Modeling Temporary Market Impact and Optimal Execution Framework

1 - Modeling Temporary Impact $g_t(x)$

Introduction

When executing a buy order of size x, one "walks the book" by consuming ask-side liquidity at successive prices. The temporary impact $g_t(x)$ is defined as the difference between the volume-weighted executed price and the prevailing best ask (or midprice) at time t. A naïve approach assumes linear slippage, $g_t(x) \approx \beta_t x$, but empirical microstructure and academic studies consistently find that marginal cost decreases with order size—i.e., the impact curve is sub-linear.

Why Linear Models Fall Short

Non-uniform depth. The limit order book (LOB) exhibits irregular depths: the best ask might have a small posted quantity, deeper levels often display larger quotes. A single slope β_t ignores this heterogeneity and over-charges large orders that tap deeper, more liquid levels while under-charging very small orders that sweep only the top of book.

Diminishing marginal impact. Suppose the best ask has only 100 shares at \$10.00, and the second ask has 500 shares at \$10.01. A 100-share order pays \$10.00; a 200-share order pays

$$\frac{100 \times 10.00 + 100 \times 10.01}{200} = 10.005.$$

The incremental cost for the second 100 shares is only \$0.005, not the first 100's \$0.00—clearly not a constant per share.

Empirical curvature. Plotting the simulated impact ΔP vs. order size x on log–log scales for our three tickers (CRWV, FROG, SOUN) reveals curves of slope < 1. A linear model ($\alpha = 1$) would force these to lie on a 45° line, which visibly misfits both small and large orders.

Power-Law Impact: $g_t(x) = \beta x^\alpha$

We therefore adopt the two-parameter power-law form:

$$g_t(x) = \beta x^\alpha, \quad 0 < \alpha < 1,$$

where:

- x = order size (shares)
- β = scale parameter (controls overall impact magnitude)
- α = curvature parameter (governs diminishing marginal costs)

Methodology

For each symbol:

LOB Reconstruction:

- Ingest raw MBP-10 events (adds, cancels, trades) from one-minute snapshots of top 10 levels.
- Maintain arrays of bid/ask prices and sizes, updating per event.

Impact Simulation: For order sizes $x \in \{100, 500, 1000, 2000, 5000\}$, "walk" the ask side:

$$\text{total cost} = \sum_i \min(\text{remaining}, \text{ask\_sz}_i) \times \text{ask\_px}_i$$

$$\Delta P = \left(\frac{\text{total cost}}{x}\right) - \text{best ask}_t$$

Parameter Estimation:

- Flatten $\{(x, \Delta P)\}$ across all timestamps.
- Fit $\Delta P = \beta x^\alpha$ via nonlinear least squares.

Fitted Parameters

Symbol	α	β	**Interpretation**
CRWV	0.2569	0.0049	Strongly diminishing impact (deep book)
FROG	0.4784	0.0059	Near "square-root law," typical of liquid assets
SOUN	0.0537	0.1542	Almost all cost at best ask; very lumpy book

$\alpha < 1$ unambiguously rejects linearity, confirming sub-linear impact.

FROG's $\alpha \approx 0.48$ sits near the canonical 0.5; CRWV is more concave; SOUN is extremely steep (book depth is concentrated at the top).

Economic Significance

Algorithmic execution. In highly liquid FROG, one can scale up orders with moderate cost growth ($\alpha \approx 0.5$). In CRWV and especially SOUN, large orders incur rapidly diminishing marginal cost savings—algos should slice even more finely.

Risk management. Sub-linear models better estimate slippage risk: a linear model would overestimate costs for FROG (discouraging beneficial liquidity consumption) and under-estimate for SOUN (exposing to hidden depth risk).

Conclusion

Modeling temporary impact with a power-law $g_t(x) = \beta x^{\alpha}$ captures both empirical sub-linearity and liquidity differences across stocks. Our three-ticker study, though small, demonstrates clear curvature ($\alpha \ll 1$) and varying scale (β), vindicating this approach over crude linear slippage. This model thus provides both theoretical fidelity and practical guidance for optimal execution in diverse microstructure environments.

2 - Execution Algorithm Framework for Allocating x_i at t_i

Problem Setup

Let S denote the total size of an order to be executed over N discrete time periods $\{t_1, t_2, ..., t_N\}$. Let x_i be the number of shares executed at time t_i, with the constraint:

$$\sum_{i=1}^{N} x_i = S.$$

Each x_i induces a temporary price impact modeled as:

$$g_i(x_i) = \beta_i x_i^{\alpha_i},$$

with known or estimable parameters β_i, α_i (from Question 1), and $0 < \alpha_i < 1$ ensuring sub-linearity.

Objective

Minimize the total impact cost over the execution horizon:

$$\min_{x_1,\ldots,x_N} \sum_{i=1}^{N} g_i(x_i) = \sum_{i=1}^{N} \beta_i x_i^{\alpha_i}, \quad \text{subject to } \sum_{i=1}^{N} x_i = S, \quad x_i \geq 0.$$

Solution Approach

Offline Case (Full Information):

Use the method of Lagrange multipliers. Define:

$$\mathcal{L} = \sum_{i=1}^{N} \beta_i x_i^{\alpha_i} - \lambda \left(\sum_{i=1}^{N} x_i - S \right).$$

Setting $\partial \mathcal{L} / \partial x_i = 0$ yields:

$$\alpha_i \beta_i x_i^{\alpha_i - 1} = \lambda.$$

Solving:

$$x_i = \left(\frac{\lambda}{\alpha_i \beta_i} \right)^{\frac{1}{\alpha_i - 1}}.$$

Substitute into $\sum x_i = S$ to find λ, then compute x_i.

Online / Causal Approximation:

If β_i and α_i are estimated dynamically using recent book data (as in our snapshot-based fitting), a rolling update strategy may be adopted:

- Estimate (β_i, α_i) at time t_i using current LOB.
- Allocate x_i proportional to the inverse marginal cost:

$$x_i \propto \left(\frac{1}{\beta_i \alpha_i}\right)^{1/(1-\alpha_i)}.$$

- Normalize to ensure $\sum x_i = S$.

Remarks

This framework accommodates both static and adaptive execution. It balances cost efficiency (via curvature α_i) with sensitivity to liquidity conditions (via β_i). Sub-linear curvature encourages temporally distributed execution (e.g., time-weighted average price), while steep impact (as in SOUN) urges finer slicing and liquidity-aware scheduling.