An Arrow–Pratt Foundation for Almgren–Chriss Execution

Local Utility, Exact Variance Kernels, and Wealth-Scaled Risk Aversion

Grant Stenger

December 2024

Abstract

Classical optimal-execution models usually begin with implementation shortfall and penalize its variance by an exogenous risk-aversion coefficient. This paper starts instead from utility over terminal wealth. For any sufficiently smooth increasing, concave utility, a local certainty-equivalent expansion gives the money-metric objective

$$J_U(Q) = M(Q) - \frac{1}{2}A(M(Q))V(Q), \qquad A(w) = -\frac{U''(w)}{U'(w)}.$$

Thus the variance penalty in execution is an Arrow–Pratt absolute risk-aversion term evaluated at expected terminal wealth.

This perspective separates utility classes that are often conflated in mean–variance execution. CARA utility recovers a wealth-independent Almgren–Chriss risk weight, while CRRA and log utility imply wealth-scaled risk aversion. Under deterministic open-loop execution with geometric Brownian motion and market impact, exact variance is nonlocal; with utility curvature frozen at a benchmark wealth level, the resulting exact-kernel problem is a strictly concave Hilbert-space optimization problem with a unique maximizer. The familiar Almgren–Chriss hyperbolic-sine schedule is then recovered as a frozen-risk benchmark, with the exogenous risk coefficient replaced by a calibrated Arrow–Pratt quantity.

The result is not a full adaptive stochastic-control solution. It is a local utility foundation for execution-risk weights, an exact-kernel open-loop formulation, and a practical calibration map connecting portfolio wealth, utility curvature, and Almgren–Chriss execution.

Contents

1 Introduction

Optimal execution asks how a trader should build or liquidate a position over a finite horizon when trading moves prices and unexecuted inventory remains exposed to market risk. The classical Almgren–Chriss framework answers this question by minimizing expected implementation shortfall plus a variance penalty. In its simplest form, the trader chooses an inventory path Q to optimize an objective of the form

$$\text{expected cost} + \lambda \, \text{variance of cost}.$$

This is powerful and tractable, but the coefficient λ is an external input. It is a knob that represents risk aversion without deriving that risk aversion from preferences over terminal wealth.

That missing link matters in practice. Two traders can face the same asset, volatility, impact model, and order size, yet rationally choose different execution schedules if the order represents a different fraction of their wealth. Under the CRRA calibration derived below,

$$\lambda_{\mathrm{CRRA}}(M_b) = \frac{\gamma \sigma^2 \bar{S}_0^2}{2M_b},$$

a trader with a 10 million wealth base uses one tenth the execution-risk weight of an otherwise identical trader with a 1 million wealth base. This is the paper's practical bite: utility turns the risk-aversion knob into a calibrated wealth-scale object.

The purpose of this paper is to rebuild that knob from a utility-theoretic primitive. Let U be a strictly increasing, strictly concave utility function over terminal wealth, and define Arrow–Pratt absolute risk aversion by

$$A(w) := -\frac{U''(w)}{U'(w)}.$$

If terminal wealth generated by an execution path Q has mean $M(Q)$ and variance $V(Q)$, then a local certainty-equivalent expansion gives

$$J_U(Q) = M(Q) - \frac{1}{2} A(M(Q)) V(Q).$$

This is the mathematical bridge from terminal-wealth utility to Almgren–Chriss-style execution. The bridge is local, but it is economically sharp: local execution risk weights are Arrow–Pratt utility-curvature objects.

The canonical utility classes then become a taxonomy of execution-risk scaling. CARA utility has $A(w) = a$, so it recovers a wealth-independent mean–variance penalty. CRRA utility has $A(w) = \gamma/w$, so it gives

$$J_{\mathrm{CRRA}}(Q) = M(Q) - \frac{\gamma}{2} \frac{V(Q)}{M(Q)}.$$

A trader with CRRA preferences does not dislike dollar variance in isolation. The trader dislikes variance relative to wealth. A dollar of execution risk has a different meaning for a trader with wealth 5 million than it does for a trader with wealth 500 million. Classical Almgren–Chriss can represent this only by manually changing λ. CRRA preferences make the wealth scaling endogenous. In the frozen-risk benchmark, this taxonomy becomes a separation theorem: wealth-independent execution-risk weights characterize CARA, while wealth-scaled execution-risk weights characterize CRRA and log utility.

The exact utility-first problem is natural:

$$\sup_Q \mathbb{E}\left[U(W_T^Q) \right],$$

where Q is the inventory path and W_T^Q is terminal wealth generated by that path. With geometric Brownian motion, market impact, and possible feedback controls, this exact problem is a serious stochastic-control problem. It is not the problem for which this paper claims a closed-form solution. Throughout the analytical sections, Q is deterministic and chosen at time zero, so the first-order conditions are open-loop pre-commitment conditions rather than feedback policies.

Relation to adjacent literature. Arrow and Pratt introduced the local curvature measures that connect utility and risk premia [3, 10]. Almgren and Chriss can also be given a utility interpretation along a mean–variance efficient frontier [2]. The distinction here is not that Almgren–Chriss lacks any utility reading, but that we derive the local execution-risk weight directly from terminal-wealth curvature. Closest to the present utility-first perspective are Schied and Schöneborn [11], who study risk aversion and optimal liquidation dynamics in illiquid markets, and Schied, Schöneborn, and Tehranchi [12], who show that optimal basket liquidation for CARA investors is deterministic in a broad class of models. Our contribution is different: we isolate the local certainty-equivalent channel through which general Arrow–Pratt curvature, and CRRA curvature in particular, produces Almgren–Chriss execution-risk weights. The exact GBM variance formula also connects to nonlocal execution problems of the kind arising under transient impact and Fredholm-integral formulations [7]. For broader optimal-execution and market-impact context, see Almgren [1], Bertsimas and Lo [4], Obizhaeva and Wang [9], and Cartea and Jaimungal [5]. The variational arguments below use standard one-dimensional Euler–Lagrange and direct-method tools [6]; the adaptive utility-control direction is conceptually closer to the continuous-time dynamic-programming tradition exemplified by Merton [8].

Contribution. The contribution is conceptual and asymptotic rather than an exact solution of the full stochastic-control problem. The local certainty-equivalent expansion is classical; the contribution is to apply it to execution-generated terminal wealth and then follow the consequences for execution-risk calibration. First, the paper derives the local execution objective J_U. Second, it specializes this objective to CRRA, log, and CARA utility, separating wealth-relative and wealth-independent execution risk. Third, it proves the CARA–CRRA separation result: wealth-independent execution-risk weights characterize CARA, while wealth-scaled execution-risk weights characterize CRRA and log utility. Fourth, it shows that the exact deterministic open-loop utility problem is concave on convex fixed-endpoint domains, so the exact problem is hard but not structurally pathological. Fifth, it turns the expansion into the practical Almgren–Chriss calibration map

$$\lambda_U(M_b) = \frac{1}{2}A(M_b)\sigma^2\bar{S}_0^2.$$

Finally, for deterministic open-loop execution with temporary and permanent linear impact, it derives the mean wealth formula, the exact nonlocal GBM variance kernel, and a frozen-curvature exact-kernel theorem: retaining the full GBM covariance kernel gives a unique nonlocal Fredholm-type optimizer. The local martingale-risk closure and the frozen-coefficient benchmark then recover the familiar Almgren–Chriss hyperbolic-sine schedule with this utility-implied execution-risk weight.

What is and is not being claimed. The paper does not claim that exact utility maximization under geometric Brownian motion and market impact admits a universal closed-form execution rule. The closed-form hyperbolic-sine schedule appears only after additional benchmark approximations. The durable message is instead:

local execution risk weights are Arrow–Pratt curvature terms.

Reader map. Four qualifiers organize the paper. “Local” refers to the certainty-equivalent expansion around expected terminal wealth. “Exact” means that no diagonal or frozen-price approximation has been made to the GBM covariance kernel. “Frozen” means that the Arrow–Pratt curvature, or the normalization terms generated by it, are held at a benchmark wealth level. “Benchmark” means the closed-form Almgren–Chriss comparison problem obtained after those frozen-coefficient approximations. The reader should keep three objects separate: the exact open-loop expected-utility problem, the exact-kernel local utility problem with frozen curvature, and the closed-form frozen-risk Almgren–Chriss benchmark.

2 A General Arrow–Pratt Foundation

The local Arrow–Pratt expansion itself is classical. The contribution here is to apply that expansion to terminal wealth generated by an execution path and then read the resulting variance penalty as an Almgren–Chriss execution-risk coefficient.

2.1 The Local Risk-Premium Expansion

Let $I \subseteq \mathbb{R}$ be an open interval and let $U \in C^3(I)$ satisfy $U' > 0$ and $U'' < 0$. Define the Arrow–Pratt coefficient of absolute risk aversion by

$$A(w) := -\frac{U''(w)}{U'(w)}. \tag{2.1}$$

For any execution path Q with terminal wealth W_T^Q, write

$$M(Q) := \mathbb{E}[W_T^Q], \qquad V(Q) := \mathrm{Var}(W_T^Q).$$

The certainty equivalent $\mathrm{CE}_U(Q)$ is defined locally by

$$U(\mathrm{CE}_U(Q)) = \mathbb{E}[U(W_T^Q)]. \tag{2.2}$$

Proposition 1 (Local Arrow–Pratt certainty equivalent). *Fix Q with $M := M(Q) \in I$, and write*

$$W_T^Q = M + Y, \qquad \mathbb{E}[Y] = 0, \qquad \mathbb{E}[Y^2] = V(Q).$$

Assume $M + Y$ remains in a compact subinterval $K \Subset I$. Then there exist constants $C_1, C_2 < \infty$, depending only on U and K, such that

$$\left|\mathrm{CE}_U(Q) - M + \frac{1}{2}A(M)V(Q)\right| \leq C_1\mathbb{E}[|Y|^3] + C_2V(Q)^2. \tag{2.3}$$

Equivalently, as terminal-wealth risk becomes locally small,

$$\mathrm{CE}_U(Q) = M(Q) - \frac{1}{2}A(M(Q))V(Q) + O\big(\mathbb{E}[|Y|^3] + V(Q)^2\big). \tag{2.4}$$

Proof. Taylor expand $U(M + Y)$ around M:

$$U(M+Y) = U(M) + U'(M)Y + \frac{1}{2}U''(M)Y^2 + R_3(Y),$$

with $|R_3(Y)| \leq C|Y|^3$ on K. Taking expectations and using $\mathbb{E}[Y] = 0$ gives

$$\mathbb{E}[U(W_T^Q)] = U(M) + \frac{1}{2}U''(M)V(Q) + R, \qquad |R| \leq C\mathbb{E}[|Y|^3].$$

Let $\Delta := \mathbb{E}[U(W_T^Q)] - U(M)$ and write $\mathrm{CE}_U(Q) = M + h$. Since $U' > 0$ on K, the inverse-function expansion around $U(M)$ gives

$$h = \frac{\Delta}{U'(M)} + \widetilde{R}, \qquad |\widetilde{R}| \leq C'\Delta^2,$$

after shrinking to a compact neighborhood if necessary. Substituting the expression for Δ yields

$$h = \frac{1}{2}\frac{U''(M)}{U'(M)}V(Q) + O(\mathbb{E}[|Y|^3] + V(Q)^2) = -\frac{1}{2}A(M)V(Q) + O(\mathbb{E}[|Y|^3] + V(Q)^2).$$

This proves both (2.3) and (2.4). □

Definition 1 (Local utility execution objective). The local utility-based execution objective associated with U is

$$J_U(Q) := M(Q) - \frac{1}{2}A(M(Q))V(Q). \tag{2.5}$$

Remark 1 (Relative-risk form). If $R(w) := wA(w)$ is relative risk aversion, then

$$\frac{1}{2}A(M(Q))V(Q) = \frac{1}{2}R(M(Q))\frac{V(Q)}{M(Q)}.$$

Thus the local penalty can be read either as absolute risk aversion times dollar variance or as relative risk aversion times variance scaled by the wealth base.

2.2 Canonical Utility Classes

Corollary 1 (CRRA, log, and CARA)**.** *For CRRA utility*

$$U(w) = \frac{w^{1-\gamma} - 1}{1-\gamma}, \qquad \gamma > 0, \quad \gamma \neq 1,$$

we have $A(w) = \gamma/w$*, and therefore*

$$J_{\mathrm{CRRA}}(Q) = M(Q) - \frac{\gamma}{2}\frac{V(Q)}{M(Q)}. \tag{2.6}$$

For log utility, $A(w) = 1/w$*, so*

$$J_{\log}(Q) = M(Q) - \frac{1}{2}\frac{V(Q)}{M(Q)}.$$

For CARA utility with absolute risk aversion $a > 0$*,* $A(w) = a$*, so*

$$J_{\mathrm{CARA}}(Q) = M(Q) - \frac{a}{2}V(Q).$$

Remark 2 (Why the coefficient is an Arrow–Pratt coefficient)*.* The money-metric CRRA penalty is $\gamma V/(2M)$. The coefficient is not $\gamma(\gamma-1)/2$. The latter may appear in other expansions of power utility, but after converting the local utility approximation into a certainty equivalent, the variance penalty is governed by the Arrow–Pratt curvature scale $A(M) = \gamma/M$.

3 Model

3.1 Unaffected Price

All prices are expressed in units of the money-market account. Under the physical probability measure relevant for utility maximization, the unaffected price $\bar{S}_t$ follows geometric Brownian motion

$$\mathrm{d}\bar{S}_t = \mu\bar{S}_t \,\mathrm{d}t + \sigma\bar{S}_t \,\mathrm{d}B_t, \qquad \bar{S}_0 > 0, \tag{3.1}$$

where $\mu \in \mathbb{R}$ is the excess drift, $\sigma \geq 0$ is volatility, and B_t is a Brownian motion.

3.2 Inventory and Execution Prices

The trader chooses a deterministic open-loop inventory path Q on $[0, T]$. The trading rate is

$$v_t := \dot{Q}_t,$$

with the sign convention that $v_t > 0$ buys and $v_t < 0$ sells. The endpoints are fixed:

$$Q(0) = Q_0, \qquad Q(T) = Q_T. \tag{3.2}$$

The classical liquidation case is $Q_0 > 0$, $Q_T = 0$, but the notation also covers accumulation, partial liquidation, and short covering.

Following the linear Almgren–Chriss convention, the impacted midprice and execution price are

$$S_t^{\mathrm{mid}} = \bar{S}_t + \theta(Q_t - Q_0), \qquad S_t^{\mathrm{exec}} = S_t^{\mathrm{mid}} + \eta v_t, \tag{3.3}$$

where $\eta > 0$ is temporary impact and $\theta \geq 0$ is permanent impact. The temporary component affects only the execution price. The permanent component shifts the midprice by cumulative order flow.

Remark 3 (Permanent impact with fixed endpoints). Because Q_T is fixed, the permanent-impact contribution to expected wealth will be a boundary term. In this model it does not create a direct pathwise timing cost. It may still affect the local utility solution indirectly by changing the wealth level at which Arrow–Pratt curvature is evaluated.

3.3 Admissible Paths

For the variational derivations, take

$$\mathcal{A} := \{Q \in C^2([0,T]) : Q(0) = Q_0,\ Q(T) = Q_T\}.$$

The formulas extend in the usual weak sense to Sobolev paths with square integrable trading rate. The C^2 convention is only to keep the Euler–Lagrange calculations uncluttered.

Assumption 1 (Open-loop deterministic execution)**.** The inventory path Q is deterministic and chosen at time 0. It is not adapted to the realized price filtration. All first-order conditions in this paper are therefore pre-commitment conditions.

Assumption 2 (Fixed endpoints and cumulative-flow permanent impact)**.** The endpoint Q_T is fixed, and permanent impact is the linear cumulative-flow term in (3.3). The boundary-term conclusion for permanent impact relies on these assumptions; it need not survive transient impact, nonlinear permanent impact, stochastic fills, or free terminal inventory.

4 Wealth Accounting

Let C_t be cash and define marked-to-mid wealth

$$W_t = C_t + Q_t S_t^{\mathrm{mid}}.$$

Self-financing cash dynamics are

$$\mathrm{d}C_t = -S_t^{\mathrm{exec}} v_t \ \mathrm{d}t.$$

Since Q has finite variation, $\mathrm{d}Q_t = v_t \ \mathrm{d}t$, and

$$\mathrm{d}S_t^{\mathrm{mid}} = \mathrm{d}\bar{S}_t + \theta v_t \ \mathrm{d}t.$$

Therefore

$$\begin{aligned} \mathrm{d}W_t &= \mathrm{d}C_t + Q_t \ \mathrm{d}S_t^{\mathrm{mid}} + S_t^{\mathrm{mid}} \ \mathrm{d}Q_t \\ &= -\eta v_t^2 \ \mathrm{d}t + Q_t \ \mathrm{d}\bar{S}_t + \theta Q_t v_t \ \mathrm{d}t. \end{aligned} \tag{4.1}$$

Integrating over $[0, T]$ yields

$$W_T^Q = W_0 - \eta \int_0^T \dot{Q}_t^2 \ \mathrm{d}t + \theta \int_0^T Q_t \dot{Q}_t \ \mathrm{d}t + \int_0^T Q_t \ \mathrm{d}\bar{S}_t. \tag{4.2}$$

The permanent-impact integral is path independent:

$$\int_0^T Q_t \dot{Q}_t \ \mathrm{d}t = \frac{1}{2}(Q_T^2 - Q_0^2). \tag{4.3}$$

Proposition 2 (Mean terminal wealth)**.** *For deterministic $Q \in \mathcal{A}$,*

$$M(Q) := \mathbb{E}[W_T^Q] = W_0 - \eta \int_0^T \dot{Q}_t^2 \ \mathrm{d}t + \mu \bar{S}_0 \int_0^T e^{\mu t} Q_t \ \mathrm{d}t + \frac{\theta}{2}(Q_T^2 - Q_0^2). \tag{4.4}$$

Proof. Take expectations in (4.2). Since $\mathbb{E}[\,\mathrm{d}\bar{S}_t] = \mu \bar{S}_0 e^{\mu t}\,\mathrm{d}t$ and Q is deterministic,

$$\mathbb{E}\left[\int_0^T Q_t \ \mathrm{d}\bar{S}_t\right] = \mu \bar{S}_0 \int_0^T e^{\mu t} Q_t \ \mathrm{d}t.$$

Use (4.3) for the permanent-impact term. □

Lemma 1 (Variation of mean terminal wealth)**.** *For $Q \in \mathcal{A}$ and a fixed-endpoint perturbation h with $h(0) = h(T) = 0$,*

$$\delta M_Q[h] = \int_0^T \Big(2\eta\,\ddot{Q}_t + \mu\bar{S}_0 e^{\mu t}\Big)\,h_t\;\mathrm{d}t. \tag{4.5}$$

Proof. Differentiate (4.4) in the direction h:

$$\delta M_Q[h] = -2\eta \int_0^T \dot{Q}_t \dot{h}_t\;\mathrm{d}t + \mu\bar{S}_0 \int_0^T e^{\mu t} h_t\;\mathrm{d}t.$$

Integrating the first term by parts and using $h(0) = h(T) = 0$ gives (4.5). □

5 Exact Utility Problem and Local Domains

5.1 The Exact Utility Problem

Fix a utility function $U : I \to \mathbb{R}$ satisfying the assumptions of Section 2. The exact open-loop utility problem is defined only on paths that keep terminal wealth in the utility domain. Set

$$\mathcal{A}_U := \Big\{ Q \in \mathcal{A} : W_T^Q \in I \text{ a.s. and } \mathbb{E}\Big[|U(W_T^Q)|\Big] < \infty \Big\}. \tag{5.1}$$

The exact open-loop problem is

$$\sup_{Q \in \mathcal{A}_U} \mathbb{E}\Big[U(W_T^Q)\Big]. \tag{5.2}$$

This is the primitive economic object. It is also generally hard: under geometric Brownian motion, terminal wealth is a nonlinear functional of the price path, and exact feedback control would require a state-dependent dynamic program. Nonemptiness of $\mathcal{A}_U$ is parameter-dependent; this paper does not attempt to characterize it.

Proposition 3 (Concavity of exact open-loop utility)**.** *Let $\mathcal{D} \subseteq \mathcal{A}_U$ be a convex set of deterministic fixed-endpoint paths. Thus every convex combination of two paths in $\mathcal{D}$ remains admissible for the exact utility objective. Then*

$$Q \mapsto \mathbb{E}[U(W_T^Q)]$$

is concave on $\mathcal{D}$. If $\eta > 0$, U is strictly increasing, and $Q^1, Q^2 \in \mathcal{D}$ are distinct paths with the same endpoints, then the objective is strictly concave along the segment joining Q^1 and Q^2. Consequently, any interior stationary point of the exact open-loop problem on such a convex domain is the unique global maximizer on that domain.

Proof. Fix $Q^1, Q^2 \in \mathcal{D}$ and $\alpha \in (0,1)$. Let $Q^\alpha := \alpha Q^1 + (1-\alpha)Q^2$. From (4.2), the stochastic price-exposure term is linear in Q, and the permanent-impact term is fixed by the common endpoints. The temporary-impact term is concave because

$$-\eta \int_0^T \Big(\alpha\dot{Q}_t^1 + (1-\alpha)\dot{Q}_t^2\Big)^2\;\mathrm{d}t \ge -\alpha\eta \int_0^T (\dot{Q}_t^1)^2\;\mathrm{d}t - (1-\alpha)\eta \int_0^T (\dot{Q}_t^2)^2\;\mathrm{d}t.$$

Thus, pathwise,

$$W_T^{Q^\alpha} \ge \alpha W_T^{Q^1} + (1-\alpha) W_T^{Q^2}.$$

Since U is increasing and concave,

$$U(W_T^{Q^\alpha}) \ge U\Big(\alpha W_T^{Q^1} + (1-\alpha)W_T^{Q^2}\Big) \ge \alpha U(W_T^{Q^1}) + (1-\alpha)U(W_T^{Q^2}).$$

Taking expectations proves concavity. If $\eta > 0$ and $Q^1 \ne Q^2$, then $\dot{Q}^1 \ne \dot{Q}^2$ on a set of positive measure, so the temporary-impact inequality is strict by strict convexity of the square. Strict monotonicity of U then gives strict concavity of the expected-utility objective along the segment. □

5.2 CRRA as the Positive-Wealth Specialization

For $\gamma > 0$, $\gamma \neq 1$, CRRA utility is

$$U_\gamma(W) = \frac{W^{1-\gamma} - 1}{1-\gamma}. \tag{5.3}$$

The log case is recovered as $\gamma \to 1$. The utility domain is $(0, \infty)$, so $\mathcal{A}_U$ becomes the positive-wealth domain

$$\mathcal{A}_+ := \left\{ Q \in \mathcal{A} : W_T^Q > 0 \text{ a.s. and } \mathbb{E}\Big[|U_\gamma(W_T^Q)| \Big] < \infty \right\}. \tag{5.4}$$

5.3 Local Approximation Domains

The exact domain $\mathcal{A}_U$ is not the same as the local moment domain used for the certainty-equivalent approximation. In the GBM execution model, terminal wealth need not satisfy the compactness assumption in Proposition 1 globally. The local objective should therefore be read as a small-risk moment approximation valid when higher moments of terminal wealth around $M(Q)$ are negligible.

Remark 4 (Lognormal tails). For lognormal price dynamics, compact support of terminal wealth deviations does not generally hold without localization or truncation. One can either work with stopped/localized wealth variables or treat the local certainty-equivalent objective as a formal small-relative-risk approximation controlled by higher moments.

Definition 2 (Local CRRA domain and objective)**.** For $\varepsilon > 0$, define the local approximation domain

$$\mathcal{A}_{\text{loc}}(\varepsilon) := \left\{ Q \in \mathcal{A} : M(Q) > 0,\ \frac{V(Q)}{M(Q)^2} \leq \varepsilon \right\}. \tag{5.5}$$

By Corollary 1, the CRRA specialization of the local utility objective is

$$J(Q) := M(Q) - \frac{\gamma}{2} \frac{V(Q)}{M(Q)}. \tag{5.6}$$

The general variational results below apply to (2.5). Equation (5.6) is the CRRA specialization used for the CRRA corollaries.

Remark 5 (Domain of the approximation). The exact domain $\mathcal{A}_U$ and the local approximation domain are not the same. In the CRRA case, exact utility requires almost-sure positivity and integrability of utility, while the local approximation requires positive mean wealth and small relative terminal wealth risk. A path may satisfy one condition without satisfying the other.

Remark 6 (Zero-volatility case). The nontrivial execution-risk case is $\sigma > 0$. If $\sigma = 0$, the variance penalty vanishes in the local risk closures and the problem reduces to a deterministic tradeoff between temporary impact and drift.

Assumption 3 (Local utility validity)**.** Whenever we optimize a local certainty-equivalent objective, the candidate paths are restricted to a regime in which the compact-neighborhood and small-risk conditions behind Proposition 1 make the neglected higher-order utility terms negligible. In the CRRA specialization, this is represented by $\mathcal{A}_{\text{loc}}(\varepsilon)$ for small ε.

6 Variance: Exact GBM and Local Closure

6.1 Exact Variance under GBM

The mean (4.4) is simple. The exact variance under geometric Brownian motion is less local. Integration by parts gives

$$\int_0^T Q_t \, \mathrm{d}\bar{S}_t = Q_T \bar{S}_T - Q_0 \bar{S}_0 - \int_0^T \dot{Q}_t \bar{S}_t \, \mathrm{d}t. \tag{6.1}$$

Define the signed measure

$$\mathcal{K}_Q(\,\mathrm{d}t) := Q_T\delta_T(\,\mathrm{d}t) - \dot{Q}_t\ \mathrm{d}t.$$

Then the random part of $\int Q\ \mathrm{d}\bar{S}$ is

$$\int_{[0,T]} \bar{S}_t\,\mathcal{K}_Q(\,\mathrm{d}t),$$

up to the deterministic constant $-Q_0\bar{S}_0$. For $s,t \in [0,T]$,

$$\mathrm{Cov}(\bar{S}_s,\bar{S}_t) = \bar{S}_0^2 e^{\mu(s+t)}\left(e^{\sigma^2\min(s,t)} - 1\right).$$

Thus the exact GBM variance is

$$V_{\mathrm{GBM}}(Q) = \iint_{[0,T]^2} \bar{S}_0^2 e^{\mu(s+t)}\left(e^{\sigma^2\min(s,t)} - 1\right)\mathcal{K}_Q(\,\mathrm{d}s)\mathcal{K}_Q(\,\mathrm{d}t). \tag{6.2}$$

Remark 7. Formula (6.2) is the exact variance for deterministic open-loop execution in the present no-execution-noise model. It is nonlocal in Q. This is why an exact GBM treatment does not naturally collapse to the elementary Almgren–Chriss hyperbolic-sine equation.

6.2 Exact-Variance Local Utility First-Order Condition

It is useful to record the first-order condition obtained by using the exact variance (6.2) inside the local utility objective, even though we do not solve the resulting nonlocal boundary-value problem. Let

$$C(s,t) := \bar{S}_0^2 e^{\mu(s+t)}\left(e^{\sigma^2\min(s,t)} - 1\right) \tag{6.3}$$

be the GBM covariance kernel. For a path Q, define

$$G_Q(t) := \int_{[0,T]} C(s,t)\,\mathcal{K}_Q(\,\mathrm{d}s). \tag{6.4}$$

Proposition 4 (Exact-variance local utility stationarity)**.** *Let $Q^\star \in \mathcal{A}$ be an interior stationary point of* (2.5) *with $V(Q) = V_{\mathrm{GBM}}(Q)$. Set*

$$M := M(Q^\star), \qquad V := V_{\mathrm{GBM}}(Q^\star), \qquad \Lambda_{U,\mathrm{GBM}} := 1 - \frac{1}{2}A'(M)V.$$

Then $Q^\star$ satisfies, in the weak sense on $(0,T)$*,*

$$\Lambda_{U,\mathrm{GBM}}\left(2\eta\,\ddot{Q}_t^\star + \mu\bar{S}_0 e^{\mu t}\right) - A(M)\partial_t G_{Q^\star}(t) = 0. \tag{6.5}$$

If $G_{Q^\star}$ is absolutely continuous, this identity holds a.e. with $\partial_t G_{Q^\star} = \mathrm{d}G_{Q^\star}/\,\mathrm{d}t$. In the CRRA case, $A(M) = \gamma/M$ and $\Lambda_{U,\mathrm{GBM}} = 1 + \gamma V/(2M^2)$.

Proof. For a fixed-endpoint perturbation h, the signed measure $\mathcal{K}_Q$ varies by

$$\delta\mathcal{K}_Q(\,\mathrm{d}t) = -\dot{h}_t\ \mathrm{d}t.$$

Since

$$V_{\mathrm{GBM}}(Q) = \iint C(s,t)\mathcal{K}_Q(\,\mathrm{d}s)\mathcal{K}_Q(\,\mathrm{d}t),$$

symmetry of C gives

$$\delta V_{\mathrm{GBM}}[h] = -2\int_0^T G_Q(t)\dot{h}_t\ \mathrm{d}t = 2\langle\partial_t G_Q, h\rangle,$$

where $\partial_t G_Q$ is the distributional derivative and the endpoint term vanishes because $h(0) = h(T) = 0$. Combining this with Lemma 1 and

$$\delta J_U[h] = \left(1 - \frac{1}{2}A'(M)V\right)\delta M[h] - \frac{1}{2}A(M)\delta V[h]$$

gives (6.5). □

Remark 8. Equation (6.5) is the exact-GBM analogue of the local Euler–Lagrange equation below. It contains an integral operator through G_Q, so the closed-form hyperbolic-sine bridge should not be expected without additional approximation. We record this as a weak stationarity condition for deterministic fixed-endpoint perturbations. The next subsection gives a complete well-posedness result for the important frozen-curvature exact-kernel problem.

6.3 Exact GBM Kernel with Frozen Utility Curvature

The local martingale-risk closure gives a tractable differential equation, but it replaces the exact GBM covariance structure by a diagonal quadratic-variation approximation. This replacement is not needed for well-posedness. If the Arrow–Pratt curvature is frozen at a benchmark wealth level M_b, while the full GBM covariance kernel is retained, the local utility execution problem is a strictly concave Hilbert-space problem with a unique optimizer.

Let

$$\mathcal{H} := \{Q \in H^1([0,T]) : Q(0) = Q_0,\ Q(T) = Q_T\}.$$

Fix $M_b \in I$ and set

$$a_b := A(M_b) > 0.$$

Define the exact-kernel frozen-curvature objective

$$J_{K,b}(Q) := M(Q) - \frac{a_b}{2} V_{\mathrm{GBM}}(Q), \tag{6.6}$$

with $M(Q)$ from (4.4) and $V_{\mathrm{GBM}}(Q)$ from (6.2).

Proposition 5 (Exact-kernel local Arrow–Pratt execution)**.** *Assume $\eta > 0$, $T > 0$, $\bar{S}_0 > 0$, and $a_b > 0$. Then $J_{K,b}$ is a continuous, strictly concave functional on $\mathcal{H}$. Moreover, $J_{K,b}$ admits a unique maximizer $Q^{K,b} \in \mathcal{H}$.*

The maximizer is characterized by the weak equation

$$-2\eta \int_0^T \dot{Q}_t^{K,b} \dot{h}_t \,\mathrm{d}t + \mu \bar{S}_0 \int_0^T e^{\mu t} h_t \,\mathrm{d}t + a_b \int_0^T G_{Q^{K,b}}(t) \dot{h}_t \,\mathrm{d}t = 0 \tag{6.7}$$

for every $h \in H_0^1([0,T])$, where G_Q is defined in (6.4). Equivalently,

$$2\eta\, \ddot{Q}_t^{K,b} + \mu \bar{S}_0 e^{\mu t} - a_b \partial_t G_{Q^{K,b}}(t) = 0 \tag{6.8}$$

in the weak sense on $(0,T)$, with

$$Q^{K,b}(0) = Q_0, \qquad Q^{K,b}(T) = Q_T.$$

For CRRA utility, $a_b = \gamma/M_b$. For CARA utility, $a_b = a$.

Proof. The kernel C is continuous, symmetric, and positive semidefinite because it is the covariance kernel of the GBM price process. Hence, for every finite signed measure α on $[0,T]$,

$$\iint C(s,t)\alpha(\mathrm{d}s)\alpha(\mathrm{d}t) \geq 0.$$

In particular, $V_{\mathrm{GBM}}(Q) \geq 0$.

The map $Q \mapsto \mathcal{K}_Q$ is affine on $\mathcal{H}$. Hence $V_{\mathrm{GBM}}(Q)$ is a continuous convex quadratic functional of Q, meaning a positive-semidefinite quadratic form plus affine and constant terms after subtracting the endpoint bridge. To see this explicitly, write $Q = \ell + u$ with $u \in H_0^1([0,T])$. Then

$$\mathcal{K}_Q(\mathrm{d}t) = \mathcal{K}_\ell(\mathrm{d}t) - \dot{u}_t \,\mathrm{d}t,$$

and therefore

$$V_{\mathrm{GBM}}(\ell + u) = V_{\mathrm{GBM}}(\ell) - 2\int_0^T G_\ell(t)\dot{u}_t \,\mathrm{d}t + \mathcal{B}_C(\dot{u}, \dot{u}),$$

where

$$\mathcal{B}_C(f,g) := \int_0^T \int_0^T C(s,t) f(s) g(t) \, \mathrm{d}s \, \mathrm{d}t.$$

The last term is nonnegative by positive semidefiniteness of C. Also, $Q' \in L^2([0,T]) \subset L^1([0,T])$ on the finite interval, so $\mathcal{K}_Q$ is a finite signed measure, and boundedness of C gives continuity with respect to the H^1 norm. The mean functional $M(Q)$ equals a strictly concave temporary-impact term $-\eta \int \dot{Q}^2$, plus a linear drift term and an endpoint-fixed permanent-impact term. Since $-a_b V_{\mathrm{GBM}}/2$ is concave, $J_{K,b}$ is strictly concave on $\mathcal{H}$.

To prove existence, write $Q = \ell + u$, where

$$\ell(t) = Q_0 + \frac{Q_T - Q_0}{T} t, \qquad u \in H_0^1([0,T]).$$

The linear drift term is bounded by $C\|u\|_{L^2}$, and Poincare's inequality gives $\|u\|_{L^2} \le C_T \|\dot{u}\|_{L^2}$. Meanwhile,

$$-\eta \int_0^T (\dot{\ell}_t + \dot{u}_t)^2 \, \mathrm{d}t$$

dominates negatively as $\|\dot{u}\|_{L^2} \to \infty$. The variance penalty is nonpositive because $a_b > 0$ and $V_{\mathrm{GBM}} \ge 0$. Thus

$$J_{K,b}(\ell + u) \to -\infty \qquad \text{as} \qquad \|u\|_{H^1} \to \infty.$$

Any maximizing sequence is therefore bounded in H_0^1. By weak compactness, it has a weakly convergent subsequence. The negative temporary-impact quadratic is weakly upper semicontinuous, the linear terms are weakly continuous, and $-V_{\mathrm{GBM}}$ is weakly upper semicontinuous because V_{GBM} is continuous and convex. Hence $J_{K,b}$ is weakly upper semicontinuous and attains its maximum. Strict concavity gives uniqueness.

For the first-order condition, let $Q_\varepsilon = Q + \varepsilon h$ with $h \in H_0^1([0,T])$. The mean variation is

$$\delta M_Q[h] = -2\eta \int_0^T \dot{Q}_t \dot{h}_t \, \mathrm{d}t + \mu \bar{S}_0 \int_0^T e^{\mu t} h_t \, \mathrm{d}t.$$

The signed measure varies as

$$\delta \mathcal{K}_Q(\mathrm{d}t) = -\dot{h}_t \, \mathrm{d}t.$$

Symmetry of C gives

$$\delta V_{\mathrm{GBM},Q}[h] = -2 \int_0^T G_Q(t) \dot{h}_t \, \mathrm{d}t.$$

Therefore

$$\delta J_{K,b,Q}[h] = -2\eta \int_0^T \dot{Q}_t \dot{h}_t \, \mathrm{d}t + \mu \bar{S}_0 \int_0^T e^{\mu t} h_t \, \mathrm{d}t + a_b \int_0^T G_Q(t) \dot{h}_t \, \mathrm{d}t.$$

At the maximizer this variation vanishes for all $h \in H_0^1([0,T])$, giving (6.7). Finally, if G_Q is interpreted distributionally,

$$\int_0^T G_Q(t) \dot{h}_t \, \mathrm{d}t = -\langle \partial_t G_Q, h \rangle,$$

because $h(0) = h(T) = 0$. This gives (6.8). □

Corollary 2 (Positive-definite Fredholm form). *Let $\ell(t) = Q_0 + (Q_T - Q_0)t/T$ and write $Q^{K,b} = \ell + u^{K,b}$, with $u^{K,b} \in H_0^1([0,T])$. Define*

$$\mathcal{B}_C(f,g) := \int_0^T \int_0^T C(s,t) f(s) g(t) \, \mathrm{d}s \, \mathrm{d}t,$$

and

$$\mathcal{K}_\ell(\mathrm{d}t) = Q_T \delta_T(\mathrm{d}t) - \dot{\ell}_t \, \mathrm{d}t, \qquad G_\ell(t) = \int_{[0,T]} C(s,t) \mathcal{K}_\ell(\mathrm{d}s).$$

Then $u^{K,b}$ *is the unique element of* $H_0^1([0,T])$ *satisfying*

$$2\eta \int_0^T \dot{u}_t^{K,b} \dot{h}_t \, \mathrm{d}t + a_b \mathcal{B}_C(\dot{u}^{K,b}, \dot{h}) = -2\eta \int_0^T \dot{\ell}_t \dot{h}_t \, \mathrm{d}t + \mu \bar{S}_0 \int_0^T e^{\mu t} h_t \, \mathrm{d}t + a_b \int_0^T G_\ell(t) \dot{h}_t \, \mathrm{d}t \tag{6.9}$$

for every $h \in H_0^1([0,T])$.

Proof. Substitute $Q = \ell + u$ into (6.7). Since

$$\mathcal{K}_Q(\,\mathrm{d}t) = \mathcal{K}_\ell(\,\mathrm{d}t) - \dot{u}_t \, \mathrm{d}t,$$

we have

$$G_Q(t) = G_\ell(t) - \int_0^T C(s,t) \dot{u}_s \, \mathrm{d}s.$$

Collecting the terms involving u on the left gives (6.9). The bilinear form on the left is bounded on $H_0^1([0,T])$. Moreover,

$$2\eta \int_0^T \dot{u}_t^2 \, \mathrm{d}t + a_b \mathcal{B}_C(\dot{u}, \dot{u}) \geq 2\eta \|\dot{u}\|_{L^2}^2,$$

because C is a covariance kernel and $a_b > 0$. The form is therefore coercive, while the right-hand side is bounded by Poincare's inequality and boundedness of G_ℓ. Lax–Milgram gives a unique solution, which is the unique maximizer from Proposition 5. ☐

Remark 9 (Implementation). The exact-kernel optimizer can be computed by one symmetric positive-definite linear solve after time discretization. The temporary-impact term contributes the discrete analogue of $2\eta D^\top D$, while the exact GBM covariance contributes the analogue of $a_b D^\top C D$, with $C_{ij} = C(t_i, t_j)$. Endpoint and drift terms enter the right-hand side. Thus the exact-kernel problem is numerically a standard concave quadratic program, even though the continuous first-order condition is nonlocal.

6.4 Local Martingale-Risk Closure

To obtain a transparent execution benchmark, we also study the local martingale-risk closure

$$V_{\mathrm{loc}}(Q) := \sigma^2 \bar{S}_0^2 \int_0^T Q_t^2 e^{(2\mu+\sigma^2)t} \, \mathrm{d}t. \tag{6.10}$$

This is the expected quadratic variation of the martingale term $\int_0^T Q_t \sigma \bar{S}_t \, \mathrm{d}B_t$. Its frozen-price specialization, $\sigma^2 \bar{S}_0^2 \int_0^T Q_t^2 \, \mathrm{d}t$, is the standard variance term in the arithmetic Brownian Almgren–Chriss model with constant dollar volatility. With the exponential weight retained, V_{loc} is a local GBM martingale-risk closure for short horizons or small relative price moves. The exact GBM variance (6.2) remains the correct object when the full covariance structure of the geometric price path is material.

Unless otherwise stated, the variational calculations in the next section use V_{loc}. This choice should be read as a moment closure, not as a claim that the exact GBM variance is local.

Assumption 4 (Local variance closure)**.** Unless explicitly stated otherwise, the variance input in the local utility problem is V_{loc}, not V_{GBM}.

Remark 10 (A constant-inventory diagnostic). The closure V_{loc} should not be interpreted as an exact small-σ approximation uniformly over long horizons with nonzero drift. If $Q_t \equiv q$, then

$$\int_0^T Q_t \, \mathrm{d}\bar{S}_t = q(\bar{S}_T - \bar{S}_0),$$

so the exact GBM variance is

$$q^2 \bar{S}_0^2 e^{2\mu T}(e^{\sigma^2 T} - 1).$$

The local closure instead gives

$$q^2\sigma^2\bar{S}_0^2\int_0^T e^{(2\mu+\sigma^2)t}\,\mathrm{d}t.$$

These coincide only in special limits, such as $\mu = 0$ or short horizons. Thus V_{loc} is best viewed as a martingale-risk closure, not as the exact GBM terminal-wealth variance.

7 Local Utility Variational Problem

In this section set

$$V(Q) := V_{\mathrm{loc}}(Q)$$

and use $M(Q)$ from (4.4). Let

$$a_0 := 2\mu + \sigma^2, \qquad \nu := \sigma^2\bar{S}_0^2.$$

Then

$$J_U(Q) = M(Q) - \frac{1}{2}A(M(Q))V(Q), \qquad V(Q) = \nu\int_0^T e^{a_0 t}Q_t^2\,\mathrm{d}t.$$

Proposition 6 (General local utility first variation)**.** *Let $Q^\star \in \mathcal{A}$ be an interior stationary point of* (2.5) *with $M := M(Q^\star)$ and $V := V(Q^\star)$. Define*

$$\Lambda_U := 1 - \frac{1}{2}A'(M)V. \tag{7.1}$$

Then $Q^\star$ satisfies the Euler–Lagrange equation

$$\Lambda_U\left(2\eta\,\ddot{Q}_t^\star + \mu\bar{S}_0e^{\mu t}\right) - A(M)\sigma^2\bar{S}_0^2e^{(2\mu+\sigma^2)t}Q_t^\star = 0, \qquad 0 < t < T, \tag{7.2}$$

with boundary conditions $Q^\star(0) = Q_0$, $Q^\star(T) = Q_T$.

Proof. For any smooth perturbation h with $h(0) = h(T) = 0$, write $Q_\varepsilon = Q^\star + \varepsilon h$. By Lemma 1, the variation of the mean is

$$\delta M[h] = \int_0^T\left(2\eta\ddot{Q}_t^\star + \mu\bar{S}_0e^{\mu t}\right)h_t\,\mathrm{d}t.$$

The variation of the local variance is

$$\delta V[h] = 2\sigma^2\bar{S}_0^2\int_0^T e^{(2\mu+\sigma^2)t}Q_t^\star h_t\,\mathrm{d}t.$$

Since

$$\delta J_U[h] = \delta M[h] - \frac{1}{2}\left(A'(M)V\delta M[h] + A(M)\delta V[h]\right),$$

we have

$$\delta J_U[h] = \left(1 - \frac{1}{2}A'(M)V\right)\delta M[h] - \frac{1}{2}A(M)\delta V[h].$$

Substituting the mean and variance variations gives (7.2). Because h is arbitrary in the interior, the integrand must vanish pointwise. □

Corollary 3 (General variable-coefficient form)**.** *If $\Lambda_U \neq 0$, equation* (7.2) *can be written as*

$$\eta\,\ddot{Q}_t^\star - \lambda_U^\star e^{(2\mu+\sigma^2)t}Q_t^\star = -\frac{\mu\bar{S}_0}{2}e^{\mu t}, \tag{7.3}$$

where

$$\lambda_U^\star := \frac{A(M)\sigma^2\bar{S}_0^2}{2\Lambda_U} = \frac{A(M)\sigma^2\bar{S}_0^2}{2\left(1 - \frac{1}{2}A'(M)V\right)}. \tag{7.4}$$

Remark 11 (Nondegenerate normalization). The scalar execution-risk-weight form (7.3) divides by Λ_U, so it requires $\Lambda_U \neq 0$. The undivided Euler–Lagrange equation (7.2) remains the primary stationarity condition. For decreasing absolute risk aversion utilities, $A'(M) \leq 0$, and hence

$$\Lambda_U = 1 - \frac{1}{2}A'(M)V \geq 1.$$

Thus the nondegeneracy condition is automatic for the standard DARA case and for CRRA.

Remark 12 (Necessary condition only). Equation (7.2) is a necessary condition for an interior stationary point of the local utility objective with the martingale-risk closure. We do not claim global optimality for the full ratio objective without additional assumptions. In the frozen-risk benchmark below, the denominator and execution-risk weights are fixed, the objective becomes a strictly concave quadratic functional, and the boundary-value problem has a unique maximizer.

Corollary 4 (CRRA variable-coefficient form). *For CRRA utility, $A(M) = \gamma/M$ and*

$$A'(M) = -\frac{\gamma}{M^2}.$$

Thus Λ_U becomes

$$\Lambda := 1 + \frac{\gamma}{2}\frac{V}{M^2}. \tag{7.5}$$

The general first-order condition (7.2) *becomes*

$$\Lambda\left(2\eta\,\ddot{Q}_t^\star + \mu\bar{S}_0 e^{\mu t}\right) - \frac{\gamma\sigma^2\bar{S}_0^2}{M}e^{(2\mu+\sigma^2)t}Q_t^\star = 0, \qquad 0 < t < T. \tag{7.6}$$

Equivalently, if $\Lambda \neq 0$,

$$\eta\,\ddot{Q}_t^\star - \lambda_\star e^{(2\mu+\sigma^2)t}Q_t^\star = -\frac{\mu\bar{S}_0}{2}e^{\mu t}, \tag{7.7}$$

where the scalar

$$\lambda_\star := \frac{\gamma\sigma^2\bar{S}_0^2}{2M\Lambda} \tag{7.8}$$

is evaluated at the stationary path.

Remark 13 (What remains nonlocal). Once M, V, and therefore Λ are evaluated at $Q^\star$, $\lambda_\star$ is a scalar. Nevertheless the coefficient of $Q_t^\star$ in (7.7) is generally time-dependent through $e^{(2\mu+\sigma^2)t}$. Moreover, the scalar itself depends on the whole path through $M(Q^\star)$ and $V(Q^\star)$. Thus the honest local-CE first-order condition is a nonlocal boundary-value problem, not generically a constant-coefficient ODE.

Remark 14 (No direct permanent-impact timing term). The parameter θ enters (7.2) only through $M(Q^\star)$, because for fixed endpoints its contribution to the mean is $\frac{1}{2}\theta(Q_T^2 - Q_0^2)$. Therefore in this linear permanent-impact model, θ does not directly create a marginal preference for faster or slower paths. It changes the expected wealth level and can thereby affect $A(M)$ and Λ_U. In the CRRA specialization this is the indirect wealth-denominator channel.

8 Frozen-Risk Almgren–Chriss Benchmark

The local first-order condition (7.3) already identifies the utility-curvature channel, but its coefficients still depend on the path and on the GBM variance weight. The benchmark in this section freezes those secondary effects to isolate the main comparison with the classical Almgren–Chriss model: replace the exogenous execution-risk weight λ_{AC} by the utility-implied execution-risk weight $\lambda_U(M_b)$. To recover the familiar closed-form Almgren–Chriss schedule, impose a drift-retaining frozen-risk benchmark:

(i) freeze the exponential variance weight, $e^{(2\mu+\sigma^2)t} \approx 1$, as in a short-horizon or arithmetic-price local-volatility approximation;

(ii) use the leading local normalization $\Lambda_U \approx 1$.

The resulting problem is a closed-form comparison benchmark, not an exact solution of the GBM utility problem. It freezes the execution-risk weight in the variance term while retaining the deterministic GBM drift profile in the mean, so it should not be interpreted as a systematic first-order expansion in μ.

Assumption 5 (Benchmark frozen coefficients)**.** For the closed-form benchmark, $e^{(2\mu+\sigma^2)t}$ is frozen at 1, Λ_U is frozen at 1, and the wealth normalization is a fixed scalar $M_b \in I$.

Let M_b be the wealth level used in this benchmark normalization and define the utility-implied frozen execution-risk weight

$$\lambda_b := \lambda_U(M_b) := \frac{1}{2}A(M_b)\sigma^2\bar{S}_0^2. \tag{8.1}$$

For CRRA utility, this becomes

$$\lambda_{\text{CRRA}}(M_b) := \lambda_{\text{eff}} := \frac{\gamma\sigma^2\bar{S}_0^2}{2M_b}. \tag{8.2}$$

For log utility,

$$\lambda_{\log}(M_b) = \frac{\sigma^2\bar{S}_0^2}{2M_b},$$

and for CARA utility with absolute risk aversion a,

$$\lambda_{\text{CARA}} = \frac{a\sigma^2\bar{S}_0^2}{2}.$$

Then (7.3) reduces to

$$\eta\,\ddot{Q}_t - \lambda_b Q_t = -\frac{\mu\bar{S}_0}{2}e^{\mu t}. \tag{8.3}$$

Proposition 7 (Benchmark sufficiency)**.** *On the affine Sobolev endpoint space $\mathcal{H}$, the benchmark functional*

$$J_b(Q) = -\eta\int_0^T \dot{Q}_t^2\,\mathrm{d}t - \lambda_b\int_0^T Q_t^2\,\mathrm{d}t + \mu\bar{S}_0\int_0^T e^{\mu t}Q_t\,\mathrm{d}t \tag{8.4}$$

plus any endpoint-fixed constants is continuous, strictly concave, and admits a unique maximizer when $\eta > 0$ and $\lambda_b > 0$. This maximizer is the weak solution of (8.3) *with the endpoint constraints and is therefore C^2.*

Proof. Write $Q = \ell + u$, where ℓ is the linear endpoint bridge and $u \in H_0^1([0,T])$. The negative quadratic terms in J_b dominate the linear term, so $J_b(\ell+u) \to -\infty$ as $\|u\|_{H^1} \to \infty$. Hence a maximizer exists by the direct method. For a perturbation h with $h(0) = h(T) = 0$, the second variation is

$$\delta^2 J_b[h,h] = -2\eta\int_0^T \dot{h}_t^2\,\mathrm{d}t - 2\lambda_b\int_0^T h_t^2\,\mathrm{d}t.$$

This is strictly negative for every nonzero admissible h, so J_b is strictly concave and the maximizer is unique. The weak Euler–Lagrange equation of (8.4) is exactly (8.3); standard one-dimensional ODE regularity gives a C^2 representative. □

Proposition 8 (Benchmark closed form)**.** *Assume $\lambda_b > 0$ and set*

$$\kappa := \sqrt{\lambda_b/\eta}.$$

If $\eta\mu^2 \neq \lambda_b$, define

$$\beta := -\frac{\mu\bar{S}_0}{2(\eta\mu^2 - \lambda_b)}. \tag{8.5}$$

Then the solution of (8.3) *with $Q(0) = Q_0$, $Q(T) = Q_T$ is*

$$Q_t = \beta e^{\mu t} + (Q_0 - \beta)\frac{\sinh(\kappa(T-t))}{\sinh(\kappa T)} + (Q_T - \beta e^{\mu T})\frac{\sinh(\kappa t)}{\sinh(\kappa T)}. \tag{8.6}$$

The trading rate is

$$v_t = \dot{Q}_t = \beta\mu e^{\mu t} - \kappa(Q_0 - \beta)\frac{\cosh(\kappa(T-t))}{\sinh(\kappa T)} + \kappa(Q_T - \beta e^{\mu T})\frac{\cosh(\kappa t)}{\sinh(\kappa T)}. \tag{8.7}$$

Proof. The homogeneous equation $\eta Q'' - \lambda_b Q = 0$ has fundamental solutions $\sinh(\kappa(T-t))$ and $\sinh(\kappa t)$. A particular solution of (8.3) is $\beta e^{\mu t}$, because

$$(\eta\mu^2 - \lambda_b)\beta e^{\mu t} = -\frac{\mu \bar{S}_0}{2} e^{\mu t}.$$

The coefficients of the two homogeneous solutions are chosen so that the particular solution is subtracted from the endpoints. Differentiating gives (8.7). ☐

Remark 15 (Resonance). If $\eta\mu^2 = \lambda_b$ and $\mu \neq 0$, the particular solution in (8.6) is replaced by

$$-\frac{\bar{S}_0}{4\eta}\, t e^{\mu t}.$$

This is the standard repeated-root particular solution. The resonance case is non-generic.

Corollary 5 (Classical hyperbolic-sine bridge)**.** *If $\mu = 0$, then $\beta = 0$ and*

$$Q_t = Q_0\frac{\sinh(\kappa(T-t))}{\sinh(\kappa T)} + Q_T\frac{\sinh(\kappa t)}{\sinh(\kappa T)}. \tag{8.8}$$

For full liquidation, $Q_T = 0$, this is the usual Almgren–Chriss inventory profile.

Corollary 6 (Constant-forcing benchmark)**.** *If, in addition to the frozen-risk approximation, the deterministic drift profile is frozen so that $e^{\mu t} \approx 1$, then the benchmark equation becomes*

$$\eta Q_t'' - \lambda_b Q_t = -\frac{\mu \bar{S}_0}{2}. \tag{8.9}$$

This is the closer analogue of the arithmetic-price limit with constant drift forcing.

Remark 16 (The one-line bridge). The benchmark replacement for the exogenous Almgren–Chriss execution-risk weight is

$$\boxed{\lambda_{\mathrm{AC}} \quad \rightsquigarrow \quad \lambda_U(M_b) = \frac{1}{2}A(M_b)\sigma^2\bar{S}_0^2.}$$

For CRRA, this becomes $\lambda_{\mathrm{eff}} = \gamma\sigma^2\bar{S}_0^2/(2M_b)$. For CARA, it becomes the wealth-independent weight $\lambda_{\mathrm{CARA}} = a\sigma^2\bar{S}_0^2/2$.

Proposition 9 (CARA–CRRA separation)**.** *Assume $\sigma\bar{S}_0 > 0$, and for the relative-risk statement work on a positive wealth interval $I \subseteq (0,\infty)$. In the frozen-risk benchmark,*

$$\lambda_U(w) = \frac{1}{2}A(w)\sigma^2\bar{S}_0^2.$$

Then $\lambda_U(w)$ is independent of wealth w if and only if U has constant absolute risk aversion, hence is CARA up to positive affine transformation. Also, $w\lambda_U(w)$ is independent of w if and only if U has constant relative risk aversion, hence is CRRA or log utility up to positive affine transformation.

Proof. Since $\lambda_U(w)$ is proportional to $A(w)$, wealth independence of λ_U is equivalent to $A(w) = a$ for a constant $a > 0$. Thus

$$-\frac{U''(w)}{U'(w)} = a,$$

so $\mathrm{d}\log U'(w)/\,\mathrm{d}w = -a$. Integrating gives $U'(w) = Ce^{-aw}$, and integrating again gives CARA utility up to positive affine transformation.

Similarly, $w\lambda_U(w)$ is constant if and only if $wA(w) = \gamma$ for a constant $\gamma > 0$. Thus

$$-\frac{U''(w)}{U'(w)} = \frac{\gamma}{w},$$

so $U'(w) = Cw^{-\gamma}$. Integrating gives power utility for $\gamma \neq 1$ and log utility for $\gamma = 1$. These are exactly the constant-relative-risk-aversion utilities, up to positive affine transformation. □

9 Practical Calibration

The frozen-risk benchmark's main practical output is a direct calibration rule for the Almgren–Chriss execution-risk weight:

$$\lambda_{\mathrm{AC}} = \lambda_U(M_b) = \frac{1}{2}A(M_b)\sigma^2\bar{S}_0^2.$$

It converts utility curvature, volatility, price scale, and the relevant wealth base into the coefficient that appears in the benchmark execution ODE. If inventory is measured in shares, price in currency per share, and time in years, then

$$[\sigma^2] = \mathrm{year}^{-1}, \qquad [A(M_b)\bar{S}_0^2] = \frac{\mathrm{currency}}{\mathrm{share}^2},$$

so

$$[\lambda_U(M_b)] = \frac{\mathrm{currency}}{\mathrm{share}^2\ \mathrm{year}}.$$

This is the same dimensional role played by the Almgren–Chriss inventory execution-risk weight: $\lambda_U(M_b)Q_t$ has units of price change per unit time, matching $\eta\ddot{Q}_t$ in (8.3).

Calibration recipe. Given a utility specification U, benchmark wealth normalization M_b, annualized volatility σ, and price level $\bar{S}_0$, set

$$\boxed{\lambda_{\mathrm{AC}} := \lambda_U(M_b) = \frac{1}{2}A(M_b)\sigma^2\bar{S}_0^2.} \tag{9.1}$$

Then use λ_{AC} in the usual Almgren–Chriss calculation, subject to the frozen-risk assumptions stated above. For CRRA, this is

$$\lambda_{\mathrm{AC}} = \frac{\gamma\sigma^2\bar{S}_0^2}{2M_b};$$

doubling γ doubles the execution-risk weight, while doubling the wealth base halves it. For CARA, $A(M_b) = a$, so the same calibration is wealth-independent.

γ	σ	M_b	λ_{eff}
2	30%	\$100,000	9.0×10^{-3}
2	30%	\$1,000,000	9.0×10^{-4}
2	30%	\$10,000,000	9.0×10^{-5}
5	30%	\$1,000,000	2.25×10^{-3}
2	60%	\$1,000,000	3.6×10^{-3}

Table 1: Illustrative annualized CRRA execution-risk weights from $\lambda_{\mathrm{eff}} = \gamma\sigma^2\bar{S}_0^2/(2M_b)$, using $\bar{S}_0 = \$100$. Units are currency per share squared per year when Q is measured in shares.

Remark 17 (Choice of wealth base). The wealth normalization M_b should match the wealth base over which the trader evaluates terminal-wealth utility. For a stand-alone execution desk this may be allocated trading capital; for a portfolio manager it may be strategy NAV; for a founder or concentrated holder it may be total liquid wealth plus marked position value. The formula is only as meaningful as this wealth-base choice.

10 Numerical Illustration

Figure 1 illustrates the frozen-risk benchmark for a full liquidation with $Q_0 = 1$, $Q_T = 0$, $\eta = 1$, $\gamma = 2$, $\sigma = 0.30$, $\bar{S}_0 = 100$, and $T = 1$. The three solid curves use $M_b \in \{200, 500, 2000\}$ and $\mu = 0$. Smaller wealth raises λ_{eff} and front-loads liquidation; larger wealth lowers λ_{eff} and produces a flatter path. The dashed curve keeps $M_b = 500$ and uses $\mu = 0.05$, showing the inventory tilt induced by expected price appreciation. The figure is generated from Proposition 8; the plotting source is listed in Appendix D.

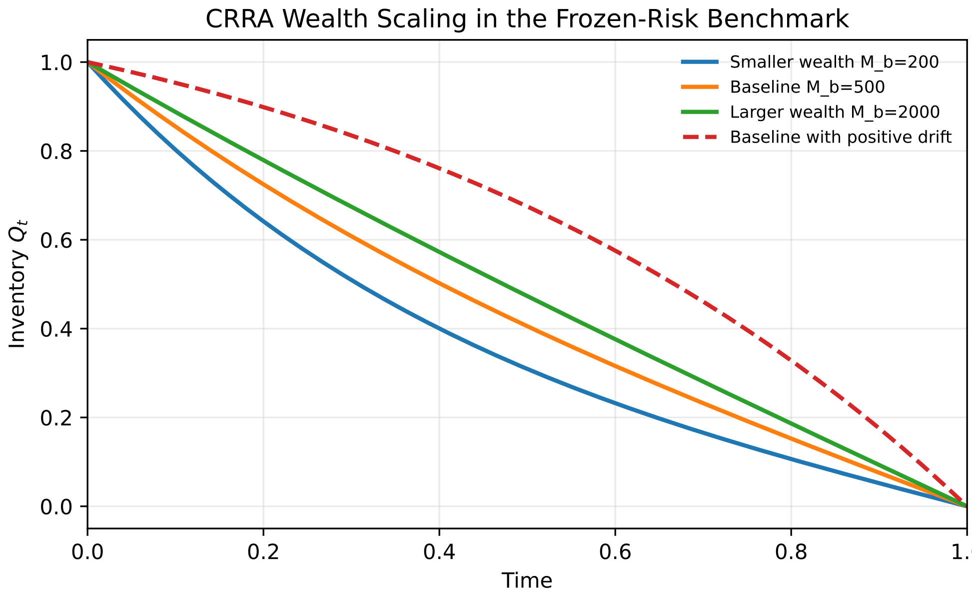

Figure 1: The same dollar inventory risk generates faster execution for smaller wealth and slower execution for larger wealth in the frozen-risk benchmark. Solid curves use $M_b = 200, 500, 2000$ with $\mu = 0$; the dashed curve uses $M_b = 500$, $\mu = 0.05$.

Figure 2 compares the exact-kernel frozen-curvature optimizer $Q^{K,b}$ from Proposition 5 with the closed-form frozen-risk benchmark Q^b from Proposition 8. Both panels use $Q_0 = 1$, $Q_T = 0$, $\eta = 1$, $\gamma = 2$, $M_b = 500$, and $\bar{S}_0 = 100$. In the baseline case, the two curves are almost indistinguishable. In the longer-horizon, higher-volatility case, the exact GBM covariance makes delayed inventory materially more expensive and the exact-kernel optimizer front-loads liquidation. The objective gain reported in Table 2 is measured in the same money-metric units as the reduced exact-kernel objective $J_{K,b}$, with endpoint-fixed constants omitted. It is positive by construction; the gain/risk column divides that gain by the benchmark path's exact-kernel variance penalty, $\frac{1}{2}a_b V_{\mathrm{GBM}}(Q^b)$, to indicate scale.

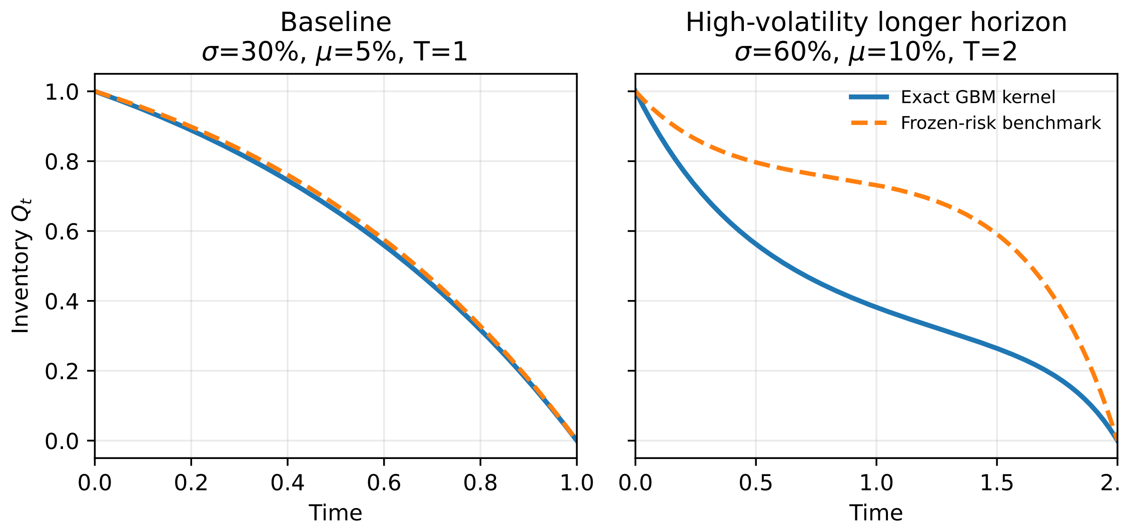

Figure 2: The exact GBM-kernel optimizer and the closed-form frozen-risk benchmark agree closely in a short baseline case but can diverge materially when horizon, volatility, and drift are larger.

Case	T	σ	μ	Max dev.	L^2 dev.	Gain	Gain/risk
Baseline	1.0	30%	5%	0.0168	0.0121	1.744×10^{-3}	0.2%
Stress	2.0	60%	10%	0.3633	0.2677	2.6594	20.3%

Table 2: Deviation between the exact-kernel optimizer $Q^{K,b}$ and the frozen-risk benchmark Q^b. The gain is $J_{K,b}(Q^{K,b}) - J_{K,b}(Q^b)$, and gain/risk divides that gain by $\frac{1}{2}a_b V_{\mathrm{GBM}}(Q^b)$, the benchmark path's exact-kernel variance penalty. All quantities are evaluated under the same finite-element midpoint discretization used to generate Figure 2; endpoint-fixed constants are omitted.

11 Interpretation and Comparative Statics

The comparative statics below are cleanest in the frozen coefficient benchmark. In the full local-utility boundary-value problem, the same economic forces are present, but monotonicity is not automatic because all normalizing quantities move with the stationary path.

Wealth scaling. The central comparative static is the wealth dependence of $A(M)$. In the frozen benchmark, $\lambda_U(M_b)$ moves one-for-one with Arrow–Pratt absolute risk aversion. Under CRRA, $A(M_b) = \gamma/M_b$, so a larger expected terminal wealth lowers the effective execution-risk penalty and a smaller wealth base raises it. Under CARA, $A(M_b) = a$, so the benchmark execution-risk weight is independent of wealth. This is the precise distinction between wealth-relative and dollar-risk execution.

Risk aversion. Holding the normalization terms fixed, larger Arrow–Pratt curvature raises the local risk penalty. In the general stationarity equation this appears through $\lambda_U^\star$ in (7.4); in the CRRA specialization it appears through $\lambda_\star$ in (7.8). In the benchmark, larger λ_b raises $\kappa = \sqrt{\lambda_b/\eta}$, which makes liquidation front-loaded for $Q_T = 0$. In the full local problem, this remains the natural directional force, but a global monotonicity theorem would require additional assumptions.

Volatility and price scale. Holding the endogenous normalization fixed, the variance scale $\sigma^2 \bar{S}_0^2$ raises the effective penalty. Higher volatility or a larger price scale means that a given inventory path produces more terminal-wealth risk, so the benchmark path reduces exposure more aggressively.

Temporary impact. Temporary impact η penalizes trading speed. In the benchmark, $\kappa = \sqrt{\lambda_b/\eta}$, so higher η flattens the hyperbolic-sine path and discourages aggressive front-loading.

Drift. Drift enters the local first-order condition as a forcing term. Positive μ makes holding inventory more attractive for a long position, while negative μ makes exposure more costly. In the benchmark solution, this effect appears through the exponential particular solution $\beta e^{\mu t}$, which tilts the path away from the symmetric zero-drift bridge.

Permanent impact. With fixed endpoints and the linear cumulative-flow permanent-impact specification, θ contributes

$$\frac{1}{2}\theta(Q_T^2 - Q_0^2)$$

to expected terminal wealth. It does not directly alter the marginal timing tradeoff among admissible paths with the same endpoints. Its effect in the local utility objective is through expected terminal wealth: changing θ changes M, and changing M changes $A(M)$ and Λ_U. In the CRRA specialization, this is the wealth-denominator channel. The sign and strength of the indirect effect depend on the endpoint configuration and the stationary path.

Open-loop nature. All results in this paper are deterministic pre-commitment results. They are not time-consistent feedback policies. A trader who re-optimizes after observing price moves faces a different stochastic-control problem.

12 Extensions

Exact utility control. The exact utility problem should be treated through a dynamic program or stochastic maximum principle, with state variables for cash, inventory, and price. The CRRA domain (5.4) is the main positive-wealth specialization in this paper. The local Arrow–Pratt analysis supplies the utility-curvature benchmark against which exact solutions can be compared.

Exact GBM variance. Using V_{GBM} from (6.2) inside the local utility objective gives a nonlocal variational problem. With frozen Arrow–Pratt curvature, Proposition 5 and Corollary 2 show that this problem is well posed and has a unique Fredholm-type optimizer. A natural extension is to study the fully path-dependent curvature problem, where $A(M(Q))$ and Λ_U are not frozen.

Adaptive controls. Allowing v_t to depend on the price filtration changes both the mean and variance calculations. It also changes the interpretation of terminal wealth risk, because controls can respond to realized price paths.

Execution noise and nonlinear impact. Temporary execution noise, transient impact, and nonlinear impact laws can be added to the same utility-first framework. In those cases, the clean Almgren–Chriss bridge may survive only as a local or perturbative result.

13 Conclusion

This paper argues that optimal execution should be derived from terminal wealth utility rather than from implementation shortfall plus an exogenous variance penalty. Starting from a general increasing, concave utility function, the local certainty-equivalent expansion yields

$$J_U(Q) = M(Q) - \frac{1}{2}A(M(Q))V(Q).$$

The penalty on execution risk is therefore an Arrow–Pratt curvature object. CARA utility recovers a wealth-independent dollar-risk penalty, while CRRA utility gives

$$J_{\mathrm{CRRA}}(Q) = M(Q) - \frac{\gamma}{2}\frac{V(Q)}{M(Q)}.$$

The same dollar variance should lead to different execution speeds for CRRA traders with different wealth levels.

For deterministic fixed-endpoint paths, the exact open-loop expected utility objective is concave on convex admissible domains; with temporary impact, it is strictly concave. Thus the exact problem may be analytically hard, but it is structurally well-behaved in the deterministic pre-commitment setting. Under deterministic execution, temporary impact penalizes speed, drift rewards or penalizes inventory exposure, and permanent impact enters expected wealth as a fixed-endpoint boundary term. Under exact geometric Brownian motion, terminal-wealth variance is nonlocal. Freezing Arrow–Pratt curvature at a benchmark wealth level turns the exact-kernel local utility problem into a strictly concave Fredholm-type Hilbert-space problem with a unique optimizer. If instead one uses the local martingale-risk closure, the resulting first-order condition becomes a tractable variable-coefficient boundary-value problem. In the full local first-order condition, the corresponding utility scalar is

$$\lambda_U^\star = \frac{A(M(Q^\star))\sigma^2\bar{S}_0^2}{2\Lambda_U},$$

with the CRRA specialization

$$\lambda_\star = \frac{\gamma\sigma^2\bar{S}_0^2}{2M(Q^\star)\Lambda}.$$

The inventory coefficient remains time-dependent through the GBM variance weight. Under further frozen-coefficient benchmark assumptions, the local closure reduces to the familiar Almgren–Chriss hyperbolic-sine schedule with

$$\lambda_U(M_b) = \frac{1}{2}A(M_b)\sigma^2\bar{S}_0^2.$$

This formula is also the paper's practical calibration recipe: a trader can map utility curvature, volatility, price scale, and wealth base into the Almgren–Chriss execution-risk weight. CRRA gives $\lambda_{\mathrm{eff}} = \gamma\sigma^2\bar{S}_0^2/(2M_b)$; CARA gives $\lambda_{\mathrm{CARA}} = a\sigma^2\bar{S}_0^2/2$.

The resulting framework is not a complete solution of adaptive stochastic control. It is a local utility foundation, a well-posed exact-kernel open-loop problem under frozen curvature, and a calibrated Almgren–Chriss benchmark. The central message does not depend on the benchmark closure: local execution risk weights are Arrow–Pratt utility-curvature terms. CRRA is the wealth-relative case; CARA is the dollar-risk case.

A Deriving the Mean Formula

Starting from (4.1),

$$\mathrm{d}W_t = -\eta\dot{Q}_t^2\,\mathrm{d}t + Q_t\,\mathrm{d}\bar{S}_t + \theta Q_t\dot{Q}_t\,\mathrm{d}t.$$

Integrating and taking expectations gives

$$\mathbb{E}[W_T^Q] = W_0 - \eta\int_0^T \dot{Q}_t^2\,\mathrm{d}t + \mathbb{E}\left[\int_0^T Q_t\,\mathrm{d}\bar{S}_t\right] + \theta\int_0^T Q_t\dot{Q}_t\,\mathrm{d}t.$$

The last term is

$$\theta \int_0^T Q_t \dot{Q}_t \, \mathrm{d}t = \frac{\theta}{2}(Q_T^2 - Q_0^2).$$

Since Q is deterministic,

$$\mathbb{E}\left[\int_0^T Q_t \, \mathrm{d}\bar{S}_t\right] = \int_0^T Q_t \, \mathbb{E}[\, \mathrm{d}\bar{S}_t] = \mu \bar{S}_0 \int_0^T e^{\mu t} Q_t \, \mathrm{d}t.$$

This proves (4.4).

B Exact GBM Variance Kernel

For $s \le t$,

$$\bar{S}_t = \bar{S}_s \exp\big((\mu - \tfrac{1}{2}\sigma^2)(t-s) + \sigma(B_t - B_s)\big).$$

Hence

$$\mathbb{E}[\bar{S}_s \bar{S}_t] = \bar{S}_0^2 e^{\mu(s+t)} e^{\sigma^2 s}.$$

By symmetry, for all s, t,

$$\mathbb{E}[\bar{S}_s \bar{S}_t] = \bar{S}_0^2 e^{\mu(s+t)} e^{\sigma^2 \min(s,t)}.$$

Since $\mathbb{E}[\bar{S}_s]\mathbb{E}[\bar{S}_t] = \bar{S}_0^2 e^{\mu(s+t)}$,

$$\operatorname{Cov}(\bar{S}_s, \bar{S}_t) = \bar{S}_0^2 e^{\mu(s+t)} \left(e^{\sigma^2 \min(s,t)} - 1\right).$$

Combining this covariance kernel with the integration-by-parts representation (6.1) gives (6.2).

C Local Euler–Lagrange Derivation

This appendix records the compact derivation behind Proposition 6. With

$$M(Q) = W_0 - \eta \int_0^T \dot{Q}_t^2 \, \mathrm{d}t + \mu \bar{S}_0 \int_0^T e^{\mu t} Q_t \, \mathrm{d}t + \frac{\theta}{2}(Q_T^2 - Q_0^2)$$

and

$$V(Q) = \sigma^2 \bar{S}_0^2 \int_0^T e^{(2\mu + \sigma^2)t} Q_t^2 \, \mathrm{d}t,$$

the local objective is

$$J_U(Q) = M(Q) - \frac{1}{2} A(M(Q)) V(Q).$$

For a perturbation h with $h(0) = h(T) = 0$,

$$\delta M[h] = \int_0^T \left(2\eta \ddot{Q}_t + \mu \bar{S}_0 e^{\mu t}\right) h_t \, \mathrm{d}t,$$

and

$$\delta V[h] = 2\sigma^2 \bar{S}_0^2 \int_0^T e^{(2\mu + \sigma^2)t} Q_t h_t \, \mathrm{d}t.$$

Thus

$$\delta J_U[h] = \left(1 - \frac{1}{2} A'(M) V\right) \delta M[h] - \frac{1}{2} A(M) \delta V[h].$$

Setting this equal to zero for all admissible h gives (7.2). For CRRA, $A(M) = \gamma/M$ and $A'(M) = -\gamma/M^2$, which recovers (7.6).

D Numerical Figure Source

Figures 1 and 2, together with Table 2, are generated by the standalone script listed in this appendix. Running it from the writing-repo root regenerates the wealth-scaling PDF, the exact-kernel comparison PDF, and the exact-kernel comparison table in `almgren_chriss/`. The wealth-scaling figure evaluates the closed-form benchmark solution from Proposition 8. The exact-kernel comparison solves the finite-dimensional positive-definite linear system obtained by discretizing Corollary 2; no Monte Carlo simulation is used.

References

[1] Almgren, R. (2003). Optimal execution with nonlinear impact functions and trading-enhanced risk. *Applied Mathematical Finance*, 10(1), 1–18.

[2] Almgren, R., and Chriss, N. (2000). Optimal execution of portfolio transactions. *Journal of Risk*, 3(2), 5–39.

[3] Arrow, K. J. (1965). *Aspects of the Theory of Risk-Bearing*. Yrjö Jahnssonin Säätiö.

[4] Bertsimas, D., and Lo, A. W. (1998). Optimal control of execution costs. *Journal of Financial Markets*, 1(1), 1–50.

[5] Cartea, A., and Jaimungal, S. (2015). *Algorithmic and High-Frequency Trading*. Cambridge University Press.

[6] Gelfand, I. M., and Fomin, S. V. (2000). *Calculus of Variations*. Dover Publications.

[7] Gatheral, J., Schied, A., and Slynko, A. (2012). Transient linear price impact and Fredholm integral equations. *Mathematical Finance*, 22(3), 445–474.

[8] Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4), 373–413.

[9] Obizhaeva, A., and Wang, J. (2013). Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1), 1–32.

[10] Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, 32(1/2), 122–136.

[11] Schied, A., and Schöneborn, T. (2009). Risk aversion and the dynamics of optimal liquidation strategies in illiquid markets. *Finance and Stochastics*, 13, 181–204.

[12] Schied, A., Schöneborn, T., and Tehranchi, M. (2010). Optimal basket liquidation for CARA investors is deterministic. *Applied Mathematical Finance*, 17(6), 471–489.