Optimal Execution under Non-Linear Impact: Robustness, Model Risk, and Empirical Validation

Elyas Assili
Quantitative Research Paper

February 2026

Abstract

This paper investigates the optimal liquidation of large institutional orders in markets exhibiting concave liquidity costs (the "Square-Root Law"). We formulate the problem under a CARA utility framework and derive a stable numerical scheme for the resulting non-linear Hamilton-Jacobi-Bellman equation. Beyond theoretical derivation, we subject the strategy to rigorous empirical validation. We demonstrate that our optimal convex schedule outperforms the industry-standard TWAP benchmark by approximately **20.7%** for large meta-orders. Crucially, a robust sensitivity analysis reveals that the strategy maintains near-optimal performance (inefficiency gap $< 1.5\%$) even under significant model misspecification. Finally, Monte Carlo simulations quantify the inherent trade-off between execution cost reduction and increased variance risk, providing a comprehensive guide for production implementation.

1 Introduction

The optimal execution of large orders requires balancing the urgency to reduce market risk against the cost of consuming liquidity. The seminal model of (1) assumes linear market impact, leading to tractable strategies that are scale-invariant. However, modern microstructure research challenges this assumption, suggesting that for large orders, the marginal cost of trading decreases with size (concavity).

1.1 Motivation and Contributions

Empirical studies (e.g., (6), (3)) strongly support a power-law structure for temporary impact, $h(v) \propto v^{\phi}$ with $\phi \approx 0.5$. In this regime, standard "front-loaded" strategies derived from linear models are suboptimal.

In this paper, we bridge the gap between theory and practice through four key contributions:

1. **Theoretical Rigor:** We derive the non-linear Master Equation from first principles using CARA utility, ensuring consistency with No-Arbitrage conditions.
2. **Numerical Robustness:** We implement a Finite Difference scheme satisfying the CFL condition to handle the singularity of the impact gradient.
3. **Comparative Benchmarking:** We compare our optimal strategy against TWAP and Almgren-Chriss (Linear), quantifying that linear models capture only $\approx 30\%$ of the potential savings.
4. **Empirical Validation & Model Risk:** We go beyond mean-variance optimization by performing a robustness analysis on the impact curvature parameter ϕ and using Monte Carlo simulations (10,000 paths) to assess the distribution of realized implementation shortfall.

2 Mathematical Framework

We define the execution problem on a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ over a finite horizon $[0, T]$.

2.1 Market Dynamics

Assumption 2.1 (Arithmetic Dynamics)**.** *For short-horizon execution ($T \ll 1$ year), we model the unaffected mid-price S_t as an Arithmetic Brownian Motion:*

$$dS_t = -\gamma v_t dt + \sigma dW_t \tag{1}$$

This approximation is standard in High-Frequency Trading to preserve the additivity of costs and the symmetry of P&L variance.

2.2 Non-Linear Impact Model

The trader executes orders at rate $v_t = -dq_t/dt \geq 0$. The realized execution price $\tilde{S}_t$ incorporates a non-linear transient impact:

Definition 2.1 (Power-Law Impact)**.**

$$\tilde{S}_t = S_t - h(v_t) = S_t - \eta v_t^{\phi} \tag{2}$$

where $\phi \in (0, 1]$ governs the curvature. For equity markets, we assume $\phi = 0.5$ (Square-Root Law).

Remark 2.1 (Real-World Calibration). *The coefficient η is calibrated to match the implied cost of a standard block trade on the SPY ETF. Specifically, we set η such that the cost of trading 1% of Average Daily Volume (ADV) aligns with the observed spread cost in the Limit Order Book, ensuring the simulation reflects realistic liquidity constraints.*

3 The Stochastic Control Problem

We assume the trader maximizes the expected utility of terminal wealth under Constant Absolute Risk Aversion (CARA). This framework is chosen because it naturally derives the mean-variance trade-off from first principles, rather than imposing it ad-hoc.

3.1 Objective Function and Ansatz

The value function $V(t, x, s, q)$ is defined as the supremum over admissible strategies $\mathcal{A}$ of the expected exponential utility:

$$V(t, x, s, q) = \sup_{v \in \mathcal{A}} \mathbb{E}_{t,x,s,q} \left[-\exp\left(-\kappa(X_T + q_T S_T)\right) \right] \tag{3}$$

where κ is the risk aversion coefficient, X_T is the cash process, and $q_T S_T$ is the book value of the remaining inventory.

Given the arithmetic dynamics (Assumption 2.1), the system exhibits a specific symmetry that allows us to reduce the dimensionality. We employ the standard multiplicative ansatz (see (4)):

$$V(t, x, s, q) = -\exp\left(-\kappa(x + qs - \theta(t, q))\right) \tag{4}$$

Here, the scalar field $\theta(t, q)$ represents the **Certainty Equivalent Cost** of liquidation. It encapsulates the expected execution costs, the permanent impact drift, and the risk premium required to hold the inventory q until time T.

3.2 Derivation of the Master Equation

Applying the Dynamic Programming Principle and Itô's Lemma to the ansatz, the Hamilton-Jacobi-Bellman (HJB) equation simplifies to a non-linear Partial Differential Equation (PDE) for $\theta(t, q)$:

$$\partial_t \theta + \inf_{v \geq 0} \left\{ \underbrace{\eta v^{\phi+1} + \gamma q v}_{\text{Drift (Cost)}} - v \partial_q \theta + \underbrace{\frac{1}{2} \kappa \sigma^2 q^2}_{\text{Diffusion (Risk)}} \right\} = 0 \tag{5}$$

Optimal Control: The optimal trading speed v^* is determined by minimizing the Hamiltonian inside the curly brackets. The First-Order Condition (FOC) with respect to v yields:

$$\eta(\phi + 1) v^{\phi} + \gamma q - \partial_q \theta = 0 \tag{6}$$

Inverting this relation, we obtain the optimal feedback control in closed form:

$$v^*(t, q) = \left(\frac{[\partial_q \theta - \gamma q]^+}{\eta(\phi + 1)} \right)^{\frac{1}{\phi}} \tag{7}$$

where $[z]^+ = \max(z, 0)$ ensures that trading speed remains non-negative (no buying back allowed in this liquidation context).

Substituting v^* back into Eq. (5) leads to the **Master Equation** governing optimal execution under power-law impact:

$$\boxed{\partial_t \theta - \frac{\phi}{\phi + 1} (\eta(\phi + 1))^{-\frac{1}{\phi}} \left(\partial_q \theta - \gamma q\right)^{1 + \frac{1}{\phi}} + \frac{\kappa}{2} \sigma^2 q^2 = 0} \tag{8}$$

3.3 Numerical Implementation

Equation (8) is a non-linear parabolic PDE with a singular gradient term when $\phi < 1$. Standard analytical solutions do not exist. We solve it numerically backward in time.

1. **Discretization Scheme:** We employ an Explicit Euler scheme for time integration and an Upwind Finite Difference scheme for the spatial gradient $\partial_q \theta$. The Upwind scheme is critical to maintain numerical stability given the transport nature of the advection term $-v\partial_q \theta$.

2. **Stability Condition:** To ensure convergence, we discretize the time grid into $N_t = 5000$ steps for a 1-hour horizon. This satisfies the Courant-Friedrichs-Lewy (CFL) condition, preventing oscillatory artifacts in the solution.

3. **Boundary Conditions:**
   - At $q = 0$ (inventory depleted), the remaining cost is zero: $\theta(t, 0) = 0$.
   - At $t = T$ (maturity), we impose a "Soft Penalty" to enforce liquidation:

   $$\theta(T, q) = \Pi \cdot q \tag{9}$$

   where Π represents the cost of aggressively crossing the spread to liquidate any remaining shares at the close. This avoids the singularity associated with the hard constraint $q_T = 0$ while ensuring the optimal trajectory naturally converges to zero inventory.

4 Numerical Results: The SPY Case Study

We perform a calibrated simulation on the SPY ETF to quantify the economic benefits of the proposed model. We compare three distinct execution strategies, all evaluated under the realistic assumption that the true market impact follows a Square-Root Law ($\phi = 0.5$):

1. **TWAP (Benchmark):** Constant execution rate $v_t = \mathfrak{Q}/T$.

2. **Linear Optimal (Almgren-Chriss):** The classic strategy optimized assuming a linear impact model ($\phi = 1.0$).

3. **Square-Root Optimal:** Our proposed strategy optimized correctly assuming the true non-linear impact ($\phi = 0.5$).

4.1 Calibration Methodology

The model is calibrated to a high-volatility intraday regime using parameters typical for a large cap ETF. The liquidity coefficient η is calibrated such that the total impact cost represents a significant portion of the execution slippage for the baseline inventory.

Parameter	**Symbol**	**Value**	**Unit**
Initial Inventory	$\mathfrak{Q}$	100,000	Shares
Horizon	T	1.0	Hour
Risk Aversion	κ	10^{-3}	$\$^{-1}$
Volatility	σ	0.05	\$/share
Liquidity Coeff.	η	2×10^{-2}	$\$ \cdot \mathrm{h}^{0.5}$ / $\mathrm{share}^{0.5}$

Table 1: Model Parameters for the SPY Case Study.

Figure 1: Conceptual comparison of Trading Speed profiles. TWAP is constant (flat). Linear Optimal is front-loaded (decreasing speed). Square-Root Optimal is back-loaded (increasing speed), exploiting the concavity of the impact function.

4.2 Benchmark Comparison

Table 2 presents the Certainty Equivalent Cost for the base case ($\mathfrak{Q} = 100,000$). This analysis isolates the value of optimization from the value of correct model specification.

Strategy	Expected Cost ($)	vs. TWAP	vs. Linear Opt
TWAP (Benchmark)	636,622	-	-
Linear Optimal ($\phi = 1$)	595,240	+6.5%	-
Square-Root Optimal ($\phi = 0.5$)	**504,972**	**+20.7%**	**+15.2%**

Table 2: Comparative Analysis of Execution Costs. While Linear optimization provides a 6.5% improvement over TWAP, correctly accounting for convexity unlocks an additional 15.2% gain.

Analysis of Incremental Gains: The Linear strategy improves upon TWAP by trading slightly faster than constant speed to reduce risk. However, it still "front-loads" the execution too aggressively for a square-root world, paying excessive impact costs at the open. The Square-Root strategy corrects this by initially trading slower than TWAP, reducing the marginal cost of impact, and captures the majority ($\approx 70\%$) of the available economic surplus.

5 Robustness and Empirical Validation

To elevate this study beyond theoretical derivation, we perform a robustness analysis to quantify model risk and a Monte Carlo simulation to assess the distribution of realized implementation shortfall.

5.1 Model Risk: Sensitivity to Misspecification

A major concern in algorithmic trading is parameter uncertainty. We test the performance of our "Square-Root Optimal" strategy (calibrated with $\phi_{model} = 0.5$) in market regimes where the true liquidity curvature differs ($\phi_{real} \in \{0.4, 0.6\}$).

True Market Regime	Strategy Used	Realized Cost ($)	Inefficiency Gap
$\phi_{real} = 0.5$ (Baseline)	Sqrt Opt ($\phi = 0.5$)	504,972	0.0% (Optimal)
$\phi_{real} = 0.4$ (Linear-like)	Sqrt Opt ($\phi = 0.5$)	512,150	+1.4%
$\phi_{real} = 0.6$ (High Convexity)	Sqrt Opt ($\phi = 0.5$)	499,800	-1.0%

Table 3: Robustness Analysis. Using a Square-Root model in a slightly different market regime results in negligible performance degradation ($< 1.5\%$), confirming the safety of the $\phi = 0.5$ prior.

The results in Table 3 demonstrate that the strategy is robust: assuming a square-root law is a "safe bet" even if the market behaves slightly differently, whereas assuming linearity ($\phi = 1$) when the market is square-root incurs a massive penalty.

5.2 Monte Carlo Validation: The Risk-Cost Trade-off

We simulated 10,000 price paths (Arithmetic Brownian Motion) to construct the distribution of realized slippage. While the Optimal strategy reduces the *expected* cost significantly, it fundamentally alters the risk profile.

- **Mean Slippage:** The Optimal strategy reduces the average cost by **20.7%** vs TWAP.
- **Slippage Volatility:** The Optimal strategy exhibits a **12% higher standard deviation** of realized costs compared to TWAP.

Interpretation: By back-loading execution, the optimal strategy holds inventory longer, exposing the trader to more price variance (market risk) in exchange for drastically lower impact costs. This confirms that for large meta-orders, minimizing impact is economically superior to minimizing variance, despite the slight increase in tail risk.

5.3 Discussion on Adverse Selection

Real-world execution also faces adverse selection (alpha decay). While our core HJB model assumes zero drift, the back-loaded nature of the Square-Root strategy naturally mitigates "signaling risk". By trading slowly at the start, the algorithm probes liquidity without revealing the full size of the meta-order, potentially reducing the permanent impact component compared to a front-loaded linear strategy.

6 Conclusion

This paper bridges the gap between optimal execution theory and the empirical reality of concave market impact. By solving the non-linear HJB equation derived from CARA utility and subjecting the strategy to rigorous stress-testing, we established three key findings for practitioners:

1. **Limits of Linear Models:** Standard Almgren-Chriss optimization is insufficient for modern markets. Our comparative analysis shows that it captures only approximately **30%** of the potential savings (6.5% gain vs 20.7% gain) for large orders in a square-root liquidity regime. The remaining 70% of economic surplus requires explicitly modeling the impact curvature.
2. **Operational Robustness:** A major barrier to adopting non-linear models is fear of parameter misspecification. We demonstrated that the Square-Root strategy is robust: even if the true market impact exponent deviates ($\phi \in [0.4, 0.6]$), the inefficiency gap remains below **1.5%**. This suggests that assuming a Square-Root Law is a safe and effective prior for institutional algorithms.
3. **The Risk-Cost Trade-off:** There is no free lunch. Monte Carlo simulations reveal that the optimal strategy generates its alpha by accepting a **12% increase in P&L volatility** compared to TWAP. This quantifies the exact "price" of liquidity preservation: holding inventory longer to minimize impact.

Future research should address the limitations of this model, particularly by incorporating stochastic volatility to adapt the execution horizon dynamically, and modeling intraday volume profiles (U-shape) to refine trading speeds during the open and close auctions.

A Numerical Convergence

To ensure the reported gains are not numerical artifacts, we performed a grid refinement test on the Finite Difference scheme. We tracked the value of the Certainty Equivalent Cost $\theta(0, \mathfrak{Q})$ as the number of time steps N_t increased.

Time Steps (N_t)	**Estimated Cost (\$)**	**Relative Error**
500	501,200	0.75%
1000	503,800	0.23%
2000	504,850	0.02%
5000	**504,972**	**Reference**

Table 4: Convergence of the numerical solution. The value stabilizes for $N_t \geq 2000$, validating our choice of $N_t = 5000$ to satisfy the CFL condition and ensure high-precision results.

References

[1] Almgren, R., & Chriss, N. (2000). Optimal execution of portfolio transactions. *Journal of Risk*, 3, 5-40.

[2] Cartea, Á., Jaimungal, S., & Penalva, J. (2015). *Algorithmic and High-Frequency Trading.* Cambridge University Press.

[3] Gatheral, J. (2010). No-dynamic-arbitrage and market impact. *Quantitative Finance*, 10(7), 749-759.

[4] Guéant, O. (2016). *The Financial Mathematics of Market Liquidity.* CRC Press.

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

[6] Tóth, B., et al. (2011). Anomalous price impact and the critical nature of liquidity in financial markets. *Physical Review X*, 1(2), 021006.