Adaptive Arrival Price

Robert Almgren* and Julian Lorenz**

April 27, 2006

Abstract

Electronic trading of equities and other securities makes heavy use of "arrival price" algorithms, that determine optimal trade schedules by balancing the market impact cost of rapid execution against the volatility risk of slow execution. In the standard formulation, mean-variance optimal strategies are static: they do not modify the execution speed in response to price motions observed during trading. We show that with a more realistic formulation of the mean-variance tradeoff, and even with no momentum or mean reversion in the price process, substantial improvements are possible for adaptive strategies that spend trading gains to reduce risk, by accelerating execution when the price moves in the trader's favor. The improvement is larger for large initial positions.

*Electronic Trading Services, Banc of America Securities LLC, New York; Robert.Almgren@bofasecurities.com.

**Institute of Theoretical Computer Science, ETH Zürich; jlorenz@inf.ethz.ch.
Partially supported by UBS AG.

Contents

1 Introduction **3**
1.1 Example 5
1.2 Trading in practice 6
1.3 Other adaptive strategies 7

2 Market Model **8**
2.1 Static trajectories 10
2.2 Nondimensionalization 10
2.3 Small-portfolio limit 12
2.4 Portfolio comparison 12

3 Single Update **13**
3.1 Mean and variance 14
3.2 Numerical results 16

4 Continuous Response **18**
4.1 Numerical results 18

5 Discussion and Conclusions **20**

A Detailed formulas **22**
A.1 Means and variances 22
A.2 Full distribution 22

1 Introduction

Algorithmic trading represents a large and growing fraction of total order flow, especially in equity markets. When the size of a requested buy or sell order is larger than the market can immediately supply or absorb, then the order must be worked across some period of time, exposing the trader to price volatility. The algorithm attempts to achieve an average execution price whose probability distribution is suited to the client's preferences. This paper proposes a way to dramatically improve this distribution.

Arrival price algorithms, which are currently the most widely used framework, take as their benchmark the pre-trade or "decision" price. The difference between the execution price and the benchmark is the "implementation shortfall" (Perold 1988), which is an uncertain quantity since order execution takes a finite amount of time. In the most straightforward version of this model, the expected value of the implementation shortfall is entirely due to market impact incurred by trading at a nonzero rate (we neglect anticipated price drift); this expected cost is minimized by trading as slowly as possible, for example, a VWAP strategy across the maximum allowed time horizon. Since market impact is assumed deterministic, the variance of the implementation shortfall is entirely due to price volatility; this variance is minimized by trading rapidly.

This risk-reward tradeoff is very familiar in finance, and a variety of criteria can be used to determine risk-averse optimal solutions. Arrival price algorithms compute the set of "efficient" strategies that minimize risk for a specified maximum level of expected cost or conversely; the set of such strategies is summarized in the "efficient frontier of optimal trading" introduced by Almgren and Chriss (2000) (see also Almgren and Chriss (1999)). The simple mean-variance approach has the advantage that the risk-reward tradeoff is independent of initial wealth, a useful property in an institutional setting.

A central question is whether the trade schedule should be *static* or *dynamic:* should the list of shares to be executed in each interval of time be computed and fixed before trading begins, or should the trade list be updated in "real time" using information revealed during execution?

The surprising observation of Almgren and Chriss (2000) is that, under very realistic assumptions about the asset price process (arithmetic random walk with no serial correlation), static strategies are *equivalent* to dynamic strategies. No value is added by considering “scaling” strategies in which the execution speed changes in response to price motions.

To be more specific, let us consider two different specifications of the trade scheduling problem:

1. For a static strategy, we require that the entire trade schedule must be fixed in advance (Huberman and Stanzl (2005) suggest that a reasonable example of this is insider trading, where trades must be announced in advance). For any candidate schedule, the mean and variance are evaluated at the initial time, and the optimal schedule is determined for a specific risk aversion level.

2. For a dynamic strategy as usually understood in dynamic programming, we allow arbitrary modification of the strategy at any time. To recalculate the trade list, we use all information available at that time and we value strategies by a mean-variance tradeoff of the remaining cost, using a constant parameter of risk aversion.

In the model of Almgren and Chriss (2000), 1 and 2 have the same solution. Liquidity and volatility are assumed known in advance, so the only information revealed is the asset price motion. Price information revealed in the first part of the execution does not change the probability distribution of future price changes. Because the mean-variance tradeoff is independent of initial wealth, the trading gains or losses incurred in the first part of the program are “sunk costs” and do not influence the strategy for the remainder.

This paper presents an alternative formulation:

3. In the new formulation, we precompute the *rule* determining the trade rate as a function of price, using a mean-variance tradeoff measured at the initial time. Once trading begins, the rule may not be modified, even if the trader’s preferences reevaluated at an intermediate time would lead him or her to choose a different strategy, as in 2 above (we call this the “Dr. Strangelove” strategy).

The optimal solution of problem 3 is generally not the same as the solution of problems 1 and 2.

As an illuminating contrast, in the well-known problem of option hedging, the optimal hedge position and hence the trade list depend on price and hence are not known until the price is observed, although the *rule* giving this hedge position is computed in advance using dynamic programming. Thus formulation 1 is dramatically suboptimal, but 3 gives the same result as 2.

For algorithmic trading, the improved results of 3 over 1 and 2 come from introducing a negative correlation between the trading gains or losses in the first part of the execution and market impact costs incurred in the second part. Trading gains and losses due to price movement are serially uncorrelated, but they can be correlated with market impact costs by a simple rule: if the price moves in your favor in the early part of the trading, then spend those gains on market impact costs by accelerating the remainder of the program. If the price moves against you, then reduce future costs by trading more slowly, despite the increased exposure to risk of future fluctuations. The result is an overall decrease in variance measured at the initial time, which can be traded for a decrease in expected cost.

In practice there are no artificial constraints on the adaptivity of trading strategies. The key observation of this paper is that the *ex ante* mean-variance optimization expressed by formulation 3 corresponds better to the way that trading results are measured in practice, via *ex post* sample mean and variance over a collection of similar programs. A simple example will make the logic clear.

1.1 Example

Suppose that two bets are available. Bet A pays 0 or 6 with equal probability; its expected value is 3 and its variance is 9. Bet B pays 1 with certainty; its expected value is 1 and its variance is zero. We consider a risk-averse investor whose coefficient of risk aversion is 1/9: he assigns *ex ante* value $E - (1/9)V$ to a random payout with expected value E and variance V. For this investor, a single play of A has value 2 and a single play of B has value 1, so he prefers A.

Now suppose that our investor will play this game two times, with independence between the outcomes. We consider three ways in which he may choose his bets.

1. In a static strategy, he must fix the sequence AA, AB, BA, or BB before the game begins. By independence, choice AA has twice the value of A and is preferred. Its value is 4.

2. In a dynamic strategy, he chooses the second bet after he learns the result of the first play. By that time, the first result will be a constant wealth offset, so he will always choose A on the second play. Knowing that that will be his future choice, he chooses A on the first bet to maximise his total value measured at the initial time. Thus the strategy and the payoff are the same as in the static case.

3. In our new formulation, the investor specifies *three* choices: his bet on the first play, his bet on the second play if he wins the first one, and his bet on the second play if he loses the first. The optimal rule is to bet A on the first play, and if then he wins to choose B, if he loses to play A again, giving payouts 0, 6, 7, and 7 with equal probability. Its value is 4.06, better than choices 1 or 2.

In this model, bet A corresponds to slow trading, with high expected value (low cost) and high variance, and B is fast trading. If the random outcome (trading gain) in the first period is positive, then the trader spends some of this gain on reducing the variance in the second period.

Now suppose that the investor plays this game many times in sequence, and wishes to optimize his sample mean and variance, combined using the same coefficient of risk aversion. If the results are reported over individual plays, then the *ex post* sample mean and variance will be close to the *ex ante* expectation and variance of a single play, and the optimal strategy will be to bet A each time, as in 1 and 2 above.

However, suppose the results are aggregated over *pairs* of plays. That is, the gains of play 1 and play 2 are added together, play 3 and play 4 are added, *etc.* Then the adaptive strategy of case 3 above will give the best results: within each pair, choose the second bet based on the result of the first one. If the results are grouped into larger sets, then a more complicated strategy will be even more optimal.

1.2 Trading in practice

As in the simple example, the question of which formulation is more realistic depends on how trading results are reported. At Banc of America Securities, and probably at other firms, clients of the agency trading desk are provided with a post-trade report daily, weekly, or monthly depending on their trading activity. This report shows sample average and standard deviation of execution price relative to the implementation shortfall benchmark, across all trades executed for that client during the reporting period. The results are further broken down into subsets across a dozen dimensions such as strategy type, primary exchange, buy or sell, trade size, market capitalization, sector, and the like.

Because of the subsets, it is difficult to identify a larger unit than the individual order. We therefore argue that the broker-dealer's goal is to design algorithms that optimize sample mean and variance at the per-order level, so that the post-trade report will be as favorable as possible. As in the simple example, this criterion translates to formulation 3 above, which is not optimized by current arrival price algorithms.

Of course, the broker also has a responsibility to design the post-trade report so that it will be maximally useful to the client; that is, so that it corresponds as closely as possible to the client's investment goals. One interpretation of the results here is that the report should show statistics with finer resolution. For example, it could show mean and variance of shortfall for each one thousand dollars of client money spent, for example. The best choice of reporting interval is an open question.

1.3 Other adaptive strategies

Our new optimal stratgies are "aggressive-in-the-money" (AIM) in the sense of Kissell and Malamut (2006): execution accelerates when the price moves in the trader's favor, and slows when the price moves adversely. A "passive-in-the-money" (PIM) strategy would react oppositely. Adaptive strategies of this form are called "scaling" strategies, and they can arise for a number of reasons beyond those considered here.

A decrease in risk tolerance following a gain, and increase following a loss, is consistent with traders' observed preferences (Shefrin and Statman 1985) and is well-known in "prospect theory" (Kahneman and Tversky 1979). Perhaps for this reason, scaling strategies often seem intuitively reasonable, though such qualitative preferences properly have no place in quantitative institutional trading. Our formulation is straightforward mean-variance optimization.

One important reason for using a AIM or PIM strategy would the expectation of serial correlation in the price process. If the price is believed to have momentum (positive serial correlation), then a PIM strategy is optimal: if the price moves favorably, one should slow down to capture even more favorable prices in the future. Conversely, if the price is believed to be mean-reverting, then favorable prices should be captured quickly before they revert. Adaptive strategies can also be optimal according to risk aversion criteria other than simple mean-variance (Kissell and Malamut 2006). Our strategies arise in a pure random walk model with no serial correlation, using pure classic mean and variance.

These models do provide an important caveat for our formulation. Our AIM strategy suggests to “cut your gains and let your losses run.” If the price process does have any significant momentum, even on a small fraction of the real orders, then this strategy can cause much more serious losses than the gains it provides. Thus we do not advocate implementing them in practice before doing extensive empirical tests.

In Section 2 we present our market and trading model, and show the general importance of the “market power” parameter. We then consider two simple “proofs of concept:” in Section 3 a single update time, and in Section 4 a continuous response function depending linearly on asset price. In Section 5 we summarize and describe ongoing work towards the full continuous-time model.

2 Market Model

We consider trading in a single asset whose price is $S(t)$, obeying the arithmetic random walk

$$S(t) = S_0 + \sigma B(t)$$

where $B(t)$ is a standard Browian motion and σ is an absolute volatility. This process has neither momentum nor mean reversion: future price changes are completely independent of past changes. The Brownian motion $B(t)$ is the only source of randomness in the problem. In the presence of intraday seasonality, we interpret t as a volume time relative to a historical profile, and we assume that volatility is constant under this transformation.

The trader has an order of X shares, which begins at time $t = 0$ and must be completed by time $t = T < \infty$. We shall suppose $X > 0$ and interpret this as a buy order. The benchmark value of this position at the start of trading is XS_0.

A *trading trajectory* is a function $x(t)$ with $x(0) = X$ and $x(T) = 0$, representing the number of shares remaining to buy at time t. For a static trajectory, $x(t)$ is determined at $t = 0$, but in general $x(t)$ may be any non-anticipating random functional of B.

The *trading rate* is $v(t) = -dx/dt$, which will generally be positive as $x(t)$ decreases to zero. With a linear market impact function for simplicity, although empirical work (Almgren, Thum, Hauptmann, and Li 2005) suggests a concave function, the actual execution price is

$$\tilde{S}(t) \;=\; S(t) \;+\; \eta\, v(t)$$

where $\eta > 0$ is the coefficient of temporary market impact. Permanent market impact is also important but has no effect on the optimal trade trajectory if it is linear. (See Almgren and Chriss (2000) for a general discussion of this model.) We assume that the model parameters are known with certainty, and thus the underlying price $S(t)$ is observable based on our execution prices $\tilde{S}(t)$ and our trade rate $v(t)$.

The *implementation shortfall* C is the total cost of executing the buy program relative to the initial value:

$$\begin{aligned} C \;&=\; \int_0^T \tilde{S}(t)\, v(t)\, dt \;-\; X S_0 \\ &=\; \sigma \int_0^T x(t)\, dB(t) \;+\; \eta \int_0^T v(t)^2\, dt. \end{aligned}$$

(We have substituted the expressions above and integrated once by parts, using $B(0) = 0$ and $x(T) = 0$.) The first term represents the trading gains or losses: since we are buying, a positive price motion gives positive cost. The second term represents the market impact cost. For an adaptive strategy, both terms are random since $x(t)$ and hence $v(t)$ are random.

Mean-variance optimization solves the problem

$$\min_{x(t)} \big(E + \lambda V\big) \tag{1}$$

for each $\lambda \ge 0$, where $E = \mathbb{E}(C)$ and $V = \mathrm{Var}(C)$ are the expected value and variance of C. As λ varies, the resulting set of points $\big(V(\lambda), E(\lambda)\big)$ trace out the efficient frontier. For adaptive strategies, C is not Gaussian, but we continue to optimize mean and variance.

2.1 Static trajectories

If $x(t)$ is fixed independently of $B(t)$, then C is a Gaussian random variable with mean and variance

$$E = \eta \int_0^T v(t)^2\, dt \qquad \text{and} \qquad V = \sigma^2 \int_0^T x(t)^2\, dt.$$

The solution of (1) is then obtained as $x(t) = X\, h(t, T, \kappa)$, where the static trajectory function is

$$h(t, T, \kappa) = \frac{\sinh(\kappa(T - t))}{\sinh(\kappa T)} \qquad \text{for } 0 \le t \le T, \tag{2}$$

and the static "urgency" parameter is

$$\kappa = \sqrt{\frac{\lambda \sigma^2}{\eta}}. \tag{3}$$

The units of κ are inverse time, and $1/\kappa$ is a desired time scale for liquidation, the "half-life" of Almgren and Chriss (2000). The static trajectory is effectively an exponential $\exp(-\kappa t)$ with adjustments to reach $x = 0$ at $t = T$. For fixed λ, the optimal time scale is independent of portfolio size X since both expected costs and variance scale as X^2.

Equivalence of the static and dynamic solutions is demonstrated by observing that

$$h(t, T, \kappa) = h(s, T, \kappa)\, h(t - s, T - s, \kappa) \qquad \text{for } 0 \le s \le t \le T.$$

That is, the trajectory recomputed at time s, using the same urgency parameter, is the same as the tail of the original trajectory.

By taking $\kappa \to 0$, we recover the linear profile $x(t) = X(T-t)/T$, which is equivalent to a VWAP profile under the volume time transformation. This profile has expected cost $E_{\text{lin}} = \eta X^2/T$ and variance $V_{\text{lin}} = \sigma^2 X^2 T/3$.

2.2 Nondimensionalization

The solution and the cost will depend on five dimensional constants: the initial shares X, the time horizon T, the volatility σ, the impact coefficient η, and the risk aversion λ. To simplify the structure of the solution, it is convenient to define scaled variables.

We measure time relative to T and shares relative to X. That is, we define the nondimensional time $\hat{t} = t/T$ and nondimensional function $\hat{x}(\hat{t}) = x(T\hat{t})/X$, so that $0 \le \hat{t} \le 1$ and $\hat{x}(0) = 1$, $\hat{x}(1) = 0$. The nondimensional velocity is $\hat{v}(\hat{t}) = v(T\hat{t})/(X/T) = -d\hat{x}/d\hat{t}$.

We scale the cost by the dollar cost of a typical move due to volatility. That is, we define $\hat{C} = C/(\sigma X\sqrt{T})$, and then we have

$$\hat{C} \;=\; \int_0^1 \hat{x}(\hat{t})\, d\hat{B}(\hat{t}) \;+\; \mu \int_0^1 \hat{v}(\hat{t})^2\, d\hat{t} \tag{4}$$

where $\hat{B}(\hat{t}) = B(T\hat{t})/\sqrt{T}$ and the "market power" parameter is

$$\mu \;=\; \frac{\eta X/T}{\sigma\sqrt{T}}.$$

Here the numerator is the price concession for trading at a constant rate, and the denominator is the typical size of price motion due to volatility over the same period. The ratio μ is a nondimensional preference-free measure of portfolio size, in terms of its ability to move the market.

To estimate realistic sizes for this parameter, we recall that Almgren, Thum, Hauptmann, and Li (2005) introduced the nonlinear model $K/\sigma = \eta(X/VT)^\alpha$, where K is temporary impact (the only kind relevant here), σ is daily volatility, X is trade size, V is an average daily volume (ADV), and T is the fraction of a day over which the trade is executed. The coefficient was estimated empirically as $\eta = 0.142$, as was the exponent $\alpha = 3/5$. Therefore, a trade of 100% ADV executed across one full day gives $\mu = 0.142$. Although this is only an approximate parallel to the linear model used here, it does suggest that for realistic trade sizes, μ will be substantially smaller than one.

Problem (1) has the scaled form $\min(\hat{E} + \mu\bar{\kappa}^2\hat{V})$, where $\hat{E} = \mathbb{E}(\hat{C})$, $\hat{V} = \mathrm{Var}(\hat{C})$, and the scaled static urgency is $\bar{\kappa} = \kappa T$ with κ from (3), or

$$\bar{\kappa}^2 \;=\; \frac{\lambda\sigma^2 T^2}{\eta}.$$

The scaled risk aversion parameter $\mu\bar{\kappa}^2$ depends on X via the factor μ, though the scaled time scale $\bar{\kappa}$ is independent of X.

We use $\bar{\kappa}$ as the parameter to trace the frontier in place of λ. The result will be a trajectory $\hat{x}(\hat{t}; \bar{\kappa}, \mu)$, with scaled cost values $\hat{E}(\bar{\kappa}, \mu)$ and $\hat{V}(\bar{\kappa}, \mu)$. For each value of $\mu \ge 0$, there will be an efficient frontier obtained by tracing $\hat{E}$ and $\hat{V}$ as functions of $\bar{\kappa}$ over $0 \le \bar{\kappa} < \infty$. The linear trajectory has scaled expected cost $\hat{E}_{\mathrm{lin}} = \mu$ and variance $\hat{V}_{\mathrm{lin}} = 1/3$.

2.3 Small-portfolio limit

We now consider the limit $\mu \to 0$, with $\bar{\kappa}$ constant. Since X appears in μ but not in $\bar{\kappa}$, and all the other dimensional variables do appear in $\bar{\kappa}$, this is equivalent to taking $X \to 0$ with T, σ, η, and λ fixed. We show that for small portfolios, static strategies are optimal.

When μ is small, then (assuming that $x(t)$ and $v(t)$ have reasonable limits) the second term in (4) is small compared to the first and the variance of nondimensional cost is approximately

$$\mathrm{Var}(\hat{C}) \sim \mathrm{Var} \int_0^1 \hat{x}(\hat{t})\, d\hat{B}(\hat{t}) = \int_0^1 \mathbb{E}\Big(\hat{x}(\hat{t})^2\Big)\, d\hat{t}, \qquad \mu \to 0.$$

That is, the uncertainty in realized price comes primarily from price volatility. Even if the strategy is adapted to the price process so that $\hat{x}(\hat{t})$ is random, the market impact cost is itself a small number and the uncertainty in that number can be neglected next to volatility.

The first term in (4) has strictly zero expected value for any nonanticipating strategy (it is an Itô integral) and hence the expectation comes entirely from the second term. Thus $\mathbb{E}(\hat{C}) = \mu \mathbb{E} \int_0^1 \hat{v}(\hat{t})^2 d\hat{t}$, and the complete risk-averse cost function is approximately

$$\hat{E} + \mu \bar{\kappa}^2 \hat{V} \sim \mu \int_0^1 \mathbb{E}\Big(\hat{v}(\hat{t})^2 + \bar{\kappa}^2 \hat{x}(\hat{t})^2 \Big)\, d\hat{t}, \qquad \mu \to 0.$$

Suppose we had a candidate adaptive strategy $\hat{x}(\hat{t})$. Since the quadratic is convex, the static strategy $\overline{x}(\hat{t}) = \mathbb{E}\hat{x}(\hat{t})$ will give a lower value of the objective function (thus $x(t)$ and $v(t)$ have limits, justifying the original assumption). When μ is not small, adaptive strategies can create negative correlation between the two terms in (4), reducing the overall variance below its value for purely static trajectories.

2.4 Portfolio comparison

In its simplest form, our goal is to determine the optimal strategy $x(t)$ for any specific set of parameters. But to understand the results, it is useful to compare strategies and costs for portfolios of different sizes.

Consider two portfolios X_1 and X_2, with $X_2 = 2X_1$ and all other parameters the same including risk aversion; thus $\mu_2 = 2\mu_1$ and $\bar{\kappa}$ is the same. Portfolio X_2 will in general cost four times as much to trade as portfolio X_1. For example, static trajectories for the two portfolios will have identical shapes, and the costs will satisfy $E_2 = 4E_1$ and $V_2 = 4V_1$.

For adaptive strategies, the larger portfolio is still more expensive to trade than the smaller portfolio, but it can take more advantage of negative correlation. Thus we will have $E_2 + \lambda V_2 < 4(E_1 + \lambda V_1)$ for each λ (it is generally not true that separately $E_2 < 4E_1$ and $V_2 < 4V_1$). The ratio of adaptive cost to static cost will be less for a large portfolio than for a small portfolio, though all costs are higher for the large portfolio. The solutions presented in this paper will of most interest to large investors.

To highlight the difference in relative costs, when we draw efficient frontiers as in Figure 1, we show expectation of cost and its variance *relative* to their values for the linear trajectory. Then the static efficient frontiers for all values of $\mu > 0$ superimpose, since the costs of all static trajectories scale precisely as X^2. This common static frontier appears as the limit of the adaptive frontiers as $\mu \to 0$. As μ increases, the adaptive frontiers move down and to the left, away from the static frontier.

For convenience, we now drop the $\hat{}$. All variables are nondimensional, the time interval is $0 \le t \le 1$, and we shall use E, V to refer to the scaled variables $\hat{E}, \hat{V}$ above.

3 Single Update

In this section, we update the urgency at a single decision time T_*, with $0 < T_* < 1$ (recall that all variables are now nondimensional). On the first trading period $0 \le t \le T_*$, we use an initial urgency κ_0; that is, the trajectory is $x(t) = h(t, 1, \kappa_0)$ with h from (2). Let $X_*(\kappa_0, T_*) = h(T_*, 1, \kappa_0)$ be the remaining shares at this time. At time T_*, we switch to one of n new urgencies $\kappa_1, \ldots, \kappa_n$: with urgency κ_i, we set $x(t) = X_*(\kappa_0, T_*)\, h(t - T_*, 1 - T_*, \kappa_i)$ for $T_* \le t \le 1$.

We choose the new urgency based on the (nondimensional) realized cost up until T_*:

$$\begin{aligned} C_0 &= \int_0^{T_*} x(t)\, dB(t) \;+\; \mu \int_0^{T_*} v(t)^2\, dt \\ &= \int_0^{T_*} \big(B(t) \;+\; \mu\, v(t) \big)\, v(t)\, dt \;+\; X_*\, B(T_*). \end{aligned}$$

To measure C_0 at time T_*, we note that in the second expression above, the first term is the total dollar cost paid to acquire the shares so far, minus the value of those shares at the pre-trade price. The second term is our estimation of the additional cost that will need to be paid on the remaining shares, relative to the pre-trade price, due to price movements observed so far. As noted in Section 2, $B(t)$ is observable if we know our execution prices, our trade rate, and the coefficient of market impact.

We partition the real line into n intervals $I_1, \dots, I_n$ and use κ_j if $C_0 \in I_j$; for large n, this approaches a continuous dependence $\kappa = f(C_0)$. The intuition in the Introduction suggests that using accumulated cost should be more effective than using the instantaneous price at time T_*.

Before trading begins, we fix the decision time T_*, the interval breakpoints, and the $n+1$ urgencies $\kappa_0, \dots, \kappa_{n+1}$. But we do not know which trajectory we shall actually execute until we observe C_0 at time T_*.

We denote by C_j the cost incurred on the second part of the trajectory, if urgency κ_j is used:

$$\begin{aligned} C_j &= \int_{T_*}^1 x(t)\, dB(t) \;+\; \mu \int_{T_*}^1 v(t)^2\, dt \\ &= \int_{T_*}^1 \big(B(t) \;+\; \mu\, v(t) \big)\, v(t)\, dt \;-\; X_*\, B(T_*). \end{aligned}$$

Then the total cost is

$$C \;=\; C_0 + C_{\mathcal{I}(C_0)} \tag{5}$$

where $\mathcal{I}(C_0) = i$ if $C_0 \in I_i$. Although this total cost is not Gaussian, we still compute the optimal frontier by mean-variance optimization.

3.1 Mean and variance

As described in Section 1, we calculate means and variances at the initial time. Each variable C_i is Gaussian with mean $E_i = \mu F_i$ and variance V_i, where F_i and V_i are integrals of the form $F_i = \int v(t)^2\, dt$ and $V_i = \int x(t)^2\, dt$ which do not depend on μ (see Appendix). We define the intervals as $I_j = \{ b_{j-1} < C_0 < b_j \}$ with $b_j = E_0 + a_j \sqrt{V_0}$ and $a_0, \dots, a_n$ fixed constants with $a_0 = -\infty$, $a_n = \infty$.

To calculate mean and variance of the composite cost C, we define the fixed nondimensional quantities

$$p_j = \Phi(a_j) - \Phi(a_{j-1}) \qquad \text{and} \qquad q_j = \phi(a_{j-1}) - \phi(a_j),$$

for $j = 1, \ldots, n$, where ϕ is the standard normal density and Φ its cumulative. Thus $\mathrm{Prob}\{C_0 \in I_j\} = p_j$ and $\mathbb{E}(C_0 \,|\, C_0 \in I_j) = E_0 + (q_j/p_j)\sqrt{V_0}$.

By linearity of expectation, we readily get

$$E \;=\; \mu\big(\, F_0 \;+\; \overline{F}\,\big) \qquad \text{with} \qquad \overline{F} = \sum p_i F_i .$$

The variance is more complicated because of the dependence of the two terms in (5). We use the conditional variance formula $\mathrm{Var}(X) = \mathbb{E}(\mathrm{Var}(X|Y)) + \mathrm{Var}(\mathbb{E}(X|Y))$ to write, with $\overline{V} = \sum p_i V_i$,

$$\begin{aligned}
\mathrm{Var}(C) \;&=\; \mathbb{E}\Big(\mathrm{Var}\big(\, C_0 + C_{\mathcal{I}(C_0)} \;|\; C_0 \,\big)\Big) \;+\; \mathrm{Var}\Big(\mathbb{E}\big(\, C_0 + C_{\mathcal{I}(C_0)} \;|\; C_0 \,\big)\Big) \\
&=\; \mathbb{E}\big(V_{\mathcal{I}(C_0)}\big) \;+\; \mathrm{Var}\big(C_0 + E_{\mathcal{I}(C_0)}\big) \\
&=\; \overline{V} \;+\; \mathrm{Var}(C_0) \;+\; 2\,\mathrm{Cov}\big(C_0, E_{\mathcal{I}(C_0)}\big) \;+\; \mathrm{Var}\big(E_{\mathcal{I}(C_0)}\big).
\end{aligned}$$

By definition, $\mathrm{Var}(C_0) = V_0$, and $\mathrm{Var}\big(E_{\mathcal{I}(C_0)}\big) = \mu^2 \sum p_i \big(F_i - \overline{F}\big)^2$. Also,

$$\begin{aligned}
&\mathrm{Cov}\big(\, C_0,\, E_{\mathcal{I}(C_0)}\,\big) \;=\; \mathbb{E}\big(\, C_0\, E_{\mathcal{I}(C_0)}\,\big) \;-\; \mathbb{E}(C_0)\,\mathbb{E}\big(E_{\mathcal{I}(C_0)}\big) \\
&=\; \sum \mathrm{Prob}\{C_0 \in I_i\}\;\mathbb{E}\big(C_0 E_{\mathcal{I}(C_0)} \,|\, C_0 \in I_i\big) - \mathbb{E}(C_0)\,\mathbb{E}\big(E_{\mathcal{I}(C_0)}\big) \;=\; \mu\sqrt{V_0}\sum q_i F_i .
\end{aligned}$$

Putting all this together, we have

$$V \;=\; \mathrm{Var}(C) \;=\; V_0 + \overline{V} \;+\; 2\mu\sqrt{V_0}\sum q_i E_i \;+\; \mu^2 \sum p_i \big(E_i - \overline{E}\big)^2 .$$

The overall objective function is $U = (E + \mu\bar{\kappa}^2 V)/\mu$, or

$$\begin{aligned}
U\big(\, \kappa_0, \kappa_1, \ldots, \kappa_n, T_*;\; \bar{\kappa}, \mu \,\big) \;&=\; F_0 + \overline{F} \;+\; \bar{\kappa}^2\big(V_0 + \overline{V}\big) \\
&\quad +\; 2\mu\bar{\kappa}^2\sqrt{V_0}\sum q_i F_i \;+\; \mu^2\bar{\kappa}^2 \sum p_i \big(F_i - \overline{F}\big)^2 .
\end{aligned}$$

The $\mathcal{O}(\mu)$ term is approximately $2\sum (C_0 - E_0) p_i E_i$, and can be made negative by making E_i negatively related to C_0, corresponding to anticorrelation between second-period impact costs and first-period trading losses.

For a given market power μ and static urgency $\bar{\kappa}$, we minimize U numerically over the urgencies $\kappa_0, \kappa_1, \ldots, \kappa_n$ and the decision time T_*. As $\bar{\kappa}$ varies, the resulting set of points (V, E) traces the efficient frontier. There is a one-parameter family of efficient frontiers, depending on μ (Figure 1); the static trajectories appear as the limit $\mu = 0$.

3.2 Numerical results

Figure 1 shows the complete set of efficient frontiers for the single-update problem. Each curve is computed by varying the static urgency parameter $\bar{\kappa}$ from 0 to ∞, for a fixed value of μ. The solution for each pair $(\bar{\kappa}, \mu)$ is computed using a fixed set of 32 equal-probability breakpoints. As described in Section 2.3, we plot E and V relative to their values for the linear trajectories, to clearly see the improvement due to adaptivity.

We use these frontiers to obtain cost distributions for adaptive strategies that are better than the cost distributions for any static strategy. In Figure 1, the point labeled "$\kappa = 8$" describes a particular static trajectory computed with parameter $\bar{\kappa} = 8$, giving a normal cost distribution. For a portfolio with $\mu = 0.1$, this distribution has expectation $E \approx 4 \times E_{\text{lin}} \approx 4 \times \mu = 0.4$ and variance $V \approx 0.2 \times V_{\text{lin}} = 0.2/3 = 0.067$. The inset shows this distribution as a black dashed line.

The pink shaded wedge in Figure 1 shows the set of values of (V, E) accessible to an adaptive strategy with $\mu = 0.1$, that are strictly preferable to the static strategy since they have lower expected cost and/or variance. On the efficient frontier for $\mu = 0.1$, these solutions are obtained by computing adaptive solutions with parameters approximately in the range $4.9 \le \bar{\kappa} \le 7.1$. There is no need to use the same value of $\bar{\kappa}$ for the adaptive strategy as for the static strategy to which it is compared.

The inset shows the cost distributions associated with these adaptive strategies. For $\bar{\kappa} = 4.9$, the adaptive distribution has lower expected cost than the static distribution, with the same variance. For $\bar{\kappa} = 7.1$, the adaptive distribution has lower variance than the static distribution, with the same mean. These distributions are the extreme points of a one-parameter family of distributions, each of which is strictly preferable to the given static strategy, regardless of the trader's risk preferences. For example, the adaptive solution for $\bar{\kappa} = 6$ has both lower expected cost and lower variance than the static solution.

These cost distributions are strongly skewed toward positive costs, suggesting that mean-variance optimization may not give the best possible solutions. Nonetheless, it is clear that these adaptive distributions are strictly preferable to the reference static strategy, since they have lower probability of high costs and higher probability of low costs.

Figure 2 shows the adaptive trading strategy for $\mu = 0.1$ and $\bar{\kappa} = 6$. The dashed line is the static optimal trajectory with urgency $\bar{\kappa} = 8$, com-

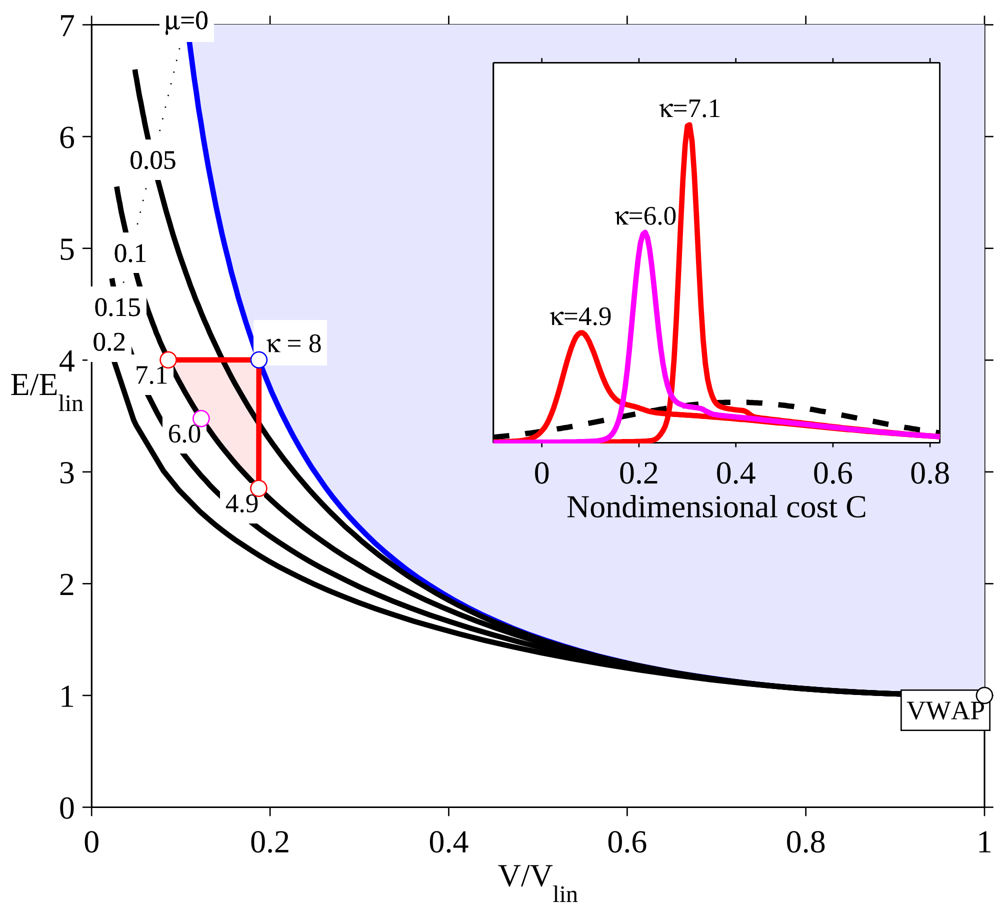

Figure 1: Adaptive efficient frontiers for different values of market power μ. The expectation of trading cost $E = \mathbb{E}(C)$ and its variance $V = \mathrm{Var}(C)$ are normalized by their values for a linear trajectory (VWAP), as described in Section 2.3. The blue shaded region is the set of values accessible to a static trajectory and the blue curve is the static frontier, which is also the limit $\mu \to 0$ with fixed static urgency $\bar{\kappa}$. The black curves are the improved values accessible to adaptive strategies; the improvement is greater for larger portfolios. The inset shows the actual distributions corresponding to the indicated points.

pared to which this adaptive strategy delivers both lower expectation of cost and lower variance. The adaptive strategy initially trades more slowly than the optimal static trajectory. At T_*, if prices have moved in the trader's favor, then the strategy accelerates, spending the investment gains on impact costs. If prices have moved against the trader, corresponding to positive values of C_0, then the strategy decelerates to save impact costs in the remaining period. The values of κ become very large when C_0 is large negative, corresponding to the instruction: "if you have gains in the first part of trading, then finish the program immediately."

4 Continuous Response

We now illustrate a simple form of *continuous response* to trading gains or losses. In general, we may specify any rule giving the trade rate $v(t)$ as a function of the price history $B(s)$ for $0 \le s \le t$. Rather than adjusting the rate $v(t)$ directly, it is more convenient to adjust the urgency $\kappa(t)$. From (2), we differentiate $x(s) = x(t)\, h(s-t, 1-t, \kappa)$ with respect to s and evaluate at $s = t$, obtaining the relationship between v and κ

$$v(t) \;=\; x(t)\,\kappa(t)\,\coth\bigl(\kappa(t)(1-t)\bigr). \tag{6}$$

For all choices of $\kappa(t)$ the trajectories hit $x = 0$ at $t = 1$.

Determining the full optimal dependence of $\kappa(t)$ on $B(s)$ for $0 \le s \le t$ is difficult (see Section 5). Here we consider only the relationship

$$\kappa(t) \;=\; a\,\exp\bigl(\,b\,B(t)\,\bigr)$$

in which the instantaneous urgency depends on the instantaneous price level. Other functional relationships for $\kappa(t)$ in terms of $B(t)$ are possible as well. Here, $\kappa(t)$ is always positive, and is monotone in $B(t)$.

From (6), we readily obtain $x(t)$ and finally the shortfall C by integration as in (4). However, because of the highly non-linear dependence of $\kappa(t)$, and thus $v(t)$ and $x(t)$, on the Brownian motion $B(t)$, analytic evaluation of this stochastic integral is beyond reach.

4.1 Numerical results

For numerical solutions, we generate a fixed collection of sample paths using a Brownian bridge construction with quasi-random variates. For

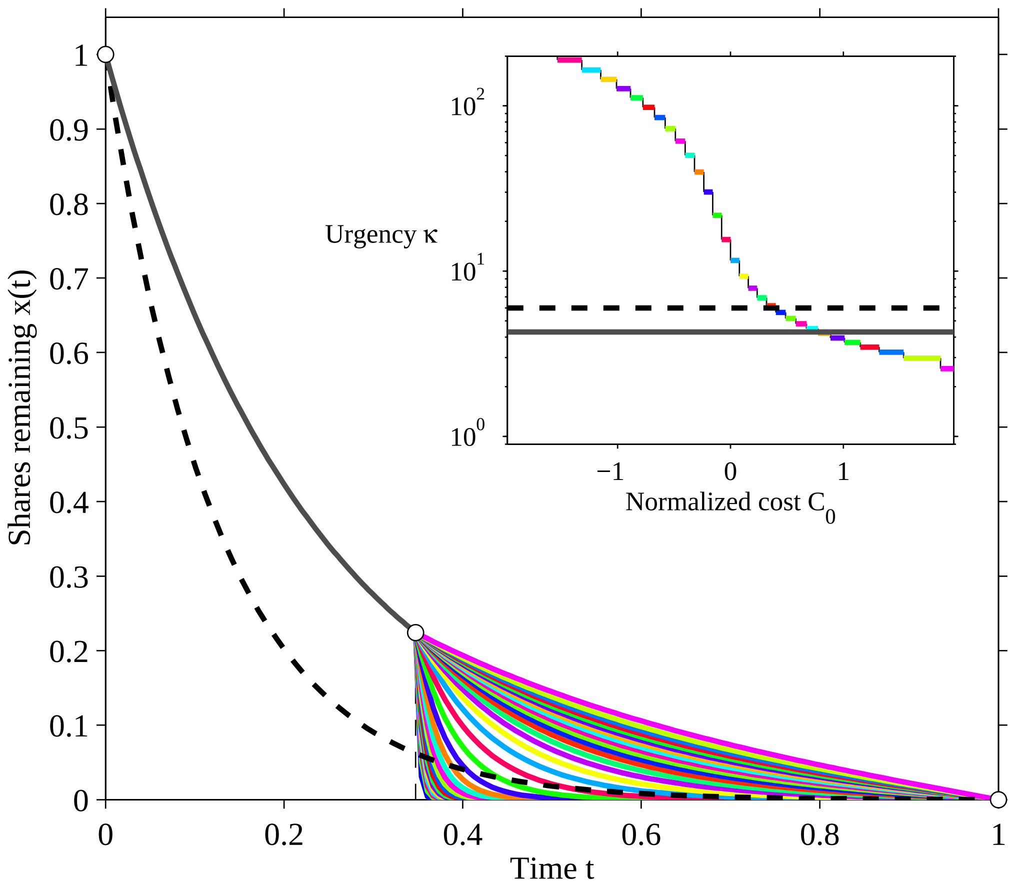

Figure 2: Adaptive trading trajectories for market power $\mu = 0.1$, matching the points on the frontiers in Figure 1. The dashed line is the static optimal trajectory with urgency $\bar{\kappa} = 8$; the adaptive strategy has $\bar{\kappa} = 6$ and 32 equal-probability paths. This adaptive strategy delivers both lower expectation of cost and lower variance than the static strategy. The inset shows the dependence of the new urgency on the initial trading cost C_0, normalized by the *ex ante* expectation and standard deviation of C_0.

any candidate values of a and b, we evaluate the stochastic integrals numerically and evaluate the sample mean E and variance V. We then minimize the objective function $E + \bar{\kappa}^2\mu^2 V$ numerically over a and b.

By solving for a series of values of $0 < \bar{\kappa} < \infty$ we can again trace the efficient frontiers for different values of μ, yielding similar results as in the single update framework in Section 3.

Again, the optimal strategies are "aggressive in the money," having $b < 0$. When the stock price goes down, we incur unexpectedly smaller shortfall and react with increasing urgency $\kappa(t)$, whereas for rising stock prices we slow down trading. Figure 3 illustrates this behaviour for two sample paths of the stock price.

5 Discussion and Conclusions

The simple update rules presented in Sections 3 and 4 demonstrate that price adaptive scaling strategies can lead to significant improvements over static trade schedules, and they illustrate the importance of the new "market power" parameter μ. However, neither of these rules is the fully optimal adaptive execution strategy. A fully optimal adaptive strategy would use stochastic dynamic programming to determine the trading rate as a general function of the continuous state variables such as number of shares remaining, time remaining, current stock price, and trading gains or losses experienced to date.

One subtlety is that the mean-variance criterion cannot be used directly in this context: it involves the square of an expectation, which is not amenable to dynamic programming techniques. However, Li and Ng (2000) have shown how to embed mean-variance optimization into a family of optimizations using a quadratic utility function. The mean-variance solution is recovered as one element of this family. The need to solve this family of problems is an addition degree of complication.

The calculation uses the tools of stochastic optimal control and requires numerical solution of a highly nonlinear Hamilton-Jacobi-Bellman partial differential equation. Proper formulation of this problem, and solution of the resulting equations, is ongoing work of the authors. The examples presented here show that even with very simple adaptive strategies, substantial improvement is possible over static strategies.

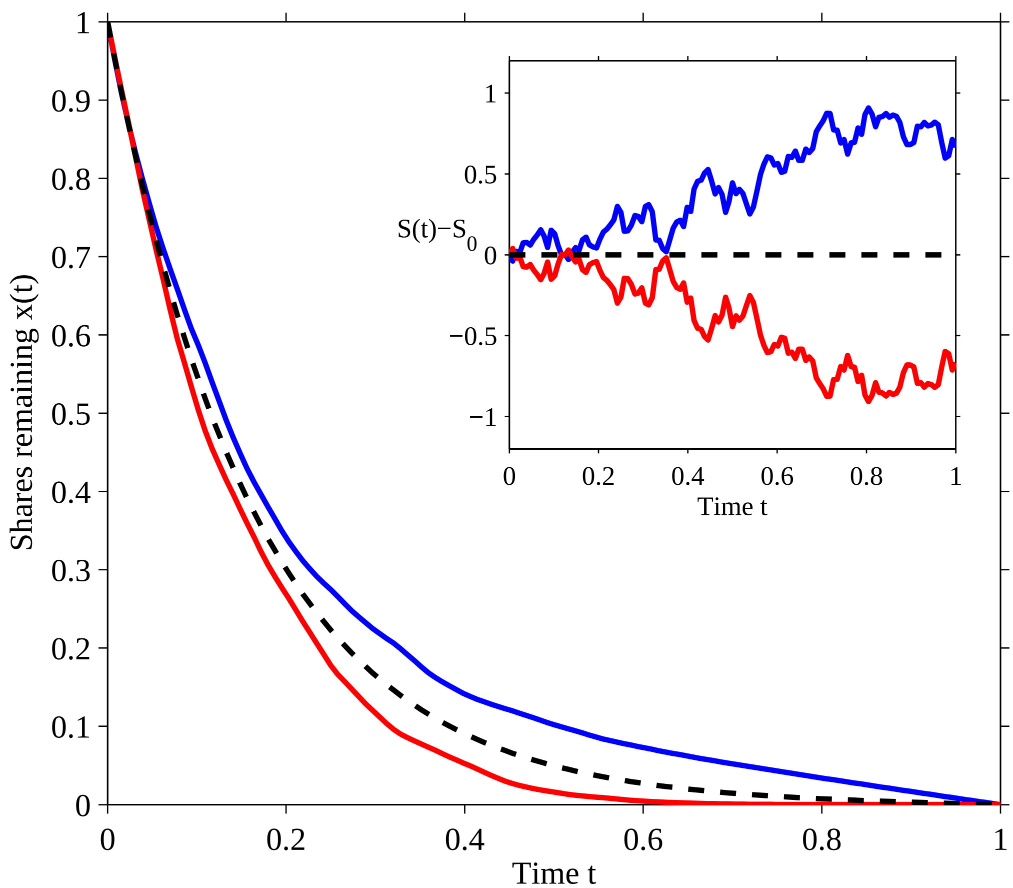

Figure 3: Optimal trading trajectories using the adaptation rule $\kappa(t) = a\exp(b\,B(t))$ with $a = 5.9$ and $b = -1.7$, for static urgency $\bar{\kappa} = 6$. As the stock price goes down (lower, red curve), trading is accelerated compared to the optimal static trajectory (dashed line), whereas for rising stock price it is slowed down.

A Detailed formulas

Here we present the detailed calculations described in Section 3.1.

A.1 Means and variances

The integrals are readily determined to be

$$
\begin{aligned}
F_0 &= \frac{\kappa_0\big(\sinh(2\kappa_0) - \sinh(2\kappa_0(1-T_*)) + 2\kappa_0 T_*\big)}{4\,\sinh^2(\kappa_0)} \\
V_0 &= \frac{\sinh(2\kappa_0) - \sinh(2\kappa_0(1-T_*)) - 2\kappa_0 T_*}{4\kappa_0\,\sinh^2(\kappa_0)}
\end{aligned}
$$

and

$$
\begin{aligned}
F_i &= \frac{\sinh^2(\kappa_0(1-T_*))}{\sinh^2(\kappa_i(1-T_*))}\,\frac{\kappa_i\big(\sinh(2\kappa_i(1-T_*)) + 2\kappa_i(1-T_*)\big)}{4\,\sinh^2(\kappa_0)} \\
V_i &= \frac{\sinh^2(\kappa_0(1-T_*))}{\sinh^2(\kappa_i(1-T_*))}\,\frac{\sinh(2\kappa_i(1-T_*)) - 2\kappa_i(1-T_*)}{4\kappa_i\,\sinh^2(\kappa_0)}
\end{aligned}
$$

for $i = 1, \ldots, n$.

A.2 Full distribution

Each C_i is Gaussian with mean E_i and variance V_i, so its density is

$$
f_i(C_i) = \frac{1}{\sqrt{2\pi V_i}} \exp\left(-\frac{(C_i - E_i)^2}{2V_i}\right), \qquad i = 0, 1, \ldots, n.
$$

The composite variable is $C = C_0 + C_i$ for $C_0 \in I_i$ where $I_i = (b_{i-1}, b_i)$ and $b_i = E_0 + a_i\sqrt{V_0}$ with $a_1, \ldots, a_{n-1}$ fixed constants. Then

$$
\begin{aligned}
f(c)\,dc &\equiv \mathrm{Prob}\{\, C \in [c, c + dc)\,\} \\
&= \sum_{i=1}^{n} \mathrm{Prob}\{\, C_0 \in I_i \text{ and } C_i \in [c - C_0, c - C_0 + dc)\,\}
\end{aligned}
$$

so

$$
\begin{aligned}
f(C) &= \sum_{i=1}^{n} \int_{I_i} f(C_0)\, f_i(C - C_0)\, dC_0 \\
&= \sum_{i=1}^{n} \frac{1}{2\pi\sqrt{V_0 V_i}} \int_{b_{i-1}}^{b_i} \exp\!\left(-\frac{(C_0 - E_0)^2}{2V_0} - \frac{(C - C_0 - E_i)^2}{2V_i} \right) dC_0 \\
&= \sum_{i=1}^{n} \frac{1}{2\pi\sqrt{V_0 V_i}}\ \exp\!\left(-\frac{1}{2}\left[\frac{E_0^2}{V_0} + \frac{(C - E_i)^2}{V_i} - \frac{\big(E_0 V_i + (C - E_i) V_0\big)^2}{V_0 V_i (V_0 + V_i)} \right] \right) \\
&\qquad\qquad \times \int_{b_{i-1}}^{b_i} \exp\!\left(-\frac{1}{2}\left[\frac{V_0 + V_i}{V_0 V_i} \left(C_0 - \frac{E_0 V_i + (C - E_i) V_0}{V_0 + V_i} \right)^2 \right] \right) dC_0 \\
&= \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi (V_0 + V_i)}}\ \exp\!\left(-\frac{\big(C - E_0 - E_i\big)^2}{2\,(V_0 + V_i)} \right) \\
&\qquad\qquad \times \left[\Phi\!\left(\frac{(C - E_i - b_{i-1}) V_0 + (E_0 - b_{i-1}) V_i}{\sqrt{V_0 V_i (V_0 + V_i)}} \right) \right. \\
&\qquad\qquad\qquad \left. -\ \Phi\!\left(\frac{(C - E_i - b_i) V_0 + (E_0 - b_i) V_i}{\sqrt{V_0 V_i (V_0 + V_i)}} \right) \right]
\end{aligned}
$$

References

Almgren, R. and N. Chriss (1999). Value under liquidation. *Risk 12*(12), 61–63. 3

Almgren, R. and N. Chriss (2000). Optimal execution of portfolio transactions. *J. Risk 3*(2), 5–39. 3, 4, 9, 10

Almgren, R., C. Thum, E. Hauptmann, and H. Li (2005). Equity market impact. *Risk 18*(7, July), 57–62. 9, 11

Huberman, G. and W. Stanzl (2005). Optimal liquidity trading. *Review of Finance 9*(2), 165–200. 4

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–291. 7

Kissell, R. and R. Malamut (2006). Algorithmic decision-making framework. *J. Trading 1*(1), 12–21. 7, 8

Li, D. and W.-L. Ng (2000). Optimal dynamic portfolio selection: Multiperiod mean-variance formulation. *Math. Finance 10*(3), 387–406. 20

Perold, A. F. (1988). The implementation shortfall: Paper versus reality. *J. Portfolio Management 14*(3), 4–9. 3

Shefrin, H. and M. Statman (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *J. Finance 40*(3), 777–790. 7