



ELSEVIER

Journal of Financial Markets 1 (1998) 1–50

---

---

Journal of  
FINANCIAL  
MARKETS

---

---

## Optimal control of execution costs

Dimitris Bertsimas, Andrew W. Lo\*

*Sloan School of Management, MIT, Cambridge, MA 02142–1347, USA*

---

### Abstract

We derive dynamic optimal trading strategies that minimize the expected cost of trading a large block of equity over a fixed time horizon. Specifically, given a fixed block  $\bar{S}$  of shares to be executed within a fixed finite number of periods  $T$ , and given a price-impact function that yields the execution price of an individual trade as a function of the shares traded and 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  $\bar{S}$  within  $T$  periods. Our analysis is extended to the portfolio case in which price impact *across* stocks can have an important effect on the total cost of trading a portfolio. © 1998 Elsevier Science B.V. All rights reserved.

*JEL classification:* G23

*Keywords:* Market microstructure; Transactions costs; Portfolio management

---

### 1. Introduction

The tremendous growth in equity trading over the past 20 years, fueled largely by the burgeoning assets of institutional investors such as mutual and pension funds, has created a renewed interest in the measurement and management of

---

\* Corresponding author. Tel.: (617) 253-0920; fax: (781) 863-9695; e-mail: alo@mit.edu.

trading costs.<sup>1</sup> Such costs – often called ‘execution costs’ because they are associated with the execution of investment strategies – include commissions, bid/ask spreads, opportunity costs of waiting, and price impact from trading (see Loeb, 1983 and Wagner, 1993 for further discussion), and they can have a substantial impact on investment performance. For example, Pérold (1988) observes that a hypothetical or ‘paper’ portfolio constructed according to the Value Line rankings outperforms the market by almost 20% per year during the period from 1965 to 1986, whereas the actual portfolio – the Value Line Fund – outperformed the market by only 2.5% per year, the difference arising from execution costs.<sup>2</sup> This ‘implementation shortfall’ is surprisingly large and underscores the importance of execution-cost control, particularly for institutional investors whose trades often comprise a large fraction of the average daily volume of many stocks.

There has also been considerable interest from the regulatory perspective in defining ‘best’ execution, especially in the wake of recent concerns about NASDAQ trading practices, the impact of tick size on trading costs, and the economic consequences of market fragmentation. Indeed, Macey and O’Hara (1996) observe that ‘... while the obligation to give customers the benefits of best-execution of orders is one of the most well-established principles of securities law, and despite the fact that the concept of best execution is continually referred to in cases, treatises, law review articles, exchange rules, and administrative proceedings, no clear definition of best execution exists’.

---

<sup>1</sup> There is a large and diverse literature on transactions costs in economics and finance. Niehaus (1987) provides an excellent review of the transactions costs literature in economics.

The impact of transactions costs on financial decision-making has been considered by Aiyagari and Gertler (1991), Bensaid et al. (1992), Boyle and Vorst (1992), Cohen et al. (1981), Constantinides (1986), Davis and Norman (1991), Dumas and Luciano (1991), Epps (1976), Garman and Ohlson (1981), Grossman and Miller (1988), Grossman and Vila (1992), Heaton and Lucas (1994, 1995), Hodges and Neuberger (1989), Litzenberger and Rolfo (1984), Leland (1985), Magill and Constantinides (1976), Tuckman and Vila (1992), Vayanos (1995), Vayanos and Vila (1995), and many others.

Empirical studies of transactions costs in financial markets include Berkowitz et al. (1988), Birinyi (1995), Brennan and Copeland (1988), Chan and Lakonishok (1993, 1995), Demsetz (1968), Hasbrouck and Schwartz (1988), Huang and Stoll (1995), Keim and Madhavan (1995a,b,c), Kraus and Stoll (1972), Schwartz and Whitcomb (1988), Stoll (1989, 1993), Tinic (1972), and Turnbull and White (1995).

And the more practical aspects of transactions cost management have been explored by Arnott and Wagner (1990), Bodurtha and Quinn (1990), Brinson et al. (1986), Brinson et al. (1991), Collins and Fabozzi (1991), Cuneo and Wagner (1975), Gammill and Pérold (1989), Loeb (1983), Pérold (1988), Treynor (1981), Wagner and Banks (1992), Wagner and Edwards (1993), and the papers in Sherrerd (1993).

<sup>2</sup> See also Bodurtha and Quinn (1990), Brinson et al. (1986), Brinson et al. (1991), Chan and Lakonishok (1993, 1995), Hasbrouck and Schwartz (1988), Keim and Madhavan (1995a,b,c), Loeb (1983), Treynor (1981), and Turnbull and White (1995).

In this paper, we provide one clear definition of best execution, based on the minimization of the expected cost of execution using stochastic dynamic programming. While dynamic optimization is certainly not new to financial economics (see, for example, Merton, 1969; Samuelson, 1969), the use of dynamic programming in defining best execution is novel. In particular, our approach explicitly recognizes the fact that trading takes time, and that the very act of trading affects not only current prices but also *price dynamics* which, in turn, affects future trading costs. Therefore, defining and controlling execution costs are fundamentally dynamic problems, not static ones, a fact recognized implicitly by Pérol (1988) and most professional portfolio managers, and developed explicitly here.

Indeed, recent studies by Chan and Lakonishok (1995) and Keim and Madhavan (1995a,b,c) show that because the typical institutional investor's trades are so large, they are almost always broken up into smaller trades executed over the course of several days. Chan and Lakonishok call such sequences 'packages', and using a sample of 1.2 million transactions of 37 large investment management firms during the period from July 1986 to December 1988, they show that only 20% of the market value of these packages are completed within a day and that over 53% are spread over four trading days or more (Chan and Lakonishok, 1995, Table 1). For this reason, best execution cannot be defined as a single number or in the context of a single trade – it is a *strategy* that unfolds over the course of several days and which ought to adapt to changing market conditions.

Dynamic optimization provides a compelling economic rationale for trading in packages: properly parceled packages minimize the expected costs of execution over a fixed time horizon. In particular, we propose and solve the following problem in this paper: given a fixed block  $\bar{S}$  of shares to be executed within a fixed finite number of periods  $T$ , and given 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 will minimize the expected cost of executing  $\bar{S}$  within  $T$  periods.<sup>3</sup>

Using stochastic dynamic programming, we obtain explicit closed-form expressions for these optimal trading strategies, which we call *best-execution strategies*, for various specifications of price dynamics. We show that best-execution strategies can sometimes be expressed as a linear combination of a naive (and not uncommon) strategy – breaking up  $\bar{S}$  shares evenly into

---

<sup>3</sup> We do not consider the selection of the particular *type* of trade – limit versus market order – in a dynamic context. For an analysis of this complementary and important issue, see Angel (1994), Bernhardt and Hughson (1994), Harris (1994), Hasbrouck and Harris (1992), Kumar and Seppi (1993), and Lo et al. (1998).

a package of  $T$  trades each of size  $\bar{S}/T$  – and a correction factor that adjusts each trade size up or down according to additional information, e.g., stock-specific private information, opportunity costs, changing market conditions, etc. In the absence of such information, we derive conditions under which the naive strategy is optimal: an arithmetic random walk for prices with linear price impact (see Section 2.3). We also show by construction that apart from these rather restrictive and empirically implausible conditions, the naive strategy is *not* optimal in general.

We also obtain the expected cost of best execution – the optimal-value function which is given recursively by the Bellman equation – as a by-product of the optimization process, which may serve as a useful benchmark for pricing principal-bid and negotiated-block transactions. The typical broker/dealer engaging in such transactions will not willingly hold large positions for long, and will seek to trade out of these positions as quickly and as cost-effectively as possible, i.e., he will seek best-execution strategies for his holdings. Of course, risk aversion, adverse selection, and inventory and opportunity costs may change the objective function to be minimized, in which case our benchmark may only be a lower bound on the fair market price of a block transaction. Nevertheless, even in these cases the problem of best execution is still a dynamic optimization problem and our approach is still applicable (although closed-form expressions for best-execution strategies may not be available).

Moreover, we show that our basic approach – described in Sections 2 and 3 – can be extended in several important ways: (1) specifying more general price-impact functions and deriving numerical solutions (Section 4); (2) trading a portfolio of stocks simultaneously (Section 5); and (3) imposing constraints such as no-sales or, in the portfolio case, a maximum dollar amount invested (Section 6). These results comprise a systematic and quantitative approach to defining and controlling execution costs, measuring the liquidity of large-block transactions, and rationalizing within an economic paradigm the kind of informal trading practices that characterize many institutional equity investors.

## 2. The basic model

Consider an investor seeking to acquire a large block of  $\bar{S}$  shares of some stock over a fixed time interval  $[0, T]$ .<sup>4</sup> Since it is well-known that the short-term demand curves for even the most actively traded equities are not perfectly elastic, a market order at date 0 for the entire block  $\bar{S}$  is clearly not an optimal

---

<sup>4</sup>For simplicity, we consider buying  $\bar{S}$  only – selling  $\bar{S}$  is symmetric and the solution follows directly from our analysis below with only minor changes in notation.

trading strategy.<sup>5</sup> A more effective strategy would be to break  $\bar{S}$  into smaller purchases distributed throughout the interval  $[0, T]$ , but how should such purchases be parceled out?

The answer depends, of course, on the degree to which a purchase affects the market price, i.e., the ‘price impact’ and the dynamics of future market prices. Given a particular price-impact function, e.g., the ordered probit model of Hausman et al. (1992), and a specification for the price dynamics, e.g., a random walk, a dynamic optimal trading strategy that minimizes the expected total cost of acquiring  $\bar{S}$  in  $[0, T]$  may be obtained by stochastic dynamic programming.

### 2.1. Defining best execution

To illustrate this approach, suppose that at time 0 the investor begins his program to acquire  $\bar{S}$  shares, and this program must be completed by time  $T$ . With little loss in generality, let time be measured in discrete intervals of unit length. Since the length of a ‘period’ is arbitrary, it can be set to accommodate even the finest trading-decision interval that is of practical relevance. For example, if the decision to acquire  $\bar{S}$  is made at the start of the day and the acquisition must be completed by the day’s end, setting  $T = 13$  yields 30 minute intervals from the 9:30 am market open to the 4:00 pm market close. If the acquisition is part of an end-of-quarter portfolio rebalancing, the trading horizon may be extended to three or four days, in which case  $T$  increases proportionally.<sup>6</sup> Although all of our results are qualitatively independent of both the time horizon and the number of trading periods (with the exception of numerical examples, of course), for concreteness the length of each period should be regarded as some fraction of a single day and related parameters should be calibrated accordingly.

Denote by  $S_t$  be the number of shares acquired in period  $t$  at price  $P_t$ , where  $t = 1, 2, \dots, T$ . Then the investor’s objective of minimizing execution costs may

---

<sup>5</sup> For theoretical models that provide economic motivation for the existence of price impact, see Allen and Gale (1992), Easley and O’Hara (1987), Foster and Viswanathan (1990), Jarrow (1992), Kyle (1985, 1989), Vayanos (1992), and Vila (1989). For empirical studies of price impact, see Barclay and Warner (1993), Bodurtha and Quinn (1990), Chan and Lakonishok (1993, 1995), Hausman et al. (1992), Holthausen et al. (1987, 1990), Keim and Madhavan (1995b,c), Kraus and Stoll (1972), and Shleifer (1986).

<sup>6</sup> If 30 minute intervals are deemed too long for a given trading horizon, either because of volatility or opportunity cost,  $T$  can be increased and the parameters of the optimization problem rescaled accordingly. Of course, a more natural way to handle the timing issue is to re-cast this problem in continuous time. However, the analysis becomes considerably more complex and closed-form solutions are difficult to obtain (see, for example, Davis and Norman, 1990; Magill and Constantinides, 1976). Moreover, for empirical applications continuous-time models must be discretized. Therefore, we take a discrete-time approach throughout this paper.

be expressed as:

$$\text{Min}_{\{S_t\}} E_1 \left[ \sum_{t=1}^T P_t S_t \right] \tag{2.1}$$

subject to the constraint

$$\sum_{t=1}^T S_t = \bar{S}. \tag{2.2}$$

We may also wish to impose a no-sales constraint, i.e.,  $S_t \geq 0$  (after all, it is difficult to justify selling stocks as part of a buy-program), but for expositional convenience we shall ignore these constraints for now and return to them in Section 6.

To complete the statement of the problem, we must specify the ‘law of motion’ for  $P_t$ . This includes two distinct components: the dynamics of  $P_t$  in the absence of our trade (the trades of others may be causing prices to fluctuate), and the impact that our trade of  $S_t$  shares has on the execution price  $P_t$ . For simplicity, suppose that the former component is given by an arithmetic random walk, and the latter component is simply a linear function of trade size so that a purchase of  $S_t$  shares may be executed at the prevailing price  $P_{t-1}$  plus an impact premium of  $\theta S_t$ ,  $\theta > 0$ .<sup>7</sup> Then the law of motion for  $P_t$  may be expressed as:

$$P_t = P_{t-1} + \theta S_t + \varepsilon_t, \quad \theta > 0, E[\varepsilon_t | S_t, P_{t-1}] = 0 \tag{2.3}$$

where  $\varepsilon_t$  is assumed to be a zero-mean independently and identically distributed (IID) random shock, i.e., white noise.

Observe that the two components – price impact and price dynamics – can be separated. A nonlinear price impact function can easily be incorporated into the random walk specification, and non-random-walk dynamics can be combined with a linear price impact function. However, these two components interact in important ways. For example, Eq. (2.3) implies that price impact has a ‘permanent’ effect on the price level because of the random-walk specification of the price dynamics. It is this interaction between price impact and price dynamics that makes execution-cost control a dynamic optimization problem. This interaction also explains the difficulties in developing a clear economic definition of best execution: such a definition requires the specification of price dynamics as well as price impact, and these vary from one stock to another, and may well vary over time (see Bertsimas et al., 1998 for further details).

---

<sup>7</sup> Although we refer to  $\theta S_t$  as a price-impact function, it is a simple matter to modify it include other components of execution costs such as commissions, a bid/ask spread, and opportunity cost. However, for expositional convenience we shall refer to  $\theta S_t$  simply as price impact throughout this paper.

Despite the fact that Eq. (2.3) has some implausible empirical implications – independent price increments, positive probability of negative prices, percentage price impact that decreases with price, permanent price impact, etc. – it provides a concrete illustration of the more general and considerably more complex analysis which we shall present in later sections. Moreover, we shall see in Section 2.3 that Eq. (2.3) is precisely the dynamics necessary to render the naive strategy of dividing  $\bar{S}$  into  $T$  trades each of size  $\bar{S}/T$  the optimal one.

The investor's problem is now well-posed: find the *sequence* of trades  $\{S_t\}$  that minimizes the expected execution costs  $E_0[\sum_{t=1}^T P_t S_t]$ , subject to the constraint that  $\sum_{t=1}^T S_t = \bar{S}$ , and given a linear price-impact function incorporated into the law of motion (2.3) for  $P_t$ . This is a classical optimal control problem which can be solved by stochastic dynamic programming, and we define the best-execution strategy as its solution.

## 2.2. The Bellman equation

The basic ingredients for any dynamic programming problem are the state of the environment at time  $t$ , the control variable, the randomness, the cost function, and the law of motion. In our context, the state at time  $t = 1, \dots, T$  consists of the price  $P_{t-1}$  realized at the previous period, and  $W_t$ , the number of shares that remain to be purchased. The state variables summarize all the information the investor requires in each period  $t$  to make his decision regarding the control. The control variable at time  $t$  is the number of shares  $S_t$  purchased. The randomness is characterized by the random variable  $\varepsilon_t$ . The objective is given by Eq. (2.1), while the law of motion is given by Eq. (2.3) and an additional state equation which measures the remaining number of shares to be traded:

$$W_t = W_{t-1} - S_{t-1}, \quad W_1 = \bar{S}, \quad W_{T+1} = 0, \quad (2.4)$$

where the boundary condition  $W_{T+1} = 0$  is equivalent to the constraint that  $\bar{S}$  must be executed by period  $T$ .

The dynamic programming algorithm is based on the observation that a solution or 'optimal control'  $\{S_1^*, S_2^*, \dots, S_T^*\}$  must also be optimal for the remaining program at every intermediate date  $t$ . That is, for every  $t$ ,  $1 < t < T$  the sequence  $\{S_t^*, S_{t+1}^*, \dots, S_T^*\}$  must still be optimal for the remaining program  $E_t[\sum_{k=t}^T P_k S_k]$ . This important property is summarized by the Bellman equation (2.5), which relates the optimal value of the objective function in period  $t$  to its optimal value in period  $t + 1$ :

$$V_t(P_{t-1}, W_t) = \text{Min}_{S_t} E_t[P_t S_t + V_{t+1}(P_t, W_{t+1})]. \quad (2.5)$$

By starting at the end (time  $T$ ) and applying the Bellman equation (2.5) and the law of motion for  $P_t$ , Eq. (2.3) and  $W_t$ , Eq. (2.4), recursively, the optimal control

can be derived as functions of the state variables that characterize the information that the investor must have to make his decision in each period.

In particular, the optimal-value function  $V_T(\cdot)$ , as a function of the two state variables  $P_{T-1}$  and  $W_T$ , is given by

$$V_T(P_{T-1}, W_T) = \text{Min}_{S_T} E_T[P_T W_T] = (P_{T-1} + \theta W_T) W_T. \quad (2.6)$$

Since this is the last period and  $W_{T+1}$  must be set to 0, there is no choice but to execute the entire remaining order  $W_T$ , hence the optimal trade size  $S_T^*$  is simply  $W_T$ . Substituting the law of motion (2.3) into  $P_T W_T$  yields  $V_T$  as a function of  $P_{T-1}$  and  $W_T$ .

In the next-to-last period  $T - 1$ , the Bellman equation is less trivial:

$$\begin{aligned} V_{T-1}(P_{T-2}, W_{T-1}) &= \text{Min}_{S_{T-1}} E_{T-1}[P_{T-1} S_{T-1} + V_T(P_{T-1}, W_T)] \quad (2.7) \\ &= \text{Min}_{S_{T-1}} E_{T-1}[(P_{T-2} + \theta S_{T-1} + \varepsilon_{T-1}) S_{T-1} + V_T(P_{T-2} + \theta S_{T-1} \\ &\quad + \varepsilon_{T-1}, W_{T-1} - S_{T-1})]. \quad (2.8) \end{aligned}$$

By substituting the right-hand side of Eq. (2.6) into Eq. (2.8), and substituting for  $W_T$  and  $P_{T-1}$  using Eqs. (2.3) and (2.4), respectively, the right-hand side of Eq. (2.8) may be cast as an explicit function of  $S_{T-1}$  which can be minimized by taking its derivative with respect to  $S_{T-1}$  and solving for its zero. This yields:

$$S_{T-1}^* = W_{T-1}/2, \quad (2.9)$$

$$V_{T-1}(P_{T-2}, W_{T-1}) = W_{T-1}(P_{T-2} + \frac{3}{4}\theta W_{T-1}), \quad (2.10)$$

where Eq. (2.10) is obtained by substituting  $S_{T-1}^*$  into Eq. (2.8).

Continuing in this fashion, the optimal trades  $S_{T-k}^*$  and the optimal-value function  $V_{T-k}(P_{T-k-1}, W_{T-k})$  may be obtained recursively as:

$$S_{T-k}^* = W_{T-k}/(k + 1) \quad (2.11)$$

$$V_{T-k}(P_{T-k-1}, W_{T-k}) = W_{T-k} \left( P_{T-k-1} + \frac{k+2}{2(k+1)} \theta W_{T-k} \right), \quad (2.12)$$

until we reach the beginning of the program and find

$$S_1^* = W_1/T \quad (2.13)$$

$$V_1(P_0, W_1) = W_1 \left( P_0 + \frac{T+1}{2T} \theta W_1 \right). \quad (2.14)$$



### 2.3. The best-execution strategy

Substituting the initial conditional  $W_1 = \bar{S}$  into Eqs. (2.13) and (2.14) then yields the optimal trade size  $S_1^*$  as an explicit function of  $\bar{S}$ , and the expected best-execution cost  $V_1$  as an explicit function of  $\bar{S}$ ,  $P_0$ , and the price-impact parameter  $\theta$ :

$$S_1^* = \bar{S}/T, \quad (2.15)$$

$$V_1(P_0, W_1) = E_1 \left[ \sum_{t=1}^T P_t S_t^* \right] = P_0 \bar{S} + \frac{\theta \bar{S}^2}{2} \left( 1 + \frac{1}{T} \right). \quad (2.16)$$

By forward substitution using Eqs. (2.15), (2.11) and (2.4), we find that

$$S_1^* = S_2^* = \dots = S_T^* = \bar{S}/T. \quad (2.17)$$

The best-execution strategy is simply to divide the total order  $\bar{S}$  into  $T$  equal ‘waves’ and trade them at regular intervals.

This remarkably simple trading strategy comes from the fact that the price impact  $\theta S_t$  does not depend on either the prevailing price  $P_{t-1}$  or the size of the unexecuted order  $W_t$ , hence the price-impact function is the same in each period and independent from one period to the next. But since each period’s execution cost  $P_t S_t$  is a convex (quadratic) function of  $S_t$ , the sum of these single-period execution costs will be minimized at the point where the marginal execution costs are equated across all periods. There is no advantage to shifting trades to one period or another – they all offer the same trade-offs to the objective function – hence the trade sizes should be set equal across all periods. Note that in this case the optimal controls  $\{S_t^*\}$  are all non-negative hence the non-negativity constraints could have been imposed trivially.

The optimal-value function at time 1,  $V_1(P_0, W_1)$ , gives the expected cost of the best-execution strategy and we see from Eq. (2.16) that this cost is the sum of two terms: the no-impact cost  $P_0 \bar{S}$  and the cumulative price impact  $\theta \bar{S}^2 (1 + (1/T))/2$ . Observe that while the impact term is a decreasing function of  $T$  – having more time to acquire  $\bar{S}$  can never increase the expected cost – the cumulative price impact does not vanish as  $T$  increases without bound. This seems counterintuitive since one might expect price impact to become negligible if there is no time limit on completing the purchase. However, observe that our law of motion for  $P_t$ , Eq. (2.3), implies that the price impact  $\theta S_t$  of an individual trade has a *permanent* effect on  $P_t$ , hence even infinitesimally small trades will have an impact on next period’s price, and the limiting sum of all these infinitesimal trades multiplied by infinitesimally increased prices is finite and non-zero:  $\theta \bar{S}^2/2$ .

These results underscore the importance of the law of motion’s specification in determining the total expected cost of executing  $\bar{S}$ . Of course, Eq. (2.3) is empirically implausible for a number of reasons. However, it serves a useful purpose in demonstrating the basic approach to best execution, as well as in

rationalizing the rather common practice of parceling a large trade into smaller pieces of equal size and submitting them at regular intervals over some fixed time span. This naive strategy is indeed optimal if the price-impact function and price dynamics of  $P_t$  are given by Eq. (2.3).

In the next section we present a closed-form solution for the best-execution strategy under a more complex price-impact function, one which depends both on the trade size and a serially-correlated state variable that proxies for information such as proprietary research or market conditions. With information, the best-execution strategy differs in important ways from the naive strategy  $S_t^* = \bar{S}/T$ . In particular, the best-execution strategy becomes a nontrivial function of the information variable and can sometimes exhibit counterintuitive trading patterns.

#### 2.4. Linear price impact with information

Suppose that the price-impact function is linear in  $S_t$  as in Eq. (2.3), but now let  $X_t$  denote a serially-correlated state variable which also affects the execution price  $P_t$  linearly, hence

$$P_t = P_{t-1} + \theta S_t + \gamma X_t + \varepsilon_t, \quad \theta > 0 \quad (2.18)$$

$$X_t = \rho X_{t-1} + \eta_t, \quad \rho \in (-1, 1) \quad (2.19)$$

where  $\varepsilon_t$  and  $\eta_t$  are independent white noise processes with mean 0 and variances  $\sigma_\varepsilon^2$  and  $\sigma_\eta^2$ , respectively.

The presence of  $X_t$  in the law of motion for  $P_t$  captures the potential impact of changing market conditions or of private information about the security. For example,  $X_t$  might be the return on the S&P 500 index, a common component in the prices of most equities. Broad market movements affect all securities to some degree, and  $\gamma$  measures the sensitivity of this particular security to such market movements.

Alternatively,  $X_t$  might represent some private information about the security, and  $\gamma$  the importance of that information for  $P_t$ . In particular,  $X_t$  may denote the output of an ‘alpha model’ which incorporates stock-specific analysis to yield an excess return not yet impounded into market prices.

In either case, the impact of  $X_t$  on the execution price, and the time series properties of  $X_t$  have important implications for the best-execution strategy. Having specified the linear price-impact function with information in Eqs. (2.18) and (2.19), the best-execution strategy and optimal-value function can be obtained by dynamic programming as before (see the Appendix), and is given by

$$S_{T-k}^* = \delta_{w,k} W_{T-k} + \delta_{x,k} X_{T-k}, \quad (2.20)$$

$$V_{T-k}(P_{T-k-1}, X_{T-k}, W_{T-k}) = P_{T-k-1} W_{T-k} + a_k W_{T-k}^2 + b_k X_{T-k} W_{T-k} + c_k X_{T-k}^2 + d_k \quad (2.21)$$

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

$$\delta_{w, k} \equiv \frac{1}{k + 1}, \quad \delta_{x, k} \equiv \frac{\rho b_{k-1}}{2a_{k-1}} \quad (2.22)$$

and

$$\begin{aligned} a_k &= \frac{\theta}{2} \left( 1 + \frac{1}{k + 1} \right), & a_0 &= \theta, \\ b_k &= \gamma + \frac{\theta \rho b_{k-1}}{2a_{k-1}}, & b_0 &= \gamma, \\ c_k &= \rho^2 c_{k-1} - \frac{\rho^2 b_{k-1}^2}{4a_{k-1}}, & c_0 &= 0, \\ d_k &= d_{k-1} + c_{k-1} \sigma_\eta^2, & d_0 &= 0. \end{aligned} \quad (2.23)$$

Since we assume  $\theta > 0$ ,  $a_k$  is positive, and  $c_k$  and  $d_k$  are negative for all  $k > 0$ . The sign of  $b_k$  can vary, but is positive for all  $k > 0$  if  $\theta$ ,  $\gamma$ , and  $\rho$  are all positive.

In contrast to the case of a linear price-impact function with no information, the best-execution strategy (2.20) varies over time as a linear function of the remaining shares  $W_{T-k}$  and the information variable  $X_{T-k}$ . In particular, the first term of Eq. (2.20) is simply the naive strategy of dividing the remaining shares  $W_{T-k}$  at time  $T - k$  evenly over the remaining  $k + 1$  periods.

The second term of Eq. (2.20) is an adjustment that arises from the presence of serially correlated information  $X_{T-k}$ . Observe that this term vanishes if  $\rho = 0$ . When  $\rho = 0$  this implies that  $X_{T-k}$  is unforecastable, and while  $X_{T-k}$  still has an impact on the current execution price, observing  $X_{T-k}$  tells us nothing about expected future execution prices hence it can no longer affect the best-execution strategy.

If  $\rho > 0$  and we assume, without loss of generality, that  $\gamma > 0$ , then  $\delta_{x,k}$  in Eq. (2.20) is also positive, implying that positive realizations of  $X_{T-k}$  increases the number of shares purchased at  $T - k$ , *ceteris paribus*. This may seem counterintuitive at first because a positive  $X_{T-k}$  necessarily increases the execution price  $P_{T-k}$  by  $\gamma X_{T-k}$ , so why trade more?

The answer may be found in the fact that  $X_{T-k}$  is positively serially correlated, hence  $X_{T-k} > 0$  implies that future realizations are likely to be positive which, in turn, implies additional expected increases in future execution prices.<sup>8</sup> Therefore, although a positive  $X_{T-k}$  makes it more costly to purchase shares in

---

<sup>8</sup> In particular, a one-unit increase in  $X_{T-k}$  implies an expected increase in  $P_{T-k+1}$  of  $\gamma\rho$ , an expected increase in  $P_{T-k+2}$  of  $\gamma\rho^2$ , and so on.

period  $T - k$ , this additional cost is more than offset by the *sequence* of expected future price increases that arise from positively serially-correlated information.

Alternatively, if  $\rho < 0$  so that  $X_{T-k}$  exhibits reversals, Eq. (2.20) shows that a positive realization of  $X_{T-k}$  *decreases* the number of shares purchased, *ceteris paribus*: it is more expensive to trade in period  $T - k$  and  $X_{T-k}$  is likely to reverse next period making it less expensive to trade then, hence it is optimal to trade less now.

The impact of an increase in  $X_{T-k}$  on expected best-execution costs may be measured explicitly by the derivative of the optimal-value function  $V_{T-k}$  with respect to  $X_{T-k}$ :

$$\partial V_{T-k} / \partial X_{T-k} = b_k W_{T-k} + 2c_k X_{T-k}. \quad (2.24)$$

Suppose  $\gamma$  and  $\rho$  are positive so that  $b_k$  is positive. Since  $c_k$  is always negative, the impact of an increase in  $X_{T-k}$  on the expected best-execution cost depends on whether  $b_k W_{T-k}$  outweighs  $2c_k X_{T-k}$ . For empirically plausible parameter values, the first term will generally dominate the second, hence increases in  $X_{T-k}$  will typically increase the expected best-execution cost, a sensible implication given that an increase in  $X_{T-k}$  increases current and all future expected prices.

It is also not surprising that  $V_{T-k}$  is an increasing function of  $W_{T-k}$  for empirically plausible parameter values – the larger is the unexecuted portion of the initial block, the higher the expected best-execution cost. In the next section, we provide a numerical example to illustrate the behavior of the best-execution strategy under several simulated scenarios.

## 2.5. A numerical example

Tables 1–3 provide illustrative numerical examples of the best-execution strategies under the linear price-impact function with information (2.18)–(2.19) for three simulated realizations of the information variable  $X_t$  and pricing shocks  $\varepsilon_t$ . The goal is to minimize the expected execution costs of a 100,000-share purchase over  $T = 20$  periods for a stock currently trading at  $P_0 = \$50$ , given the following parameter values:

$$\theta = 5 \times 10^{-5}, \quad \gamma = 5.0, \quad \rho = 0.50, \quad \sigma_\varepsilon^2 = (0.125)^2, \quad \sigma_\eta^2 = 0.001.$$

To develop some intuition for these parameters, observe that the no-impact cost of acquiring  $\bar{S}$  is  $100,000 \times P_0 = \$5\text{M}$ , and the expected full-impact cost is

$$100,000 \times \text{E}[P_0 + \theta \bar{S} + \gamma X_1 + \varepsilon_1] = 100,000 \times \$55 = \$5.5 \text{ million}$$

since  $\text{E}[X_t] = 0$  from Eq. (2.19), hence  $\theta$  is calibrated to yield an impact of \$500,000 on a 100,000-share block purchase. From Eq. (2.19) it also

Table 1

Best-execution strategies for acquiring a 100,000-share block of stock within 20 periods, for a simulated realization of the information variable  $\{X_t\}$  and price shock  $\{\varepsilon_t\}$ , assuming an initial price  $P_0 = \$50$  and linear price-impact function  $\theta S_t + \gamma X_t$ , where  $P_t = P_{t-1} + \theta S_t + \gamma X_t + \varepsilon_t$  and  $\theta = 5 \times 10^{-5}$ ,  $\gamma = 5.0$ ,  $X_t = \rho X_{t-1} + \eta_t$ ,  $X_0 = 0$ ,  $\sigma_\eta^2 = 0.001$ ,  $\sigma_\varepsilon^2 = (0.125)^2$ , and  $\rho = 0.50$ . Note: The non-negativity restriction was not imposed and was not binding in this realization.

<i>Performance summary</i>							
Metric	Naive strategy (\$1000s)		Optimal strategy (\$1000s)			Improvement (cents/share)	
Expected cost	5255.2		5251.4			3.7899	
Actual cost	5314.9		5296.6			18.2906	
Period	$P_t$	$S_t^*$	$\delta_{w,T-t}W_t$	$\delta_{x,T-t}X_t$	$\gamma X_t$	$V_t/\$1000$	Cost-to-go
1	49.994	4307	5000	- 693	- 0.04	5251.4	5296.6
2	50.187	5470	5036	434	0.02	5026.0	5081.3
3	50.666	6383	5012	1371	0.08	4752.7	4806.7
4	50.572	2114	4932	- 2817	- 0.16	4405.4	4483.3
5	50.783	5680	5108	572	0.03	4312.7	4376.4
6	51.091	6250	5070	1180	0.07	4023.2	4087.9
7	51.673	8527	4985	3542	0.21	3720.6	3768.7
8	52.169	8423	4713	3710	0.22	3289.4	3328.0
9	52.923	8230	4404	3827	0.23	2852.6	2888.6
10	53.547	8227	4056	4171	0.25	2434.1	2453.0
11	54.213	6916	3639	3278	0.20	1996.7	2012.5
12	54.414	3182	3275	- 92	- 0.01	1620.6	1637.5
13	54.935	6578	3286	3292	0.22	1458.9	1464.4
14	55.086	3040	2816	224	0.02	1093.9	1103.0
15	55.615	4743	2779	1965	0.15	930.0	935.6
16	56.095	4305	2386	1919	0.16	670.3	671.8
17	56.132	1465	1906	- 441	- 0.04	428.8	430.3
18	56.402	2621	2053	569	0.07	347.5	348.1
19	56.475	1617	1768	- 152	- 0.03	199.8	200.2
20	56.722	1920	1920	0	- 0.02	108.9	108.9

follows that:

$$\sqrt{\text{Var}[\gamma X_t]} = \frac{\gamma \sigma_\eta}{\sqrt{1 - \rho^2}} = 0.183$$

hence the standard deviation of the information component is approximately 18 cents (per period). Finally, the standard deviation of  $\varepsilon_t$  is calibrated to be 12.5 cents or one ‘tick’ (per period).

Table 2

Best-execution strategies for acquiring a 100,000-share block of stock within 20 periods, for a simulated realization of the information variable  $\{X_t\}$  and price shock  $\{\varepsilon_t\}$ , assuming an initial price  $P_0 = \$50$  and linear price-impact function  $\theta S_t + \gamma X_t$ , where  $P_t = P_{t-1} + \theta S_t + \gamma X_t + \varepsilon_t$  and  $\theta = 5 \times 10^{-5}$ ,  $\gamma = 5.0$ ,  $X_t = \rho X_{t-1} + \eta_t$ ,  $X_0 = 0$ ,  $\sigma_\eta^2 = 0.001$ ,  $\sigma_\varepsilon^2 = (0.125)^2$ , and  $\rho = 0.50$ . Note: The non-negativity restriction was not imposed and was not binding in this realization.

<i>Performance summary</i>							
Metric	Naive strategy (\$1000s)	Optimal strategy (\$1000s)	Improvement (cents/share)				
Expected cost	5231.0	5226.9	4.0849				
Actual cost	5248.3	5252.4	– 4.1143				
Period	$P_t$	$S_t^*$	$\delta_{w,T-t}W_t$	$\delta_{x,T-t}X_t$	$\gamma X_t$	$V_t/\$1000$	Cost-to-go
1	49.927	2014	5000	– 2986	– 0.17	5226.9	5252.4
2	50.508	7927	5157	2770	0.15	5169.8	5151.8
3	50.779	6310	5003	1307	0.07	4771.9	4751.5
4	51.390	5070	4926	144	0.01	4436.7	4431.1
5	51.693	5547	4917	630	0.04	4210.4	4170.5
6	51.669	4227	4875	– 648	– 0.04	3915.5	3883.7
7	51.844	6208	4922	1287	0.08	3694.8	3665.3
8	51.767	1322	4823	– 3501	– 0.21	3331.2	3343.3
9	51.644	2794	5115	– 2320	– 0.14	3261.7	3274.9
10	51.890	5978	5325	653	0.04	3121.8	3130.6
11	51.901	2880	5260	– 2380	– 0.15	2790.2	2820.4
12	52.045	4950	5525	– 575	– 0.04	2645.1	2670.9
13	52.840	10082	5597	4486	0.30	2408.5	2413.3
14	53.639	10378	4956	5423	0.38	1888.3	1880.6
15	54.224	6980	4052	2928	0.22	1329.5	1323.9
16	54.367	3273	3466	– 193	– 0.02	948.1	945.4
17	54.237	1280	3515	– 5534	– 0.21	765.7	767.4
18	54.436	3643	4259	– 617	– 0.07	697.1	698.0
19	54.479	3979	4568	– 588	– 0.12	499.1	499.7
20	54.871	5156	5156	0	0.03	282.9	282.9

For the first simulated realization the expected best-execution cost is \$5,251,395, which compares favorably with the expected cost of \$5,255,185 for the naive strategy,<sup>9</sup> implying an improvement of 3.8 cents/share on average. However, since the information variable  $X_t$  is stochastic, and since

<sup>9</sup> The expected cost of the naive strategy under Eqs. (2.18) and (2.19) is given by

$$E_1 \left[ \sum_{t=1}^T P_t \bar{S}/T \right] = P_0 \bar{S} + \theta \left( \frac{\bar{S}}{T} \right)^2 \frac{T(T+1)}{2} + \frac{\bar{S}}{T} \frac{\gamma X_1}{1-\rho} \left( T - \frac{\rho - \rho^{T+1}}{1-\rho} \right).$$

Table 3

Best-execution strategies for acquiring a 100,000-share block of stock within 20 periods, for a simulated realization of the information variable  $\{X_t\}$  and price shock  $\{\varepsilon_t\}$ , assuming an initial price  $P_0 = \$50$  and linear price-impact function  $\theta S_t + \gamma X_t$ , where  $P_t = P_{t-1} + \theta S_t + \gamma X_t + \varepsilon_t$  and  $\theta = 5 \times 10^{-5}$ ,  $\gamma = 5.0$ ,  $X_t = \rho X_{t-1} + \eta_t$ ,  $X_0 = 0$ ,  $\sigma_\eta^2 = 0.001$ ,  $\sigma_\varepsilon^2 = (0.125)^2$ , and  $\rho = -0.50$ . Note: The non-negativity restriction was not imposed and was not binding in this realization.

Performance summary							
Metric	Naive strategy (\$1000s)		Optimal strategy (\$1000s)			Improvement (cents/share)	
Expected cost	5256.3		5255.6			0.6	
Actual cost	5267.2		5265.0			2.1	
Period	$P_t$	$S_t^*$	$\delta_{x,T-t}W_t$	$\delta_{x,T-t}X_t$	$\gamma X_t$	$V_t/\$1000$	Cost-to-go
1	50.373	5586	5000	586	-0.09	5255.7	5265.0
2	50.467	6341	4969	1372	-0.21	4976.1	4983.6
3	51.081	3429	4893	-1464	0.23	4662.4	4663.6
4	51.140	5077	4979	98	-0.02	4512.0	4488.5
5	51.330	5246	4973	273	-0.04	4234.4	4228.8
6	51.828	3846	4955	-1109	0.17	3970.5	3959.6
7	52.099	4579	5034	-455	0.07	3788.7	3760.3
8	52.279	6033	5069	964	-0.15	3542.8	3521.7
9	52.674	4204	4989	-785	0.12	3231.4	3206.3
10	52.463	7732	5060	2672	-0.43	2999.4	2984.8
11	52.821	3652	4793	-1140	0.18	2583.3	2579.2
12	53.161	4839	4919	-80	0.01	2393.2	2386.3
13	53.332	5708	4929	779	-0.13	2136.5	2129.0
14	53.222	7144	4818	2326	-0.39	1821.8	1824.6
15	53.782	2006	4431	-2425	0.41	1442.7	1444.4
16	53.947	4886	4916	-30	0.01	1340.0	1336.5
17	53.929	6253	4923	1330	-0.24	1071.0	1072.9
18	54.296	3827	4480	-652	0.13	732.0	735.7
19	54.803	3450	4806	-1355	0.27	527.2	527.9
20	54.989	6161	6161	0	0.04	338.8	338.8

the best-execution strategy  $S_t^*$  is a function of  $X_t$ , the actual execution cost of the best-execution strategy will not typically be equal to its expected value. For this realization, the actual cost of the best-execution strategy is \$5,296,615 and the actual cost of the naive strategy is \$5,314,906, yielding an improvement of 18.3 cents/share.

Of course, while the expected best-execution cost is guaranteed to be lower than the expected cost of all other feasible strategies, there is no assurance that the best-execution strategy will *always* yield the lowest execution cost – Table 2 provides another simulated realization for which the *actual* cost of the naive

strategy is lower than that of the best-execution strategy, despite the fact that the best-execution strategy possesses a lower *expected* cost.

Tables 1 and 2 also demonstrate the adaptive nature of the best-execution strategy. For example, Table 1 shows that in periods 7–10, unusually large trades – over 8000 shares in each of these four periods – were executed (recall that the naive strategy executes 5000 shares in each period). The columns labelled ' $\delta_{w,T-t}W_t$ ' and ' $\delta_{x,T-t}X_t$ ' report the breakdown of these trades into two components, one due to the non-information-related motive for trade, and the other due to the impact of the information contained in the state-variable  $X_t$  for the expected price. In periods 7–10, there were unusually large realizations for  $X_t$  implying higher expected prices in the future (since  $X_t$  is positively autocorrelated), and this information component is responsible for the additional shares executed. In period 12, an unusually low realization for  $X_t$  occurs, yielding a negative information component, hence the shares executed in that period is lower than usual. Figs. 1 and 2 provide a clearer illustration of the adaptive nature of the best-execution strategy.

Because information is positively serially correlated, positive realizations of  $X_t$  imply not only higher trading costs now, but higher trading costs for several periods thereafter (on average), hence the best-execution strategy calls for larger-than-usual trades during these high-cost periods, a somewhat counterintuitive but optimal trading pattern. When the information component exhibits persistence, trading 'into' expected price increases can be more economical than the natural tendency to wait for a more favorable price.

Of course, the best-execution strategy depends intimately on the specific dynamics of the information variables that enter the price-impact function. In particular, if in our example  $\rho$  were negative, the opposite patterns would be observed – a large positive realization of  $X_t$  would portend negative realizations of  $X_t$  in the near future (on average), and this would cause trading to be shifted from today into the future. Table 3 and Fig. 3 illustrate this intuition explicitly. In contrast to the best-execution strategies of Tables 1 and 2, here a large positive realization  $X_t$  – in period 3, for example – reduces the trade size for that period, and vice-versa for large negative realizations of  $X_t$ . Instead of trading into price rises, the fact that  $X_t$  exhibits reversals implies that it is optimal to trade out of them.

### 3. Linear-percentage temporary price impact

The closed-form best-execution strategies derived in Section 2 were primarily illustrative examples, meant to introduce the overall approach and provide some intuition for the determinants and properties of best-execution strategies in a simple context. But from a practical perspective, these illustrative examples have several important limitations.



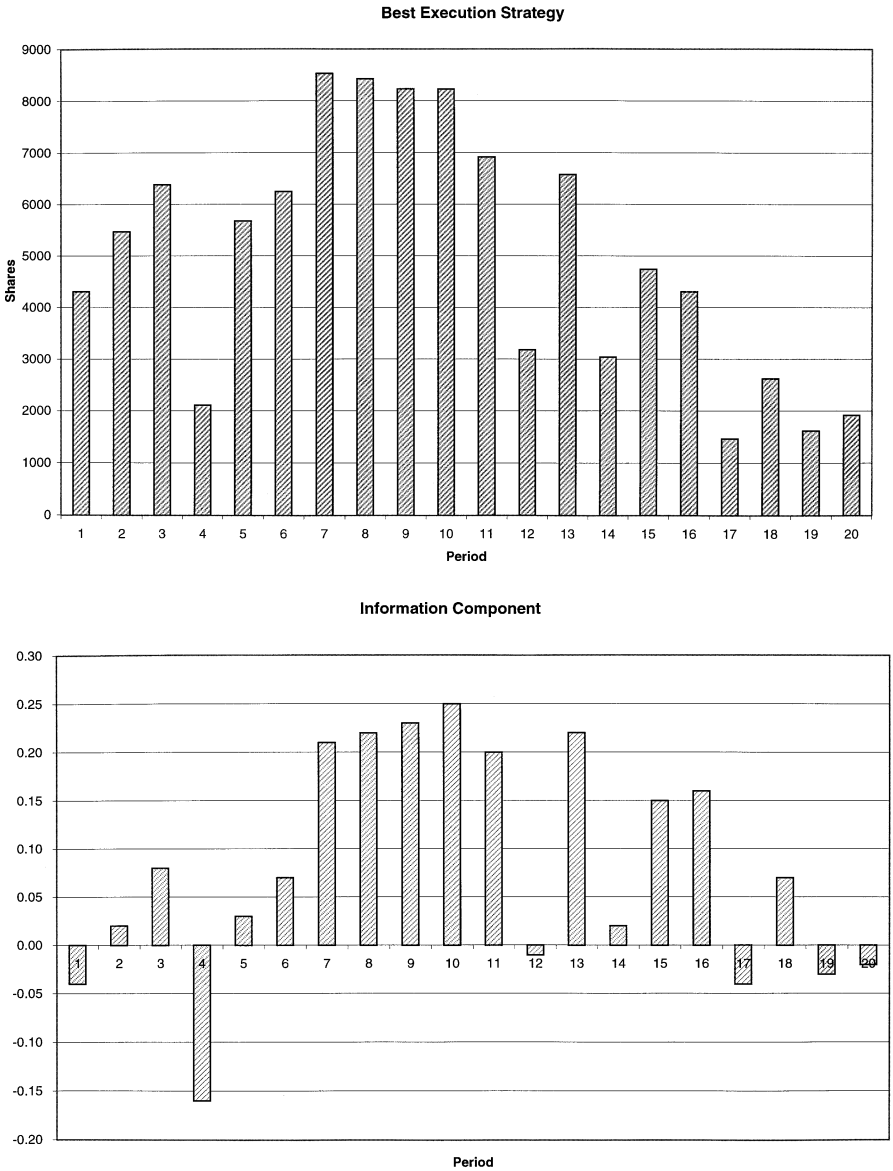


Fig. 1. Best-execution strategy and information-based component for Table 1.

In particular, the fact that, aside from price impact, prices  $P_t$  are assumed to follow arithmetic random walks in Eqs. (2.3) and (2.18) implies a positive probability of negative prices. Also, the random walk specification implies that both price impact and information have *only* permanent effects on prices, which

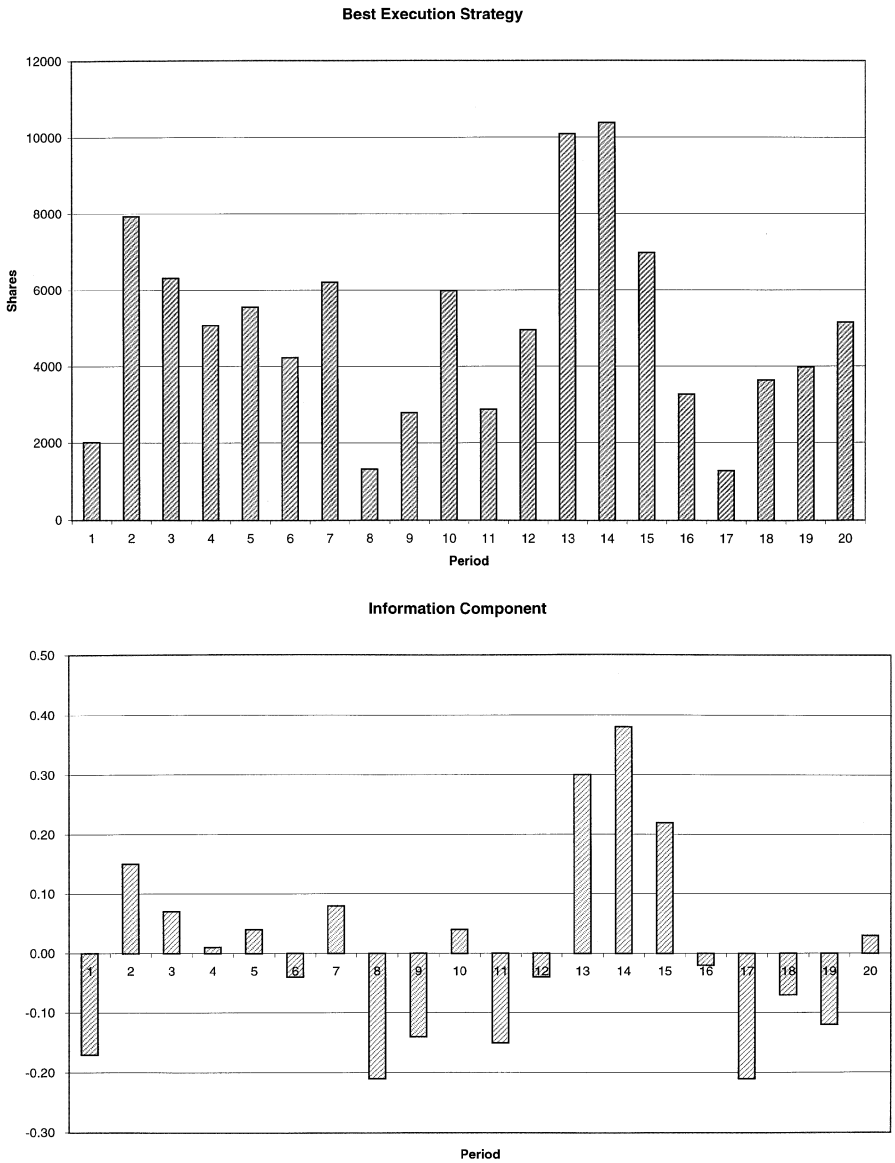


Fig. 2. Best-execution strategy and information-based component for Table 2.

contradicts several recent empirical studies that suggest some combination of permanent and temporary effects (see, for example, Barclay and Litzenberger, 1988; Barclay and Warner, 1993; Chan and Lakonishok, 1993, 1995; Holthausen; Leftwich, and Mayers, 1987, 1990). And finally, because of the linearity

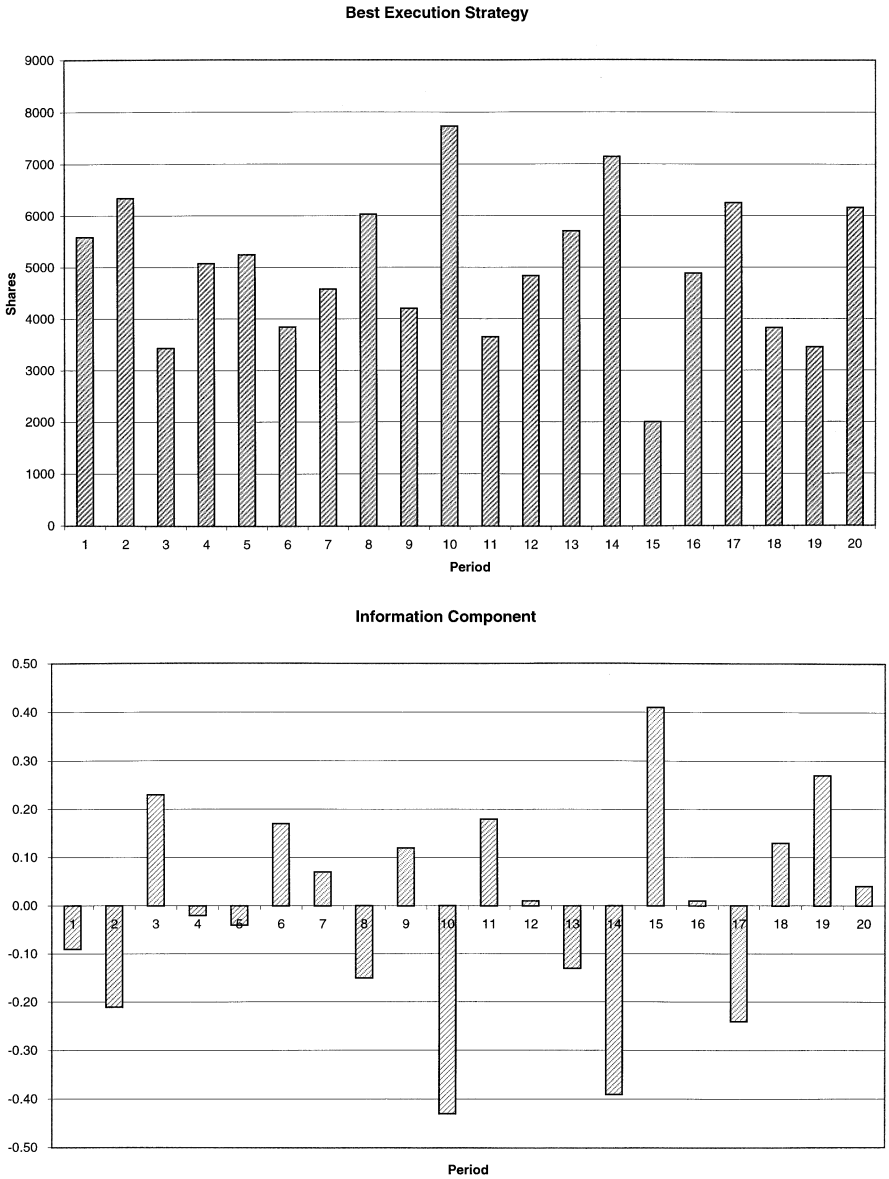


Fig. 3. Best-execution strategy and information-based component for Table 3.

in Eqs. (2.3) and (2.18), the *percentage* price impact – as a percentage of the execution price – is a decreasing function of the price level, which is also counterfactual (see, for example, Birinyi, 1995; Leinweber, 1993, 1994; Loeb, 1983).

### 3.1. LPT law of motion

For these reasons, we propose a more plausible alternative to Eqs. (2.3) and (2.18) which we call the ‘linear-percentage temporary’ (LPT) law of motion. Specifically, let the execution price at time  $t$  be comprised of two components, a no-impact price  $\tilde{P}_t$ , and the price impact  $\Delta_t$ :

$$P_t = \tilde{P}_t + \Delta_t. \quad (3.1)$$

The no-impact price may be viewed as the price which would prevail in the absence of any market impact – a plausible and observable proxy for such a price is the midpoint of the bid/offer spread (although it can be arbitrary as long as it is unaffected by the trade size  $S_t$ ).<sup>10</sup> The fact that  $\tilde{P}_t$  may be observable is important for the empirical implementation of our best-execution strategy, and is explored more fully in Bertsimas et al. (1998). For convenience, and to ensure non-negative prices, we shall adopt a geometric Brownian motion specification for the price dynamics of  $\tilde{P}_t$ :

$$\tilde{P}_t = \tilde{P}_{t-1} \exp(Z_t) \quad (3.2)$$

where  $Z_t$  is an IID normal random variable with mean  $\mu_z$  and variance  $\sigma_z^2$ .

The price impact  $\Delta_t$  captures the effect of trade size  $S_t$  on the transaction price (including the bid/offer spread), and we shall assume that as a percentage of the no-impact price  $\tilde{P}_t$ , it is a linear function of the trade size  $S_t$  and  $X_t$  where, as before,  $X_t$  proxies for private information or market conditions, hence

$$\Delta_t = (\theta S_t + \gamma X_t) \tilde{P}_t, \quad (3.3)$$

$$X_t = \rho X_{t-1} + \eta_t \quad (3.4)$$

where  $\eta_t$  is white noise with mean 0 and variance  $\sigma_\eta^2$ . Once again we have set  $X_t$  to be an AR(1) process which allows us to capture varying degrees of predictability in information or market conditions. The parameters  $\theta$  and  $\gamma$  measure the sensitivity of price impact to trade size and market conditions.

This LPT specification of the law of motion for  $P_t$  resolves a number of problems facing the linear price impact specification of Section 2. First,  $\tilde{P}_t$  is guaranteed to be non-negative, and  $P_t$  is also guaranteed to be non-negative under mild restrictions on  $\Delta_t$ .

---

<sup>10</sup> Indeed, the electronic crossing system POSIT – long recognized as a means of trading equities without market impact – uses the average of the bid and offer prices in crossing its orders.

Second, by separating the transaction price  $P_t$  into a no-impact component  $\tilde{P}_t$  and the impact component  $\Delta_t$ , the price impact of a trade is temporary, moving 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 more plausible than having absolute price impact increasing linearly with trade size (see, for example, Loeb, 1983 (Table 2)).

Fourth, the LPT law of motion implies a natural decomposition of execution costs, decoupling market-microstructure effects from price dynamics, which is closely related to P erold’s (1988) notion of implementation shortfall. To see this, observe that under Eq. (3.1) the optimization problem is given by

$$\text{Min}_{\{S_t\}} E_1 \left[ \sum_{t=1}^T P_t S_t \right] = \text{Min}_{\{S_t\}} E_1 \left[ \sum_{t=1}^T \tilde{P}_t (1 + \theta S_t + \gamma X_t) S_t \right] \tag{3.5}$$

$$= \text{Min}_{\{S_t\}} \left\{ E_1 \left[ \sum_{t=1}^T \tilde{P}_t S_t \right] + E_1 \left[ \sum_{t=1}^T \Delta_t S_t \right] \right\} \tag{3.6}$$

subject to the laws of motion (3.1)–(3.4) and (2.4). Because Eq. (3.1) implies that price impact is a temporary phenomenon, affecting only the current trade price  $P_t$  and not the no-impact price level  $\tilde{P}_t$ , the objective function in Eq. (3.6) 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 P erold (1988) in his definition of implementation shortfall, but now applied to executing  $\bar{S}$ . 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  $\bar{S}$ .

And finally, we shall see in the following section that the LPT specification 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.

### 3.2. The best-execution strategy

As in Section 2, we consider the unconstrained problem here and derive the best-execution strategy via dynamic programming. The constrained case is discussed in Section 6 – dynamic programming cannot easily accommodate constraints because of its recursive nature, and we refer readers to some approximation methods that may be more useful in the presence of constraints. The best-execution problem is given by

$$\text{Min}_{\{S_t\}} E_1 \left[ \sum_{t=1}^T P_t S_t \right] \tag{3.7}$$

subject to the LPT law of motion (3.1)–(3.4) and (2.4).

By solving the usual Bellman equation recursively, we obtain the following best-execution strategy and optimal-value function:

$$S_{T-k}^* = \delta_{x,k} X_{T-k} + \delta_{w,k} W_{T-k} + \delta_{1,k}, \quad (3.8)$$

$$V_{T-k}(\tilde{P}_{T-k-1}, X_{T-k}, W_{T-k}) = q\tilde{P}_{T-k-1}[a_k + b_k X_{T-k} + c_k X_{T-k}^2 + d_k X_{T-k} W_{T-k} + e_k W_{T-k} + f_k W_{T-k}^2], \quad (3.9)$$

where  $q \equiv E[\exp(Z_i)] = \exp(\mu_z + \sigma^2/2)$  and  $\{\delta_{x,k}, \delta_{w,k}, \delta_{1,k}\}$  and  $\{a_k, \dots, f_k\}$  are fixed coefficients given in the Appendix.

These expressions are qualitatively similar to those of the linear price impact case: the best-execution strategy is a linear function of the two state-variables  $X_t$  and  $W_t$ , but now with an added constant term, and the optimal-value function is a quadratic function of  $X_t$  and  $W_t$ , but is now proportional to  $\tilde{P}$ .

### 3.3. A Monte Carlo analysis of best execution

To gauge the practical relevance of the best-execution strategy under the LPT law of motion, we perform several Monte Carlo simulation experiments for various combinations of parameter values. Specifically, we return to the numerical example of Section 2.5 in which the goal is to minimize the expected execution costs of a 100,000-share purchase over  $T = 20$  periods for a stock currently trading at  $P_0 = \$50$ , where the parameters of the LPT law of motion are calibrated as:

$$\mu = 0, \quad \sigma_z = 0.02/\sqrt{13}, \quad \theta = 5 \times 10^{-7}, \quad \eta_t \sim N(0, 1 - \rho^2),$$

$$\gamma = 0.0000, 0.0010, 0.0025, 0.0050, 0.0100,$$

$$\rho = -0.50, -0.25, 0.00, 0.25, 0.50.$$

The first two parameters imply that the continuously compounded return  $\log \tilde{P}_t/\tilde{P}_{t-1}$  has zero mean and a 2% daily standard deviation (recall that each period is meant to be a 30 minute interval and there are 13 such intervals in a typical trading day for the New York Stock Exchange, hence the divisor  $\sqrt{13}$  for the standard deviation). The parameter  $\theta$  is calibrated to yield a percentage price impact of 5% for a 100,000-share trade.

Observe that the variance of the error term  $\eta_t$  for the information variable  $X_t$  is a function of the autocorrelation coefficient  $\rho$  – this specification yields a unit variance for  $X_t$  for any value of  $\rho$ . Such a normalization is always

possible regardless of the natural units of  $X_t$  (simply divide  $X_t$  by its standard deviation), and is a convenient normalization since it implies that the information-sensitivity parameter  $\gamma$  is measured in units of standard deviation of  $X_t$ .

For each combination of parameter values (we considered five different values for  $\gamma$  and five different values for  $\rho$ , yielding 25 combinations), we simulated an independent set of 50,000 sample paths of  $\{\tilde{P}_t\}$  and  $\{X_t\}$ , implemented the best-execution strategy (3.8) on each path, and recorded the realized best-execution cost as well as the realized cost of the naive strategy  $\bar{S}/T$ .

The results of these simulations are summarized in Table 4 which is organized according to the values of  $\gamma$  (rows) and  $\rho$  (columns). For example, the first three rows of the table correspond to simulation results for  $\gamma = 0.0000$  – the first row reports the expected cost (in cents/share above the no-impact cost  $P_0\bar{S}$ ) of either the best execution strategy (the ‘S\*’ column) or the naive strategy (the ‘ $\bar{S}/T$ ’ column), the second row reports the average realized cost of either strategy (averaged over 50,000 replications), and the third row contains in parentheses the standard errors of the average realized cost reported in the second row. The five column-subpanels correspond to the five different values for  $\rho$  and the five row-subpanels correspond to the five different values for  $\gamma$ .

Consider the first row-subpanel and third column-subpanel, which corresponds to the simulation for  $\gamma = 0.0000$  and  $\rho = 0.00$ . In this case, the expected cost of best execution is 13.3058 cents/share above the no-impact cost  $P_0\bar{S}$ , and is close but not identical to the expected cost of the naive strategy, 13.3098 (the ‘Diff.’ column reports the difference of the two). This is not surprising because in this case the information variable  $X_t$  plays no role in determining price impact, hence the best-execution strategy and the naive strategy yield similar costs. However, the costs are not identical because the optimal strategy does account for the drift in the stock price, and while these effects are minimal for the  $\gamma = 0.0000$  case, they are not zero. Observe that some slight differences between the best-execution and naive strategies are observed in the second row’s entries. Recall that these entries are the average realized execution costs for the two strategies, and these averages are subject to sampling variation (even for 50,000 replications). However, they are reasonably close to their theoretical counterparts, and would be even closer for a larger number of replications.

As we move down to the next row-subpanel, where  $\gamma = 0.0010$ , the differences between the best-execution and naive strategies become more pronounced. For example, when  $\rho = -0.50$  the best-execution strategy has an expected cost of 12.8778 cents/share while the naive strategy’s expected cost is still fixed at 13.3098 (the expected cost of the naive strategy does not depend on  $\gamma$  or  $\rho$ ).

When  $\gamma$  increases, implying that information  $X_t$  has a larger effect on price impact  $\Delta_t$ , the best-execution strategy performs even better. For  $\gamma = 0.0050$  and

Table 4

Monte Carlo simulations of execution costs for purchasing  $\bar{S} = 100,000$  shares of stock with an initial price  $P_0 = \$50$  over  $T = 20$  periods – in cents/share above the no-impact cost  $P_0\delta$  – under the optimal strategy  $S^*$  and the naive strategy  $\bar{S}/T$ . The three rows that correspond to each value of  $\gamma$  contain the theoretical values of the best execution costs (first row), the means (each averaged over 50,000 replications) of the Monte Carlo simulations of the actual realized costs (second row), and the standard errors of those means (third row), respectively. A linear-percentage price-impact function  $P_t = \bar{P}_t(1 + \alpha S_t + \gamma X_t)$  is assumed, where  $\bar{P}_t$  is the zero-impact price,  $S_t$  is the number of shares executed, and  $X_t$  is an information variable that measures current market conditions. The data-generating process for  $\bar{P}_t$  and  $X_t$  are given by:  $\bar{P}_t = \bar{P}_{t-1} \exp(Z_t)$ ,  $Z_t \sim N(\mu_z, \sigma_z^2)$ , and  $X_t = \rho X_{t-1} + \eta_t$ ,  $\eta_t \sim N(0, \sigma_\eta^2)$ . The simulation parameters were calibrated to approximate half-hour intervals:  $\alpha = 5 \times 10^{-7}$ ,  $\mu_z = 0$ ,  $\sigma_z^2 = (0.02)^2/13$ ,  $\sigma_\eta^2 = 1 - \rho^2$ ,  $X_1 = 0$ ,  $\bar{P}_0 = 50$

$\gamma$	$\rho = -0.50$			$\rho = -0.25$			$\rho = 0.00$			$\rho = 0.25$			$\rho = 0.50$		
	$S^*$	$\bar{S}/T$	Diff.	$S^*$	$\bar{S}/T$	Diff.	$S^*$	$\bar{S}/T$	Diff.	$S^*$	$\bar{S}/T$	Diff.	$S^*$	$\bar{S}/T$	Diff.
0.0000	13.3058	13.3098	0.0040	13.3058	13.3098	0.0040	13.3058	13.3098	0.0040	13.3058	13.3098	0.0040	13.3058	13.3098	0.0040
	13.3207	13.3260	0.0053	13.2289	13.2329	0.0040	13.2538	13.2575	0.0038	12.9591	12.9584	-0.0007	13.1136	13.1163	0.0027
	(0.3325)	(0.3349)	(0.0032)	(0.3301)	(0.3325)	(0.0031)	(0.3302)	(0.3325)	(0.0031)	(0.3307)	(0.3331)	(0.0031)	(0.3309)	(0.3333)	(0.0031)
0.0010	12.8778	13.3098	0.4320	12.8933	13.3098	0.4165	12.9195	13.3098	0.3903	12.9590	13.3098	0.3508	13.0228	13.3098	0.2870
	12.9780	13.4086	0.4305	13.0345	13.4556	0.4212	12.9273	13.3059	0.3786	12.9476	13.3056	0.3580	13.3972	13.6819	0.2847
	(0.3308)	(0.3331)	(0.0069)	(0.3306)	(0.3329)	(0.0083)	(0.3319)	(0.3342)	(0.0098)	(0.3292)	(0.3314)	(0.0113)	(0.3308)	(0.3330)	(0.0129)
0.0025	10.6307	13.3098	2.6791	10.7276	13.3098	2.5822	10.8911	13.3098	2.4187	11.1381	13.3098	2.1717	11.5370	13.3098	1.7728
	10.9115	13.5869	2.6754	10.8182	13.4205	2.6023	10.6637	13.1275	2.4638	11.4047	13.5947	2.1900	11.1427	12.8976	1.7550
	(0.3315)	(0.3338)	(0.0166)	(0.3309)	(0.3329)	(0.0197)	(0.3311)	(0.3327)	(0.0236)	(0.3330)	(0.3343)	(0.0275)	(0.3326)	(0.3338)	(0.0319)
0.0050	2.6054	13.3098	10.7044	2.9929	13.3098	10.3169	3.6468	13.3098	9.6630	4.6348	13.3098	8.6750	6.2305	13.3098	7.0793
	2.9075	13.6040	10.6965	2.8609	13.2108	10.3499	3.6529	13.3177	9.6647	4.7800	13.5062	8.7262	6.5647	13.6405	7.0758
	(0.3308)	(0.3317)	(0.0370)	(0.3323)	(0.3331)	(0.0421)	(0.3330)	(0.3322)	(0.0488)	(0.3357)	(0.3333)	(0.0567)	(0.3398)	(0.3363)	(0.0650)
0.0100	-29.4961	13.3098	42.8059	-27.9460	13.3098	41.2558	-25.3304	13.3098	38.6402	-21.3783	13.3098	34.6881	-14.9956	13.3098	28.3054
	-29.4341	13.3893	42.8234	-28.2093	13.1683	41.3776	-25.3868	13.2366	38.6233	-21.3424	13.4958	34.8382	-15.2465	13.0046	28.2511
	(0.3454)	(0.3343)	(0.1051)	(0.3453)	(0.3344)	(0.1052)	(0.3512)	(0.3368)	(0.1129)	(0.3578)	(0.3393)	(0.1273)	(0.3679)	(0.3432)	(0.1445)



$\rho = -0.50$ , the best-execution strategy outperforms the naive strategy by 10.7044 cents/share, with an expected execution cost of only 2.6054 cents/share.

When  $\gamma = 0.0100$  the expected execution costs become negative – in this case, there is so much value in exploiting  $X_t$  that the best-execution strategy performs better than the no-impact benchmark  $P_0\bar{S}$ . Although this case is unrealistic, it does highlight the important role that information can play in determining execution costs.

Table 4 shows that under the LPT law of motion, the best-execution strategy yields a significant improvement over the naive strategy, approximately 3–5 cents/share for empirically plausible parameter values, which amounts to a 25–40% reduction in execution costs. Of course, these figures are highly parameter specific and also depend intimately on the particular law of motion we have chosen. Nevertheless these Monte Carlo results seem to suggest that in some cases, best-execution strategies are quite effective in controlling execution costs.

#### 4. The general formulation

In this section we describe the general approach to minimizing expected execution costs. Despite the fact that closed-form solutions may not exist for many cases of empirical interest, in almost all cases it is possible to obtain numerical solutions to well-posed dynamic programming problems. We begin with the same objective function:

$$\text{Min}_{\{S_t\}} E_1 \left[ \sum_{t=1}^T P_t S_t \right]$$

subject to the same constraint:

$$\sum_{t=1}^T S_t = \bar{S}$$

but with a more general law of motion:

$$P_t = f_t(P_{t-1}, \mathbf{X}_t, S_t, \varepsilon_t), \tag{4.1}$$

$$\mathbf{X}_t = g_t(\mathbf{X}_{t-1}, \eta_t), \tag{4.2}$$

$$W_t = W_{t-1} - S_{t-1}, \quad W_1 = \bar{S}, \quad W_{T+1} = 0. \tag{4.3}$$

The price-impact function is now implicitly contained in the law of motion (4.1) for  $P_t$ , which is a general nonlinear and possibly time-varying function  $f_t$  of  $P_{t-1}$ , a state-vector  $\mathbf{X}_t$ , the control  $S_t$ , and a random shock  $\varepsilon_t$ .<sup>11</sup> Since  $\mathbf{X}_t$  is a vector of

---

<sup>11</sup> We follow the convention that all vectors are column vectors unless they are explicitly transposed, and that boldface Roman letters denote vectors and matrices.

arbitrary dimension, it can accommodate multiple factors as in the ordered probit model of Hausman et al. (1992), and the fact that  $g_t(\cdot)$  is a function of only a single lag is not as restrictive as it may seem because additional lags may be captured by increasing the dimension of  $\mathbf{X}_t$ , i.e., expansion of the states. Moreover, since  $g_t(\cdot)$  is not restricted to be linear, it is possible to capture complex dynamic behavior of the state variables parsimoniously, e.g., threshold autoregressions, Markov switching processes, etc.

The Bellman equation and optimal trade at the end of the fixed horizon is given by

$$V_T(P_{T-1}, \mathbf{X}_T, W_T) = \underset{S_T}{\text{Min}} E_T[P_T S_T] = E_T[f_T(P_{T-1}, \mathbf{X}_T, W_T, \varepsilon_T)W_T], \quad (4.4)$$

$$S_T^* = W_T. \quad (4.5)$$

In period  $T - 1$ , the Bellman equation becomes

$$V_{T-1}(P_{T-2}, \mathbf{X}_{T-1}, W_{T-1}) = \underset{S_{T-1}}{\text{Min}} E_{T-1}[P_{T-1} S_{T-1} + V_T(P_{T-1}, \mathbf{X}_T, W_T)] \quad (4.6)$$

$$\begin{aligned} &= \underset{S_{T-1}}{\text{Min}} E_{T-1}[f_{T-1}(P_{T-2}, \mathbf{X}_{T-1}, S_{T-1}, \varepsilon_{T-1})S_{T-1} \\ &\quad + V_T(f_{T-1}(\cdot), g_T(\cdot), W_{T-1} - S_{T-1})]. \end{aligned} \quad (4.7)$$

Performing the one-period minimization in Eq. (4.7) – subject to any additional constraints that might be present – yields the optimal trade  $S_{T-1}^*$  as a function of the state variables  $P_{T-2}$ ,  $\mathbf{X}_{T-1}$ , and  $W_{T-1}$ . Hence

$$S_{T-1}^* = h_{T-1}(P_{T-2}, \mathbf{X}_{T-1}, W_{T-1}). \quad (4.8)$$

Proceeding recursively in this fashion, the optimal-value function and optimal trade at time  $T - k$  is then

$$\begin{aligned} V_{T-k}(P_{T-k-1}, \mathbf{X}_{T-k}, W_{T-k}) &= \underset{S_{T-k}}{\text{Min}} E_{T-k}[P_{T-k} S_{T-k} \\ &\quad + V_{T-k+1}(P_{T-k}, \mathbf{X}_{T-k+1}, W_{T-k+1})] \end{aligned} \quad (4.9)$$

$$\begin{aligned} &= \underset{S_{T-k}}{\text{Min}} E_{T-k}[f_{T-k}(P_{T-k-1}, \mathbf{X}_{T-k}, S_{T-k}, \varepsilon_{T-k})S_{T-k} \\ &\quad + V_{T-k+1}(f_{T-k}(\cdot), g_{T-k+1}(\cdot), W_{T-k} - S_{T-k})], \end{aligned} \quad (4.10)$$

$$S_{T-k}^* = h_{T-k}(P_{T-k-1}, \mathbf{X}_{T-k}, W_{T-k}) \quad (4.11)$$

and when we reach the starting point of the trading horizon:

$$V_1(P_0, \mathbf{X}_1, W_1) = \text{Min}_{S_1} E_1[P_1 S_1 + V_2(P_1, \mathbf{X}_2, W_2)] \tag{4.12}$$

$$= \text{Min}_{S_1} E_1[f_1(P_0, \mathbf{X}_1, S_1, \varepsilon_1)S_1 + V_2(f_1(\cdot), g_2(\cdot), W_1 - S_1)] \tag{4.13}$$

$$S_1^* = h_1(P_0, \mathbf{X}_1, W_1), \quad W_1 = \bar{S} \tag{4.14}$$

the initial conditions will enable us to obtain the entire sequence of optimal trades as functions of the state variables, which yields the best-execution strategy:

$$\begin{aligned} S_1^* &= h_1(P_0, \mathbf{X}_1, \bar{S}), \\ S_2^* &= h_2(P_1, \mathbf{X}_2, \bar{S} - S_1^*), \\ &\vdots \\ S_k^* &= h_k\left(P_{k-1}, \mathbf{X}_k, \bar{S} - \sum_{t=1}^{k-1} S_t^*\right), \\ &\vdots \\ S_{T-1}^* &= h_{T-1}\left(P_{T-2}, \mathbf{X}_{T-1}, \bar{S} - \sum_{t=1}^{T-1} S_t^*\right), \\ S_T^* &= \bar{S} - \sum_{t=1}^{T-1} S_t^*. \end{aligned} \tag{4.15}$$

Moreover, the optimal-value function at the start of the trading horizon,  $V_1(P_0, \mathbf{X}_1, \bar{S})$ , is an estimate of the best-execution costs, and depends only on the initial price, the initial state-vector, and the total number of shares to be traded.

Of course, for certain specifications of the law of motion, computing the optimal control explicitly may be intractable because a closed-form expression for the optimal-value function  $V_{T-k+1}(\dots)$  is not available (see Eq. (4.10)). Therefore, in the next two sections we propose alternatives that address the computational challenges of determining best execution.

#### 4.1. A discretization approach

Perhaps the most common method for solving dynamic programming problems numerically is to discretize the state space and control space and perform the dynamic optimization via grid search at each stage.

In particular, we discretize the possible prices  $P_t$  as a multiple of some constant,  $d$ . A natural choice for  $d$  is  $\frac{1}{8}$  since the minimum variation of prices on most US stocks is \$0.125. Suppose that the horizon  $T$  is such that the range of

possible prices is concentrated on a relatively small interval and let  $K$  be the number of possible values in that interval. Let the trade sizes  $S_t$  also be discrete, varying in fixed increments of  $s$  shares – an obvious choice for  $s$  is 100 shares since most institutional trades are round lots. This implies that  $W_t$ , the remaining shares to be executed at time  $t$ , is also denominated in multiples of  $s$ . Let  $J = \bar{S}/s$  denote the number of round lots that need to be executed initially. Finally, let the information vector  $\mathbf{X}_t$  take on a finite number  $N$  of possible values.

Under these assumptions, at each time  $t$  the optimal-value function  $V_t(P_{t-1}, \mathbf{X}_t, W_t)$  must be evaluated numerically for  $KJN$  possible values. Note that these values need to be stored only temporarily, as they will be used in step  $t - 1$  of the dynamic programming algorithm, but not at step  $t - 2$  (at this step the values  $V_{t-1}(P_{t-2}, \mathbf{X}_{t-1}, W_{t-1})$  are needed). As a result, the total memory requirements are of the order  $O(KJN)$ .

In order to develop some intuition for the computational requirements of the discretization, consider the following example. Suppose that 100,000 shares, currently trading at \$50, must be executed over the next 20 periods. We assume that within this time period, the price will be within \$45 and \$55, meaning that there are  $K = 80$  possible prices to consider (multiples of  $\frac{1}{8}$ ). If we choose  $s = 100$  then  $J = 1000$ . We also discretize the information variable into  $N = 10$  states. Then at each step we need to compute the optimal-value function for  $KJN = 800,000$  values of the state and control variables. With 20 periods, this implies a total of 16 million evaluations of the optimal-value function.

Given the speed of today's workstations, such computations can be performed in seconds. For example, if we assume that the computation of each optimal-value function  $V_t(P_{t-1}, X_t, W_t)$  takes  $10^{-6}$  seconds, then the total computation will be 16 seconds.

Of course, if we require a more refined discretization, the computational requirements increase dramatically. For example, if we allow trade size to vary in increments of 1 share instead of 100 shares, and if the information variable takes on 100 discrete values instead of 10, the computational demands increase by a factor of 1000.

#### 4.2. Approximate dynamic programming

For certain applications – stocks with high volatility, longer horizons, or a large number of information variables – the discretization approach may be computationally infeasible. In such cases, we propose a recently developed technique known as ‘approximate dynamic programming’ in which the optimal-value function is approximated at each stage by a quadratic function (see, for example, Bertsekas (1995)). In contrast to the discretization method which is numerical (but exact), approximate dynamic programming always yields an *analytical* (but approximate) solution.

Specifically, let  $\mathbf{Y}_t = (P_{t-1}, \mathbf{X}_t, W_t)$  denote the state vector at time  $t$ . Beginning at the end, we must first compute  $V_T(\mathbf{Y}_T)$  (note that this requires no minimization). We approximate this function with  $\hat{V}_T(\mathbf{Y}_T)$  where

$$\hat{V}_T(\mathbf{Y}_T) \equiv \mathbf{Y}'_T \mathbf{Q}_T \mathbf{Y}_T + \mathbf{b}'_T \mathbf{Y}_T \tag{4.16}$$

and the matrix  $\mathbf{Q}_T$  and vector  $\mathbf{b}_T$  are selected to minimize:

$$\int_{\mathbf{y}_T} (V_T(\mathbf{Y}_T) - \hat{V}_T(\mathbf{Y}_T))^2 d\mathbf{Y}_T. \tag{4.17}$$

This is a least-squares problem which can always be solved in closed form. At step  $T - k$  we apply the Bellman recursion but with one important difference: we use  $\hat{V}_{T-k+1}(\mathbf{Y}_{T-k+1})$  in place of  $V_{T-k+1}(\mathbf{Y}_{T-k+1})$  for the minimization:

$$\begin{aligned} V_{T-k}(\mathbf{Y}_{T-k}) = \text{Min}_{S_{T-k}} E_{T-k} [ & f_{T-k}(P_{T-k-1}, \mathbf{X}_{T-k}, S_{T-k}, \varepsilon_{T-k}) S_{T-k} \\ & + \hat{V}_{T-k+1}(f_{T-k}(\cdot), g_{T-k+1}(\cdot), W_{T-k} - S_{T-k}) ]. \end{aligned} \tag{4.18}$$

This minimization can be easily performed because the approximating optimal-value function is quadratic. After computing  $V_{T-k}(\mathbf{Y}_{T-k})$ , again we approximate this function with a quadratic function:

$$\hat{V}_{T-k}(\mathbf{Y}_{T-k}) = \mathbf{Y}'_{T-k} \mathbf{Q}_{T-k} \mathbf{Y}_{T-k} + \mathbf{b}'_{T-k} \mathbf{Y}_{T-k} \tag{4.19}$$

where, as before, the matrix  $\mathbf{Q}_{T-k}$  and the vector  $\mathbf{b}_{T-k}$  are selected to minimize:

$$\int_{\mathbf{y}_{T-k}} (V_{T-k}(\mathbf{Y}_{T-k}) - \hat{V}_{T-k}(\mathbf{Y}_{T-k}))^2 d\mathbf{Y}_{T-k}. \tag{4.20}$$

There are several compelling reasons for using a quadratic approximation to  $V_{T-k}(\mathbf{Y}_{T-k})$ . First, in all of cases we have considered in Sections 2 and 3, the optimal-value functions are quadratic, hence in these cases approximate dynamic programming is exact. Second, a quadratic approximation can capture a variety of nonlinearities parsimoniously. For example, if the vector of information variables contains three scalars – quite a challenge for the discretization approach – this implies that  $\hat{V}_{T-k}(\mathbf{Y}_{T-k})$  is uniquely determined by the  $(5 \times 5)$ -matrix  $\mathbf{Q}_t$  and the  $(5 \times 1)$ -vector  $\mathbf{b}_t$ . Finally, the minimization that must be performed at each stage of the dynamic program is considerably more tractable when the optimal-value function is quadratic.

Of course, an unresolved issue is the magnitude of the approximation error. Because approximate dynamic programming is still in its infancy, very little is known about the magnitude of the approximation error. However, some preliminary studies seem to indicate that the approximations are quite good – see

Bertsekas (1995, Chapter 6) and the accompanying references for further discussion.

## 5. Best execution for portfolios

In the previous sections we have focused our attention on the best-execution strategies for a single stock, but in typical applications institutional investors trade many stocks at once, i.e., ‘list’ trading. In this section we extend our approach to the multivariate setting in which a portfolio of  $n$  stocks must be executed within  $T$  periods.

The portfolio case contains several interesting features not captured by our single-stock analysis. Perhaps the most important feature is the ability to capture cross-stock relations such as the cross-autocorrelations reported in Lo and MacKinlay (1990). In particular, one might expect price movements in one stock to 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, 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.

Fortunately, much of the analysis in Sections 2 and 3 – including closed-form solutions in the linear price impact and LPT law of motion cases – extends to the portfolio case with only minor changes in notation. Of course, for more complex laws of motion, especially those which require numerical solutions, the portfolio problem can quickly become computationally intractable. However, in such cases the approximate dynamic programming approach we discuss in Section 4.2 would be a suitable alternative – see Bertsimas et al. (1998) for a more detailed analysis.

Let  $\bar{\mathbf{S}} \equiv [\bar{S}_1 \dots \bar{S}_n]'$  denote the vector of stocks to be purchased or sold within  $T$  periods, where a negative entry indicates a sale and a positive entry indicates a purchase. Denote by  $\mathbf{P}_t$ ,  $\mathbf{S}_t$ , and  $\mathbf{W}_t$  the  $n$ -vector of prices, shares executed, and remaining shares to be executed, respectively, at time  $t$ . Then the best-execution problem for the portfolio  $\bar{\mathbf{S}}$  may be expressed compactly as

$$\text{Min}_{\{\mathbf{S}_t\}} E_1 \left[ \sum_{t=1}^T \mathbf{P}'_t \mathbf{S}_t \right] \quad (5.1)$$

Subject to

$$\sum_{t=1}^T \mathbf{S}_t = \bar{\mathbf{S}}, \tag{5.2}$$

$$\mathbf{W}_t = \mathbf{W}_{t-1} - \mathbf{S}_t \tag{5.3}$$

and the law of motion for  $\mathbf{P}_t$  and accompanying information variables  $\mathbf{X}_t$ . In the next two sections, we present analytical solutions to Eq. (5.1) which illustrate the complexities (and opportunities) that the portfolio case has to offer.

*5.1. The multivariate linear price impact case*

Suppose the price vector  $\mathbf{P}_t$  follows the multivariate version of Eq. (2.18)

$$\mathbf{P}_t = \mathbf{P}_{t-1} + \mathbf{A}\mathbf{S}_t + \mathbf{B}\mathbf{X}_t + \boldsymbol{\varepsilon}_t \tag{5.4}$$

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

$$\mathbf{X}_t = \mathbf{C}\mathbf{X}_{t-1} + \boldsymbol{\eta}_t \tag{5.5}$$

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  $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ , and which is independent of  $\boldsymbol{\varepsilon}_t$ .

Under the law of motion (5.4)–(5.5), the portfolio problem (5.1)–(5.3) can be solved via Bellman’s equation, which yields the following best-execution strategy and optimal-value function:

$$\mathbf{S}_{T-k}^* = (\mathbf{I} - \frac{1}{2} \mathbf{A}_k^{-1} \mathbf{A}') \mathbf{W}_{T-k} + \frac{1}{2} \mathbf{A}_k^{-1} \mathbf{B}'_{k-1} \mathbf{C} \mathbf{X}_{T-k}, \tag{5.6}$$

$$\begin{aligned} V_{T-k}(\mathbf{P}_{T-k-1}, \mathbf{X}_{T-k}, \mathbf{W}_{T-k}) &= \mathbf{P}'_{T-k-1} \mathbf{W}_{T-k} + \mathbf{W}'_{T-k} \mathbf{A}_k \mathbf{W}_{T-k} \\ &\quad + \mathbf{X}'_{T-k} \mathbf{B}_k \mathbf{W}_{T-k} + \mathbf{X}'_{T-k} \mathbf{C}_k \mathbf{X}_{T-k} + d_k \end{aligned} \tag{5.7}$$

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

$$\begin{aligned} \mathbf{A}_k &= \mathbf{A} - \frac{1}{4} \mathbf{A} \mathbf{A}_{k-1}^{-1} \mathbf{A}', \quad \mathbf{A}_0 = \mathbf{A}, \\ \mathbf{B}_k &= \frac{1}{2} \mathbf{C}' \mathbf{B}_{k-1} (\mathbf{A}'_{k-1})^{-1} \mathbf{A}' + \mathbf{B}', \quad \mathbf{B}_0 = \mathbf{B}', \\ \mathbf{C}_k &= \mathbf{C}' \mathbf{C}_{k-1} \mathbf{C} - \frac{1}{4} \mathbf{C}' \mathbf{B}_{k-1} (\mathbf{A}'_{k-1})^{-1} \mathbf{B}_{k-1} \mathbf{C}, \quad \mathbf{C}_0 = \mathbf{0}, \\ d_k &= d_{k-1} + E[\boldsymbol{\eta}'_{T-k} \mathbf{C}_{k-1} \boldsymbol{\eta}_{T-k}], \quad d_0 = 0. \end{aligned} \tag{5.8}$$

Observe that Eq. (5.6) is qualitatively similar to the single-stock best-execution strategy – it is linear in the two state variables  $\mathbf{W}_{T-k}$  and  $\mathbf{X}_{T-k}$ . However, 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  $(\mathbf{I} - \frac{1}{2}\mathbf{A}_k^{-1}\mathbf{A}')$  multiplying  $\mathbf{W}_{T-k}$  in Eq. (5.6) 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 Eq. (5.4)), 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 is examined in Bertsimas et al. (1998).

### 5.2. The multivariate LPT case

The multivariate version of the linear-percentage temporary price impact case is given by the following law of motion:

$$\mathbf{P}_t = \tilde{\mathbf{P}}_t + \Delta_t, \quad (5.9)$$

$$\tilde{\mathbf{P}}_t = \exp(\mathbf{Z}_t)\tilde{\mathbf{P}}_{t-1}, \quad \text{vec}(\mathbf{Z}_t) \sim \mathbf{N}(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z), \quad (5.10)$$

$$\Delta_t = \text{diag}[\tilde{\mathbf{P}}_t](\mathbf{A}\mathbf{S}_t + \mathbf{B}\mathbf{X}_t), \quad (5.11)$$

$$\mathbf{X}_t = \mathbf{C}\mathbf{X}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathbf{WN}(0_z, \boldsymbol{\Sigma}_\eta), \quad (5.12)$$

where  $\boldsymbol{\eta}_t$  is vector white noise with mean  $\mathbf{0}$  and covariance matrix  $\boldsymbol{\Sigma}_\eta$ , the  $\text{vec}(\cdot)$  is the vectorization operator (which maps its matrix argument into a column vector composed of stacked columns of the matrix), and  $\text{diag}(\cdot)$  is the diagonalization operator (which maps its vector argument into a diagonal matrix with the vector as the diagonal). The  $(n \times n)$ -matrix  $\mathbf{A}$  is assumed to be positive definite,  $\mathbf{B}$  is an arbitrary  $(n \times m)$ -matrix, and  $\mathbf{C}$  is an  $(m \times m)$ -matrix with eigenvalues less than unity in modulus.

The specification (5.9) has the same motivation as the single-stock case (3.1) in which price impact is assumed to have only a temporary effect on market prices. The form of the price impact function  $\Delta_t$  in Eq. (5.11) is also similar to that of the single-stock case (3.3), except now the percentage price impact function for each stock  $i$ ,  $\Delta_{it}/P_{it}$  is a linear function  $\mathbf{A}\mathbf{S}_t$  of the trade sizes of *all*  $n$  stocks, not just of  $S_{it}$ . As in Section 5.1, if  $\mathbf{A}$  is diagonal the portfolio problem reduces to  $n$  independent single-stock problems.

Under the law of motion (5.9)–(5.12), the portfolio problem (5.1)–(5.3) can be solved via Bellman's equation, which yields the following best-execution strategy:

$$\mathbf{S}_{T-k}^* = \mathbf{L}_k \mathbf{W}_{T-k} + \mathbf{G}_k \mathbf{X}_{T-k} + c_k \quad (5.13)$$



where the fixed parameters  $\mathbf{L}_k$ ,  $\mathbf{G}_k$  and  $\mathbf{c}_k$  are calculated recursively.<sup>12</sup> As in the single-stock case (3.8), the best-execution strategy is linear in the two state variables  $\mathbf{W}_{T-k}$  and  $\mathbf{X}_{T-k}$  and has a constant term  $\mathbf{c}_k$  as well.

## 6. Imposing constraints

In 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.<sup>13</sup> 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. For example, Table 5 reports the frequency and magnitude of sells generated by the best-execution strategy for the buy-program in Section 3.3 under the LPT law of motion. The last row shows that when  $\gamma = 0.0100$ , i.e., when information has an enormous effect on price impact – over 25% of the best-execution trades are sells, and the average total sell size is between 25% and 30% of the total number of shares  $\bar{S}$  to be executed.

Of course, if we seek numerical solutions to the best-execution problem as discussed in Section 4.1, non-negativity constraints can be imposed almost trivially. Moreover, approximations to dynamic optimization such as the static optimization approach of Bertsimas and Lo (1998) may accommodate constraints more readily than a recursive algorithm like dynamic programming. Section 6.1 contains a more detailed discussion of the difficulties of imposing non-negativity constraints in our context. In Section 6.2 we present a rare example in which a closed-form best-execution strategy is available even when non-negativity constraints are imposed. Although certainly not generic, this example provides some insight into the nature of such constraints and their effect on best execution.

---

<sup>12</sup> We omit these formulae for the sake of brevity – they offer no particular insights or intuition and would lengthen this paper by several pages. Interested readers may contact the authors for the formulae.

<sup>13</sup> Other 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.

Table 5

Frequency and magnitude of sells generated by the best-execution strategy  $\{S_t^*\}$  in Monte Carlo simulations of execution costs for purchasing  $\bar{S} = 100,000$  shares of stock with an initial price  $P_0 = \$50$  over  $T = 20$ . Entries in the 'Prob. (%)' column are averages, across 50,000 replications, of the percentage of  $T$  trades that are sells, hence these are estimates of the probability that any particular trade will be a sell. Entries in the 'Size (%)' column are averages, across 50,000 replications, of the total sell position of each replication as a percentage of  $\bar{S}$ . Standard errors are given in parentheses. A linear-percentage price-impact function  $P_t = \bar{P}_t(1 + \alpha S_t + \gamma X_t)$  is assumed, where  $P_t$  is the zero-impact price,  $S_t$  is the number of shares executed, and  $X_t$  is an information variable that measures current market conditions. The data-generating process for  $\bar{P}_t$  and  $X_t$  are given by:  $\bar{P}_t = \bar{P}_{t-1} \exp(Z_t)$ ,  $Z_t \sim N(\mu_z, \sigma_z^2)$ , and  $X_t = \rho X_{t-1} + \eta_t$ ,  $\eta_t \sim N(0, \sigma_\eta^2)$ . The simulation parameters were calibrated to approximate half-hour intervals:  $\alpha = 5 \times 10^{-7}$ ,  $\mu_z = 0$ ,  $\sigma_z^2 = (0.02)^2/13$ ,  $\sigma_\eta^2 = 1 - \rho^2$ ,  $X_1 = 0$ ,  $\bar{P}_0 = 50$

$\gamma$	$\rho = -0.50$		$\rho = -0.25$		$\rho = 0.00$		$\rho = 0.25$		$\rho = 0.50$	
	Prob. (%)	Size (%)	Prob. (%)	Size (%)	Prob. (%)	Size (%)	Prob. (%)	Size (%)	Prob. (%)	Size (%)
0.0000	0.00 (0.00)	—	0.00 (0.00)	—	0.00 (0.00)	—	0.00 (0.00)	—	0.00 (0.00)	—
0.0010	0.00 (0.00)	—	0.00 (0.00)	—	0.00 (0.00)	—	0.00 (0.00)	—	0.00 (0.00)	—
0.0025	1.71 (0.01)	1.04 (0.01)	1.55 (0.01)	0.94 (0.01)	1.29 (0.01)	0.88 (0.01)	0.95 (0.01)	0.80 (0.01)	0.48 (0.01)	0.71 (0.01)
0.0050	13.81 (0.03)	7.00 (0.02)	13.43 (0.03)	6.59 (0.02)	12.60 (0.03)	5.92 (0.02)	11.45 (0.03)	5.11 (0.02)	9.21 (0.03)	3.91 (0.02)
0.0100	28.38 (0.03)	34.21 (0.06)	28.09 (0.03)	33.09 (0.05)	27.48 (0.03)	31.06 (0.05)	26.42 (0.03)	28.05 (0.05)	24.53 (0.03)	22.91 (0.05)

### 6.1. Why imposing constraints is difficult

Although there are well-known techniques for performing constrained optimization 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 price-impact model of Section 2.1. 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 (see Eq. (2.21)). 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 solution is given by

$$S_{T-1}^* = \begin{cases} 0 & \text{if } a_1 W_{T-1} + b_1 X_{T-1} < 0, \\ a_1 W_{T-1} + b_1 X_{T-1} & \text{if } 0 < a_1 W_{T-1} + b_1 X_{T-1} < W_{T-1}, \\ W_{T-1} & \text{if } a_1 W_{T-1} + b_1 X_{T-1} > W_{T-1}. \end{cases} \tag{6.1}$$

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).

### 6.2. A closed-form solution with non-negativity constraints

To develop further intuition for how non-negativity constraints affect best-execution strategies, in this section we present a specification of the law of motion under which a closed-form best-execution strategy is available under non-negativity constraints.

Consider the single-stock case of Section 2 in which some information  $X_t$  affects price impact. However, in contrast to the linear price impact specification of Section 2.4, suppose that  $X_t$  affects price impact in a multiplicative fashion, so that the price impact of a trade of  $S_t$  shares is  $\theta X_t S_t$ .

Instead of increasing or decreasing the price level,  $X_t$  now affects the price elasticity. Let the price and information dynamics be given by the following law of motion:

$$P_t = P_{t-1} + \theta X_t S_t + \varepsilon_t \quad (6.2)$$

$$\log X_t = \log X_{t-1} + \eta_t \quad (6.3)$$

where  $\varepsilon_t$  and  $\eta_t$  are independent white noise processes with mean 0 and variances  $\sigma_\varepsilon^2$  and  $\sigma_\eta^2$ , respectively. The price dynamics (6.2) imply that price impact is permanent, and the information dynamics (6.3) is an AR(1) in the logarithm of  $X_t$  to ensure that  $X_t$  is always positive so that the price elasticity  $\theta X_t$  cannot change sign (if it does, the optimization problem is no longer well-defined because the objective function may not be convex).

Under the law of motion (6.2)–(6.3) and (2.4), and with non-negativity constraints  $S_t \geq 0$ , the investor's minimization problem is well-posed. Once again, we apply Bellman's equation recursively to characterize the best-execution strategy, this time performing a *constrained* minimization at each stage:

$$\begin{aligned} V_{T-k}(P_{T-k-1}, X_{T-k}, W_{T-k}) = & \text{Min}_{0 \leq S_{T-k} \leq W_{T-k}} E_{T-k}[P_{T-k} S_{T-k} \\ & + V_{T-k+1}(P_{T-k}, X_{T-k+1}, W_{T-k+1})]. \end{aligned} \quad (6.4)$$

It can be shown that the best-execution strategy and optimal-value function is given by

$$S_{T-k}^* = a_k W_{T-k}, \quad (6.5)$$

$$V_{T-k}(P_{T-k-1}, X_{T-k}, W_{T-k}) = P_{T-k-1} W_{T-k} + \theta b_k X_{T-k} W_{T-k}^2 \quad (6.6)$$

for  $k = T - 1, \dots, 1$  where the constants  $(a_k, b_k)$  are computed recursively as follows:

$$(a_k, b_k) = \begin{cases} \left(1 - \frac{1}{2\kappa b_{k-1}}, 1 - \frac{1}{4\kappa b_{k-1}}\right) & \text{if } b_{k-1} \geq \frac{1}{2\kappa} \\ (0, \kappa b_{k-1}) & \text{if } b_{k-1} < \frac{1}{2\kappa} \end{cases} \quad (6.7)$$

for  $k = 1, \dots, T - 1$ , where  $\kappa \equiv E[\exp(\eta_t)] = E[X_t/X_{t-1}]$  is the gross expected growth rate of  $X_t$ , and  $(a_0, b_0) = (1, 1)$ . By forward substitution using the law of motion, explicit expressions for the optimal controls  $S_1^*, \dots, S_T^*$  may be

obtained as

$$\begin{aligned}
 S_1^* &= a_{T-1} \bar{S}, \\
 S_2^* &= a_{T-2} (1 - a_{T-1}) \bar{S}, \\
 S_3^* &= a_{T-3} (1 - a_{T-2} - a_{T-1}) \bar{S}, \\
 &\vdots \\
 S_k^* &= a_{T-k} (1 - a_{T-k-1} - \dots - a_{T-1}) \bar{S}, \\
 &\vdots \\
 S_{T-1}^* &= a_1 (1 - a_2 - \dots - a_{T-1}) \bar{S}, \\
 S_T^* &= \bar{S} - \sum_{t=1}^{T-1} S_t^*. \tag{6.8}
 \end{aligned}$$

From Eqs. (6.5) and (6.7), a clear structure for the best-execution strategy emerges. If  $\kappa$  lies in the interval  $(0, \frac{1}{2}]$ , so that the expected growth rate of the price elasticity  $\theta X_t$  is between  $-100\%$  and  $-50\%$ , then it follows from Eq. (6.7) that  $a_0 = 1$  and  $a_1 = a_2 = \dots = a_{T-1} = 0$ . This implies that the best-execution strategy is to trade nothing in the first  $T - 1$  periods and trade everything in the last period. With such a negative expected growth rate for the price elasticity, it pays to wait until the very end before trading.

If  $\kappa$  lies in the interval  $(\frac{1}{2}, \frac{3}{4}]$ , implying an expected growth rate between  $-50\%$  and  $-25\%$ , then  $a_0 = 1$ ,  $a_1 = 1 - 1/2\kappa$ , and  $a_2 = \dots = a_{T-1} = 0$ . In this case, the best-execution strategy is to trade nothing in the first  $T - 2$  periods, trade a fraction  $1 - 1/2\kappa$  of  $\bar{S}$  in the second-to-last period  $T - 1$ , and all of the remaining shares in period  $T$ .

As  $\kappa$  increases, increasing the opportunity cost of delayed trades, the best-execution strategy begins its trading sooner and sooner. When  $\kappa = 1$ , implying a zero expected growth rate for the price elasticity, Eqs. (6.5) and (6.7) show that the best-execution growth rate for the price elasticity, Eqs. (6.5) and (6.7) show that the best-execution strategy reduces to that of the linear price impact model with no information (see Section 2.3):  $S_1^* = \dots = S_T^* = \bar{S}/T$ . When the expected growth of  $X_t$  is 0, the law of motion for the multiplicative case has virtually identical implications for expected execution costs as the linear price impact case, hence the best-execution strategy is the same.

When  $\kappa$  exceeds unity, the best-execution strategy tilts trades towards the earlier periods for obvious reasons, and as  $\kappa$  increases without bound,  $a_{T-1}$  approaches unity from below, and in the limit the best-execution strategy is to trade the entire block  $\bar{S}$  in the first period.

### 7. Limitations, extensions, and open questions

We have argued above that our general approach to defining and implementing best execution is quite general indeed, nevertheless, there are several

important limitations, extensions, open research questions that bear further discussion.

### 7.1. Order types and optimal order-submission strategies

In modelling the price impact of a single trade as a simple function of the number of shares traded, we have subsumed the choice of order type – limit versus market order – the investor will use to execute the trade. This choice is no simple matter, and requires a certain sophistication as well. Market orders are executed immediately but incur substantial price impact. Limit orders incur no price impact but may not be executed immediately, if at all. Therefore, the trade-off between limit and market orders is a subtle one that generates yet another dynamic optimization problem, one that is clearly beyond the scope of this paper but which has been the subject of several recent studies (see, for example, Angel, 1994; Harris and Hasbrouck, 1992; Harris, 1994; Kumar and Seppi, 1993; and Lo et al. (1998)). In particular, this optimization problem requires an explicit measure of the investor's need for immediacy or 'urgency' which can then be balanced against the price impact of a market order.

Of course, a mathematically more complete specification of the best-execution optimization problem would include order type as a control variable and urgency in the objective function. However, without some sort of simplification or approximation, this problem is computationally intractable. The best-execution strategies proposed in this paper can be viewed as a second-best solution: first optimize the number of shares to be traded within each 30 minute interval, and then perform a second optimization within this 30 minute interval to decide the proportion of market and limit orders to use. Although such a two-stage optimization will almost certainly not coincide with the globally optimal execution strategy, it may be an adequate approximation and will be the subject of future investigations.

### 7.2. The risks of best execution

As Table 2 clearly illustrates, the best-execution strategy minimizes *expected* execution costs, not *actual* execution costs. Of course, without perfect foresight, minimizing actual execution costs is impossible. But in the same way that investors evaluate the risks surrounding a given expected return, we can also evaluate the uncertainty around expected best-execution costs.

There are at least four sources of uncertainty surrounding the expected best-execution costs  $V_1$ : (1) the realized cost is typically different from the expected cost, and although over time and trades the average of the former should equal the latter, this may be small comfort if the 'tracking errors' are wild and unpredictable; (2) the expected cost  $V_1$  is itself a function of random initial conditions ( $X_1$  in particular), and will vary from program to program even if the

same block of shares were to be executed each time and at the same initial price; (3) estimation errors of the parameter estimates for the law of motion for  $P_t$  and  $X_t$  will be propagated recursively through Bellman's equation to  $V_1$ , so that even small uncertainties in the parameter values may lead to large uncertainty in the expected execution cost; and (4) the law of motion for  $P_t$  and  $X_t$  may suffer from the kind of non-stationarities and time-variation that plague all economic models – factors that affect price impact and price dynamics last year need not be relevant next year, and vice versa.

The third and fourth sources of uncertainty – parameter estimation errors and parameter instabilities – are problems that affect all quantitative models in financial economics. Moreover, there are no simple solutions to these problems and their impact must be assessed on a case-by-case basis. Therefore, we can only alert readers to these potentially important issues and leave them aside for now.

To address the first and second sources of uncertainty, we simulate 50,000 sample paths of  $\{P_t, X_t\}$  for the LPT specification of Section 3.1 and for each sample path, we calculate the expected best-execution cost  $V_1$  and the actual cost that the best-execution strategy generated. Moreover, we can decompose both the expected and actual cost into two components according to Eq. (3.6): a 'fundamental value' component (the first term of (3.6)) and a 'price-impact' component (the second term of (3.6)). The standard deviation of these two components over the 50,000 replications measures the risk of best-execution due to fundamentals and price-impact.

Table 6 presents the Monte Carlo simulation results for various combinations of parameter values for  $\gamma$  and  $\rho$ , and where the other parameter values are calibrated as in Table 4. The two rows that correspond to each value of  $\gamma$  describe the behavior of the fundamental component (the first row) and the price-impact component (the second row). The 'Expected Cost' entry is the theoretical best-execution cost, and the 'Simulation Mean' and 'Simulation Std. Dev.' is the mean and standard deviation, respectively, of the actual cost of the best-execution strategy across the 50,000 replications.

The entries in Table 6 show that under the LPT specification there is considerable uncertainty in execution costs, but that almost all of the uncertainty is due to variability in fundamentals (the first term of Eq. (3.6)), not price impact. For example, in the  $(\gamma, \rho) = (0.0000, 0.00)$  case, the standard deviation of the fundamental component is 73.6114 cents/share, but the standard deviation of the price-impact component is only 0.1789 cents/share. Of course, the standard deviation of the price-impact component increases as the information component  $X_t$  becomes more important, i.e., as  $\gamma$  increases. For example, in the extreme case  $(\gamma, \rho) = (0.0050, 0.25)$ , the mean of the price-impact component is 3.8326 cents/share but the standard deviation is 7.7815. This is not surprising because as  $\gamma$  increases, information plays a more important role in the best-execution strategy (and in the price-impact component specifically), hence the variability of information will manifest itself in the variability of the price-impact component.

Table 6

Monte Carlo simulations of the two components of execution costs – fundamental and price-impact components (both in cents/share) – for purchasing  $\bar{S} = 100,000$  shares of stock with an initial price  $P_0 = \$50$  over  $T = 20$  periods under the optimal strategy  $S^*$ . A linear-percentage temporary price-impact function  $P_t = \bar{P}_t + \Delta_t$  is assumed, where  $\Delta_t = (\theta S_t + \gamma X_t) \bar{P}_t$ ,  $\bar{P}_t$  is the zero-impact price,  $S_t$  is the number of shares executed, and  $X_t$  is an information variable that measures current market conditions. The data-generating process for  $\bar{P}_t$  and  $X_t$  are given by:  $\bar{P}_t = \bar{P}_{t-1} \exp(Z_t)$ ,  $Z_t \sim N(\mu_z, \sigma_z^2)$ , and  $X_t = \rho X_{t-1} + \eta_t$ ,  $\eta_t \sim N(0, \sigma_\eta^2)$ . The simulation parameters were calibrated to approximate half-hour intervals:  $\theta = 5 \times 10^{-7}$ ,  $\mu_z = 0$ ,  $\sigma_z^2 = (0.02)^2/13$ ,  $\sigma_\eta^2 = 1 - \rho^2$ ,  $X_1 = 0$ ,  $\bar{P}_0 = \$50$ . The two rows that correspond to each value of  $\gamma$  describe the behavior of the fundamental component (first row) and the price-impact component (second row). Within each row and for each of the three subpanels corresponding to  $\rho = -0.25, 0.00, 0.25$ , the expected cost is reported in the first column, the mean and standard deviation of the simulated cost (over 50,000 replications) is reported in the second and third columns, respectively, of each subpanel.

$\gamma$	$\rho = -0.25$						$\rho = 0.00$						$\rho = 0.25$					
	Expected Cost		Simulation Mean		Simulation Std. Dev.		Expected Cost		Simulation Mean		Simulation Std. Dev.		Expected Cost		Simulation Mean		Simulation Std. Dev.	
0.0000	5000.8000	5001.0883	73.6337	0.1789	5000.8000	12.5058	5000.8000	12.5058	5000.8000	12.5058	5000.8000	12.5058	5000.8000	12.5058	5000.8000	12.5046	5000.8000	12.5046
0.0010	5000.8000	5000.4702	73.5890	0.8721	5000.8000	12.1195	5000.8000	12.1195	5001.0150	73.7902	1.0957	5000.8000	12.1590	5000.8000	5000.9525	73.9691	5000.8000	73.9691
0.0050	5000.8000	5000.7811	74.5058	5.8138	5000.8000	2.8468	5000.8000	2.8468	4999.9618	74.3045	6.5070	5000.8000	3.8348	5000.8000	5000.8120	74.9754	5000.8000	74.9754



But even in this case, the standard deviation of the price-impact component is still an order of magnitude smaller than the standard deviation of the fundamental component.

These simulations suggest that while the overall risk of the best-execution strategy is substantial, virtually all of it is due to the variability of prices, not to the variability of price impact. This implies that most trading strategies, regardless of their optimality for execution costs, will exhibit similar risks. While risk preferences are crucial for determining optimal portfolios (see Section 7.3), they may not be as important for the narrower goal of minimizing execution costs.

### 7.3. Other objective functions

Expected execution costs is only one of many possible objective functions that investors may wish to minimize. In particular, a natural alternative to expected costs is an objective function that also incorporates some measure of risk, e.g., variance of costs and other higher moments. Investors may be willing to accept a higher expected cost of execution if doing so will ensure a much lower variance in the cost of execution. Using the first moment as the objective function implicitly assumes that the investor is risk neutral with respect to execution costs, and while this may be a plausible assumption for broker/dealers and other institutional block-equity traders, we might expect such traders to charge some sort of premium above minimum expected execution costs due to adverse selection and other strategic considerations (see, for example, Glosten and Milgrom, 1985).

While the inclusion of higher moments into the objective function is conceptually straightforward, analytical solutions for the best-execution strategy are not available except in a few very specialized cases.<sup>14</sup> Nevertheless, numerical solutions are readily obtained in these cases, hence the risk of best execution can be managed in much the same way that static portfolio theory balances risk and expected return.

However, this raises the larger issue of what the proper objective function should be. Specifically, while we have focused exclusively on execution costs in this paper, investors are ultimately interested in maximizing the expected utility of their wealth. Therefore, the most natural approach to execution costs is to maximize the investor's expected utility of wealth subject to the law of motion for  $P_t$  and  $X_t$ , i.e., combine the portfolio optimization problem with the execution problem. This is the approach taken by Davis and Norman (1990), Grossman and Vila (1992), Heaton and Lucas (1996), Magill and Constantinides (1976), Vayanos (1995), and Vayanos and Vila (1995).

There is no doubt that such an approach is the 'right' one – our implicit two-stage optimization cannot yield a better outcome (in terms of the investor's expected utility) than a combined optimization. But as with many financial optimization

---

<sup>14</sup>In particular, we have derived approximate results for a mean-variance objective function in the linear additive case of Section 2.4

problems, the maximization of expected utility with transactions costs is analytically and computationally intractable in all but the most stylized examples. And although such examples do provide important insights into the economics of transactions costs, they have little to say about minimizing transactions costs in practice.

#### 7.4. *Partial versus general equilibrium*

Our analysis assumes that the law of motion for  $P_t$  and  $X_t$  is given. In particular, the parameters and functional form of the law of motion are unaffected by the investor's trades. Although this partial-equilibrium approach may be appropriate in some circumstances, one can imagine other circumstances when an equilibrium model is more appropriate. For example, if a small number of large investors dominate the market, then strategic considerations become more significant. In such cases, the law of motion for  $P_t$  and  $X_t$  will be directly influenced by each of the large investors' trades, hence the best-execution strategy will depend on an infinite regress of expectations of other investors' expectations. These strategic interactions and their equilibrium implications have been explored in some simple contexts by Jarrow (1992), Kyle (1985, 1989), and Vayanos (1995).

Despite the potential importance of equilibrium considerations, our approach is still appropriate even in the presence of strategic behavior if the law of motion is properly calibrated and the environment relatively stable. In particular, if the law of motion is estimated from historical data (see, for example, Bertsimas et al., 1998), then it represents a 'reduced-form' description of price impact and price dynamics, which includes the effects of strategic interactions among large investors. Of course, the standard criticism of reduced-form models is also relevant in this context: structural changes cannot be anticipated by reduced-form models. Nevertheless, given the level of anonymity imposed by current market structure, it is difficult to see how strategic interactions can be easily incorporated into the best-execution problem. However, this is an important open question that deserves further investigation.

## 8. Conclusion

With the advent of more accurate models and measures of execution costs, and given the dramatic increase in institutional trading in recent years, the optimal control of execution costs has never been more important. We have argued that this is a dynamic optimization problem because trading takes time, the demand for financial securities is not perfectly elastic, and the price impact of current trades, even small trades, can affect the course of future prices.

Using stochastic dynamic programming, we have derived trading strategies that minimize the expected cost of execution – called best-execution strategies – which adapt to changing market conditions and price changes in an optimal

fashion. We have shown that the common approach of breaking up a large trade into a number of smaller trades of equal size is optimal only in one very special case: when price impact is linear in the trade size, permanent in its affect on future prices, and when prices follow an arithmetic random walk. For empirically more plausible price dynamics and price-impact functions, the best-execution strategy varies through time as a function of several state variables which measure market conditions and the remaining shares to be executed. Monte Carlo simulations indicate that for at least one specification of price dynamics and price impact, on average the execution cost of the best-execution strategy is 25% to 40% less than that of the naive strategy of trading in equal-size lots.

A by-product of the dynamic programming algorithm is the expected cost of best-execution, which is given by the optimal-value function. This may be a useful benchmark for pricing principal-bid and negotiated-block transactions since dealers engaging in such ‘basket’ trades will be pricing their bids with an eye towards trading out of these positions as quickly and as cheaply as possible, i.e., they will seek best-execution strategies for the basket.

We have extended our analysis to the portfolio case in which trades in several securities must be executed simultaneously. The results are qualitatively similar to the single-security case, but they differ in one important respect: the portfolio case captures potentially important cross-effects between prices and over time, and these cross-effects are optimally exploited by the portfolio best-execution strategy. In particular, unless the collection of securities are independent in terms of price impact – trades in one security have no impact on other securities’ prices – the components of the best-execution strategy for the portfolio will not, in general, be identical to the best-execution strategy for individual securities. Whether or not the portfolio execution costs are greater or less than the sum of individual securities’ execution costs depends of course on the signs and magnitudes of the cross-effects, a subject of ongoing investigation in Bertsimas et al. (1998).

We have also discussed extensions of our results to more general price-impact functions and price dynamics, and to the case with constraints. In many cases numerical solutions are available, but in some cases – those with many constraints – alternatives to dynamic programming must be considered.

Of course, the challenging problem of building transactions costs directly into optimal consumption/investment decisions and dynamic equilibrium models still remains. Our more modest goal in this paper is to minimize expected execution costs, a far simpler problem, but one which has considerable relevance and for which we can provide a fairly complete set of solutions.

## **Acknowledgements**

We thank Robert Ferstenberg for stimulating our interest in this area, and Petr Adamek and Martin Haugh for excellent research assistance. We

are grateful to Dave Cushing, Chris Darnell, Rohit D’Souza, John Heaton, Holly Horrigan, Leonid Kogan, Bruce Lehmann, Greg Peterson, Jiang Wang, a referee and seminar participants at Columbia, Harvard, London Business School, MIT, Northwestern, the 1996 NNCM Conference, the NYSE Conference on Best Execution, the Society of Quantitative Analysts, and Yale for helpful comments and discussion. This research was partially supported by the MIT Laboratory for Financial Engineering, an Alfred P. Sloan Research Fellowship, an NSF grant (SBR-9709976), and a Presidential Young Investigator Award (DDM-9158118) with matching funds from Draper Laboratory.

## Appendix A.

In this appendix we derive the best-execution strategies for the linear and linear-percentage price-impact functions of Sections 2 and 3.

### A.1. Linear price impact

We show that Eqs. (2.20) and (2.21) is the optimal control and optimal-value function, respectively, by induction on  $k$ . For  $k = 0$ , the optimal control is  $S_T^* = W_T$  and the optimal-value function is

$$\begin{aligned} V_T(P_{T-1}, X_T, W_T) &\equiv \text{Min}_{S_T} E_T[P_T S_T] = (P_{T-1} + \theta S_T + \gamma X_T) S_T \\ &= P_{T-1} W_T + \theta W_T^2 + \gamma X_T W_T \end{aligned} \quad (\text{A.1})$$

where we have substituted  $a_0 = \theta$ ,  $b_0 = \gamma$ ,  $c_0 = 0$ , and  $d_0 = 0$  in Eq. (2.21) to obtain Eq. (A.1). Assuming that Eq. (2.21) holds for  $k$ , we will prove it for  $k + 1$ . Applying Bellman’s equation yields

$$\begin{aligned} V_{T-k-1}(P_{T-k-2}, X_{T-k-1}, W_{T-k-1}) &= \text{Min}_{S_{T-k-1}} E_{T-k-1}[P_{T-k-1} S_{T-k-1} \\ &\quad + V_{T-k}(P_{T-k-1}, X_{T-k}, W_{T-k})]. \end{aligned} \quad (\text{A.2})$$

Using the induction hypothesis we substitute Eq. (2.21) into Eq. (A.2) and obtain

$$\begin{aligned} V_{T-k-1}(P_{T-k-2}, X_{T-k-1}, W_{T-k-1}) &= \text{Min}_{S_{T-k-1}} E_{T-k-1}[P_{T-k-1} S_{T-k-1} \\ &\quad + P_{T-k-1} W_{T-k} + a_k W_{T-k}^2 \\ &\quad + b_k X_{T-k} W_{T-k} + c_k X_{T-k}^2 + d_k]. \end{aligned} \quad (\text{A.3})$$

Using the laws of motion (2.18), (2.19), and (2.4) we obtain

$$\begin{aligned}
 &V_{T-k-1}(P_{T-k-2}, X_{T-k-1}, W_{T-k-1}) \\
 &= \text{Min}_{S_{T-k-1}} [(P_{T-k-2} + \theta S_{T-k-1} + \gamma X_{T-k-1})W_{T-k-1} \\
 &\quad + a_k(S_{T-k-1} - W_{T-k-1})^2 + b_k \rho X_{T-k-1}(W_{T-k-1} - S_{T-k-1}) \\
 &\quad + c_k(\rho^2 X_{T-k-1}^2 + \sigma_{\eta}^2) + d_k]. \tag{A.4}
 \end{aligned}$$

Note that the function to be minimized in Eq. (A.4) is quadratic in  $S_{T-k-1}$ . Moreover, from the induction hypothesis  $a_k \geq 0$  (if  $\theta \geq 0$ ) and therefore the function is convex. It is straightforward to verify that the optimal solution is attained at

$$S_{T-k-1}^* = \left(1 - \frac{\theta}{2a_k}\right) W_{T-k-1} + \frac{\rho b_k}{2a_k} X_{T-k-1}. \tag{A.5}$$

Substituting Eq. (A.5) into Eq. (A.4) then yields Eq. (2.21) for  $k + 1$ . Observe that  $a_k$  satisfies the following recursion:

$$a_k = \theta \left(1 - \frac{\theta}{4a_{k-1}}\right), \quad a_0 = \theta. \tag{A.6}$$

Solving this recursion yields

$$a_k = \frac{\theta}{2} \left(1 + \frac{1}{k+1}\right). \tag{A.7}$$

Therefore, the optimal control becomes

$$S_{T-k-1}^* = \frac{W_{T-k-1}}{k} + \frac{\rho b_k}{2a_k} X_{T-k-1} \tag{A.8}$$

and the induction is complete.

### *A.2. Linear percentage price impact*

As usual, we begin at the end by characterizing the optimal-value function  $V_T$  and then proceed recursively according to Bellman's equation:

$$V_T(\tilde{P}_{T-1}, X_T, W_T) = \text{Min}_{S_T} E_T[P_T S_T] = E_T[P_T W_T] \tag{A.9}$$

$$= q \tilde{P}_{T-1}(W_T + \gamma X_T W_T + \theta W_T^2) \tag{A.10}$$

where

$$q \equiv E_{T-1}[\exp(Z_T)] = E[\exp(Z_T)] = \exp\left(\mu_z + \frac{\sigma_z^2}{2}\right) \quad (\text{A.11})$$

since, by assumption,  $Z_t$  is IID normal with mean  $\mu_z$  and variance  $\sigma_z^2$ . In the next-to-last period, the optimal-value function satisfies:

$$V_{T-1} = \text{Min}_{S_{T-1}} E_{T-1}[P_{T-1}S_{T-1} + V_T(\tilde{P}_{T-1}, X_T, W_T)] \quad (\text{A.12})$$

$$\begin{aligned} &= \text{Min}_{S_{T-1}} \{ \tilde{P}_{T-1}(1 + \theta S_{T-1} + \gamma X_{T-1})S_{T-1} + q\tilde{P}_{T-1} \\ &\quad \times (W_{T-1} - S_{T-1})[1 + \gamma\rho X_{T-1} + \theta(W_{T-1} - S_{T-1})] \}. \end{aligned} \quad (\text{A.13})$$

The best execution at  $T - 1$  is then

$$S_{T-1}^* = \delta_{x,1}X_{T-1} + \delta_{w,1}W_{T-1} + \delta_{1,1} \quad (\text{A.14})$$

where

$$\delta_{x,1} = \frac{\gamma(q\rho - 1)}{2\theta(q + 1)}, \quad \delta_{w,1} = \frac{q}{q + 1}, \quad \delta_{1,1} = \frac{q - 1}{2\theta(q + 1)}. \quad (\text{A.15})$$

Therefore, the optimal-value function  $V_{T-1}$  is given by

$$\begin{aligned} V_{T-1}(\tilde{P}_{T-2}, X_{T-1}, W_{T-1}) &= q\tilde{P}_{T-2}[a_1 + b_1X_{T-1} + c_1X_{T-1}^2 \\ &\quad + d_1X_{T-1}W_{T-1} + e_1W_{T-1} + f_1W_{T-1}^2] \end{aligned} \quad (\text{A.16})$$

where

$$a_1 = \delta_{1,1}(1 + \theta\delta_{1,1}) - q\delta_{1,1}(1 - \theta\delta_{1,1}), \quad (\text{A.17a})$$

$$b_1 = (1 - q)\delta_{x,1}, \quad (\text{A.17b})$$

$$c_1 = \delta_{x,1}(\theta\delta_{x,1} + \gamma) - q\delta_{x,1}(\gamma\rho - \theta\delta_{x,1}), \quad (\text{A.17c})$$

$$d_1 = \gamma(1 + \rho)\delta_{w,1}, \quad (\text{A.17d})$$

$$e_1 = 2\delta_{w,1}, \quad (\text{A.17e})$$

$$f_1 = \theta\delta_{w,1}. \quad (\text{A.17f})$$

Continuing in this fashion, we arrive at the following recursions:

$$V_{T-k} = \text{Min}_{S_{T-k}} E_{T-k}[P_{T-k}S_{T-k} + V_{T-k+1}(\tilde{P}_{T-k}, X_{T-k+1}, W_{T-k+1})] \quad (\text{A.18})$$

which yields the best-execution strategy:

$$S_{T-k}^* = \delta_{x,k} X_{T-k} + \delta_{w,k} W_{T-k} + \delta_{1,k} \quad (\text{A.19})$$

where

$$\delta_{x,k} = \frac{q\rho d_{k-1} - \gamma}{2(\theta + qf_{k-1})}, \quad \delta_{w,k} = \frac{qf_{k-1}}{\theta + qf_{k-1}}, \quad \delta_{1,k} = \frac{qe_{k-1} - 1}{2(\theta + qf_{k-1})}. \quad (\text{A.20})$$

Hence

$$\begin{aligned} V_{T-k}(\tilde{P}_{T-k-1}, X_{T-k}, W_{T-k}) &= q\tilde{P}_{T-k-1}[a_k + b_k X_{T-k} + c_k X_{T-k}^2 \\ &\quad + d_k X_{T-k} W_{T-k} + e_k W_{T-k} + f_k W_{T-k}^2] \end{aligned} \quad (\text{A.21})$$

where

$$a_k = \delta_{1,k}(1 + \theta\delta_{1,k}) + q(a_{k-1} + \sigma_\eta^2 c_{k-1}) - q\delta_{1,k}(e_{k-1} - \delta_{1,k}f_{k-1}), \quad (\text{A.22a})$$

$$b_k = q\rho b_{k-1} - \delta_{x,k}(qe_{k-1} - 1), \quad (\text{A.22b})$$

$$c_k = \delta_{x,k}(\theta\delta_{x,k} + \gamma) + q\rho^2 c_{k-1} - q\delta_{x,k}(\rho d_{k-1} - \delta_{x,k}f_{k-1}), \quad (\text{A.22c})$$

$$d_k = \gamma\delta_{w,k} + q\rho d_{k-1}(1 - \delta_{w,k}), \quad (\text{A.22d})$$

$$e_k = \delta_{w,k} + q(1 - \delta_{w,k})e_{k-1}, \quad (\text{A.22e})$$

$$f_k = \theta\delta_{w,k}. \quad (\text{A.22f})$$

This set of recursions completely characterize the best-execution strategy  $\{S_t^*\}$ , and the expected cost of best execution is given by

$$\begin{aligned} V_1(\tilde{P}_0, X_1, W_1) &= q\tilde{P}_0[a_{T-1} + b_{T-1}X_1 + c_{T-1}X_1^2 + d_{T-1}X_1W_1 \\ &\quad + e_{T-1}W_1 + f_{T-1}W_1^2] \end{aligned} \quad (\text{A.23})$$

$$= \text{Min}_{\{S_t\}} E_1 \left[ \sum_{t=1}^T P_t S_t \right]. \quad (\text{A.24})$$

## References

- Aiyagari, R., Gertler, M., Asset returns with transactions costs and uninsured individual risk. *Journal of Monetary Economics* 27, 311–332.
- Allen, F., Gale, D., 1992. Stock-price manipulation. *Review of Financial Studies* 5, 503–529.
- Angel, J., 1994. Limit versus market orders. Working Paper No. FINC-1377-01-293, School of Business Administration, Georgetown University.
- Arnott, R., Wagner, W., 1990. The measurement and control of trading costs. *Financial Analyst Journal* 46, 73–80.
- Barclay, M., Litzenberger, R., 1988. Announcement effects of new equity issues and the use of intraday price data. *Journal of Financial Economics* 21, 71–100.

- Barclay, M., Warner, J., 1993. Stealth trading and volatility: which trades move prices? *Journal of Financial Economics* 34, 281–306.
- Bensaid, B., Lesne, J., Pages, H., Scheinkman, J., 1992. Derivative asset pricing with transaction costs. *Mathematical Finance* 2, 63–86.
- Berkowitz, S., Logue, D., Noser, E. Jr., 1988. The total cost of transactions on the NYSE. *Journal of Finance* 43, 97–112.
- Bernhardt, D., Hughson, E., 1994. Splitting orders. Working Paper, Department of Economics, Queen's University.
- Bertsekas, D., 1995. *Dynamic programming and optimal control*, vol. I. Athena Scientific, Belmont, MA.
- Bertsimas, D., Lo, A., 1998. A static optimization alternative to dynamic programming with constraints. Working paper, Sloan School of Management, MIT.
- Bertsimas, D., Hummel, P., Lo, A., 1998. An empirical analysis of best execution. Working paper, Sloan School of Management, MIT.
- Birinyi, L., 1995. What does institutional trading cost?. Birinyi Associates, Greenwich, CT.
- Bodurtha, S., Quinn, T., 1990. Does patient program trading really pay? *Financial Analysts Journal* 46, 35–42.
- Boyle, P., Vorst, T., 1992. Option replication in discrete time with transaction costs. *Journal of Finance* 47, 271–294.
- Brennan, M., Copeland, T., 1988. Stock splits, stock prices, and transaction costs. *Journal of Financial Economics* 22, 83–101.
- Brinson, G., Hood, R., Beebower, G., 1986. Determinants of portfolio performance. *Financial Analysts Journal* 42, 39–44.
- Brinson, G., Singer, B., Beebower, G., 1991. Determinants of portfolio performance II: an update. *Financial Analysts Journal* 47, 40–48.
- Chan, L., Lakonishok, J., 1993. Institutional trades and intra-day stock price behavior. *Journal of Financial Economics* 33, 173–199.
- Chan, L., Lakonishok, J., 1995. The behavior of stock prices around institutional trades. *Journal of Finance* 50, 1147–1174.
- Cohen, K., Maier, S., Schwartz, R., Whitcomb, D., 1981. Transaction costs, order placement strategy and existence of the bid-ask spread. *Journal of Political Economy* 89, 287–305.
- Collins, B., Fabozzi, F., 1991. A methodology for measuring transaction costs. *Financial Analysts Journal* 47, 27–36.
- Constantinides, G., 1986. Capital market equilibrium with transaction costs. *Journal of Political Economy* 94, 842–862.
- Cuneo, L., Wagner, W., 1975. Reducing the cost of stock trading. *Financial Analysts Journal* 26, 35–44.
- Davis, M., Norman, A., 1990. Portfolio selection with transactions costs. *Mathematics of Operations Research* 15, 676–713.
- Demsetz, H., 1968. The cost of transacting. *Quarterly Journal of Economics* 82, 33–53.
- Dumas, B., Luciano, E., 1991. An exact solution to a dynamic portfolio choice problem under transactions costs. *Journal of Finance* 46, 577–596.
- Easley, D., O'Hara, M., 1987. Price, trade size, and information in securities markets. *Journal of Financial Economics* 19, 69–90.
- Epps, T.W., 1976. The demand for brokers' services: the relation between security trading volume and transaction cost. *Bell Journal of Economics* 7, 163–196.
- Foster, D., Viswanathan, S., 1990. A theory of the interday variations in volume, variance, and trading costs in securities markets. *Review of Financial Studies* 3, 593–624.
- Gammill, J., Perold, A., 1989. The changing character of stock market liquidity. *Journal of Portfolio Management* 15, 13–18.
- Garman, M., Ohlson, J., 1981. Valuation of risky assets in arbitrage-free economies with transactions costs. *Journal of Financial Economics* 9, 271–280.



- Glosten, L., Milgrom, P., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14, 71–100.
- Grossman, S., Miller, M., 1988. Liquidity and market structure. *Journal of Finance* 43, 617–633.
- Grossman, S., Vila, J., 1992. Optimal dynamic trading with leverage constraints. *Journal of Financial and Quantitative Analysis* 27, 151–168.
- Harris, L., 1994. Optimal dynamic order submission strategies in some stylized trading problems. Working Paper No. 94–8, School of Business Administration, University of Southern California.
- Harris, L., Hasbrouck, J., 1992. Market vs. limit orders: the superdot evidence on order submission strategy, NYSE working paper No. 92–02.
- Hasbrouck, J., Schwartz, R., 1988. Liquidity and execution costs in equity markets. *Journal of Portfolio Management* 14, 10–16.
- Hausman, J., Lo, A., MacKinlay, C., 1992. An ordered probit analysis of transaction stock prices. *Journal of Financial Economics* 31, 319–379.
- Heaton, J., Lucas, D., 1996. Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy* 104, 443–487.
- Hodges, S., Neuberger, A., 1989. Optimal replication of contingent claims under transactions costs. *Review of Futures Markets* 8, 222–239.
- Holthausen, R., Leftwich, R., Mayers, D., 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., Mayers, D., 1990. Large block transactions, the speed of response, and temporary and permanent stock-price effects. *Journal of Financial Economics* 26, 71–95.
- Huang, R., Stoll, H., 1995. Dealer versus auction markets: a paired comparison of execution costs on NASDAQ and the NYSE, Working paper 95–16, Financial Markets Research Center, Owen Graduate School of Management, Vanderbilt University.
- Jarrow, R., 1992. Market manipulation, bubbles, corners, and short squeezes. *Journal of Financial and Quantitative Analysis* 27, 311–336.
- Keim, D., Madhavan, A., 1995a. The anatomy of the trading process. *Journal of Financial Economics* 37, 371–398.
- Keim, D., Madhavan, A., 1995b. The upstairs market for large-block transactions: analysis and measurement of price effects. *Review of Financial Studies* to appear.
- Keim, D., Madhavan, A., 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., Stoll, H., 1972. Price impacts of block trading on the New York stock exchange. *Journal of Finance* 27, 569–588.
- Kumar, P., Seppi, D., 1993. Limit and market orders with optimizing traders. Working paper, Graduate School of Industrial Administration, Carnegie-Mellon University.
- Kyle, A., 1985. Continuous auctions and insider trading. *Econometrica* 53, 1315–1336.
- Kyle, A., 1989. Informed speculation with imperfect competition. *Review of Economic Studies* 56, 317–356.
- Leinweber, D., 1993. Using information from trading in trading and portfolio management. In: Sherrerd, K. (Ed.), *Execution Techniques, True Trading Costs, and the Microstructure of Markets*. Association for Investment Management and Research, Charlottesville, VA.
- Leinweber, D., 1994. Careful structuring reins in transaction costs. *Pensions and Investments* July 25, 19.
- Leland, H., 1985. Option pricing and replication with transactions costs. *Journal of Finance* 40, 1283–1301.
- Litzenberger, R., Rollo, J., 1984. Arbitrage pricing, transaction costs and taxation of capital gains: a study of government bonds with the same maturity date. *Journal of Financial Economics* 13, 337–351.

- Lo, A., MacKinlay, C., 1990. When are contrarian profits due to stock market overreaction? *Review of Financial Studies* 3, 175–206.
- Lo, A., MacKinlay, C., Zhang, J., 1998. Econometric models of limit-order executions, MIT Laboratory for Financial Engineering working paper No. LFE-1031-97.
- Loeb, T., 1983. Trading cost: the critical link between investment information and results. *Financial Analysts Journal* 39, 39–44.
- Macey, J., O'Hara, M., 1996. The law and economics of best execution. Working Paper, Cornell University, unpublished.
- Magill, M., Constantinides, G., 1976. Portfolio selection with transactions costs. *Journal of Economic Theory* 13, 245–263.
- Merton, R., 1969. Lifetime portfolio selection under uncertainty: the continuous-time case. *Review of Economics and Statistics* 51, 247–257.
- New York Stock Exchange Fact Book: 1994 Data, New York Stock Exchange, April 1995.
- Niehans, J., 1987. Transaction costs. In J. Eatwell, M. Milgate, and Newman, P. (Eds.), *The New Palgrave: A dictionary of economics*. The MacMillan Press, London, U.K.
- Pérolod, A., 1988. The implementation shortfall: paper versus reality. *Journal of Portfolio Management* 14, 4–9.
- Samuelson, P., 1969. Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics* 51, 239–246.
- Schwartz, R., Whitcomb, D., 1988. Transaction costs and institutional investor trading strategies. Monograph Series in Finance and Economics 1988–2/3, Salomon Brothers Center for the Study of Financial Institutions, New York University, New York.
- Sherrerd, K. (Ed.), 1993. Execution techniques, true trading costs, and the microstructure of markets. Association for Investment Management and Research, Charlottesville, VA.
- Shleifer, A., 1986. Do demand curves for stocks slope down? *Journal of Finance* 41, 579–590.
- Stoll, H., 1989. Inferring the components of the bid–ask spread: theory and empirical tests. *Journal of Finance* 44, 115–134.
- Stoll, H., 1993. Equity trading costs. Association for Investment Management and Research, Charlottesville, VA.
- Tinic, S., 1972. The economics of liquidity services. *Quarterly Journal of Economics* 86, 79–93.
- Treynor, J., 1981. What does it take to win the trading game? *Financial Analysts Journal* 37, 55–60.
- Tuckman, B., Vila, J., 1992. Arbitrage with holding costs: a utility-based approach. *Journal of Finance* 47, 1283–1302.
- Turnbull, A., White, R., 1995. Trade type and the costs of making markets. Working Paper, Western Business School University of Western Ontario.
- Vayanos, D., 1992. A dynamic model of an imperfectly competitive bid–ask market. Unpublished working paper.
- Vayanos, D., 1995. Transaction costs and asset prices: a dynamic equilibrium model. Working paper, Graduate School of Business, Stanford University.
- Vayanos, D., Vila, J., 1995. Equilibrium interest rate and liquidity premium under proportional transaction costs. Working paper, Graduate School of Business, Stanford University.
- Vila, J., 1989. Simple games of market manipulation. *Economics Letters* 29, 21–26.
- Wagner, W., 1993. Defining and measuring trading costs. In Sherrerd, K. (ed.), *Execution Techniques, True Trading Costs, and the Microstructure of Markets*. Association for Investment Management and Research, Charlottesville, VA.
- Wagner, W., Banks, M., 1992. Increasing portfolio effectiveness via transaction cost management. *Journal of Portfolio Management* 19, 6–11.
- Wagner, W., Edwards, M., 1993. Best execution. *Financial Analyst Journal* 49, 65–71.