

Optimal Control of Execution Costs for Portfolios*

Dimitris Bertsimas[†], Paul Hummel[‡], and Andrew W. Lo[§]

This Draft: September 9, 1999

Abstract

We derive dynamic optimal trading strategies that minimize the expected cost of trading blocks of securities over a fixed time horizon. Given fixed blocks \bar{s}_i of shares of stock i to be traded within a finite number of periods T , $i = 1, \dots, n$, and given price-impact functions that yield the execution price of an individual trade as a function of the shares of stock i traded and current market conditions, we obtain the optimal *sequence* of trades as a function of market conditions—closed-form expressions in some cases—that minimizes the expected cost of executing \bar{s}_i within T periods. We also propose an approximation algorithm for incorporating constraints, a particularly important extension in practice. To illustrate the practical relevance of our methods, we apply them to a hypothetical portfolio of 25 stocks by estimating their price-impact functions using historical trade data from 1996 and deriving the optimal execution strategies for the portfolio, and by performing several Monte Carlo simulation experiments.

*This research is part of the MIT Laboratory for Financial Engineering's Transactions Costs Project. We are grateful to Investment Technology Group, the National Science Foundation (Grant Nos. DMI-9610486 and SBR-9709976), and the sponsors of the LFE for financial support. We thank Dave Cushing, Chris Darnell, Robert Ferstenberg, Rohit D'Souza, John Heaton, Leonid Kogan, Bruce Lehmann, Greg Peterson, Jiang Wang, and seminar participants at Columbia, Harvard, London Business School, MIT, Northwestern, the NYSE Conference on Best Execution, the Society of Quantitative Analysts, and Yale for helpful comments and discussion.

[†]Boeing Professor of Operations Research, MIT Sloan School of Management and Operations Research Center, Massachusetts Institute of Technology, E53-359, Cambridge, MA 02142-1347.

[‡]Portfolio Strategist, Long Term Capital Management, One East Weaver St., Greenwich, CT 06831.

[§]Harris & Harris Group Professor, MIT Sloan School of Management and Operations Research Center, Cambridge, MA 02142-1347.

Contents

1	Introduction	1
2	The State Equations	4
2.1	Linear Percentage Price-Impact	4
2.2	Other Specifications	6
3	The Dynamic Programming Solution	7
3.1	Linear-Percentage Price-Impact	8
3.2	Linear Price-Impact	9
4	Imposing Constraints	10
4.1	Why Constraints Are Problematic	10
4.2	A Static Approximation Method	11
5	An Empirical Example	13
5.1	The Data	14
5.2	The Estimation Procedure	14
5.3	No-Arbitrage Constraints	19
5.4	Monte Carlo Analysis	20
6	Conclusion	22
A	Appendix	23
	References	28

1 Introduction

The rapid growth in equity investing, driven by the increasing popularity of mutual funds and defined-contribution retirement plans, has led to a rising concentration of assets among institutional money managers. A typical portfolio manager now oversees a large portfolio of several hundred securities, with individual positions that might comprise a significant fraction of the average daily volume of the security. Both active managers and passive indexers must frequently rebalance their portfolios, either to include new stock picks, sell stocks that are out of favor, or to improve the tracking of a given index or benchmark. This generates sizeable orders across many stocks that must be executed within a relatively short time horizon, and that must be executed together so as to maintain the risk/reward characteristics of the portfolio. The transactions costs associated with trading such “lists” of securities—often called *execution* costs—can be substantial.

Execution costs are comprised of several components: explicit costs such as commissions and bid/ask spreads, and costs that are harder to quantify such as the opportunity cost of waiting and the price impact from trading. Opportunity costs arise because market prices are moving constantly and can move favorably or unfavorably without warning, generating unexpected profits or lost opportunities while a portfolio manager hesitates. Price impact is the typically unfavorable effect on prices that the very act of trading creates, not unlike the turbulence that is generated by a ship’s own wake. The seller of a security will, by the very act of selling, push down the security’s price, yielding lower proceeds from the sale, and similarly for the buyer. Moreover, the larger the order, the more heavily is the price affected by the trade. For portfolios that turn over frequently or with large positions to trade, these costs can add a significant drag on the fund’s overall performance (see, for example, Loeb (1983)).

Recent studies by Chan and Lakonishok (1995) and Keim and Madhavan (1995a–c) show that institutional investors often break up their larger trades into smaller “packages” that they execute over the course of several days. There is a compelling economic rationale for “package” trading. Trading is fundamentally a dynamic, path-dependent, stochastic problem insofar as it takes time to trade and the very act of trading affects price and price dynamics which, in turn, affect execution costs. Controlling execution costs of large blocks

of stock must be accomplished by trading over a number of time periods, a fact recognized by Bertsimas and Lo (1998) who used stochastic dynamic programming techniques to derive optimal or *best-execution* strategies.

In this paper, we extend the single-asset framework outlined by Bertsimas and Lo (1998) to construct best-execution strategies for portfolio problems. Specifically, we solve the following problem: Given fixed blocks of shares in n stocks, $\bar{\mathbf{s}} = [\bar{s}_1 \ \bar{s}_2 \ \cdots \ \bar{s}_n]'$,¹ to be purchased within a fixed finite number of periods T , and given a set of price dynamics that capture price impact, i.e., the execution price of an individual trade as a function of the shares traded and other “state” variables, find the optimal *sequence* of trades (as a function of the state variables) that minimizes the expected execution costs.² Since it is well-known that the short-term demand curves for even the most actively traded equities are not perfectly elastic (see, for example, Shleifer (1986)), a market order at date 0 for the entire block $\bar{\mathbf{s}}$ is clearly not an optimal trading strategy.

Denote by $\mathbf{s}_t = \{s_{1t}, s_{2t}, \dots, s_{nt}\}$ the number of shares of each stock acquired in period t at prices $\mathbf{p}_t = \{p_{1t}, p_{2t}, \dots, p_{nt}\}$, where $t = 1, \dots, T$. Then the investor’s objective may be expressed as:

$$\text{Min}_{\{\mathbf{s}_t\}} \text{E}_1 \left[\sum_{t=1}^T \mathbf{p}'_t \mathbf{s}_t \right] \quad (1.1)$$

subject to:

$$\mathbf{p}_t = f(\mathbf{p}_{t-1}, \mathbf{x}_t, \mathbf{s}_t, \boldsymbol{\epsilon}_t) \quad (1.2a)$$

$$\mathbf{x}_t = g(\mathbf{x}_{t-1}, \boldsymbol{\eta}_t) \quad (1.2b)$$

$$\sum_{t=1}^T \mathbf{s}_t = \bar{\mathbf{s}} \quad (1.2c)$$

where \mathbf{x}_t is a vector of state variables, $\boldsymbol{\epsilon}_t$ is vector white noise, and $f(\cdot)$ and $g(\cdot)$ are the state equations or laws of motion that incorporates both price dynamics of \mathbf{p}_t , the price impact of

¹We follow the common convention that all vectors are *column* vectors unless they are explicitly transposed, and boldface Roman and Greek letters denote vectors and matrices.

²For simplicity and without loss in generality, we consider the case of *purchasing* $\bar{\mathbf{s}}$ only. Selling $\bar{\mathbf{s}}$ and a combination of buying certain stocks and selling others can easily be accommodated with the appropriate sign conventions (positive numbers for purchases, negative for sales).

trading \mathbf{s}_t , and the dynamics of the state variables. We may also wish to impose additional constraints, e.g., a no-sales constraint, $\mathbf{s}_t \geq \mathbf{0}$,³ or other conditions that are placed upon the portfolio manager by institutional restrictions, tax considerations, or other aspects of his investment process.

The portfolio case contains several interesting features not captured by the single-stock analysis of Bertsimas and Lo (1998) and others. Perhaps the most important feature is the ability to capture cross-stock relations such as the cross-autocorrelations reported in Lo and MacKinlay (1990). Price movements in one stock can induce similar movements in the price of another, either because of common factors driving both, or because of linked trading strategies, e.g., pairs trading, index arbitrage, risk arbitrage, etc. In such cases, the price impact of trading a portfolio may be larger than the sum of the price impact of trading the individual stocks separately. Alternatively, if some stocks are negatively correlated (perhaps because of portfolio substitution effects), or if the portfolio to be executed includes both purchases and sales, then the portfolio execution cost may be *lower* than the sum of the individual stocks' execution costs due to a kind of diversification effect in which the trades of one stock lower the price impact of trades in another. Whether execution costs are magnified or mollified in the portfolio case is, of course, an empirical issue that turns on the law of motion for the vector of prices and state variables. In either case, it is clear that the portfolio setting is considerably more complex than the single-stock case.

In Section 2 specify the state equations and explore in greater detail one particular specification that captures cross-stock interactions for a portfolio of stocks, allows for predictability in asset returns, and incorporates a “temporary” price-impact effect. The best-execution strategy is derived in closed form in Section 3 using stochastic dynamic programming, and in Section 4 we propose an approximation algorithm that provides a near-optimal execution strategy for the case with additional constraints. In Section 5 we evaluate the practical relevance of our approach by applying it to a portfolio of 25 stocks using historical transactions date from 1996 and estimating the execution costs for such a portfolio under several different

³If a portfolio manager is attempting to acquire a block of securities, it is difficult to justify selling the same securities during the acquisition period (unless, of course, the manager has extremely accurate negative information regarding the security's price, which is somewhat inconsistent with the original premise that he is a buyer). Indeed, in many cases, it is illegal because it is viewed as a violation of the fiduciary trust that portfolio managers have to act in the best interests of their investors.

trading scenarios. We conclude in Section 6.

2 The State Equations

In this section we present a specification for the state equations (1.2)–(1.2) that incorporates a multivariate price-impact function with cross-stock interactions (Section 2.1), and discuss other possible specifications afterwards (Section 2.2).

2.1 Linear Percentage Price-Impact

Let the execution price p_t be the sum of two components:

$$\mathbf{p}_t = \tilde{\mathbf{p}}_t + \boldsymbol{\delta}_t \tag{2.1}$$

where $\tilde{\mathbf{p}}_t$ may be viewed as a “no-impact” price—the price that would prevail in the absence of any market impact—and the impact $\boldsymbol{\delta}_t$. A plausible and observable proxy for the no-impact price is the midpoint of the bid/offer spread, although it can be arbitrary as long as it is unaffected by the trade size \mathbf{s}_t . For convenience, and to ensure non-negative prices, we model $\tilde{\mathbf{p}}_t$ as vector-geometric-Brownian-motion:

$$\tilde{\mathbf{p}}_t = \exp(\mathbf{Z}_t)\tilde{\mathbf{p}}_{t-1} \tag{2.2}$$

where \mathbf{Z}_t is a diagonal matrix whose diagonal is a normal random vector \mathbf{z}_t with mean $\boldsymbol{\mu}_z$ and covariance matrix $\boldsymbol{\Sigma}_z$. The $\exp(\cdot)$ operator denotes the matrix exponential which, in this case, reduces to the element-wise exponential of the diagonal entries in \mathbf{Z}_t .

For the price-impact term, $\boldsymbol{\delta}_t$, we set:

$$\boldsymbol{\delta}_t = \tilde{\mathbf{P}}_t(\mathbf{A}\tilde{\mathbf{P}}_t\mathbf{s}_t + \mathbf{B}\mathbf{x}_t) \tag{2.3}$$

where $\tilde{\mathbf{P}}_t = \text{diag}[\tilde{\mathbf{p}}_t]$ and $\text{diag}(\cdot)$ is the diagonalization operator which maps its vector argument into a diagonal matrix with the vector as its diagonal. This specification captures the impact of trading \mathbf{s}_t shares on the transaction prices \mathbf{p}_t , and implies that as a percentage of

the no-impact price, $\tilde{\mathbf{p}}_t$, the price impact is a linear function of the dollar-value of the trade and other state variables \mathbf{x}_t . The form of the price impact (2.3) differs from the single-stock case in that the percentage price-impact function for each stock i is a linear function, $\mathbf{A}\tilde{\mathbf{P}}_t\mathbf{s}_t$, of the dollar-values of the trades of *all* n stocks, not just of stock i . In the special case where \mathbf{A} is diagonal, the portfolio problem reduces to n independent single-stock problems solved by Bertsimas and Lo (1998).

This specification of the dynamics of \mathbf{p}_t has a number of advantages over other specifications (see Section 2.2). First, $\tilde{\mathbf{p}}_t$ is guaranteed to be non-negative, and \mathbf{p}_t is also guaranteed to be non-negative under mild restrictions on $\boldsymbol{\delta}_t$. Second, by separating the transaction price \mathbf{p}_t into a no-impact component $\tilde{\mathbf{p}}_t$ and the impact component $\boldsymbol{\delta}_t$, the price impact of a trade is temporary, affecting the current transaction price but having no effect on future prices. Third, the percentage price impact increases linearly with the trade size, which is empirically plausible (see, for example, Loeb (1983, Table II), Birinyi (1995), or Leinweber (1993, 1994)). Fourth, (2.1) implies a natural decomposition of execution costs, decoupling market-microstructure effects from price dynamics, which is closely related to Pérold’s (1988) notion of implementation shortfall.⁴ And finally, we shall see in Section 3 that (2.1) admits a closed-form solution in which the best-execution strategy is a simple linear function of the state variables and the optimal-value function is quadratic.

The presence of the vector \mathbf{x}_t in (2.3) captures the potential influence of changing market conditions or private information about the securities on the price impact $\boldsymbol{\delta}_t$. For example, \mathbf{x}_t might be the return on the S&P 500 index, a common component in the price of most equities. We model \mathbf{x}_t as a vector with r elements, allowing for multiple sources of information to influence execution prices (or several lags of a single state variable). To complete our specification of the state equation, we must specify the dynamics of \mathbf{x}_t , and for simplicity we let:

$$\mathbf{x}_t = \mathbf{C}\mathbf{x}_{t-1} + \boldsymbol{\eta}_t \tag{2.4}$$

⁴Because (2.1) implies that price impact is a temporary phenomenon, affecting only the current trade price \mathbf{p}_t and not the no-impact price level $\tilde{\mathbf{p}}_t$, the objective function (1.1) separates into two terms. The first term is the no-impact cost of execution and the second term is the total impact cost. This decomposition is precisely the one proposed by Pérold (1988) in his definition of implementation shortfall, but now applied to executing $\bar{\mathbf{s}}$. In particular, the “paper” return or execution cost is given by the first term, the actual cost is given by the sum of the two terms, hence the second term is the implementation shortfall in executing $\bar{\mathbf{s}}$.

where $\boldsymbol{\eta}_t$ is vector white noise with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_\eta$. The fact that \boldsymbol{x}_t is a vector autoregressive process with one lag (an “AR(1)”) allows us to capture varying degrees of predictability in information or market conditions. The matrices \mathbf{A} and \mathbf{B} measure the sensitivity of price impact to trade size and market conditions. Matrix \mathbf{A} must be positive definite, \mathbf{B} is arbitrary, and must \mathbf{C} have eigenvalues less than unity in modulus (to ensure stationarity of \boldsymbol{x}_t).

2.2 Other Specifications

The specification for the state equation proposed in Section 2.1 is only one of many possible specifications. For example, Bertsimas and Lo (1998) propose a natural multivariate extension of their “linear price impact” specification in which the state equation is given by:

$$\boldsymbol{p}_t = \boldsymbol{p}_{t-1} + \mathbf{A}\boldsymbol{s}_t + \mathbf{B}\boldsymbol{x}_t + \boldsymbol{\epsilon}_t \quad (2.5)$$

where \mathbf{A} is a positive definite ($n \times n$)-matrix, \mathbf{B} is an arbitrary ($n \times m$)-matrix, \boldsymbol{x}_t is an m -vector of information variables, and $\boldsymbol{\epsilon}_t$ is n -vector white noise with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_\epsilon$. As before, we assume that \boldsymbol{x}_t follows a stationary AR(1) process, hence:

$$\boldsymbol{x}_t = \mathbf{C}\boldsymbol{x}_{t-1} + \boldsymbol{\eta}_t \quad (2.6)$$

where \mathbf{C} is an ($m \times m$)-matrix with eigenvalues all less than unity in modulus, and $\boldsymbol{\eta}_t$ is m -vector white noise with mean $\mathbf{0}$, covariance matrix $\boldsymbol{\Sigma}_\eta$, and which is independent of $\boldsymbol{\epsilon}_t$.

This specification differs significantly from the linear-percentage price impact of Section 2.1 in a number of respects. Perhaps the most important difference is the fact that (2.5) implies that the price impact has a “permanent” effect on prices, due to the random-walk nature of the price dynamics. Also, the price impact is linear in price, implying that a 10,000-share trade will have the same dollar impact on \$1-stock as it will on a \$100-stock. Finally, unless some rather unnatural restrictions are placed on $\boldsymbol{\epsilon}_t$ in (2.5), there is a positive probability that \boldsymbol{p}_t can take on negative values, clearly an unrealistic prospect.

Of course, Bertsimas and Lo (1998) consider (2.5) primarily for analytic tractability, not because of supporting empirical evidence. Nevertheless, it is instructive to compare this

specification with the linear-percentage specification of Section 2.1 to develop some intuition for the more practical issues in specifying the state equation. In practice, the state equation must be estimated empirically. For example, several empirical studies seem to point to both permanent and temporary price-impact in US equity data, e.g., Kraus and Stoll (1972), Holthausen, Leftwich, and Mayers (1987, 1990), Barclay and Litzenberger (1988), Barclay and Warner (1993), and Chan and Lakonishok (1993, 1995). However, given the ever-changing nature of financial markets, it is crucial to re-estimate the state equation for each application using the most recent datasets available.

3 The Dynamic Programming Solution

We use a stochastic dynamic programming algorithm to solve the optimal execution problem (1.1). Denote by \mathbf{w}_t the vector of shares remaining to be bought (or sold) at time t :

$$\mathbf{w}_t = \mathbf{w}_{t-1} - \mathbf{s}_{t-1} \quad , \quad \mathbf{w}_1 = \bar{\mathbf{s}} \quad , \quad \mathbf{w}_{T+1} = \mathbf{0} . \quad (3.7)$$

The condition $\mathbf{w}_{T+1} = \mathbf{0}$ ensures that all $\bar{\mathbf{s}}$ shares are executed by time T . The complete statement of the problem is then:

$$\text{Min}_{\{\mathbf{s}_t\}} \text{E}_1 \left[\sum_{t=1}^T \mathbf{p}'_t \mathbf{s}_t \right]$$

subject to:

$$\begin{aligned} \mathbf{p}_t &= f(\mathbf{p}_{t-1}, \mathbf{x}_t, \mathbf{s}_t, \boldsymbol{\epsilon}_t) \\ \mathbf{x}_t &= g(\mathbf{x}_{t-1}, \boldsymbol{\eta}_t) \\ \mathbf{w}_t &= \mathbf{w}_{t-1} - \mathbf{s}_{t-1} \\ \mathbf{w}_1 &= \bar{\mathbf{s}} \\ \mathbf{w}_{T+1} &= \mathbf{0} \end{aligned}$$

3.1 Linear-Percentage Price-Impact

As with all dynamic programming solutions, we begin at the end. Denote by V_T the optimal value function at the end of our trading horizon, period T . By definition:

$$V_T(\tilde{\mathbf{p}}_T, \mathbf{x}_T, \mathbf{w}_T) = \underset{\{\mathbf{s}_T\}}{\text{Min}} \mathbb{E}_T \left[\mathbf{p}'_T \mathbf{s}_T \right] = \mathbb{E}_T \left[\mathbf{p}'_T \mathbf{w}_T \right] \quad (3.8)$$

$$= \left[\tilde{\mathbf{P}}_T (\mathbf{e}_n + \mathbf{A} \tilde{\mathbf{P}}_T \mathbf{w}_T + \mathbf{B} \mathbf{x}_T) \right]' \mathbf{w}_T. \quad (3.9)$$

Since this is the last period and \mathbf{w}_{T+1} must be set to zero, there is no choice but to execute the remaining order \mathbf{w}_T , hence the optimal trade size $\mathbf{s}_T^* = \mathbf{w}_T$. Observe that $\tilde{\mathbf{P}}_T = \tilde{\mathbf{P}}_T'$, hence (3.9) can be re-expressed as:

$$V_T(\tilde{\mathbf{p}}_T, \mathbf{x}_T, \mathbf{w}_T) = \mathbf{e}'_n \tilde{\mathbf{P}}_T \mathbf{w}_T + \mathbf{w}'_T \tilde{\mathbf{P}}_T \mathbf{A}' \tilde{\mathbf{P}}_T \mathbf{w}_T + \mathbf{x}'_T \mathbf{B}' \tilde{\mathbf{P}}_T \mathbf{w}_T \quad (3.10)$$

which shows that the optimal value function is linear in \mathbf{x}_T and linear-quadratic in \mathbf{w}_T . By continuing recursively in this fashion and applying Bellman's principle of optimality (see Bertsekas (1995) for example), we find that the optimal value function V_{t-k} is given by:

$$V_{T-k} = \underset{\{\mathbf{s}_{T-k}\}}{\text{Min}} \mathbb{E}_{T-k} \left[\mathbf{p}'_{T-k} \mathbf{s}_{T-k} + V_{T-k+1}(\tilde{\mathbf{p}}_{T-k+1}, \mathbf{x}_{T-k+1}, \mathbf{w}_{T-k+1}) \right] \quad (3.11)$$

$$\begin{aligned} &= \mathbf{e}'_n \mathbf{D}_{n,k} \mathbf{e}_n + \mathbf{e}'_r \mathbf{D}_{r,k} \mathbf{e}_r + \mathbf{e}'_n \mathbf{E}_k \mathbf{x}_{T-k} + \mathbf{x}'_{T-k} \mathbf{F}_k \mathbf{e}_n + \\ &\quad \mathbf{x}'_{T-k} \mathbf{G}_k \mathbf{x}_{T-k} + \mathbf{x}'_{T-k} \mathbf{H}_k \tilde{\mathbf{w}}_{T-k} + \\ &\quad \tilde{\mathbf{w}}'_{T-k} \mathbf{J}_{T-1} \mathbf{x}_{T-k} + \mathbf{e}'_n \mathbf{K}_k \tilde{\mathbf{w}}_{T-k} + \tilde{\mathbf{w}}'_{T-k} \mathbf{L}_k \mathbf{e}_n + \\ &\quad \tilde{\mathbf{w}}'_{T-k} \mathbf{N}_k \tilde{\mathbf{w}}_{T-k} \end{aligned} \quad (3.12)$$

where explicit expressions for $\mathbf{D}_{n,k}$, $\mathbf{D}_{r,k}$, \mathbf{E}_k , \mathbf{F}_k , \mathbf{G}_k , \mathbf{H}_k , \mathbf{J}_k , \mathbf{K}_k , \mathbf{L}_k , and \mathbf{N}_k are given in the Appendix. This yields the best-execution strategy:

$$\tilde{\mathbf{s}}_{T-k}^* = \mathbf{\Delta}_{x,k} \mathbf{x}_{T-k} + \mathbf{\Delta}_{w,k} \tilde{\mathbf{w}}_{T-k} + \mathbf{\Delta}_{1,k} \mathbf{e}_n \quad (3.13)$$

where $\mathbf{\Delta}_{x,k}$, $\mathbf{\Delta}_{w,k}$, and $\mathbf{\Delta}_{1,k}$, are also given in the Appendix. The recursion (3.12) and best-execution strategy (3.13) completely characterizes the solution to our original problem, and

also yields the expected cost of best execution, V_{T-k} , as a by-product.

3.2 Linear Price-Impact

Under the law of motion (2.5)–(2.6), Bertsimas and Lo (1998) show that the portfolio problem (1.1) can be solved via Bellman’s equation, which yields the following best-execution strategy and optimal-value function:

$$\mathbf{s}_{T-k}^* = (\mathbf{I} - \frac{1}{2}\mathbf{A}_{k-1}^{-1}\mathbf{A}')\mathbf{w}_{T-k} + \frac{1}{2}\mathbf{A}_{k-1}^{-1}\mathbf{B}'_{k-1}\mathbf{C}\mathbf{x}_{T-k} \quad (3.14)$$

$$\begin{aligned} V_{T-k}(\mathbf{p}_{T-k-1}, \mathbf{x}_{T-k}, \mathbf{w}_{T-k}) &= \mathbf{p}'_{T-k-1}\mathbf{w}_{T-k} + \mathbf{w}'_{T-k}\mathbf{A}_k\mathbf{w}_{T-k} + \\ &\quad \mathbf{x}'_{T-k}\mathbf{B}_k\mathbf{w}_{T-k} + \mathbf{x}'_{T-k}\mathbf{C}_k\mathbf{x}_{T-k} + d_k \end{aligned} \quad (3.15)$$

for $k = 0, \dots, T-1$, where:

$$\begin{aligned} \mathbf{A}_k &= \mathbf{A} - \frac{1}{4}\mathbf{A}\mathbf{A}_{k-1}^{-1}\mathbf{A}' & , & & \mathbf{A}_0 &= \mathbf{A} \\ \mathbf{B}_k &= \frac{1}{2}\mathbf{C}'\mathbf{B}_{k-1}(\mathbf{A}'_{k-1})^{-1}\mathbf{A}' + \mathbf{B}' & , & & \mathbf{B}_0 &= \mathbf{B}' \\ \mathbf{C}_k &= \mathbf{C}'\mathbf{C}_{k-1}\mathbf{C} - \frac{1}{4}\mathbf{C}'\mathbf{B}_{k-1}(\mathbf{A}'_{k-1})^{-1}\mathbf{B}_{k-1}\mathbf{C} & , & & \mathbf{C}_0 &= \mathbf{0} \\ d_k &= d_{k-1} + \mathbb{E}[\eta'_{T-k}\mathbf{C}_{k-1}\eta_{T-k}] & , & & d_0 &= 0 . \end{aligned} \quad (3.16)$$

The best-execution strategy (3.14) is qualitatively similar to the optimal single-stock strategy that Bertsimas and Lo (1998) derive, but there is one key difference: in the portfolio case, unless the matrix \mathbf{A} is diagonal, the best-execution strategy for one stock will depend on the parameters and state variables of all the other stocks. To see this, observe that the matrix coefficient $(\mathbf{I} - \frac{1}{2}\mathbf{A}_{k-1}^{-1}\mathbf{A}')$ multiplying \mathbf{w}_{T-k} in (3.14) will generally not be a diagonal matrix unless \mathbf{A} is itself diagonal. Of course, if \mathbf{A} is diagonal this means that trading in one stock has no price impact on any other stocks (see (2.5)), hence the portfolio problem essentially reduces to n independent single-stock problems.

For this reason, whether or not the portfolio best execution cost is greater or less than the sum of the individual stocks’ best execution costs depends wholly on the values in \mathbf{A} and is an empirical issue which we shall consider in detail in Section 5.

4 Imposing Constraints

For most practical applications, there will be constraints on the kind of execution strategies that institutional investors can follow. For example, if a block of shares is to be purchased within T periods, it is very difficult to justify selling the stock during these T periods even if such sales are warranted by the best-execution strategy.⁵ Therefore, in practice buy-programs (sell-programs) will almost always be accompanied by non-negativity (non-positivity) constraints. Such constraints are often binding for best-execution strategies, particularly when the information variable has a large effect on price impact. In Section 4.1 we illustrate the difficulties of imposing constraints, and propose an approximate solution method in Section 4.2.

4.1 Why Constraints Are Problematic

Although there are well-known techniques for solving constrained-optimization problems in a static setting, corresponding techniques for *dynamic* optimization problems have not yet been developed. To see why this is such a difficult task, consider the simplest case of imposing non-negativity restrictions $s_t \geq 0$ in the linear percentage price-impact model of section 2.1 with only one asset (scalar equations). Without any constraints, the optimal-value function V_{T-k} is quadratic in the state variable w_{T-k} , hence the Bellman equation can be easily solved in closed form. But if non-negativity constraints $s_t \geq 0$ are imposed, V_{T-k} becomes a piecewise-quadratic function, with 3^k pieces.

To see how this arises, observe that for $k = 0$ the optimal control is $s_T^* = w_T$ and V_T is a quadratic function of w_T . In the next stage, $k = 1$, we calculate the optimal control s_{T-1}^* by minimizing a quadratic function of s_{T-1} subject to the constraints $0 \leq s_{T-1} \leq w_{T-1}$. The

⁵Other common constraints include sector-balance constraints, turnover constraints, tax-motivated constraints, and, in the portfolio case, dollar-balance constraints. This last type of constraint—the dollar-value of the portfolio at the end of trading lies within some fixed interval—is one of the most difficult to impose because the constraint is a function of the entire vector of prices which is stochastic. See Bertsimas and Lo (1998) for a probabilistic method of imposing such constraints.

solution is given by:

$$s_{T-1}^* = \begin{cases} 0 & \text{if } s_{u,T-1} < 0 \\ s_{u,T-1} & \text{if } 0 < s_{u,T-1} < w_{T-1} \\ w_{T-1} & \text{if } s_{u,T-1} > w_{T-1} \end{cases} \quad (4.1)$$

where:

$$s_{u,T-1} = \frac{1}{\tilde{p}_{T-1}} [\Delta_{w,1} \tilde{p}_{T-1} w_{T-1} + \Delta_{x,1} x_{T-1} + \Delta_{1,1}] . \quad (4.2)$$

This partitions the range of w_{T-1} into three intervals; over each interval there is a different optimal control s_{T-1}^* , and within each interval V_{T-1} is a continuous quadratic function of w_{T-1} . At the next stage, $k = 2$, each of these three intervals is partitioned into another three intervals, each with a different optimal control s_{T-2}^* , and so on, the number of intervals growing exponentially with k . Therefore, even in this simple case, calculating s_{T-k}^* and V_{T-k} exactly is only feasible for a very small number of periods T (for example, when $T = 20$ there are $3^{20} = 3,486,784,401$ intervals at the last stage of the dynamic program!).

Faced with these difficulties we propose in the next section an approximation method to address the optimal control problem with constraints.

4.2 A Static Approximation Method

The dynamic optimization algorithm we presented in Section 3 for the case without non-negativity constraints gives the best-execution strategy $\tilde{\mathbf{s}}_{T-k}^*$ (see (3.13)) as a function of the state vector $(\mathbf{x}_{T-k}, \tilde{\mathbf{w}}_{T-k}, \tilde{\mathbf{p}}_{T-k})$ at time $T-k$. At time $t = 1$ the expected execution cost $E[J]$ is given by:

$$\begin{aligned} E[J] &= E \left[\sum_{k=1}^T \mathbf{p}'_k \mathbf{s}_k \right] \\ &= E \left[\sum_{k=1}^T [\tilde{\mathbf{P}}_k \mathbf{e}_n + \tilde{\mathbf{P}}_k (\mathbf{A} \tilde{\mathbf{P}}_k \mathbf{s}_k + \mathbf{B} \mathbf{x}_k)]' \mathbf{s}_k \right] \\ &= E \left[\mathbf{e}'_n \tilde{\mathbf{P}}_1 \mathbf{s}_1 + \mathbf{s}'_1 \tilde{\mathbf{P}}_1 \mathbf{A}' \tilde{\mathbf{P}}_1 \mathbf{s}_1 + \mathbf{x}'_1 \mathbf{B}' \tilde{\mathbf{P}}_1 \mathbf{s}_1 + \mathbf{e}'_n \tilde{\mathbf{Z}}_2 \tilde{\mathbf{P}}_1 \mathbf{s}_2 + \right. \\ &\quad \left. \mathbf{s}'_2 \tilde{\mathbf{P}}_1 \tilde{\mathbf{Z}}_2 \mathbf{A}' \tilde{\mathbf{Z}}_2 \tilde{\mathbf{P}}_1 \mathbf{s}_2 + (\mathbf{C} \mathbf{x}_1 + \boldsymbol{\eta}_2)' \mathbf{B}' \tilde{\mathbf{Z}}_2 \tilde{\mathbf{P}}_1 \mathbf{s}_2 + \right. \end{aligned}$$

$$\begin{aligned}
& e'_n \tilde{\mathbf{Z}}_2 \tilde{\mathbf{Z}}_3 \tilde{\mathbf{P}}_1 \mathbf{s}_3 + s'_3 \tilde{\mathbf{P}}_1 \tilde{\mathbf{Z}}_3 \tilde{\mathbf{Z}}_2 \mathbf{A}' \tilde{\mathbf{Z}}_2 \tilde{\mathbf{Z}}_3 \tilde{\mathbf{P}}_1 \mathbf{s}_3 + \\
& [\mathbf{C}(\mathbf{C}\mathbf{x}_1 + \boldsymbol{\eta}_2) + \boldsymbol{\eta}_3]' \mathbf{B}' \tilde{\mathbf{Z}}_2 \tilde{\mathbf{Z}}_3 \tilde{\mathbf{P}}_1 \mathbf{s}_3 + \\
& \cdots + \\
& e'_n \tilde{\mathbf{Z}}_2 \tilde{\mathbf{Z}}_3 \cdots \tilde{\mathbf{Z}}_T \tilde{\mathbf{P}}_1 \mathbf{s}_T + \\
& s'_T \tilde{\mathbf{P}}_1 \tilde{\mathbf{Z}}_T \tilde{\mathbf{Z}}_{T-1} \cdots \tilde{\mathbf{Z}}_2 \mathbf{A}' \tilde{\mathbf{Z}}_2 \tilde{\mathbf{Z}}_3 \cdots \tilde{\mathbf{Z}}_T \tilde{\mathbf{P}}_1 \mathbf{s}_T + \\
& [\mathbf{C}^T \mathbf{x}_1 + \mathbf{C}^{T-1} \boldsymbol{\eta}_2 + \dots + \mathbf{C} \boldsymbol{\eta}_T + \boldsymbol{\eta}_T]' \times \\
& \left. \mathbf{B}' \tilde{\mathbf{Z}}_2 \tilde{\mathbf{Z}}_3 \cdots \tilde{\mathbf{Z}}_T \tilde{\mathbf{P}}_1 \mathbf{s}_T \right] \tag{4.3}
\end{aligned}$$

where $\tilde{\mathbf{P}}_k = \text{diag}[\mathbf{p}_k]$ and $\tilde{\mathbf{Z}}_k = \exp(\mathbf{Z}_k)$. Taking the expectation of the cost function yields:

$$\begin{aligned}
\mathbb{E}[J] &= e'_n \tilde{\mathbf{P}}_1 \mathbf{s}_1 + s'_1 \mathbf{A}' \tilde{\mathbf{P}}_1 \mathbf{s}_1 + \mathbf{x}'_1 \mathbf{B}' \tilde{\mathbf{P}}_1 \mathbf{s}_1 + \\
& e'_n \mathbf{Q} \tilde{\mathbf{P}}_1 \mathbf{s}_2 + s'_2 (\mathbf{A}' \cdot \mathbf{R}) \tilde{\mathbf{P}}_1 \mathbf{s}_2 + \mathbf{x}'_1 \mathbf{C}' \mathbf{B}' \mathbf{Q} \tilde{\mathbf{P}}_1 \mathbf{s}_2 + \\
& e'_n \mathbf{Q}^2 \tilde{\mathbf{P}}_1 \mathbf{s}_3 + s'_3 (\mathbf{A}' \cdot \mathbf{R} \cdot \mathbf{R}) \tilde{\mathbf{P}}_1 \mathbf{s}_3 + \mathbf{x}'_1 (\mathbf{C}^2)' \mathbf{B}' \mathbf{Q}^2 \tilde{\mathbf{P}}_1 \mathbf{s}_3 + \\
& \dots + \\
& e'_n \mathbf{Q}^T \tilde{\mathbf{P}}_1 \mathbf{s}_T + s'_T (\mathbf{A}' \cdot \mathbf{R} \cdot \mathbf{R} \cdot \dots \cdot \mathbf{R} \cdot \mathbf{R}) \tilde{\mathbf{P}}_1 \mathbf{s}_T + \\
& \mathbf{x}'_1 (\mathbf{C}^T)' \mathbf{B}' \mathbf{Q}^T \tilde{\mathbf{P}}_1 \mathbf{s}_T \tag{4.4}
\end{aligned}$$

where \mathbf{Q} is a $(n \times n)$ diagonal matrix with entries

$$q_i = \exp(\mu_{z,i} + \frac{1}{2} \Sigma_{z,ii}), \tag{4.5}$$

\mathbf{R} is an $(n \times n)$ symmetric matrix with elements:

$$r_{ij} = \exp[\mu_{z,i} + \mu_{z,j} + \frac{1}{2}(\Sigma_{z,ii} + \Sigma_{z,jj} + 2\Sigma_{z,ij})], \tag{4.6}$$

and the matrix dot operator ' \cdot ' denotes an element-wise matrix multiplication, i.e., $\mathbf{A} \cdot \mathbf{B} \equiv [a_{ij} b_{ij}]$.

The expression (4.4) depends on the entire sequence of controls, $\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_T\}$ and the observed states at time $t=1$, $\tilde{\mathbf{p}}_1$ and \mathbf{x}_1 . In general, each control variable \mathbf{s}_t depends on the state at time t . Under the *static approximation approach*, we will restrict the class of

controls to those that \mathbf{s}_t depend *only* on the state at time $t = 1$, i.e., they only depend on prices $\tilde{\mathbf{p}}_1$ and information vector \mathbf{x}_1 . Under this approximation the problem reduces to the following quadratic optimization problem:

$$\begin{aligned}
\text{Minimize} \quad & \mathbf{e}'_n \tilde{\mathbf{P}}_1 \mathbf{s}_1 + \mathbf{s}'_1 \mathbf{A}' \tilde{\mathbf{P}}_1 \mathbf{s}_1 + \mathbf{x}'_1 \mathbf{B}' \tilde{\mathbf{P}}_1 \mathbf{s}_1 + \mathbf{e}'_n \mathbf{Q} \tilde{\mathbf{P}}_1 \mathbf{s}_2 + \\
& \mathbf{s}'_2 (\mathbf{A}' \cdot \mathbf{R}) \tilde{\mathbf{P}}_1 \mathbf{s}_2 + \mathbf{x}'_1 \mathbf{C}' \mathbf{B}' \mathbf{Q} \tilde{\mathbf{P}}_1 \mathbf{s}_2 + \mathbf{e}'_n \mathbf{Q}^2 \tilde{\mathbf{P}}_1 \mathbf{s}_3 + \\
& \mathbf{s}'_3 (\mathbf{A}' \cdot \mathbf{R} \cdot \mathbf{R}) \tilde{\mathbf{P}}_1 \mathbf{s}_3 + \mathbf{x}'_1 (\mathbf{C}^2)' \mathbf{B}' \mathbf{Q}^2 \tilde{\mathbf{P}}_1 \mathbf{s}_3 + \cdots + \\
& \mathbf{e}'_n \mathbf{Q}^T \tilde{\mathbf{P}}_1 \mathbf{s}_T + \mathbf{s}'_T (\mathbf{A}' \cdot \mathbf{R} \cdot \mathbf{R} \cdot \cdots \cdot \mathbf{R} \cdot \mathbf{R}) \tilde{\mathbf{P}}_1 \mathbf{s}_T + \\
& \mathbf{x}'_1 (\mathbf{C}^T)' \mathbf{B}' \mathbf{Q}^T \tilde{\mathbf{P}}_1 \mathbf{s}_T
\end{aligned} \tag{4.7}$$

$$\begin{aligned}
\text{subject to} \quad & \bar{\mathbf{s}} = \sum_{t=1}^T \mathbf{s}_t \\
& \mathbf{0} \leq \mathbf{s}_t, \quad t = 1, \dots, T
\end{aligned}$$

We solve (4.7) at time $t = 1$ and find the “optimal” controls $\mathbf{s}_1^1, \dots, \mathbf{s}_T^1$, where the superscript indicates that this is the period-1 solution of (4.7). However, we only implement the control \mathbf{s}_1^1 . After the state vector in period $t = 2$ is observed, we re-solve (4.7) for time $t = 2$, find a new set of controls $\mathbf{s}_2^2, \dots, \mathbf{s}_T^2$, but only implement the control \mathbf{s}_2^2 . We continue in this fashion, at each step solving a convex quadratic optimization problem which can be handled efficiently using commercially available packages, e.g., CPLEX or MINOS.

Although the static approximation method may not yield adequate approximations in all cases, in many of the examples we have explored, the technique performs admirably (see, for example, the empirical analysis of Section 5). Of course, it is difficult to derive accurate bounds on the approximation error in the most interesting cases because these are the cases for which the optimal solutions are unknown. We hope to explore the theoretical properties of the static approximation method in future research.

5 An Empirical Example

In this section we implement the best-execution strategies developed in Section 3 and 4 for a hypothetical portfolio of 25 stocks using real historical data. Specifically, we estimate

the parameters of the linear-percentage model of Section 2.1 for each of the 25 stocks, construct several portfolio rebalancing scenarios, and compare the best-execution strategy with a “naive” strategy of trading equal-size lots in each time period.

5.1 The Data

Our empirical analysis draws on three sources of data. Our primary data source is a proprietary record of trades performed over the NYSE DOT system by the trading desk at Investment Technologies Group (ITG) on every trading day between January 2, 1996 and December 31, 1996. Each trade is cataloged with the following information: order submission date and time, order execution date and time, whether it is a buy or sell order, size in shares, execution price, and order type (e.g. market order, limit order). The 25 stocks chosen for this study were those that had the greatest number of market orders over the year-long interval (see Table 1).

Because of our selection rule, our sample consists of companies with large market capitalizations, which ensures that we will have enough data to fit the model and arrive at reasonably accurate estimates of the parameters. But such a sample will tend to exhibit a lower-than-average price impact because stocks that trade very frequently are, by definition, very liquid and have much smaller price impact. Such a bias in our sampling procedure by no means invalidates the relevance of our example. In fact, if we can demonstrate that our best-execution strategy is beneficial for highly liquid stocks, then the value of our approach is likely to be even greater for less liquid stocks where price impact is significantly higher.

The ITG database provides valuable trade information but we must augment our analysis with the NYSE TAQ data in order to extract quotes prevailing at the time of ITG trades. The TAQ database is a complete history of all trades and quotes on the NYSE, AMEX and NASDAQ exchanges.

Finally, we will also use S&P 500 tick data provided by Tick Data Inc. to get intraday levels for the S&P 500 index during 1996.

5.2 The Estimation Procedure

Our estimation procedure is composed of three separate steps. First we estimate the parameters, μ_z and Σ_z of the no-impact price dynamics (2.2) for each stock. Given the geometric

Brownian motion specification, we know that the continuously compounded returns z_{it} are IID normal random variates:

$$z_{t,i} = \log\left(\frac{\tilde{p}_{t,i}^b}{\tilde{p}_{t-1,i}^b}\right) \sim \mathcal{N}(\mu_i, \sigma_i^2) \quad (5.1)$$

for each stock i , where $i = 1, \dots, 25$. The no-impact price is taken to be the midpoint of the prevailing bid and offer prices at time t (hence the need for quotes):

$$\tilde{p}_{it} = \frac{\tilde{p}_{it}^b + \tilde{p}_{it}^a}{2} \quad (5.2)$$

where \tilde{p}_{it}^b and \tilde{p}_{it}^a are the bid and ask prices for stock i at time t . For each of our 25 stocks we collect TAQ quotes at every half hour over the course of the trading year of 1996 and calculate the midpoint to construct the no-impact prices, \tilde{p}_t , thus the time index, t , ranges over half hours, $t = 1, 2, \dots, N_h$ where N_h is the total number of half hours in the 1996 trading year (approximately 250 days times 13 periods per day).

We then form log-returns according to (5.1) and discard any overnight returns to eliminate non-synchronous trading effects. This gives us a sample of 2,069 observations of \mathbf{z}_t during the 1996 calendar year from which we can estimate $\boldsymbol{\mu}_z$ and $\boldsymbol{\Sigma}_z$ in the standard way. Summary results are given in Table 2 (we report estimates only for the first five stocks of Table 1 to conserve space). The drift and volatility are expressed in percent per year, and are scaled up from the half-hourly units by assuming each of the 250 trading days per year consists of 13 half-hour trading intervals. The drift and volatility estimates are consistent with intuition and agree reasonably well with other data sources such as BARRA.

Our second task is to estimate the parameters of the market information process in (2.4). The variable \mathbf{x}_t captures the potential impact of changing market conditions or private information about the security. For example, a short-term excess returns model might be constructed for this purpose. In our example, we let \mathbf{x}_t denote the half-hourly return on the S&P 500 index, a common factor that influences the prices of most securities. For this specification, the AR(1) coefficients, \mathbf{C} , and the covariance matrix of the noise, $\boldsymbol{\Sigma}_\eta$, reduce to scalar quantities, C and σ_η . Using S&P 500 tick data from January 2 to December 31, 1996, we construct half-hourly returns of the index which we denote by x_t , where t ranges

over half hours, $t = 1, 2, \dots, N_h$ ranges over half-hour intervals and N_h is the total number of half hours in the trading year. We rescale the returns by subtracting out the mean and dividing by standard deviation. This leaves us with a zero mean, unit-standard-deviation process:

$$\tilde{x}_t = \frac{x_t - \mu_x}{\sigma_x} \quad (5.3)$$

and assuming $|C| < 1$ we can rewrite (2.4) as:

$$\tilde{x}_t = C\tilde{x}_{t-1} + \eta_t . \quad (5.4)$$

The maximum likelihood estimator of the AR(1) coefficient C is:

$$\hat{C} = \frac{\frac{1}{T_1} \sum_{t=2}^{T_1} \tilde{x}_t \tilde{x}_{t-1}}{\frac{1}{T_2} \sum_{t=1}^{T_2} \tilde{x}_t^2} \quad (5.5)$$

To avoid non-synchronous trading effects we discard all products $\tilde{x}_t \tilde{x}_{t-1}$ that straddle an overnight period in the numerator of (5.5). Similarly, we exclude overnight return terms from the denominator of (5.5). The constants, T_1 and T_2 , are the number of terms that are included in calculating the numerator and denominator and are 1,902 and 2,078, respectively. Our estimate of \hat{C} is 0.0354.⁶

Given \hat{C} , an estimator for the standard deviation of η_t follows immediately:

$$\hat{\sigma}_\eta = \sqrt{1 - \hat{C}^2} . \quad (5.6)$$

Our estimate is 0.999. The two parameters, \hat{C} and $\hat{\sigma}_\eta$ fully characterize the AR(1) process that describes the S&P 500 returns.

Our final task is to estimate the parameters, \mathbf{A} and \mathbf{B} of the price-impact equation (2.2).

⁶Not surprisingly, the level of serial correlation in the S&P 500 index is quite low. If not, profitable trading strategies would be possible that would quickly drive the predictability of index returns back to a level where the predictability was small.

We can recast the vector equation as 25 separate linear regressions by rearranging terms:

$$\frac{p_{it} - \tilde{p}_{it}}{\tilde{p}_{it}} = \tilde{p}_{it} \mathbf{s}_t \mathbf{a}'_i + \mathbf{x}_t \mathbf{b}'_i \quad (5.7)$$

where \mathbf{a}_i and \mathbf{b}_i are the i th rows of \mathbf{A} and \mathbf{B} , respectively. This expression shows that the percentage price impact to the i th security is a linear function of the dollar volume we intend to trade in the i th security, the dollar volumes that we and others are currently trading in the other 24 stocks, and the S&P 500 return over the preceding half-hour.

That there should be a price impact on p_{it} from trading in stock i is obvious. But less obvious is the role that trading in other stocks might play in determining the price impact on p_{it} . There are several economic sources for such cross effects. One stock may be a close substitute for another, hence a high price impact for one would imply the same for the other. Another motivation is that in a market with sharply rising (or falling) prices and high volume, the overall market impact may be expected to increase as liquidity providers demand higher premiums above posted quotes for large market orders.

To estimate \mathbf{A} and \mathbf{B} , we use a combination of ITG proprietary data, TAQ data and SPX tick data. For each executed market order in a given stock, i , the ITG database gives complete information about the market order except for the prevailing quote. We search the TAQ database to find the quote. As before, we form the no-impact price, \tilde{p}_{it} , as the average of the bid and offer (see (5.2)) and then construct the dependent variables:

$$\frac{p_{it} - \tilde{p}_{it}}{\tilde{p}_{it}} \quad (5.8)$$

for each trade. The ITG database provides one independent variable, namely the dollar volume of stock i : $\tilde{p}_{it} s_{it}$. The sign convention for shares is positive for buys and negative for sells.

A difficulty arises in constructing other dollar-volume-related independent variables (i.e., $\tilde{p}_{jt} s_{jt}$ for $i \neq j$). The ITG data is too sparse to find nearby trades in time, hence we must turn to the TAQ data to resolve this observability problem. One possible solution is to use the nearest TAQ trade that occurs prior to an ITG market order. Unfortunately, the time alignment of the TAQ and ITG datasets can be imprecise due to recording lags by either

party. To reduce the impact of this type of error, we define a proxy for the closest trade by forming a 30-second window before each market order and computing an average dollar-volume within it. Although this averaging procedure tends to smooth the data and reduces its information content, nevertheless it ensures that temporal sequencing is not violated.

Specifically, we find all N_k trades in stock j that occur within that window. Each trade is executed at price p_{ik} where $k = 1, \dots, N_k$. Trades that are executed above or at the midpoint of their quotes are classified as buys and the rest as sells. An average dollar volume within the window for stock j is then computed as:

$$\tilde{p}_{jt} s_{jt} = \frac{1}{N_k} \sum_{k=1}^{N_k} \tilde{p}_{jk} s_{jk} . \quad (5.9)$$

Finally, for the S&P 500 return, x_t , we split the trading day into 13 half-hour intervals and compute the return in the half-hour prior to the trade.

We now have a complete set of data with which to estimate the parameters of the price-impact portion of our model. The regressions were performed in SAS and contained no intercept term because price impact should be zero if no stocks are being traded. Table 3 gives a summary of the first five of the 25 regressions performed on the data. For each regression, the parameter estimates for the 26 regressors—the lagged returns for the 25 stocks and the S&P 500 lagged return—and their t -statistics are reported, along with the R^2 and the sample size at the bottom of each column.⁷

To develop some intuition for the coefficients, consider the estimated price impact for American Home Products (AHP) in $\hat{\mathbf{A}}$ due to trading in AHP, which is 4.97×10^{-10} according to Table 3. If we were to trade a 100,000-share block of AHP at its beginning-of-year price of \$64.0625 with no impact, our total cost would be $100,000 \times \$64.0625 = \$6,406,250$. But according to Table 3, the full-impact cost would be

$$\begin{aligned} 100,000 \times (\tilde{p}_t + \delta_t) &= 100,000 \times (\$64.0625 + \delta_t) \\ \delta_t &= \tilde{p}_t (4.97 \times 10^{-10} \times \tilde{p}_t \times 100,000) = 0.203969 \end{aligned}$$

⁷Diagnostics were performed on the residuals to test for the presence of heteroskedasticity and autocorrelation. The Durbin-Watson test indicated low levels of positive serial correlation, with statistics ranging from 1.12 to 1.69 for the 25 regressions. The test of first and second moment specification indicated a very weak presence of heteroskedasticity as the p -values were, in general, very low.

$$100,000 \times (\tilde{p}_t + \delta_t) = \$6,426,647$$

which implies a price impact of approximately 20 cents per share (ignoring the other factors in the regression). This is an unacceptably large impact and no professional trader would submit such a large order except in the most desperate of circumstances.

Further inspection of the regression diagnostics shows that the R^2 ranges from 0.052 to 0.440 for the 25 regressions, indicating that the regressions have varying degrees of explanatory power. However, 30% (193 of 625) of all the t -statistics are significant at the 5% level, implying the importance of factors other than own-stock trading in determining price impact. Also 18 of the 25 own-stock price-impact terms, i.e., $\hat{a}_{i,i}$ (the i -th diagonal entry of $\hat{\mathbf{A}}$), are statistically significant. We expect these these terms to be the most dominant ones in determining price impact and our regression confirm this conjecture.

Although some of the regressions have low explanatory power, recall that we have proposed a rather naive specification for these regressions, omitting many other variables that proprietary traders and other professional portfolio managers have at their disposal. But even with our naive specification, we still achieve R^2 's as high as 0.440 (for Merck, not shown in Table 3) which is quite substantial considering the high-frequency nature of the data we are analyzing.

5.3 No-Arbitrage Constraints

There is one additional aspect of the estimation procedure that must be considered which is whether or not the parameter estimates yield a well-posed optimization problem (1.1)–(1.2). In particular, for certain parameter values the optimization problem is not convex, hence the objective function can be made arbitrarily negative. The economic interpretation for such circumstances is an *arbitrage opportunity* (also known as a “free lunch”), a situation in which riskless profits can be manufactured out of thin air. Ordinarily, this would be a welcome state of affairs for investment professionals, but in this case the arbitrage is more likely a spurious side-effect of sampling variation in our parameter estimates.

Therefore, a no-arbitrage restriction should be imposed on the estimation procedure to avoid these false arbitrage opportunities. For the linear-percentage price-impact model, this is accomplished by constraining both $\hat{\mathbf{A}}$ and $\hat{\mathbf{A}}' \cdot \hat{\mathbf{R}}$ to be positive definite matrices which,

in turn, involves estimating a *constrained* linear regression model. The results of such a procedure are reported in Table 4. The two most significant differences between Tables 3 and 4 are the lower R^2 's and the higher significance of the own-stock coefficients in Table 4. The former is not surprising, since any constraint is bound to decrease the explanatory power of the regression, though it is worth noting that for AN and DD, the decline is rather small. The higher significance of own-stock coefficients follows from the definition of positive definiteness, i.e., $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all vectors \mathbf{x} . The diagonal elements of \mathbf{A} are coefficients of squared terms of the \mathbf{x} values in the matrix product $\mathbf{x}'\mathbf{A}\mathbf{x}$. Therefore, by making the squared terms sufficiently large relative to the cross terms, we arrive at a positive definite matrix.

Table 5 shows the ratio of the total sum of squared errors of the constrained regression to the unconstrained regression for all 25 stocks. The increase in squared errors is approximately 5% overall, a rather modest increase that provides some support for imposing the restriction. More importantly, if the no-arbitrage condition were not imposed, the dynamic optimization algorithms described in Sections 3 and 4 may yield nonsensical results.

5.4 Monte Carlo Analysis

Having calibrated the state equation in Section 5.2 for the linear percentage case of Section 2.1, we now investigate the performance of the best-execution strategy via Monte Carlo simulation experiments. Specifically, we consider minimizing the execution costs of purchasing \bar{s} shares of each of the 25 stocks in Table 1 over T periods, under the price dynamics (2.1)–(2.4) where \mathbf{A} , and \mathbf{B} are the estimates $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ from the constrained regression (see Section 5.3) and \mathbf{C} , $\boldsymbol{\mu}_z$, $\boldsymbol{\Sigma}_\eta$ are estimated as in in Section 5.2. We assume that the baseline covariance matrix of the no-impact price is given by $\boldsymbol{\Sigma}_z$, and that the initial no-impact prices are the prices given in Table 1 (these are closing prices selected from a random trading day in 1996).

To gauge the sensitivity of execution costs to the parameters of the model, we vary the time horizon, T , the number of shares traded, \bar{s} (assumed to be the same for each stock) and the no-impact price volatility. The price volatility is modified by scaling the variances by a constant while keeping the correlation structure fixed. The results for 10,000 replications are given in Table 4 which reports the expected execution cost in cents/share and standard error of the estimate for the best-execution strategy, the strategy under no-sales constraints

computed via the static approximation approach (see Section 4.2), and the naive strategy denoted by \mathbf{s}^* , \mathbf{s}_c^* and $\bar{\mathbf{s}}/T$, respectively.

Some general patterns emerge from the simulations. First, as T increases, execution costs fall. Because there are more time periods over which the trading can be spread, and because we have the flexibility to be more patient and wait for particularly opportune times to trade, expected costs decline. Second, as $\bar{\mathbf{s}}$ decreases, execution costs also decrease. In fact, with small enough trade sizes the expected price impact is negative! This is due to the fact that price impact consists of two terms, the impact of shares traded, \mathbf{s}_t , which is quadratic in the share size, and the impact of information, which is linear in the information variable, \mathbf{x}_t . When we trade small enough quantities, the quadratic term is negligible and the information term dominates. Our strategy makes optimal use of information so as to trade when it is least expensive, and for sufficiently significant pieces of information, trading can be quite profitable (not a new insight to proprietary traders). Finally, increasing volatility seems to increase execution costs slightly.

In all but two cases, the optimal strategy outperforms the naive on average. In the two anomalous cases, the confidence interval of the difference between the two strategies is so wide that this outcome could easily have occurred purely by chance. If we were to increase the number of replications to 100,000, these two anomalies would no doubt disappear.

Another anomalous result is the fact that for some simulations, the execution cost for the constrained strategy is less than that of the unconstrained counterpart. While the point estimates are indeed reversed in these cases, the sampling variation is so great (consider their standard errors) that it is difficult to make accurate inferences about their relative magnitudes. For the cases we consider, the sales constraint seems to have relatively little impact on the performance of the best-execution strategy, except in those cases where execution costs are negative. To achieve negative execution costs, the sales constraint must be violated, hence imposing them increases the costs dramatically.

6 Conclusion

The dramatic growth in assets managed by institutions, as well as the recent advent of internet trading and electronic brokerage for retail investors, has led to a surge in the size and volume of trading. At the same time, competition in the asset management industry has increased to the point where mere basis points can separate the top funds from those in the next tier. In this environment, portfolio managers have begun to explore active management of trading costs as a means of boosting returns. Controlling execution cost can be viewed as a stochastic dynamic optimization problem because trading takes time, stock prices exhibit random fluctuations, and execution prices depend on trade size, order flow, and market conditions.

In this paper, we apply stochastic dynamic programming to derive trading strategies that minimize the expected cost of executing a portfolio of securities over a fixed period of time. The portfolio problem raises a number of new challenges not found in the single-stock case, and we have developed a specification for this case that is both empirically plausible and computationally tractable to implement. The closed-form solution provides insight into the nature of trading portfolios, and may prove to be a useful benchmark for pricing principal-bid and negotiated-block transactions, since broker/dealers providing these services will seek to minimize their inventory risk by trading out of these positions as rapidly as possible.

To quantify the potential cost savings from using our best-execution strategies, we have fit the parameters of the linear percentage price-impact model using historical data on 25 large-cap NYSE stocks. Our Monte Carlo simulations, which are based on these estimated parameters, indicate substantial reductions in the expected execution costs for our hypothetical portfolio, though these costs must be estimated on a case-by-case basis using the most current data and after re-calibrating the parameters of the state equations.

The remaining challenge is to integrate these best-execution strategies directly into the investment process, which requires solving the portfolio optimization problem subject to transactions costs. This is a formidable challenge that is both theoretically and computationally intensive, and we plan to turn to these problems in future research.

A Appendix

Recall that the optimal value function in the last period can be expressed as:

$$V_T(\tilde{\mathbf{p}}_T, \mathbf{x}_T, \mathbf{w}_T) = \mathbf{e}'_n \tilde{\mathbf{P}}_T \mathbf{w}_T + \mathbf{w}'_T \tilde{\mathbf{P}}_T \mathbf{A}' \tilde{\mathbf{P}}_T \mathbf{w}_T + \mathbf{x}'_T \mathbf{B}' \tilde{\mathbf{P}}_T \mathbf{w}_T .$$

Bellman's equation for the optimal value function in the next-to-last period is given by:

$$V_{T-1} = \underset{\{\mathbf{s}_{T-1}\}}{\text{Min}} \mathbb{E}_{T-1} \left[\mathbf{p}'_{T-1} \mathbf{s}_{T-1} + V_T(\tilde{\mathbf{p}}_T, \mathbf{x}_T, \mathbf{w}_T) \right] \quad (\text{A.1})$$

$$\begin{aligned} &= \underset{\{\mathbf{s}_{T-1}\}}{\text{Min}} \mathbb{E}_{T-1} \left[[\tilde{\mathbf{P}}_{T-1}(\mathbf{e}_n + \mathbf{A}' \tilde{\mathbf{P}}_{T-1} \mathbf{s}_{T-1} + \mathbf{B}' \mathbf{x}_{T-1})]' \mathbf{s}_{T-1} + \right. \\ &\quad \mathbf{e}'_n \tilde{\mathbf{Z}}_{T-1} \tilde{\mathbf{P}}_{T-1} (\mathbf{w}_{T-1} - \mathbf{s}_{T-1}) + \\ &\quad (\mathbf{w}_{T-1} - \mathbf{s}_{T-1})' \tilde{\mathbf{Z}}_{T-1} \tilde{\mathbf{P}}_{T-1} \mathbf{A}' \tilde{\mathbf{Z}}_{T-1} \tilde{\mathbf{P}}_{T-1} (\mathbf{w}_{T-1} - \mathbf{s}_{T-1}) + \\ &\quad \left. (\mathbf{C}' \mathbf{x}_{T-1} + \boldsymbol{\eta}'_T)' \mathbf{B}' \tilde{\mathbf{Z}}_{T-1} \tilde{\mathbf{P}}_{T-1} (\mathbf{w}_{T-1} - \mathbf{s}_{T-1}) \right] \quad (\text{A.2}) \end{aligned}$$

where we define $\tilde{\mathbf{Z}}_{T-1} \equiv \exp(\mathbf{Z}_{T-1})$ and observe that $\text{diag}[\exp(\mathbf{Z}_{T-1}) \tilde{\mathbf{P}}_{T-1}] = \tilde{\mathbf{Z}}_{T-1} \tilde{\mathbf{P}}_{T-1}$.

Dropping the $T-1$ subscript on all time-dependent quantities and factoring (A.2) then yields:

$$\begin{aligned} V_{T-1} &= \underset{\{\mathbf{s}\}}{\text{Min}} \mathbb{E} \left[\mathbf{s}' [\tilde{\mathbf{P}}(\mathbf{A}' + \mathbf{A}' \tilde{\mathbf{Z}}) \tilde{\mathbf{P}}] \mathbf{s} + \right. \\ &\quad [\mathbf{e}'(\mathbf{I} - \tilde{\mathbf{Z}}) \tilde{\mathbf{P}} - \mathbf{w}' \tilde{\mathbf{P}} \mathbf{A}' \tilde{\mathbf{Z}} \tilde{\mathbf{P}} + \mathbf{x}'(\mathbf{B}' - \mathbf{C}' \mathbf{B}' \tilde{\mathbf{Z}}) \tilde{\mathbf{P}} - \boldsymbol{\eta}' \mathbf{B}' \tilde{\mathbf{Z}} \tilde{\mathbf{P}}] \mathbf{s} - \\ &\quad \mathbf{s}' \tilde{\mathbf{P}} \mathbf{A}' \tilde{\mathbf{Z}} \tilde{\mathbf{P}} \mathbf{w} + \mathbf{w}' \tilde{\mathbf{P}} \mathbf{A}' \tilde{\mathbf{Z}} \tilde{\mathbf{P}} \mathbf{w} + \mathbf{x}' \mathbf{C}' \mathbf{B}' \tilde{\mathbf{Z}} \tilde{\mathbf{P}} \mathbf{w} + \\ &\quad \left. \boldsymbol{\eta}' \mathbf{B}' \tilde{\mathbf{Z}} \tilde{\mathbf{P}} \mathbf{w} + \mathbf{e}'_n \tilde{\mathbf{Z}} \tilde{\mathbf{P}} \mathbf{w} \right] \quad (\text{A.3}) \end{aligned}$$

where \mathbf{I} is the $(n \times n)$ identity matrix. Taking the expectation of this expression, reintroducing time subscripts, and observing that it is a convex quadratic function of the vector \mathbf{s} , we obtain by simple differentiation that the optimal solution expressed in terms of dollars-to-trade, $\tilde{\mathbf{s}} \equiv \tilde{\mathbf{P}} \mathbf{s}$, is given by:

$$\tilde{\mathbf{s}}^*_{T-1} = \boldsymbol{\Delta}_{x,1} \mathbf{x}_{T-1} + \boldsymbol{\Delta}_{w,1} \tilde{\mathbf{w}}_{T-1} + \boldsymbol{\Delta}_{1,1} \mathbf{e}_n \quad (\text{A.4})$$

where

$$\mathbf{\Delta}_{x,1} \equiv [(\mathbf{A}' + \mathbf{A}' \cdot \mathbf{R}) + (\mathbf{A}' + \mathbf{A}' \cdot \mathbf{R})']^{-1}(\mathbf{QBC} - \mathbf{B}) \quad (\text{A.5a})$$

$$\mathbf{\Delta}_{w,1} \equiv [(\mathbf{A}' + \mathbf{A}' \cdot \mathbf{R}) + (\mathbf{A}' + \mathbf{A}' \cdot \mathbf{R})']^{-1}(\mathbf{A}' \cdot \mathbf{R} + \mathbf{A} \cdot \mathbf{R}') \quad (\text{A.5b})$$

$$\mathbf{\Delta}_{1,1} \equiv [(\mathbf{A}' + \mathbf{A}' \cdot \mathbf{R}) + (\mathbf{A}' + \mathbf{A}' \cdot \mathbf{R})']^{-1}(\mathbf{Q} - \mathbf{I}) , \quad (\text{A.5c})$$

\mathbf{Q} is an $(n \times n)$ diagonal matrix with entries:

$$q_i = \exp(\mu_{z,i} + \frac{1}{2}\Sigma_{z,ii}) , \quad (\text{A.6})$$

\mathbf{R} is an $(n \times n)$ symmetric matrix with elements:

$$r_{ij} = \exp[\mu_{z,i} + \mu_{z,j} + \frac{1}{2}(\Sigma_{z,ii} + \Sigma_{z,jj} + 2\Sigma_{z,ij})] , \quad (\text{A.7})$$

and the matrix dot operator, ' \cdot ', denotes an element-wise matrix multiplication, i.e., $\mathbf{A} \cdot \mathbf{B} = [a_{ij}b_{ij}]$. The optimal value function V_{T-1} can now be rewritten as:

$$\begin{aligned} V_{T-1}(\tilde{\mathbf{p}}_{T-1}, \mathbf{x}_{T-1}, \tilde{\mathbf{w}}_{T-1}) &= \mathbf{e}'_n \mathbf{D}_{n,1} \mathbf{e}_n + \mathbf{e}'_n \mathbf{E}_1 \mathbf{x}_{T-1} + \mathbf{x}'_{T-1} \mathbf{F}_1 \mathbf{e}_n + \\ &\quad \mathbf{x}'_{T-1} \mathbf{G}_1 \mathbf{x}_{T-1} + \mathbf{x}'_{T-1} \mathbf{H}_1 \tilde{\mathbf{w}}_{T-1} + \tilde{\mathbf{w}}'_{T-1} \mathbf{J}_1 \mathbf{x}_{T-1} + \\ &\quad \mathbf{e}'_n \mathbf{K}_1 \tilde{\mathbf{w}}_{T-1} + \tilde{\mathbf{w}}'_{T-1} \mathbf{L}_1 \mathbf{e}_n + \\ &\quad \tilde{\mathbf{w}}'_{T-1} \mathbf{N}_1 \tilde{\mathbf{w}}_{T-1} \end{aligned} \quad (\text{A.8})$$

where:

$$\mathbf{D}_{n,1} \equiv [\mathbf{\Delta}'_{1,1}(\mathbf{A}' + \mathbf{A}' \cdot \mathbf{R}) + (\mathbf{I} - \mathbf{Q})] \mathbf{\Delta}_{1,1} \quad (\text{A.9a})$$

$$\mathbf{E}_1 \equiv [\mathbf{\Delta}'_{1,1}(\mathbf{A}' + \mathbf{A}' \cdot \mathbf{R}) + (\mathbf{I} - \mathbf{Q})] \mathbf{\Delta}_{x,1} \quad (\text{A.9b})$$

$$\mathbf{F}_1 \equiv [\mathbf{\Delta}'_{x,1}(\mathbf{A}' + \mathbf{A}' \cdot \mathbf{R}) + (\mathbf{B}' - \mathbf{C}' \mathbf{B}' \mathbf{Q})] \mathbf{\Delta}_{1,1} \quad (\text{A.9c})$$

$$\mathbf{G}_1 \equiv [\mathbf{\Delta}'_{x,1}(\mathbf{A}' + \mathbf{A}' \cdot \mathbf{R}) + (\mathbf{B}' - \mathbf{C}' \mathbf{B}' \mathbf{Q})] \mathbf{\Delta}_{x,1} \quad (\text{A.9d})$$

$$\begin{aligned} \mathbf{H}_1 &\equiv [\mathbf{\Delta}'_{x,1}(\mathbf{A}' + \mathbf{A}' \cdot \mathbf{R}) + (\mathbf{B}' - \mathbf{C}' \mathbf{B}' \mathbf{Q})] \mathbf{\Delta}_{w,1} - \\ &\quad \mathbf{\Delta}'_{x,1} \mathbf{A}' \cdot \mathbf{R} + \mathbf{C}' \mathbf{B}' \mathbf{Q} \end{aligned} \quad (\text{A.9e})$$

$$\mathbf{J}_1 \equiv [\Delta'_{w,1}(\mathbf{A}' + \mathbf{A}' \cdot \mathbf{R}) - \mathbf{A}' \cdot \mathbf{R}] \Delta_{x,1} \quad (\text{A.9f})$$

$$\mathbf{K}_1 \equiv [\Delta'_{1,1}(\mathbf{A}' + \mathbf{A}' \cdot \mathbf{R}) + (\mathbf{I} - \mathbf{Q})] \Delta_{w,1} - \Delta'_{1,1} \mathbf{A}' \cdot \mathbf{R} + \mathbf{Q} \quad (\text{A.9g})$$

$$\mathbf{L}_1 \equiv [\Delta'_{w,1}(\mathbf{A}' + \mathbf{A}' \cdot \mathbf{R}) - \mathbf{A}' \cdot \mathbf{R}] \Delta_{1,1} \quad (\text{A.9h})$$

$$\mathbf{N}_1 \equiv [\Delta'_{w,1}(\mathbf{A}' + \mathbf{A}' \cdot \mathbf{R}) - \mathbf{A}' \cdot \mathbf{R}] \Delta_{w,1} - \Delta'_{w,1} \mathbf{A}' \cdot \mathbf{R} + \mathbf{A}' \cdot \mathbf{R} . \quad (\text{A.9i})$$

Similarly, the Bellman equation at time $t-k$ is given by:

$$V_{T-k} = \underset{\{\mathbf{s}_{T-k}\}}{\text{Min}} \mathbb{E}_{T-k} \left[\mathbf{p}'_{T-k} \mathbf{s}_{T-k} + V_{T-k+1}(\tilde{\mathbf{p}}_{T-k+1}, \mathbf{x}_{T-k+1}, \mathbf{w}_{T-k+1}) \right] \quad (\text{A.10})$$

which yields the best-execution strategy:

$$\tilde{\mathbf{s}}_{T-k}^* = \Delta_{x,k} \mathbf{x}_{T-k} + \Delta_{w,k} \tilde{\mathbf{w}}_{T-k} + \Delta_{1,k} \mathbf{e}_n \quad (\text{A.11})$$

where

$$\Delta_{x,k} \equiv [(\mathbf{A}' + \mathbf{N}_{k-1} \cdot \mathbf{R}) + (\mathbf{A}' + \mathbf{N}_{k-1} \cdot \mathbf{R})']^{-1} \times [\mathbf{Q}(\mathbf{H}'_{k-1} + \mathbf{J}_{k-1})\mathbf{C} - \mathbf{B}] \quad (\text{A.12a})$$

$$\Delta_{w,k} \equiv [(\mathbf{A}' + \mathbf{N}_{k-1} \cdot \mathbf{R}) + (\mathbf{A}' + \mathbf{N}_{k-1} \cdot \mathbf{R})']^{-1} \times [(\mathbf{N}_{k-1} \cdot \mathbf{R} + \mathbf{N}'_{k-1} \cdot \mathbf{R}')] \quad (\text{A.12b})$$

$$\Delta_{1,k} \equiv [(\mathbf{A}' + \mathbf{N}_{k-1} \cdot \mathbf{R}) + (\mathbf{A}' + \mathbf{N}_{k-1} \cdot \mathbf{R})']^{-1} \times [\mathbf{Q}(\mathbf{K}'_{k-1} + \mathbf{L}_{k-1}) - \mathbf{I}] \quad (\text{A.12c})$$

and the expected best-execution cost with k periods until the end of the horizon is given by:

$$\begin{aligned} V_{T-k}(\tilde{\mathbf{p}}_{T-k}, \mathbf{x}_{T-k}, \mathbf{w}_{T-k}) &= \mathbf{e}'_n \mathbf{D}_{n,k} \mathbf{e}_n + \mathbf{e}'_r \mathbf{D}_{r,k} \mathbf{e}_r + \mathbf{e}'_n \mathbf{E}_k \mathbf{x}_{T-k} + \\ &\quad \mathbf{x}'_{T-k} \mathbf{F}_k \mathbf{e}_n + \mathbf{x}'_{T-k} \mathbf{G}_k \mathbf{x}_{T-k} + \mathbf{x}'_{T-k} \mathbf{H}_k \tilde{\mathbf{w}}_{T-k} + \\ &\quad \tilde{\mathbf{w}}'_{T-k} \mathbf{J}_{T-1} \mathbf{x}_{T-k} + \mathbf{e}'_n \mathbf{K}_k \tilde{\mathbf{w}}_{T-k} + \tilde{\mathbf{w}}'_{T-k} \mathbf{L}_k \mathbf{e}_n + \\ &\quad \tilde{\mathbf{w}}'_{T-k} \mathbf{N}_k \tilde{\mathbf{w}}_{T-k} \end{aligned} \quad (\text{A.13})$$

where

$$\begin{aligned} \mathbf{D}_{n,k} \equiv & [\Delta'_{1,k}(\mathbf{A}' + \mathbf{N}_{k-1} \cdot \mathbf{R}) + (\mathbf{I} - \mathbf{K}_{k-1} \mathbf{Q})] \Delta_{1,k} - \\ & \Delta'_{1,k} \mathbf{Q} \mathbf{L}_{k-1} + \mathbf{D}_{n,k-1} \end{aligned} \quad (\text{A.14a})$$

$$\mathbf{D}_{r,k} \equiv \Sigma_\eta \cdot \mathbf{G}_{k-1} \quad (\text{A.14b})$$

$$\begin{aligned} \mathbf{E}_k \equiv & [\Delta'_{1,k}(\mathbf{A}' + \mathbf{N}_{k-1} \cdot \mathbf{R}) + (\mathbf{I} - \mathbf{K}_{k-1} \mathbf{Q})] \Delta_{x,k} - \\ & \Delta'_{1,k} \mathbf{Q} \mathbf{J}_{k-1} \mathbf{C} + \mathbf{E}_{k-1} \mathbf{C} \end{aligned} \quad (\text{A.14c})$$

$$\begin{aligned} \mathbf{F}_k \equiv & [\Delta'_{x,k}(\mathbf{A}' + \mathbf{N}_{k-1} \cdot \mathbf{R}) + (\mathbf{B}' - \mathbf{C}' \mathbf{H}_{k-1} \mathbf{Q})] \Delta_{1,k} - \\ & \Delta'_{x,k} \mathbf{Q} \mathbf{L}_{k-1} + \mathbf{C}' \mathbf{F}_{k-1} \end{aligned} \quad (\text{A.14d})$$

$$\begin{aligned} \mathbf{G}_k \equiv & [\Delta'_{x,k}(\mathbf{A}' + \mathbf{N}_{k-1} \cdot \mathbf{R}) + (\mathbf{B}' - \mathbf{C}' \mathbf{H}_{k-1} \mathbf{Q})] \Delta_{x,k} - \\ & \Delta'_{x,k} \mathbf{Q} \mathbf{J}_{k-1} \mathbf{C} + \mathbf{C}' \mathbf{G}_{k-1} \mathbf{C} \end{aligned} \quad (\text{A.14e})$$

$$\begin{aligned} \mathbf{H}_k \equiv & [\Delta'_{x,k}(\mathbf{A}' + \mathbf{N}_{k-1} \cdot \mathbf{R}) + (\mathbf{B}' - \mathbf{C}' \mathbf{H}_{k-1} \mathbf{Q})] \Delta_{w,k} - \\ & \Delta'_{x,k} \mathbf{N}_{k-1} \cdot \mathbf{R} + \mathbf{C}' \mathbf{H}_{k-1} \mathbf{Q} \end{aligned} \quad (\text{A.14f})$$

$$\begin{aligned} \mathbf{J}_k \equiv & [\Delta'_{w,k}(\mathbf{A}' + \mathbf{N}_{k-1} \cdot \mathbf{R}) - \mathbf{N}_{k-1} \cdot \mathbf{R}] \Delta_{x,k} + \\ & (\mathbf{I} - \Delta'_{w,k}) \mathbf{Q} \mathbf{J}_{k-1} \mathbf{C} \end{aligned} \quad (\text{A.14g})$$

$$\begin{aligned} \mathbf{K}_k \equiv & [\Delta'_{1,k}(\mathbf{A}' + \mathbf{N}_{k-1} \cdot \mathbf{R}) + (\mathbf{I} - \mathbf{K}_{k-1} \mathbf{Q})] \Delta_{w,k} - \\ & \Delta'_{1,k} \mathbf{N}_{k-1} \cdot \mathbf{R} + \mathbf{K}_{k-1} \mathbf{Q} \end{aligned} \quad (\text{A.14h})$$

$$\begin{aligned} \mathbf{L}_k \equiv & [\Delta'_{w,k}(\mathbf{A}' + \mathbf{N}_{k-1} \cdot \mathbf{R}) - \mathbf{N}_{k-1} \cdot \mathbf{R}] \Delta_{1,k} + \\ & (\mathbf{I} - \Delta'_{w,k}) \mathbf{Q} \mathbf{L}_{k-1} \end{aligned} \quad (\text{A.14i})$$

$$\begin{aligned} \mathbf{N}_k \equiv & [\Delta'_{w,k}(\mathbf{A}' + \mathbf{N}_{k-1} \cdot \mathbf{R}) - \mathbf{N}_{k-1} \cdot \mathbf{R}] \Delta_{w,k} + \\ & (\mathbf{I} - \Delta'_{w,k}) \mathbf{N}_{k-1} \cdot \mathbf{R} . \end{aligned} \quad (\text{A.14j})$$

This set of recursions completely characterizes the best execution strategy $\{\mathbf{s}_t^*\}$ and the expected best-execution cost is given by:

$$\begin{aligned} V_1(\tilde{\mathbf{p}}_1, \mathbf{x}_1, \mathbf{w}_1) = & \mathbf{e}'_n \mathbf{D}_{n,T-1} \mathbf{e}_n + \mathbf{e}'_r \mathbf{D}_{r,T-1} \mathbf{e}_r + \mathbf{e}'_n \mathbf{E}_{T-1} \mathbf{x}_1 + \mathbf{x}'_1 \mathbf{F}_{T-1} \mathbf{e}_n + \\ & \mathbf{x}'_1 \mathbf{G}_{T-1} \mathbf{x}_1 + \mathbf{x}'_1 \mathbf{H}_{T-1} \tilde{\mathbf{w}}_1 + \tilde{\mathbf{w}}'_1 \mathbf{J}_{T-1} \mathbf{x}_1 \\ & \mathbf{e}'_n \mathbf{K}_{T-1} \tilde{\mathbf{w}}_1 + \tilde{\mathbf{w}}'_1 \mathbf{L}_{T-1} \mathbf{e}_n + \tilde{\mathbf{w}}'_1 \mathbf{N}_{T-1} \tilde{\mathbf{w}}_1 \end{aligned} \quad (\text{A.15})$$

$$= \text{Min}_{\{\mathbf{s}_t\}} \mathbb{E}_1 \left[\sum_{t=1}^T \mathbf{p}'_t \mathbf{s}_t \right]. \quad (\text{A.16})$$

References

- Barclay, M. and R. Litzenberger, 1988, "Announcement Effects Of New Equity Issues And The Use Of Intraday Price Data", *Journal of Financial Economics* 21, 71–100.
- Barclay, M. and J. Warner, 1993, "Stealth Trading And Volatility: Which Trades Move Prices?", *Journal of Financial Economics* 34, 281–306.
- Bertsekas, D., 1995, *Dynamic Programming and Optimal Control*, Vol. I. Belmont, MA: Athena Scientific.
- Bertsimas, D. and A. Lo, 1998, "Optimal Control of Execution Costs", *Journal of Financial Markets* 1, 1–50.
- Birinyi, L., 1995, *What Does Institutional Trading Cost?* Greenwich, CT: Birinyi Associates, Inc.
- Chan, L. and J. Lakonishok, 1993, "Institutional Trades and Intra-Day Stock Price Behavior", *Journal of Financial Economics* 33, 173–199.
- Chan, L. and J. Lakonishok, 1995, "The Behavior of Stock Prices Around Institutional Trades", *Journal of Finance* 50, 1147–74.
- Holthausen, R., Leftwich, R. and D. Mayers, 1987, "The Effect of Large Block Transactions on Security Prices: A Cross-Sectional Analysis", *Journal of Financial Economics* 19, 237–267.
- Holthausen, R., Leftwich, R. and D. Mayers, 1990, "Large Block Transactions, the Speed of Response, and Temporary and Permanent Stock-Price Effects", *Journal of Financial Economics* 26, 71–95.
- Keim, D. and A. Madhavan, 1995a, "The Anatomy of the Trading Process", *Journal of Financial Economics* 37, 371–398.
- Keim, D. and A. Madhavan, 1995b, "The Upstairs Market for Large-Block Transactions: Analysis and Measurement of Price Effects", to appear in *Review of Financial Studies*.
- Keim, D. and A. Madhavan, 1995c, "Execution Costs and Investment Performance: An Empirical Analysis of Institutional Equity Trades", working paper, School of Business Administration, University of Southern California.
- Kraus, A. and H. Stoll, 1972, "Price Impacts of Block Trading on the New York Stock Exchange", *Journal of Finance* 27, 569–588.
- Leinweber, D., 1993, "Using Information From Trading in Trading and Portfolio Management", in K. Sherrerd, ed.: *Execution Techniques, True Trading Costs, and the Microstructure of Markets*. Charlottesville, VA: Association for Investment Management and Research.
- Leinweber, D., 1994, "Careful Structuring Reins In Transaction Costs", *Pensions and Investments* July 25, 19.
- Lo, A. and C. MacKinlay, 1990, "When Are Contrarian Profits Due To Stock Market Overreaction?", *Review of Financial Studies* 3, 175–206.

- Loeb, T., 1983, "Trading Cost: The Critical Link Between Investment Information and Results", *Financial Analysts Journal* 39, 39–44.
- Pérol, A., 1988, "The Implementation Shortfall: Paper Versus Reality", *Journal of Portfolio Management* 14, 4–9.
- Shleifer, A., 1986, "Do Demand Curves For Stocks Slope Down?", *Journal of Finance* 41, 579–590.

Table 1

Ticker symbols, CUSIPs, company names, and closing prices on a randomly selected day in 1996 for 25 stocks that comprise the sample portfolio for the empirical implementation of the best-execution strategy.

Ticker	CUSIP	Company Name	Closing Price
AHP	02660910	AMER HOME PRODS	64.0625
AN	03190510	AMOCO	70.5000
BLS	07986010	BELLSOUTH	37.2500
CHV	16675110	CHEVRON	62.6250
DD	26353410	DUPONT	88.9375
DIS	25468710	WALT DISNEY	63.4375
DOW	26054310	DOW CHEMICAL	80.6875
F	34537010	FORD MOTOR	31.3125
FNM	31358610	FANNIE MAE	35.0625
GE	36960410	GENERAL ELECTRIC	90.9375
GM	37044210	GM	48.1250
HWP	42823610	HEWLETT PACKARD	48.8125
IBM	45920010	IBM	25.8750
JNJ	47816010	JOHNSON & JOHNSON	51.4375
KO	19121610	COCA COLA	50.8125
MCD	58013510	MCDONALDS	47.7500
MO	71815410	PHILIP MORRIS	90.1875
MOB	60705910	MOBIL	15.9375
MRK	58933110	MERCK & CO	70.1250
PEP	71344810	PEPSICO	28.3750
PG	74271810	PROCTER & GAMBLE	97.4375
S	81238710	SEARS ROEBUCK	44.8750
T	00195710	AT&T	51.9375
WMT	93114210	WAL MART STORES	26.3125
XON	30229010	EXXON	83.5625

Table 2

Parameter estimates and correlations for the no-impact price process \tilde{p}_t for five stocks using 2,069 half-hourly observations from January 2, 1996 to December 31, 1996. The first and second rows give the annual drift and volatility parameters (percent/year) scaled up from half-hourly estimates by assuming 250 trading days with 13 half-hour periods per day. The last five rows report the correlation coefficients for the half-hourly returns of the five stocks.

	AHP	AN	BLS	CHV	DD
$\hat{\mu}$	0.200	-0.103	-0.387	-0.167	0.283
$\hat{\sigma}$	0.268	0.196	0.283	0.222	0.225
AHP	1.000	0.196	0.284	0.226	0.345
AN	0.196	1.000	0.173	0.408	0.283
BLS	0.284	0.173	1.000	0.259	0.328
CHV	0.226	0.408	0.259	1.000	0.314
DD	0.345	0.283	0.328	0.314	1.000

Table 3

Coefficients of the unconstrained price-impact regressions for five stocks, based on market orders from January 2, 1996 to December 31, 1996. All coefficients have been multiplied by 10^{10} except for the SPX coefficients, which have been multiplied by 10^5 . Sample size T and R^2 coefficients are contained in the last two rows, and t -statistics are reported in parentheses below the coefficients.

Variable	AHP	AN	BLS	CHV	DD	Variable	AHP	AN	BLS	CHV	DD
SPX	3.74 (0.92)	2.29 (0.43)	0.27 (0.04)	11.73 (1.94)	2.38 (0.56)	JNJ	0.81 (0.40)	1.60 (2.20)	3.88 (2.15)	1.37 (1.43)	1.65 (1.26)
AHP	4.97 (3.36)	-2.04 (-2.11)	-1.89 (-1.12)	0.80 (0.69)	-1.78 (-1.65)	KO	0.58 (0.53)	-0.20 (-0.28)	2.43 (1.46)	-0.36 (-0.52)	-0.30 (-0.56)
AN	-1.49 (-1.62)	5.86 (5.96)	-2.27 (-1.49)	1.38 (1.14)	-1.89 (-1.81)	MCD	2.56 (1.20)	-5.83 (-3.75)	5.80 (1.56)	-3.34 (-2.96)	-4.00 (-2.82)
BLS	4.46 (2.05)	0.31 (0.18)	0.85 (0.29)	4.37 (2.20)	-1.84 (-1.07)	MO	0.17 (0.28)	-1.78 (-1.53)	0.69 (0.77)	3.06 (1.93)	-0.01 (-0.05)
CHV	2.02 (1.38)	-0.25 (-0.32)	4.90 (2.54)	8.72 (2.84)	3.24 (3.30)	MOB	0.70 (0.69)	1.36 (1.32)	-2.02 (-1.16)	-0.48 (-0.56)	-0.32 (-0.40)
DD	-1.08 (-0.81)	-0.70 (-0.62)	5.31 (3.17)	1.29 (0.89)	7.21 (5.53)	MRK	-1.20 (-2.33)	0.65 (0.70)	0.31 (0.24)	-3.23 (-2.06)	1.03 (1.37)
DIS	2.35 (2.26)	3.04 (3.23)	0.01 (0.03)	-1.54 (-1.28)	1.17 (1.35)	PEP	-0.55 (-0.32)	5.71 (5.82)	2.01 (0.83)	-0.37 (-0.17)	1.12 (1.62)
DOW	-1.12 (-1.12)	0.12 (0.27)	-3.53 (-2.04)	-2.43 (-1.89)	-2.52 (-3.44)	PG	-0.28 (-0.24)	-2.15 (-1.99)	0.91 (0.50)	2.92 (2.19)	2.85 (2.47)
F	1.75 (0.70)	0.19 (0.15)	0.01 (0.02)	-1.01 (-0.47)	2.19 (1.16)	S	-0.01 (-0.04)	-2.69 (-1.95)	-9.44 (-2.80)	0.21 (0.08)	1.43 (0.96)
FNM	-5.10 (-1.44)	-4.87 (-2.63)	4.59 (1.19)	2.33 (0.95)	2.81 (1.23)	T	1.78 (2.63)	0.47 (0.05)	-0.49 (-0.32)	-1.49 (-2.82)	0.45 (0.82)
GE	0.14 (0.12)	0.78 (1.30)	0.11 (0.12)	0.96 (1.13)	-0.29 (-0.67)	WMT	1.58 (0.89)	-0.41 (-0.24)	3.89 (1.17)	-8.31 (-2.35)	3.69 (1.83)
GM	-3.74 (-2.32)	-0.67 (-0.99)	-0.30 (-0.50)	-0.16 (-0.21)	-3.30 (-4.80)	XON	1.66 (2.56)	-1.02 (-1.98)	-0.35 (-0.28)	0.20 (0.38)	1.04 (2.02)
HWP	-1.20 (-0.74)	1.45 (1.28)	-5.55 (-2.32)	3.78 (2.95)	3.26 (4.30)	T	726	887	768	791	759
IBM	-0.64 (-0.68)	0.51 (0.80)	-0.01 (-0.05)	-0.17 (-0.16)	-0.50 (-1.80)	R^2	0.11	0.23	0.08	0.09	0.21

Table 4

Coefficients of the constrained price-impact regressions for five stocks, based on market orders from January 2, 1996 to December 31, 1996. All coefficients have been multiplied by 10^{10} except for the SPX coefficients, which have been multiplied by 10^5 . Sample size T and R^2 coefficients are contained in the last two rows, and t -statistics are reported in parentheses below the coefficients.

Variable	AHP	AN	BLS	CHV	DD	Variable	AHP	AN	BLS	CHV	DD
SPX	3.74 (0.90)	2.28 (0.41)	0.26 (0.04)	11.70 (1.89)	2.38 (0.54)	JNJ	-1.39 (-0.67)	1.26 (1.70)	0.13 (0.07)	2.07 (2.11)	2.40 (1.78)
AHP	12.40 (8.17)	-1.69 (-1.70)	-1.99 (-1.15)	1.04 (0.88)	-1.07 (-0.96)	KO	1.11 (0.99)	0.61 (0.83)	0.69 (0.40)	-0.42 (-0.60)	1.44 (2.58)
AN	-1.32 (-1.40)	10.10 (9.98)	-1.96 (-1.25)	1.09 (0.88)	-1.63 (-1.53)	MCD	3.38 (1.55)	-3.73 (-2.34)	-0.24 (-0.06)	-2.48 (-2.15)	-1.08 (-0.74)
BLS	3.49 (1.56)	0.83 (0.48)	14.40 (4.82)	2.26 (1.11)	-2.45 (-1.39)	MO	-0.63 (-1.02)	-0.84 (-0.71)	1.87 (2.05)	-1.75 (-1.08)	0.25 (0.22)
CHV	2.09 (1.39)	-0.17 (-0.22)	2.34 (1.19)	21.20 (6.72)	2.84 (2.83)	MOB	0.36 (0.34)	-0.12 (-0.11)	-0.98 (-0.55)	-0.92 (-1.03)	0.41 (0.48)
DD	-0.93 (-0.68)	0.66 (0.57)	6.18 (3.60)	1.55 (1.05)	11.70 (8.72)	MRK	-0.75 (-1.41)	0.55 (0.58)	-0.29 (-0.22)	-0.18 (-0.11)	1.30 (1.67)
DIS	1.98 (1.85)	2.73 (2.84)	-0.30 (-0.09)	-2.59 (-2.09)	0.92 (1.04)	PEP	-0.85 (-0.48)	1.75 (1.74)	0.53 (0.21)	-1.93 (-0.90)	-0.28 (-0.39)
DOW	-0.79 (-0.77)	0.18 (0.40)	-3.88 (-2.19)	-0.23 (-0.18)	-1.59 (-2.13)	PG	-1.48 (-1.25)	-1.94 (-1.75)	0.89 (0.47)	3.51 (2.56)	2.70 (2.29)
F	-0.23 (-0.09)	0.66 (0.50)	3.29 (0.82)	-0.65 (-0.29)	4.24 (2.19)	S	0.01 (0.00)	-1.76 (-1.25)	-6.90 (-2.00)	1.13 (0.43)	-0.30 (-0.20)
FNM	0.01 (0.00)	-0.92 (-0.48)	-0.43 (-0.11)	3.27 (1.31)	0.39 (0.17)	T	1.84 (2.67)	0.70 (0.77)	-1.18 (-0.74)	-0.20 (-0.36)	0.59 (1.04)
GE	0.77 (0.66)	-0.45 (-0.73)	0.53 (0.59)	0.03 (0.04)	-0.41 (-0.92)	WMT	3.70 (2.02)	2.01 (1.15)	2.36 (0.69)	-3.65 (-1.01)	0.96 (0.46)
GM	-2.83 (-1.72)	-0.86 (-1.23)	0.23 (0.37)	0.28 (0.37)	-2.99 (-4.25)	XON	1.45 (2.18)	-0.14 (-0.26)	0.77 (0.61)	-0.16 (-0.29)	0.09 (0.17)
HWP	-1.66 (-1.00)	0.53 (0.46)	-2.73 (-1.11)	0.73 (0.55)	2.35 (3.03)	T	726	887	768	791	759
IBM	-0.09 (-0.09)	1.00 (1.53)	-0.77 (-0.47)	-1.33 (-1.28)	-0.24 (-0.84)	R^2	0.06	0.19	0.03	0.04	0.17

Table 5

Ratios of the sum of squared residuals of the unconstrained and constrained price-impact regressions for 25 stocks based on market orders from January 2, 1996 to December 31, 1996.

Ticker	Ratio	Ticker	Ratio	Ticker	Ratio	Ticker	Ratio	Ticker	Ratio
AHP	1.0501	DIS	1.0501	GM	1.0501	MCD	1.0501	PG	1.0500
AN	1.0498	DOW	1.0507	HWP	1.0416	MO	1.0499	S	1.0500
BLS	1.0500	F	1.0500	IBM	1.0508	MOB	1.0479	T	1.0281
CHV	1.0501	FNM	1.0497	JNJ	1.0501	MRK	1.0496	WMT	1.0500
DD	1.0501	GE	1.0500	KO	1.0494	PEP	1.0500	XON	1.0516