we extend to allow multiple assets. We start with the two-asset case and show that interactions between assets justify the need for a dynamic two-sided order book model. We then provide the general multi-asset case (\(M\) assets).

**A. Single Asset**

We break down the price impact process into two phases: In phase 1, the manager submits an order which is executed against available inventory of orders, creating an immediate change in the limit order book. The order book updates itself and displaces the asset’s mid-price accordingly, creating both a temporary price impact (TPI) and a permanent price impact (PPI). In phase 2, new limit orders arrive in the books, gradually

\(^{12}\)We refer to Alfonsi et al. (2010) for a discussion about general types of density functions and to Predoiu et al. (2011) for an equivalence between discrete and continuous models. A queuing-based approach can be found in Cont et al. (2010).

\(^{13}\)While the random walk assumption implies a non-zero probability of negative prices, it is not a concern in our framework given the short-term horizon of optimal execution problems in practice. As such, this assumption is commonly used in the price impact literature.
absorbing the temporary price impact and collapsing the bid-ask spread towards its new steady state. We then describe how these dynamics could be affected in a two-sided model.

Consider a market order arriving at time \( n \) to buy \( x = x_{i,n}^+ > 0 \) units in an arbitrary asset \( i \).\(^{14}\) Figure 2 shows possible dynamics that \( i \) can face after getting hit by the order.\(^{15}\) At time \( n - 1 \), we illustrate \( i \) in its steady state (see Figure 2(a)). At the next period in time \( n \) (see Figure 2(b)), the incoming order is executed against available inventory on the ask side of \( i \)’s order book, starting from the best available price and rolling up \( i \)’s supply curve towards less-favorable prices. This instantaneously drives \( i \)’s best ask price from \( a_{i,n} \) to \( a_{i,n}^* \), where the superscript denotes the moment immediately following an executed order. This results in a displacement of \( a_{i,n}^*(x) - a_{i,n} \). Given a density shape \( q_{i,a}(p) \) the amount of units executed over a small increment in price is simply \( dx = q_{i,a}(p) dp \). An executed buy order of size \( x \) therefore shifts the best ask price according to:

\[
\int_{a_{i,n}}^{a_{i,n}^*(x)} q_{i,a}(p) dp = x.
\]

Combining the above expression with Assumption 1 we have the following Lemma:

**Lemma 1 (Impact of Trading on Order Book)** An incoming market order to buy (sell) \( x_{i,n}^+ \) \((x_{i,n}^-)\) shares at time \( n \) will instantaneously displace the ask (bid) price of asset \( i \) according to

\[
a_{i,n}^* = a_{i,n} + \frac{x_{i,n}^+}{q_{i,a}} \quad \text{and} \quad b_{i,n}^* = b_{i,n} - \frac{x_{i,n}^-}{q_{i,b}}.
\]

\(^{14}\)We focus on a single buy order, implying \( x_{i,n}^- = 0 \), but the results are directly applicable to sell orders as well.

\(^{15}\)We do not illustrate the impact of the random walk here to keep the figures clear. In order words, we are holding \( u_{i,n} \) constant.
Clearly, the corresponding displacements in the best bid/ask limit order prices are linear in the order size: 

\[ a_{i,n}^* - a_{i,n} = \frac{x_{i,n}}{q_i} \text{ and } b_{i,n}^* - b_{i,n} = -\frac{x_{i,n}}{q_i}. \]

The immediate cost the manager incurs in this phase can then simply be calculated by integrating the price over the total amount of units executed:

\[ \int_0^x a_{i,n}^*(u)du. \]

Next, as shown in Figure 2(c), we assume that the current and future supply/demand will adjust accordingly. In particular, we will assume that trading gives rise to a permanent impact on asset prices, which is proportional to the cumulative trade size.\(^{16}\) In order to capture the permanent price impact, we introduce what we will call the “steady-state” mid-price \(v_{i,n}\) (\(i = 1, \ldots, M\)), before the trade arrives at \(n\), which is given by

\[ v_{i,n} = v_{i,n-1} + \lambda_{ii} \left( x_{i,n-1}^+ - x_{i,n-1}^- \right) + \epsilon_{i,n} + \lambda_{ii} \sum_{k=0}^{n-1} \left( x_{i,k}^+ - x_{i,k}^- \right), \quad (8) \]

where the second term gives the permanent price impact of trades up to and including the previous period \((n - 1)\), and \(\lambda_{ii}\) is the permanent price impact for each unit of trading in asset \(i\) itself. Hence, if the manager doesn’t submit any trades after \(n\), the best ask and bid prices of asset \(i\) will eventually converge to \(v_{i,n+1} + \frac{1}{2}s_i\) and \(v_{i,n+1} - \frac{1}{2}s_i\), respectively. For convenience, we introduce the “steady-state” best ask and bid prices:

**Assumption 3 (Steady-State Prices)** Asset \(i\)’s best ask and bid prices have steady-state levels, before the trade arrives at \(n\), which are given by

\[ a_{i,n}^\infty = v_{i,n} + \frac{1}{2}s_i, \quad b_{i,n}^\infty = v_{i,n} - \frac{1}{2}s_i, \quad (9) \]

where the steady-state mid-price is given by equation (8).

The best available ask and bid prices may generally differ from their steady-state levels.

After the order is executed at \(n\), the replenishment process (phase 2) begins (see Figure 2(d) for an illustration). During this phase, we assume that supply/demand is replenished as new limit orders arrive to refill the order books. Replenishment is spread over time, and the order books might remain in transitory state over an extended period of time. In the absence of any new market orders after \(n\), the incoming limit orders will gradually push the best bid/ask prices towards their new steady states \(a_{i,n+1}^\infty\) and \(b_{i,n+1}^\infty\). The rate at which this happens depends on the dislocation size, the inherent properties of the asset and the behavior of market participants.

We follow Obizhaeva & Wang (2013) in describing the order book replenishment process. For convenience, we define the order book displacement functions to keep track of the difference between the best ask and bid prices and their steady state levels, i.e.:

\[ a_{i,n}^a = a_{i,n} - a_{i,n}^\infty, \quad d_{i,n}^a = b_{i,n}^\infty - b_{i,n}. \quad (10) \]

\(^{16}\)The linearity assumption on the permanent price impact function is consistent with Theorem 1 of Huberman & Stanzl (2004), which provides conditions under which the price impact model does not admit arbitrage and price manipulation strategies.
The order book replenishment process is given as follows:

**Assumption 4 (Order Book Replenishment)** The limit order demand and supply are replenished exponentially, with constant decay parameters $\rho_i^a$ and $\rho_i^b$, for the ask and bid prices, respectively. Specifically, over period $\tau$, the order book displacements are given by

\[
d_{i,n+1}^a = \left( d_{i,n}^a + \left[ \frac{x_{i,n}^+}{q_i^a} - \lambda_{ii}(x_{i,n}^+ - x_{i,n}^-) \right] \right) e^{-\rho_i^a \tau}, \tag{11a}
\]
\[
d_{i,n+1}^b = \left( d_{i,n}^b + \left[ -\frac{x_{i,n}^-}{q_i^b} - \lambda_{ii}(x_{i,n}^+ - x_{i,n}^-) \right] \right) e^{-\rho_i^b \tau}. \tag{11b}
\]

Clearly, as $\rho_i^a$ and $\rho_i^b \to \infty$, the asset is highly liquid; the displacements are null, and the order books are replenished instantaneously after each trade. As $\rho_i^a$ and $\rho_i^b \to 0$, the asset is highly illiquid; no new limit orders arrive, and the displacements are permanent (i.e., they do not decay over time).\textsuperscript{17}

From the order book replenishment process described in equation (11) and the steady-state bid and ask prices in equation (9), the dynamics of the actual best bid and ask prices at any time are simply given by equation (10).

**B. Two Assets**

Adding a second asset to the problem introduces several new features. We need to take into account the correlation between the stochastic processes driving the mid-prices but also the cross-impact that a trade in one asset can have on the supply/demand curves of the other. These two features are distinct. Correlation is exogenous whereas cross-impact is a direct result of the manager’s action. While the former is straightforward, we provide an example of the latter in Figure 3.

Consider a portfolio composed of two assets, and an incoming order to buy $x_{1,n}^+$ shares in the first asset – the second asset being “inactive”. Let $\lambda_{21} > 0$ be the cross-impact parameter of asset 1 on asset 2. We illustrate how the buy order affects the mid-price of the inactive asset via the term $\lambda_{21} x_{1,n}^+$, as shown in Figure 3(b2). Given the resulting price change, the portfolio value could be significantly affected. Furthermore, the cross-impact will have a secondary effect on the supply/demand curves of the inactive asset. As is shown in Figure 3(c2), the change in the second asset’s mid-price defines a new steady state, initiating a response in the bid/ask books. Specifically, new buy orders arrive to replenish demand while existing ask orders are canceled as prices converge towards the new steady states. Thus, if any orders are later submitted in the inactive asset, these would be executed at prices which could diverge from the initial state. This effect is further exacerbated as the number of assets in the portfolio increases, since a trade in one could affect the prices of all others. A numerical study is provided in Section 4.

Analytically, for both assets, $i = 1, 2$, the steady-state mid-prices and best bid/ask prices are still given by equations (8) and (9), with only the following modification required on the steady-state mid-prices to incorporate the effect of cross-asset price impact:

\textsuperscript{17}Assumption 4 could be relaxed with alternative functional form specifications. The exponential form has the advantage of only requiring a single parameter to describe the replenishment process, keeping the problem tractable. Further, this form has been adopted in previous literature and is in line with several empirical findings on the order book replenishment process. See e.g. Biais et al. (1995) for a detailed empirical study.
Assumption 5 (Cross-Asset Price Impact) When there is trading in both assets, the steady-state mid-price remains linear in the trade size and is given by

\[ v_{i,n} = u_{i,n} + \sum_{j=1,2} \lambda_{ij} \sum_{k=1}^{n} (x_{j,k-1}^+ - x_{j,k-1}^-), \quad i = 1, 2. \] (12)

The order book replenishment dynamics for both assets are still given by equation (11) with only a slight modification required to adjust the permanent price impact term for both ask and bid sides:

\[ d_{i,n+1}^{a,b} = \left( d_{i,n}^{a,b} + \left[ \pm \frac{x_{i,n}^\pm}{q_i} \lambda_{ij} \left( x_{j,n}^+ - x_{j,n}^- \right) \right] \right) e^{-\rho_{i,n}^a}, \]

C. Multiple Assets

Once the two-asset case is understood, the generalization to the M-asset case is straightforward. In particular, we can describe the dynamics of assets' best ask and bid prices as follows:

Lemma 2 (Bid/Ask Price Processes) Following Assumptions 1-5 and Lemma 1, the best bid/ask prices available in the order books at time n, are given by

\[ a_{i,n} = u_{i,n} + \frac{1}{2} s_{i,n} + \Lambda (z_{i,0} - z_{i,n}) + d_{i,n}^{a}, \] (13a)

\[ b_{i,n} = u_{i,n} - \frac{1}{2} s_{i,n} + \Lambda (z_{i,0} - z_{i,n}) + d_{i,n}^{b}, \] (13b)

where \( s_{i,n} \) is the steady-state bid-ask spread \( M \times 1 \) vector, \( \Lambda = [\lambda_{ij}], M \times M \) is a matrix of PPI factors and \( z_{i,n} \) and \( d_{i,n} \) are \( M \times 1 \) state vectors which keep track of the order book dynamics. The state vectors depend on the manager's previous orders submitted up to time n.
The vector $z_n$ was defined in Section 2.1 as the amount of shares left to be purchased at time $n$. Recursively, $z_n = z_{n-1} - \Delta'_M x_{n-1}$. The vectors $d^a_n$ and $d^b_n$ keep track of the replenishment process for the ask and bid sides. Focusing first on the ask side, $d^a_n$ can be recursively written as

$$d^a_n = (d^a_{n-1} + \kappa^a x^+_{n-1}) e^{-\rho^a \tau}, \quad (14)$$

where $\kappa^a = 2 Q^a \Delta'_M a - \Lambda \Delta'_M$ is a displacement matrix keeping track of the difference between temporary and permanent impacts, $Q^a = \text{diag}(\frac{1}{2q^a_1}, \ldots, \frac{1}{2q^a_M})$ is a temporary price impact matrix, $\Delta'_M$ is a matrix operator defined by $\Delta'_M x_{n-1} = x^+_{n-1}$, and $e^{-\rho^a \tau} = \text{diag}(e^{-\rho^a_1 \tau}, \ldots, e^{-\rho^a_M \tau})$ is the order book replenishment matrix. Similarly, for the bid side, we have $d^b_n = (d^b_{n-1} + \kappa^b x^-_{n-1}) e^{-\rho^b \tau}$ and we define the aggregate vector as $d_n = [d^a_n; d^b_n]$. 

### 3 Optimal Execution Problem

#### 3.1 Dynamic Programming Formulation

Having detailed the liquidity model in Section 2, the next step is to derive the manager’s execution costs, as a function of his trading strategy. Using Lemma 2, we can calculate the total costs and revenues resulting from an order $x_n$ submitted at time $n$.

**Lemma 3 (Costs and Revenues)** An incoming order to execute $x_n$ shares at time $n$ will have associated total costs ($c_n$) and revenues ($r_n$), given by

$$c_n = x^+_n (a_n + Q^a x^+_n), \quad (15a)$$

$$r_n = x^-_n (b_n - Q^b x^-_n). \quad (15b)$$

Let $\pi_n$ be the manager’s reward function at $n$ which can be written as the difference between his total revenues (from his selling orders) and his total costs (from his purchasing orders). It follows that

$$\pi_n = r_n - c_n = x^-_n (b_n - Q^b x^-_n) - x^+_n (a_n + Q^a x^+_n). \quad (16)$$

Having defined the reward at each time step, we can formulate the manager’s DP. To capture the trade-off between liquidity and risk, we will assume an exponential utility function with risk-aversion coefficient $\alpha$, over the manager’s total terminal wealth $W_N = \sum_{n=0}^N \pi_n$. This choice is motivated by several factors: First, it allows us to focus exclusively on the utility derived from execution, irrespective of the manager’s initial wealth – a well-known property of constant absolute risk aversion (CARA) utility functions. Second, in our framework, the exponential objective is equivalent to a mean-variance objective – a common modeling choice in the existing portfolio management and price impact literature. Lastly, this form leads to a tractable optimization problem which can be solved in polynomial time. Letting $J_n(\cdot)$ be the value function at time $n$, we have

$$J_0 = \max_{0 \leq x \in S} \mathbb{E}_0[-e^{-\alpha W_N}], \quad (17)$$
at time 0, and for any time \( n > 0 \),

\[
J_{n-1}(z_{n-1}, d_{n-1}, W_{n-2}) = \max_{z_{n-1} \geq 0} E_{n-1}[J_n].
\]  

Here, \( E_{n-1} \) denotes the conditional expectation given \( F_{n-1} \). The boundary conditions are \( z_N = x_N^+ - x_N^- \), and \( d_i = 0 \) (i.e., the order books are initially in their steady states). In the Appendix A.2 we show that three state variables suffice to describe the system at each period \( n \). These are: 1) the remaining shares to be traded: \( z_n \), 2) the order book state: \( d_n \), and 3) the previous period’s cumulative wealth: \( W_{n-1} \). We also show that the optimal policy which solves the problem (18) is path-independent, i.e., it does not depend on the filtration \( F_n \). This statement is formalized in Proposition 1.

3.2 Equivalent Quadratic Program

**Proposition 1 (Path-Independence)** The optimal trading policy \( x^* \) which solves the problem (18) is path-independent with respect to \( (F_n) \). The optimal trades at time \( n \), \( x^*_n \), are a deterministic function of the state variables of the problem, \( z_n \) and \( d_n \), and do not depend on \( W_n \).

Proposition 1 allows us to reformulate the problem (18) as a static QP. To this end, we first introduce an equivalent static form for the stochastic wealth function, by reformulating the equations in Section 3.1.

**Lemma 4 (Equivalent Formulation of Wealth)** The manager’s wealth, post execution, \( W_N = \sum_0^N \pi_n \), can be formulated as a quadratic function of the controls given by

\[
W_N = -(x'Dx + \tilde{c}'x).
\]  

The stochastic linear terms are \( \tilde{c} = [\tilde{c}^a, -\tilde{c}^b] \), where \( \tilde{c}^a = u + \frac{1}{2}s \), \( \tilde{c}^b = u - \frac{1}{2}s \), \( u = [u_1; \ldots; u_M] \), and \( s = [s_1; \ldots; s_M] \). The \((2K \times 2K)\) matrix \( D \) can be written in terms of lower dimensional square matrices as follows:

\[
D = \begin{bmatrix} D^a & -D^{ab} \\ -D^{ba} & D^b \end{bmatrix}.
\]

The matrices \( D^a \) and \( D^b \) represent the impact terms from executing orders independently against the ask and respectively bid sides, while \( D^{ab} \) and \( D^{ba} \) account for cross-impacts between the two sides. These can further be expressed in terms of \((N+1) \times (N+1)\) building-block matrices containing the order book parameters. The explicit forms are given in the Appendix A.1.

From Proposition 1 we can treat the optimal controls as deterministic variables. It follows that the only source of uncertainty in the problem is the random walk, implying that the manager’s total post-execution

\footnote{While this result is sensitive to the random walk assumption, the subsequent solution methodology we develop can also handle cases where predictability is added, in the form of a deterministic drift. Although this would be an interesting extension, we leave this for future work. In contrast, more complex views on the behavior of asset prices (such as when serial correlation is considered) will generally lead to path-dependent optimal policies.}
wealth is normally distributed. More specifically, using the expressions from Lemma 4, we have that $W_N \sim \mathcal{N}(\mu_{W_N}, \sigma^2_{W_N})$, where

$$\mu_{W_N} = \mathbb{E}[W_N] = -(x^\prime D x + \mathbb{E}[\tilde{c}^\prime x]),$$

with $\mathbb{E}[\tilde{c}^\prime x] = u'_0 1' \Delta_{K+}^t x + \frac{1}{2} s' \Delta_{K+}^t x$ and

$$\sigma^2_{W_N} = \text{Var}[W_N] = x' \Delta_{K} \Sigma_u \Delta_{K} x,$$

where $u'_0 = [u_{1,0}, \ldots, u_{M,0}] = \nu'_0$ is the vector of initial asset mid-prices and $\Sigma_u$ is the covariance matrix of $u$ across time and assets (see Appendix A.1 for explicit forms). A consequence of this property is that we can establish equivalence between the manager’s exponential utility and the mean-variance objective often used in the execution literature. This follows directly from the identity $\mathbb{E}[e^{\alpha W}] = e^{\mathbb{E}[\alpha W] + \frac{1}{2} \alpha^2 \text{Var}[W]}$, for any normally distributed $W$, and from the monotonicity of the exponential. The manager’s original optimization problem over his exponential utility can thus be equivalently written as

$$\max_{0 \leq x \in S} \mu_{W_N} - \frac{1}{2} \alpha \sigma^2_{W_N}.$$  

Using this equivalent form and the equations (20) and (21), we can reformulate the original utility maximization problem as a standard QP minimization problem over the manager’s risk-adjusted execution shortfall.$^{19}$

**Proposition 2 (Quadratic Program)** The original dynamic maximization problem (18) is equivalent to the following static quadratic program which minimizes risk-adjusted execution shortfall:

$$\min_{x \geq 0} \frac{1}{2} x' \bar{D} x + c' x$$

subject to $1' \Delta_{K+}^t x = z_0$,

where $c' = \frac{1}{2} s' \Delta_{K+}^t$ and $\bar{D}$ is the Hermitian matrix $\bar{D} = (D + D') + \alpha \Delta_{K} \Sigma_u \Delta_{K}^t$.

### 3.3 Discussion

We compare the static optimal policy described in Proposition 1 to other types of policies found in the literature: Bertsimas et al. (1999) develop a static approximation algorithm, allowing the manager to re-optimize his objective at every period, and show that their solution is close to optimal. Basak & Chabakauri (2010) compare static pre-commitment strategies with adaptive strategies in the context of the portfolio composition problem and argue that the manager can be better off by pre-committing in certain cases. In contrast, Lorenz & Almgren (2012) develop an adaptive execution model and show that the gain in trading flexibility can indeed be valuable for the manager.

In our framework, a static solution is optimal without exogenously enforcing pre-commitment – a result which is sensitive to the random walk assumption, but which also significantly simplifies the problem.$^{19}$

---

$^{19}$The execution shortfall (i.e., net execution cost) is the difference between the pre-execution market value of the portfolio ($W_0$), and the expected wealth obtained post-execution ($\mu_{W_N}$), i.e., it is equal to $W_0 - \mu_{W_N}$, where $W_0 = -u'_0 z_0 = -u'_0 (1' \Delta_{K+}^t x)$. 

15
Intuitively, this result states that the generated filtration provides no useful information for the optimal control in our framework. This implies that the manager has nothing to gain by utilizing adaptive trading strategies in the CARA framework, under the random walk assumption. Alfonsi et al. (2008) develop a comparable static solution methodology in the context of an optimal liquidation problem for a single asset and a risk-neutral investor. Similarly, Huberman & Stanzl (2005) find a comparable static solution in their framework with a mean-variance objective.

The static formulation has several advantages. While the required inequality constraints will restrict the availability of general closed-form solutions, the QP can efficiently handle these types of constraints numerically. Specifically, in the case where $\mathbf{D}$ is positive-definite, the problem is convex and is thus solvable in polynomial time.\(^{20}\)

The static formulation can be extended to include additional deterministic linear or quadratic constraints one may want to impose on the set of feasible strategies. This feature is of consequence to practitioners. For instance, in many large-scale portfolio execution programs, managers may want to exercise particular control over certain assets. We provide an example in Section 5.2. Further, the model can easily incorporate agency trading constraints which some execution desks may face when trading on behalf of their clients. For example, an execution desk liquidating an agency position may not be allowed to trade counter-directionally and conduct any purchasing orders. This constraint could be captured in our model by setting $x^+ = 0$. A more detailed discussion on agency trading can be found in Moallemi & Saglam (2013b).

Another advantage is that the formulation can handle time-dependent parameters (relaxing the Assumptions 1, 3 and 4). Time-dependence can be critical in many situations, for instance, when markets are in turmoil and liquidity variations are expected to occur in the future (see Brown et al. (2010) for a detailed treatment with uncertain liquidity shocks). In our framework, expected liquidity variations during the execution window could be integrated into the model by adjusting the values of the density $q$, the replenishment rate $\rho$ and the steady-state bid-ask spread $s$, at the desired periods. Similarly, one could capture expected intra-day fluctuations in volume of trade (thus accounting for the well-known intra-day “smile” effect). Details are provided in Appendix A.3.

The liquidity model described above can capture various forms of transaction costs observed in the market, including fixed, proportional and quadratic costs. The proportional (linear) trading costs are captured by the constant bid-ask spread $s_i$. The quadratic trading costs are captured by the linear price impact assumed in the liquidity model. The fixed trading costs are not directly modeled but reflected implicitly in our setting. In particular, we assume a discrete and finite number of trading periods in part to reflect the fixed cost in trading. Presumably, the number of trading periods $N$ is connected to the fixed cost. Although in our model $N$ is taken as given, we can easily endogenize it as an optimal choice in the presence of fixed trading costs at say $c_0$. Clearly, larger $N$ would decrease execution costs by allowing the manager more flexibility in spreading trades. But it would also increase total fixed costs, which would be $Nc_0$. An optimal choice of $N$ will result from this trade-off. See, for example, He & Mamaysky (2005) for a more detailed discussion on this issue.

\(^{20}\)If $\mathbf{D}$ is indefinite, then the problem is NP-hard. In general, sufficient convexity bounds on the parameters can be obtained for simplified versions of the problem (e.g., 2 assets and 2 trading periods), however, the complexity of testing for positive definiteness scales with $M$ and $N$.  

16
4 Optimal Execution Policy

This section presents several case studies which illustrate our main results. We highlight cases where advanced execution strategies are optimal. These strategies constructively utilize order book cross-elasticities to improve execution efficiency. In what follows, we set the steady-state bid-ask spread to zero to simplify the exposition.\footnote{Note, this does not imply that the actual bid-ask spread is zero during the execution process. Unsurprisingly, increasing the steady-state bid-ask spread leads to higher overall execution costs, reducing the applicability of advanced trading strategies. A detailed analysis is provided in the Appendix A.3.} Furthermore, we only consider the problem of liquidating assets. The asset purchasing problem is fully equivalent (by interchanging “buy” and “sell” labels). The model can also treat mixed buy and sell objectives without any modifications.

4.1 Base Case (No Correlation, No Cross-Impact)

Our base case consists of a portfolio with two identical assets, but with no correlation in their risks ($\gamma = 0$) nor cross-impact ($\lambda_{12} = \lambda_{21} = 0$) in their liquidity. The manager needs to liquidate his position in the first asset, but has no initial and final position, or pre-defined objective in the second. We refer to the first asset as the active asset (with boundary conditions $z_{1,0} \neq 0$ and $z_{1,N} = 0$), while the second is inactive (with boundary conditions $z_{2,0} = z_{2,N} = 0$). Consider a long position in the active asset, consisting of $z_{1,0} = 100$ shares that need to be liquidated over $N = 100$ periods (i.e., $z_{1,100} = 0$). The horizon $T = 1$ day. The mid-price is $v_{1,0} = $1 at time 0, implying a pre-liquidation market value of $100$.\footnote{The rest of the parameters used in this Section are: the volatilities $\sigma_1 = \sigma_2 = 0.05$, the order book densities $q_1 = q_2 = 1500$, the replenishment rates $\rho_1 = \rho_2 = 5$, and the permanent impact parameters $\lambda_{11} = \lambda_{22} = 1/(3q_1)$. These parameters are used to generate all the figures, unless otherwise specified.}

Figure 4 displays the manager’s optimal execution policy (OEP), in the form of his net position over time, comparing the risk-neutral (RN) case to the risk-averse (RA) case. Unsurprisingly, in the absence of correlation and cross-impact between the two assets, it is never optimal to trade the inactive asset (dashed line). Doing so, would increase overall execution costs without any risk reduction. It is useful to provide some intuition on the resulting OEP of the active asset (solid line).

In the RN case (Figure 4(a)), the OEP consists of placing two large orders at times 0 and $N$, and
splitting the rest of the order evenly across time. The slope of the execution curve represents the manager’s trading rate. The steeper the slope, the faster he is executing shares. The slope is related to the order book replenishment process. The faster the order book inventory gets replenished after each executed order, the more sell orders the manager can submit per unit time. The liquidity spikes on the boundaries are related to the replenishment and boundary conditions of the order book. Assuming that the order books are initially “full”, it is natural that the first order should be large. In essence, one can obtain “cheaper” liquidity at the start. Similarly, the last order should also be large because one cannot constructively utilize order book dynamics after the execution horizon \( N \). These spikes fade as the inventory recovery rate increases and disappear at the limit when liquidity is infinite and inventories are instantaneously replenished after each executed order, \( \rho \to \infty \) (we omit the plot).

In the RA case (Figure 4(b)), the manager consumes greater liquidity early on in the liquidation process. This dampens the adverse impact of future price uncertainty, reducing execution risk, but at a cost. To understand this result, it is helpful to consider extreme values of \( \alpha \). When \( \alpha \to \infty \), the manager is only sensitive to the variance of his costs and the solution is trivial: execute everything at time zero (we omit the plot). This strategy is effectively risk free, as it guarantees zero standard deviation in execution costs. But it also understandably the worst-case scenario from a cost perspective.

![Figure 5: Dynamics of the expected ask and bid prices of the active asset, responding to the manager’s execution policy. Both prices are initially equal to $1 at time 0. Correlation (\( \gamma = 0 \)) and cross-impact (\( \lambda_{ij} = 0 \)) are turned off implying that it is never optimal to trade in the inactive asset. Prices are plotted beyond \( T = 1 \), to illustrate the convergence process towards a new steady state.](image)

The impact the OEP has on the expected ask and bid prices of the active asset is shown in Figure 5. The ask and bid prices are initially equal to $1 before the first sell order is placed. The liquidation process pushes the ask and bid prices of the active asset down over time. As there is no correlation or cross-impact between assets, the price of the inactive asset remains unaffected at $1 (we omit the plot). At any fixed time \( t \), the gap between the ask and bid prices defines the instantaneous bid-ask spread, the dynamics of which depend on two opposing forces: On the one hand, each executed order widens the bid-ask spread (as the manager is consuming liquidity in the order books). On the other hand, new limit orders arrive over time collapsing the bid-ask spread back towards its steady state. The mid-point of the bid-ask spread is the instantaneous mid-price (this is not shown on the plot).

The trade sizes and utility implications of the aforementioned strategies are reported in cases 1 and 5, of Tables 1 and 2. Table 1 shows traded volume (in shares). Table 2 reports the expected execution costs, the execution risk (stated in terms of the standard deviation of the execution costs), the manager’s execution
1) No correl., no cross-imp.
   Figure 4(a) | 14.7 | 100 | 0 | -
   Figure 4(b) | 47.6 | 100 | 0 | -

2) Effect of correlation
   Figure 6(a) | 14.7 | 100 | - | -
   Figure 6(b) | 43.8 | 100 | 12.7 | 25.4

3) Effect of cross-impact
   Figure 8(a) | 14.7 | 100 | - | -
   Figure 8(b) | 49.1 | 100 | 6.4 | 12.8
   Figure 8(c) | 14.1 | 100 | 9.6 | 33.3

4) Effect of correl. and cross-imp.
   Figure 10 | 46.4 | 100 | 18.0 | 36.0

5) Execute everything at time 0
   (not plotted) | 100 | 100 | - | -

Table 1: Shares executed in each asset.

<table>
<thead>
<tr>
<th>Case</th>
<th>Active Asset</th>
<th>Inactive Asset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st Trade</td>
<td>Total Volume</td>
</tr>
<tr>
<td>1) No correl., no cross-imp.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 4(a)</td>
<td>14.7</td>
<td>100</td>
</tr>
<tr>
<td>Figure 4(b)</td>
<td>47.6</td>
<td>100</td>
</tr>
<tr>
<td>2) Effect of correlation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 6(a)</td>
<td>14.7</td>
<td>100</td>
</tr>
<tr>
<td>Figure 6(b)</td>
<td>43.8</td>
<td>100</td>
</tr>
<tr>
<td>3) Effect of cross-impact</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 8(a)</td>
<td>14.7</td>
<td>100</td>
</tr>
<tr>
<td>Figure 8(b)</td>
<td>49.1</td>
<td>100</td>
</tr>
<tr>
<td>Figure 8(c)</td>
<td>14.1</td>
<td>100</td>
</tr>
<tr>
<td>4) Effect of correl. and cross-imp.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figure 10</td>
<td>46.4</td>
<td>100</td>
</tr>
<tr>
<td>5) Execute everything at time 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(not plotted)</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 2: Expected execution costs (Exp. Costs), standard deviation of costs (Std. Dev.), certainty equivalent (Cert. Eq.) and execution Sharpe ratio (Exec. Sharpe Ratio) defined as the cost savings achieved over the most inefficient case 5, divided by the standard deviation of the costs. The higher the Sharpe ratio, the more efficient the execution is. Costs are provided in ($) terms. As a comparison, the portfolio pre-liquidation market value is $100.
certainty equivalent, and a measure of execution efficiency which we refer to as the execution Sharpe ratio. The latter is defined as the ratio of the cost savings achieved over the most costly, risk-free, execution strategy ($\alpha \to \infty$), divided by the standard deviation of those costs. The higher the execution Sharpe ratio, the more efficient the execution.

4.2 Effect of Correlation in Risk

Next, we build on the base case by introducing correlation between the two assets, while maintaining cross-impact at zero (see Tables, case 2).

Figure 6 compares the impact of correlation ($\gamma = 0.7$) between the RN and RA cases. Unsurprisingly, if the manager is RN ($\alpha = 0$, Figure 6(a)), there are no trades in the inactive asset. On the other hand, risk aversion ($\alpha = 0.5$, Figure 6(b)), combined with correlation, leads to a complex strategy in the inactive asset. In particular, it becomes optimal to 1) go short the inactive asset at time zero, 2) hold the short position for some time, and 3) start covering the short at variable rates towards the end of the execution window, to satisfy the boundary conditions. The resulting asset price dynamics for the inactive asset are shown in Figure 7 while the number of shares traded with this strategy are reported in case 2 of Table 1.

The results in Table 2 show that on the one hand, the RA strategy trades off higher execution costs for risk reduction, compared to the RN case. In other words, in order to reduce execution risk, one generally has to be willing to incur higher expected execution costs. On the other hand, the RA strategy has a higher execution Sharpe ratio compared to the RN strategy, implying more efficient execution.

More importantly, a comparison between cases 1 and 2 in Table 2 suggests that shorting the positively correlated inactive asset during the execution process (see Figure 6(b)) allows the manager to reduce both his total execution costs (by $0.05$) and his execution risk (by $0.08$), compared to the single-asset strategy in Figure 4(b). The risk reduction is due to the fact that the short position acts as a hedge, dampening future price volatility. Note that despite the assumed positive correlation, the trades at time zero involve selling shares in both assets simultaneously. This may seem counterintuitive given that positive correlation is generally associated with offsetting trades (buy and sell) in the classical portfolio composition framework. To understand why, consider the following scenario: assume the price of the active asset randomly decreases...
in the future, implying that subsequent sell orders generate less revenue for the manager. In this case, his positively correlated short position in the inactive asset will also accrue in value, thus compensating him for the decreased revenues. An analogous argument holds for the opposite case of a random price increase in the inactive asset.

Beyond a reduction in risk, we emphasize that expected costs are also reduced over the single-asset case, despite that one is trading more shares and incurring additional price impact in the inactive asset. To understand why, consider the risk-reduction/cost-reduction trade-off mentioned previously. In the RA case, one can decrease execution costs at the expense of higher risk, and vice versa. Trading the inactive asset as a hedge leads to more efficient risk reduction compared to the un-hedged strategy. In turn, this implies that one does not have to give up as much “upside” in execution costs, to achieve a desired risk level.

Figure 7: Dynamics of the expected ask and bid prices of both assets in the case with RA (α = 0.5) and correlation (γ = 0.7), as depicted in Figure 6(b). Cross-impact (λij = 0) is turned off.

Figure 7(b) shows the evolution of the bid and ask prices of the inactive asset, resulting from the OEP portrayed in Figure 6(b). The figure clearly demonstrates why one cannot generally model bid and ask sides independently of one another, when considering cross-asset effects. As one is required to sell and subsequently purchase back shares in the same asset, it is necessary to keep track of the price impact that each order has on both sides of the book, over time.

4.3 Effect of Cross-Impact in Liquidity

Here, we remain with the previous liquidation scenario, removing correlation between the two assets and focusing instead on the effect of cross-impact. In contrast to correlation which is assumed exogenous, cross-impact accounts for the impact an order in one asset has on the price and order book supply/demand dynamics of the other (see Figure 3 for an illustration). Moreover, this impact does not need to be symmetric between the two assets: An order in stock A may impact stock B in one way, while changing the order and trading in stock B first, ceteris paribus, may impact stock A differently. Figure 8 illustrates this idea by comparing a case with symmetric cross-impact (λ12 = λ21 in Figure 8(b)) to a case with asymmetric cross-impact (λ12 = −λ21 in Figure 8(c)).
Figure 8: OEPs with cross-impact. In (a) the lack of RA implies no trades in the inactive asset. Introducing RA in (b) triggers trades in the inactive asset. In (c), asymmetric cross-impact triggers trades in the inactive asset, even in the RN case. Correlation ($\gamma = 0$) is turned off. Other parameter values: $\sigma_1 = \sigma_2 = 0.05$, $q_1 = q_2 = 1500$, $\rho_1 = \rho_2 = 5$, $\lambda_{11} = \lambda_{22} = 1/(3q_1)$.

A. Effect of cross-impact on the liquidation strategy

Symmetric cross-impact without RA (Figure 8(a)) does not result in any trades in the inactive asset and the costs over the base case remain unchanged. Adding RA (Figure 8(b)) triggers trades in the inactive asset, comparable to the ones observed in the case with correlation. Therefore, if RA is considered, symmetric cross-impact and correlation can have similar implications for the manager’s OEP. The resulting trades are reported in case 3 of Table 1, while the resulting price dynamics are reported in Figure 9.

Figure 8(c) presents a case that clearly differentiates correlation from cross-impact. We consider asymmetric cross-impact between two assets and show that even a RN manager could be better off by trading in both the active and the inactive asset. This is in stark contrast to the previous example of correlations which become irrelevant for a RN manager.

B. Effect of cross-impact on execution utility

The results in case 3 of Table 2 suggest that the effect of symmetric cross-impact on execution costs and risk reduction is comparable to that of correlation. Trading the inactive asset during the liquidation (Figure 8(b)) allows the manager to slightly reduce his total execution costs (by $0.02$) and his execution risk (by $0.01$) over the optimal single-asset trading strategy (Figure 4(b)).

In contrast, asymmetric cross-impact allows the manager to achieve the greatest cost reduction of all cases (although, this comes with increased execution risk).

Introducing symmetric cross-impact has implications on the price dynamics of the inactive asset that are not observed when only considering correlation. Figure 9(b) shows the cross-impact the active asset has on the inactive one ($\lambda_{21}$). The liquidation of the active asset pushes the price of the inactive asset down over time. This effect may seem favorable to the manager. He could short the inactive asset at time 0 and buy it back later at a lower price. However, putting in place the initial short position in the inactive asset also adversely impacts the price of the active asset via the cross-impact term ($\lambda_{12}$). This implies that the manager would be liquidating his active position at lower prices. This effectively restricts the manager’s ability to take advantage of the favorable price dynamics expected in the inactive asset. This trade-off is fully internalized in the OEP.
Figure 9: Dynamics of the expected ask and bid prices of both assets in the case with RA ($\alpha = 0.5$) and cross-impact, as depicted in Figure 8(b). Correlation is turned off. The manager’s trades in the active asset affect the bid and ask side prices of the inactive asset, and vice versa.

C. Asymmetric cross-impact and arbitrage

Huberman & Stanzl (2004), Section 5, illustrates an example of asymmetric cross-impact that can lead to price manipulation and arbitrage. The authors derive sufficient no-arbitrage conditions in the multi-asset setting. Namely: 1) cross-impact symmetry between assets and 2) lack of temporary impact costs. As the authors state, these conditions are sufficient, but they are not necessary. Case 3 of Table 2 shows that asymmetric cross-impact does not necessarily lead to arbitrage opportunities when considering positive temporary impact costs. While some cost benefits can be achieved under these scenarios over the RN base case, net execution costs remain positive at $1.45$. To understand this result, observe that every executed order has both a permanent and temporary impact component, and while asymmetric cross-impact can understandably reduce costs on the permanent component, the manager is also consistently incurring costs from the temporary component during trade (i.e., he is “rolling” up or down the supply/demand curves getting executed at increasingly costly limit price levels). This trade-off between temporary and permanent price impacts is fully internalized in the OEP. Thus, similar to Proposition 3 in Huberman & Stanzl (2004), absence of arbitrage will hold if temporary impact costs are sufficiently larger than permanent impact costs. Formally, arbitrage is precluded in our model if the execution shortfall in problem (23) is non-negative for any feasible trading strategy in the set $\mathcal{S}$. Equivalently, arbitrage is precluded if the matrix $\bar{D}$ is positive semidefinite.

4.4 Joint Effect of Correlation in Risk and Cross-Impact in Liquidity

Here, we consider both correlation and cross-impact simultaneously. The results reported in Figure 10 and in case 4 of Tables 1 and 2, suggest that the cost benefits obtained exhibit positive convexity. In other words, correlation and cross-impact work constructively, providing benefits that are greater than the sum of the individual contributions each of them brings independently.

The results in case 4 of Table 2 suggest that expected costs can be reduced to $2.05$, while risk can be reduced to $1.02$, the lowest of all cases. The execution Sharpe ratio obtained is the greatest of all cases, at 1.25. To achieve these benefits, the OEP requires trading a significant volume in the inactive asset, equal to approximately $1/3$ of the total volume of the active asset. The price dynamics of the inactive asset observed
in Figure 11 combine the cross-impact and correlation effects we described previously. In this case, the manager is generally selling “high” and buying “low” in the inactive asset, while also benefiting from a reduced initial order size in the active asset, and limit order mean-reversion dynamics.

Figure 11: Dynamics of the expected ask and bid prices of both assets in the scenario depicted in Figure 10.

5 Mixed Liquidity Portfolios

This section illustrates additional results that are of consequence to practitioners.

5.1 Portfolio Overshooting

Execution objectives are typically richer than the ones illustrated in the previous section. Portfolio managers often need to liquidate or acquire positions in multiple assets with different risk and liquidity characteristics. This Section illustrates the optimal liquidation of a portfolio composed of 2 assets with different liquidity levels. The first asset is considered liquid, with limit order replenishment rate $\rho_1 = 10$ and limit order density $q_1 = 3000$, while the second is (comparatively) illiquid, with rate $\rho_2 = 1$ and density $q_2 = 300$. \footnote{For instance, while the price of a single-name stock may be highly correlated with the price of its derivatives (e.g., call option), significant liquidity differentials will generally exist between the two. In particular, liquidity in the OTM single-name option market can be scarce, even if the underlying stock is fairly liquid.} \footnote{The portfolio is initially equally weighted, consisting of 100 shares of each asset. Unless otherwise specified, the rest of the parameters used in this Section are: $v_{1,0} = v_{2,0} = 1$, $\sigma_1 = \sigma_2 = .05$, $\lambda_{11} = \lambda_{22} = 1/(3q_1)$, and $\lambda_{12} = \lambda_{21} = 0$.}
Figure 12 shows the OEPs obtained for different RA and correlation assumptions. The results suggest that liquid assets will generally be executed more smoothly throughout the horizon, while illiquid assets tend to corner solutions (i.e., it is optimal to execute two larger trades at times 0 and 1). The intuition here is simple: illiquid assets have order books with low replenishment rates leading to asset prices with low mean-reversion. The lack of replenishment implies that one cannot take advantage of order book dynamics in any meaningful way and thus, the optimal solutions tend to be trivial. On the other hand, liquid assets with high replenishment have more interesting dynamics that can be utilized towards the execution problem, leading to richer optimal strategies.

![Graphs showing OEPs for different parameter values](image)

Figure 12: OEPs in the case of a portfolio with mixed liquidity. Parameter values: $v_1, 0 = v_2, 0 = 1, \sigma_1 = \sigma_2 = .05, \lambda_{11} = \lambda_{22} = 1/(3q_1)$, and $\lambda_{12} = \lambda_{21} = 0$.

When the two assets are correlated (see Figure 12(c)), further advanced strategies become optimal. We obtain two-sided buy and sell strategies, despite the simple unidirectional liquidation objective. The position in the liquid asset becomes negative near time $0.25$, implying overshooting. The excess shares sold are gradually purchased back in order to meet the boundary conditions as the horizon approaches. From a hedging perspective, the transient short position in the liquid asset dampens future price uncertainty and reduces execution risk. We emphasize that while overshooting was also observed in the example of Section 4.2 (because any trades in the inactive asset could be considered as overshooting trades), here, this effect is entirely driven by the assumed liquidity differences between the two assets.

### 5.2 Synchronization Risk

The results in the previous Section suggest that liquid and illiquid assets will be executed at different speeds. The manager could therefore be left over/under-exposed to individual assets during the execution process, i.e., he could be facing synchronization risk. To highlight this more clearly, we plot the weight of each asset (expressed as the ratio of net shares held in each asset over total shares held in the portfolio) over time, in Figure 13, for the same cases that were presented in Figure 12.

Assume that the manager’s initial optimal portfolio allocation is 50/50, and that there is some underlying benefit (such as portfolio diversification) to preserve this optimal split during the execution window. All three cases in Figure 13 show that the manager could be left over-exposed to the illiquid asset during the execution window, as its weight can move above the optimal 0.5 line.

A simple way to mitigate the undesirable exposure is to constrain the admissible order quantities at each
trading period. For instance, one can restrict each asset’s weight to an interval, \( w_{i,n} \in [w_i^* - \xi, w_i^* + \xi] \), where \( w_i^* \) is the desired weight targeted in asset \( i \) and \( \xi \in [0, \infty) \) controls the desired margin of error. The parameter \( \xi \) is chosen by the manager and can be thought of as the degree of tolerance to weight variability. Figures 14 and 15 show the impact of different tolerance parameters on the OEPs and weight profiles. As \( \xi \to 0 \), both asset weights converge to the 0.5 line, and the OEPs converge to a single strategy. Interestingly, the unique optimal strategy is a weighted combination of the two individual unconstrained OEPs of each asset. Further, it is in the strict interior of the two.

Adding these types of constraints to the problem reduces the set of feasible execution strategies, leading
to increased execution costs over the unconstrained global optimum. This raises the question of how costly it is to synchronize the portfolio in this fashion. We define the synchronization cost as the expense one would have to incur over the most efficient (lowest cost) outcome, in order to maintain a targeted weight profile during the execution process.

In our examples, the synchronization cost for $\xi = 10\%$ is equal to 18bps, while in the worst-case scenario ($\xi \to 0$), it is equal to 38 bps. The latter represents the maximum amount the manager would expect to pay, in excess of the most efficient outcome, in order to remain fully in line with the optimal targeted weight allocation throughout the entire execution window.

## 6 Conclusion

Controlling price impact is of central importance in portfolio management, and is particularly crucial in practical situations where managers need to execute large positions in multiple assets. We have studied the multi-asset execution problem demonstrating that it is far from being a simple extension to the single-asset case. Assets can interact in complex ways and these interactions can have a substantial impact on the aggregate portfolio execution cost and risk. Understanding the exact nature of these interactions requires an extensive market microstructure model that can adequately capture coupled supply and demand dynamics at the order book level.

Our results suggest that managing execution at the portfolio level needs to take account of links in both risk and liquidity across assets. In the presence of such links, we find that managers can improve execution efficiency by engaging in a series of non-trivial buy and sell trades in multiple assets simultaneously. The trades are non-trivial in the sense that they may require temporarily trading positively correlated assets in the same direction, or even overshooting one’s portfolio target during the execution window.

These results extend to portfolios with heterogeneous liquidity across assets. There, the liquidity differential between assets can lead to complex strategies which utilize the liquid asset to improve execution efficiency at the portfolio level. However, we also find that these advanced strategies can leave managers overexposed to illiquid assets during the execution. This synchronization risk can be mitigated by introducing constraints that can synchronize the portfolio trades, at the cost of reduced execution efficiency. This led to the concept of synchronization cost – a measure which allows managers to trade off these two factors, based on their individual preferences.

Perhaps an even more compelling takeaway is that advanced strategies can be optimal for simple and common execution objectives (such as the liquidation of a single asset in the portfolio). This implies that it may be crucial for managers to systematically take into account cross-asset interactions in risk and liquidity in their risk-management and trading decisions. It also implies that market regulators should be aware of the increased liquidity needs this can lead to, if deployed on a large-scale basis.

## A Appendix

In this appendix, we first define some notational convention. Next, we provide several partial proofs. Detailed proofs can be found in the online supplemental file. Finally, we present several additional results.
A.1 Explicit Matrix Forms

The components of the matrix $D$ are given by

$$
D_a = [D_{ij}^a]_{K \times K}, \quad D_{ab} = [D_{ij}^{ab}]_{K \times K},
$$
$$
D_{ba} = [D_{ij}^{ba}]_{K \times K}, \quad D_b = [D_{ij}^b]_{K \times K}.
$$

Building blocks: The self-impact matrices on the diagonals are given by $D_{ii}^l = Q_i^l + L_{ii} + \Gamma_i^l, \forall i \in I_M, \forall l \in \{a, b\}$. The cross-asset impact matrices take the form $D_{ij}^l = L_{ij} - L_{ij} \circ \Gamma_i^l, \forall j \in I_M, \forall l \in \{a, b\}$, where $\circ$ is the Hadamard product. The components of the matrix $A$. The difference operators are defined as follows. Let $I_k$ be the identity matrix of size $k$, where $k \in \{M, K\}$. The $2k \times k$ difference operators are defined as $\Delta_k = \begin{bmatrix} I_k \\ -I_k \end{bmatrix}$, $\Delta_{k+} = \begin{bmatrix} I_k \\ I_k \end{bmatrix}$, $\Delta_{ka} = \begin{bmatrix} I_k \\ 0 \end{bmatrix}$, and $\Delta_{kb} = \begin{bmatrix} 0 \\ -I_k \end{bmatrix}$.
A.2 Proofs

Lemma 1 (Temporary Price Impact)

A buy order of size \( x \) being executed against \( i \)'s ask-side inventory \( q^a_i \), displaces the best ask price from \( a_{i,n} \rightarrow a^*_i,n \), according to

\[
\int_{a_{i,n}}^{a^*_i,n(x)} q_i^a 1_{p \geq a_{i,n}} \, dp = x.
\]

Combining this expression with Assumption 1, we have

\[
\int_{a_{i,n}}^{a^*_i,n(x)} q_i^a \, dp = q_i^a (a^*_i,n(x) - a_{i,n}) \Rightarrow a^*_i,n(x) = a_{i,n} + \frac{x}{q_i^a}.
\]

Therefore for \( x = x^+_i,n \), we have \( a^*_i,n = a_{i,n} + x^+_i,n/q_i^a \) and the temporary price impact displacement is defined as \( a^*_i,n - a_{i,n} = x^+_i,n/q_i^a \). The derivation for a sell order follows similar steps. ■

Lemma 2 (Best Prevailing Bid/Ask-Prices)

We present below an outline of the derivation for the best ask price dynamics; the bid price dynamics are derived in a similar way. The best available ask price for asset \( i \) at time \( n \) is given by

\[
a_{i,n} = u_{i,n} + \frac{1}{2} s_i + \text{PPI} + \text{TPI},
\]

where the first term accounts for the random walk driving the mid-price, the second term accounts for the steady-state bid-ask spread, the third term accounts for the aggregate PPI of all orders up to (but excluding) \( n \), and the fourth term is the order book state vector which accounts for the TPI of all orders up to (and including) \( n \). Following equation (12) and the definition of the vector \( z_{i,n} \) in Section 2.1, the aggregate PPI for asset \( i \) can be written as

\[
\sum_{k=1}^{n} \sum_{j \in I_M} \lambda_{ij} (x^+_{j,k-1} - x^-_{j,k-1}) = [\Lambda (z_{i,0} - z_{i,n})]_i,
\]

where \( [\cdot]_i \) returns the \( i \)-th line of a matrix. Following Assumption 4, the aggregate TPI can recursively be written as

\[
d^a_{i,n} = (d^a_{i,n-1} + \kappa^a_i (x^+_{i,n-1}) \epsilon^{-\rho\tau}) e^{-\rho\tau},
\]

where \( \kappa^a_i (x^\pm_{i,n}) \) is the net displacement in the ask-side order book resulting from a buy trade in asset \( i \) at time \( n \). The net displacement is given by the difference between the TPI and PPI at time \( n \): \( \kappa^a_i (x^\pm_{i,n}) = \frac{x^\pm_{i,n}}{q_i^a} - \sum_{j \in I_M} \lambda_{ij} (x^+_{j,n} - x^-_{j,n}) \). Using these expressions, and removing the recursion in \( d^a_{i,n-1} \), the aggregate TPI can be written as

\[
d^a_{i,n} = \sum_{k=1}^{n} \left( \frac{x^+_{i,k-1}}{q^a_i} - \sum_{j \in I_M} \lambda_{ij} (x^+_{j,k-1} - x^-_{j,k-1}) \right) e^{-\rho\tau(n-k+1)}.
\]

Note, the recursive vector form of the aggregate TPI across all assets given in equation (14) follows immediately from the previous expressions, in particular, \( d^a_{i,n} = [(d^a_{i,n-1} + \kappa^a_i x_{i,n-1}) \epsilon^{-\rho\tau}]_i \) and thus \( d^a_{i,n} = (d^a_{i,n-1} + \kappa^a_i x_{i,n-1}) e^{-\rho\tau} \).

Combining the aggregate PPI and TPI terms, and repeating similar steps for the bid side, we obtain the
following expressions for the best available ask and bid prices of asset $i$ at each time $n$:

$$a_{i,n} = u_{i,n} + s_i/2 + \sum_{k=1}^{n} \sum_{j \in I_M} \lambda_{ij} \delta x_{j,k-1} + \left(\sum_{k=1}^{n} \frac{x_{i,k-1}^+ - \sum_{j \in I_M} \lambda_{ij} \delta x_{j,k-1}^-}{q_i^a} \right) e^{-\rho_i^a(n-k+1)} ,$$

$$b_{i,n} = u_{i,n} - s_i/2 + \sum_{k=1}^{n} \sum_{j \in I_M} \lambda_{ij} \delta x_{j,k-1} + \left(\sum_{k=1}^{n} -\frac{x_{i,k-1}^- - \sum_{j \in I_M} \lambda_{ij} \delta x_{j,k-1}^+}{q_i^b} \right) e^{-\rho_i^b(n-k+1)} ,$$

where $\delta x_{j,k-1} = (x_{j,k-1}^+ - x_{j,k-1}^-)$. Extending the above steps to all assets, and using vector notation, we can obtain the final vector forms for the best available ask and bid prices in equations (13a) and (13b).

Lemma 3 (Execution Costs/Revenues)

Following an executed order, the associated costs/revenues can simply be calculated by integrating the best available bid/ask prices over the total amount of units executed $x$. It follows that

$$c_{i,n}(x) = \int_0^x a_{i,n}^*(u)du \quad \text{and} \quad r_{i,n}(x) = \int_0^x b_{i,n}^*(u)du,$$

where $a_{i,n}^*(x)$ and $b_{i,n}^*(x)$ are given in (7). Specifically, for an incoming order $x = x_{i,n}^+$ or $x = x_{i,n}^-$, we find

$$c_{i,n}(x_{i,n}^+) = \left(\alpha_{i,n} + \frac{x_{i,n}^+}{2q_i^a}\right) x_{i,n}^+ \quad \text{and} \quad r_{i,n}(x_{i,n}^-) = \left(\beta_{i,n} - \frac{x_{i,n}^-}{2q_i^b}\right) x_{i,n}^- .$$

Equivalently, in vector notation: $c_n = x_{\pi,n}^T (\alpha_n + Q^c x_{\pi,n}^+)$ and $r_n = x_{\pi,n}^T (\beta_n - Q^b x_{\pi,n}^-)$.

Proposition 1 (Path Independence)

While the path-independence property follows from standard induction arguments, the details can be cumbersome. To illustrate the main intuition, a partial proof of path-independence at times $N$ and $N - 1$ is provided below, for a simplified unconstrained version of the problem. A more detailed proof for all times $n$ can be found in the online supplement of the paper.

First, note that the optimal trade at the last period $N$ does not depend on the filtration and is simply given by $x_{\pi,N}^+ \Delta_M = z_{\pi,N}$. Therefore, path independence holds trivially for time $N$. Next, we check whether path independence is preserved at the previous period $N - 1$. To this end, note that the reward function at any time $n$ can be decomposed in two terms: $\pi_n = \theta_n - x_{\pi,n}^T \Delta_M u_n$. The first term, by construction, is a deterministic function, which is quadratic in the state vectors and the controls at time $n$. The second term is the only stochastic term in the reward function. Using this decomposition and removing the dependence on the state at time $N$ (by using the recursive state equations that link the states at times $N$ and $N - 1$), we can show that the value function at time $N - 1$ is given by

$$J_{N-1} = \max_{x_{\pi,N-1}} \mathbb{E}_{N-1} \left[ -e^{-\alpha(W_{N-2} + \theta_{N-1} + \phi_{N-1} - x'_{N-1} u_{N-1} - (x'_{N-1} - x'_{N-1} \Delta_M) e_{N-1})} \right] .$$
In the above expression, we introduced $\phi_{N-1}$, a deterministic function of the state vectors at $N-1$ with the same quadratic properties as $\theta_N$.\textsuperscript{25} At this stage, we highlight two important properties: First, the expectation is conditional on the filtration, implying that the only stochastic term is $\epsilon_{s,N}$, which is normally distributed with mean 0. Second, the path-dependent term $x_{N-1}'\Delta_M u_{N-1}$ cancels out and is no longer present in the expression. These properties have important consequences on the path-independent structure of the optimal policy at $N-1$. To see this, we first need to compute the conditional expectation above. Then, using the monotonicity of the exponential, we can show that the optimal policy at $N-1$ can be obtained by solving the equivalent deterministic optimization problem given by

$$x_{N-1}^* = \arg \max_{x_{N-1}} \theta_{N-1} + \phi_{N-1} - \frac{1}{2}\alpha(x_{N-1}' - x_{N-1}')\Delta_M(x_{N-1} - \Delta_M x_{N-1}) - \frac{1}{2}\alpha x_{N-1}'\Delta_M \Sigma u \Delta_M x_{N-1}.$$  \hfill (24)

Here, observe that all the terms are deterministic and none of them depend on the filtration at time $N-1$. Therefore, path-independence is preserved at time $N-1$. In addition, if the above expression is concave in $x_{N-1}$, the optimal trade can be obtained by first order conditions.

\textbf{Proposition 2 (Equivalent Quadratic Program)}

Combining the equations for the mean (20) and variance (21), the problem (18) is written as:

$$\max_{x \geq 0} - (x' D + u_0' 1' \Delta_K x) + \frac{1}{2}s' \Delta_K x + \frac{1}{2}\alpha x' \Delta K \Sigma u \Delta K x$$

subject to $1' \Delta_K x = z_0$.

The above maximization problem over total wealth can be turned into a minimization problem over risk-adjusted execution shortfall by taking the difference between pre-execution and post-execution expected wealth $W_0 - \mu_{W_N}$, where the constant $W_0 = -u_0' z_0 = -u_0' 1' \Delta_K x$ (this term therefore cancels out in the objective function). Let $c' = \frac{1}{2}s' \Delta_K$ and $D^* = D + \frac{1}{2}\alpha \Delta K \Sigma u \Delta_K$. Since we have $x'D'x = x' \left( \frac{D^* + D'}{2} \right) x$, we set the symmetric form $\frac{1}{2}\bar{D} = \left( \frac{D^* + D'}{2} \right)$. So finally, the problem (18) is equivalent to

$$\min_{x \geq 0} \frac{1}{2}x' \bar{D} x + c'x$$

subject to $1' \Delta_K x = z_0$.

where $\bar{D} = \left( (D + \frac{1}{2}\alpha \Delta K \Sigma u \Delta_K') + (D + \frac{1}{2}\alpha \Delta K \Sigma u \Delta_K') \right)$.

\textbf{A.3 Additional Results}

\textbf{Effect of the Equilibrium Bid-Ask Spread / Proportional Transaction Costs}

Figure 16 highlights the sensitivity of the inactive asset OEP, to its steady-state bid-ask spread, in the example from Section 4.2. The OEP is plotted with values $s_2 = 0, 50, 100$ and 200 bps. The reference point is

\textsuperscript{25}More specifically, $\theta_N = \theta_N(z_N, d_N, x_N') = \theta_N(z_{N-1} - \Delta_M x_{N-1}, (d_{N-1} + \kappa x_{N-1})e^{-\rho T}, z_{N-1} - \Delta_M x_{N-1}) = \phi_{N-1}(z_{N-1}, d_{N-1}, x_{N-1}) = \phi_{N-1}$.
the inactive asset’s initial mid-price, \( v_{2,0} = \$1 \), so that 100 bps corresponds to a bid-ask spread of 1 cent. At a spread of 200 bps, any trading activity in the inactive asset is halted completely. The associated costs and total volume traded in the inactive asset are provided in Table 3.

![Graph showing the effect of the steady-state bid-ask spread on the inactive asset OEP from Figure 6(b).](image)

Table 3: Expected costs and volume traded for different values of \( s \).

<table>
<thead>
<tr>
<th>Bid-Ask Spread ( s ) (bps)</th>
<th>0</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected Costs ($)</td>
<td>2.12</td>
<td>2.41</td>
<td>2.68</td>
<td>3.17</td>
</tr>
<tr>
<td>Total Volume in Inactive Asset (Shares)</td>
<td>25.4</td>
<td>17.0</td>
<td>9.2</td>
<td>-</td>
</tr>
</tbody>
</table>

**Intra-Day Liquidity Variations and Time-Dependent Parameters**

![Graph showing RN OEP with an order book density decrease in (a) and a density increase in (b).](image)

Figure 17 shows the sensitivity of a single-asset OEP for a RN manager (\( \alpha = 0 \)) with a time-varying view on the order book densities.\(^{26}\) The fact that liquidity can predictably change at different time scales has been empirically documented (see e.g., Chordia, Roll & Subrahmanyam (2001)). We plot the OEP for respective changes in the value of \( q \), both lower (Figure 17(a)), and higher (Figure 17(b)), in the interval \([N/2, N]\).

There is a significant change of trading velocity both immediately preceding and following the change in liquidity. Furthermore, temporary “dead-zones” emerge around the time of the change in liquidity, where it becomes optimal to halt all trading activity. Intuitively, these indicate that the manager should wait until

\(^{26}\)A similar analysis is available on the order book replenishment rates upon request.
the liquidity changes are fully absorbed by the order books and supply/demand converges to the new regime before finishing off the remaining orders.

References


